Young drivers experience: the results of a second phase training on higher order skills

Saskia de Craen, Jan Vissers (Traffic Test), Maura Houtenbos & Divera Twisk

R-2005-8

INSTITUTE FOR
ROAD SAFETY RESEARCH

Young drivers experience: the results of a second phase training on higher order skills

Evaluation study in the framework of the European project NovEV

R-2005-8
Saskia de Craen, Jan Vissers (Traffic Test), Maura Houtenbos & Divera Twisk
Leidschendam, 2005
SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number:	R-2005-8
Title:	Young drivers experience: the results of a second phase training on higher order skills
Subtitle:	Evaluation study in the framework of the European project NovEV
Author(s):	Saskia de Craen, Jan Vissers (Traffic Test), Maura Houtenbos & Divera Twisk
Project leader:	Divera Twisk
Project number SWOV:	39.401
Keywords:	Recently qualified driver, experience (human), driver training, education, skill (road user), risk, perception, situation awareness, driving (veh), control, behaviour, before and after study, evaluation (assessment), Netherlands.
Contents of the project:	To diminish the high accident risk of young drivers, new methods for accident prevention are being investigated. This study in the framework of the European project NovEV evaluates the effects of a post-license training on higher order skills. This second phase driver training consisted of an on-road feedback drive, a training on closed track, a group discussion and a post-test on-road feedback drive. Using a before-and-after design, the effects of the training were evaluated using a control group.
Number of pages:	72 + 20
Price:	€ 19,-
Published by:	SWOV, Leidschendam, 2005

SWOV Institute for Road Safety Research
P.O. Box 1090
2290 BB Leidschendam
The Netherlands
Telephone +31 70 317 33 33
Telefax +31 70 320 12 61
E-mail info@swov.nl
Internet www.swov.nl

Summary

A new approach to tackle the high accident risk of young, novice drivers is post-license training. In this study the effect of a second phase driver training is evaluated using a before-and-after design with a control group.

Second phase driver training

In the basic (pre-licence) training phase, drivers are trained with respect to vehicle control and the mastery of traffic situations. What is essential in this phase is the faultless and automatic application of such driving routines. After completing his/her basic driving course, and passing the exam, the novice driver gains experience, but also is exposed to new risks. After six months of independent driving, he has driven in 'unfamiliar' situations, encountered new traffic situations, has started to develop his own driving style and to regard car driving as a means to an end (e.g. to go to a party to have fun) rather than as a meaningful activity in itself. These new developments in the novice driver's career call for a second phase in driver training.

Participants

After an appeal by mail and telephone, 376 young novice drivers agreed to participate in the project. Unfortunately, during the course of the project, many of the participants dropped out. Out of 376 young drivers that initially agreed to participate, only 127 (33%) completed all parts of the project.

The participants who did not want to participate, those who dropped out, and those who finished all parts of the project were compared for a number of variables. This led to the conclusion that there was no major problem with selective drop-out. Naturally, the groups did differ on at least one aspect, namely for one reason or another some completed the project and others did not.

Training programme and objectives

The second phase training consisted of the following modules:

– *An on-road feedback drive*
  The objective of the feedback drive was to present the driver with feedback about his driving performance. It was different from instruction drives, as the instructor confronted the driver with his 'expert' observations in order to make the participant 'think' and reflect. So he did not tell the participant what to do, but encouraged him to draw his own conclusions. During the first feedback drive the participant and instructor were accompanied by a second participant who rode along as a passenger. The drive was followed by a discussion between instructor, passenger and driver.

– *Training on a closed track*
  The objective of the track training was for participants to experience the

limits of their skills in vehicle control and to share these experiences with other group members.

– *A group discussion*
The objective of the group discussion was to stimulate recognition of potentially hazardous situations in rather 'normal' social situations. The discussion was based on video sketches, depicting typical situations (incidents rather than accidents) involving young drivers (men and women). The moderator encouraged the youngsters to reflect on the events.

– *An evaluation on-road feedback drive (about a month later)*
The objective of this second feedback drive was the same as the first feedback drive, that is to present the driver with feedback about his driving performance.

Evaluation design and data collection methods

	Training programme		Instruments
	Experimental	Control	
December 2003 *Pre-test* One month before training	Questionnaire	Questionnaire	*Questionnaire* Contained items on risk awareness, self-assessment of skill, and situation judgements
January 2004 *Training day*	Pre-test feedback drive	Pre-test feedback drive	*On-road observation form* An assessment tool to describe the driving performance of a driver. The driver himself and the driving instructor completed these forms after the feedback drive. *Driving Assessment* Assessment by the instructor of the quality of driving in three fields: vehicle control, driving skills, and calibration skills
	Track Exercises		
	Group discussion		
February 2004 *Post-test* One month after training	Questionnaire	Questionnaire	*Questionnaire* Contained items on risk awareness, self-assessment of skill, and situation judgements
	Post-test feedback drive	Post-test feedback drive	*On-road observation form* An assessment tool to describe the driving performance of a driver. The driver himself and the driving instructor completed these forms after the feedback drive. *Driving Assessment* Assessment by the instructor of the quality of driving in three fields: vehicle control, driving skills, and calibration skills. *Satisfaction questionnaire* This questionnaire contained questions on how satisfied participants were about the different components of the training day and the feedback drives.

Evaluation design and data collection methods.

The effect of the track training and group discussion was studied using a before-and-after design with a control group. Participants were randomly assigned to the control or the experimental (treatment) group. The *control group* participated in both feedback drives. In addition to the feedback

drives, the *experimental group* also participated in track training and in a group discussion.

Results and conclusions by instrument

Satisfaction questionnaire
Young drivers were not very motivated to participate on a voluntary basis in a second phase training. However, once after the course, novice drivers were enthusiastic about the training day. Within the training day, the group discussion was rated as the least attractive part, while the feedback drive was considered about as attractive and useful as the track training. The message of the second-phase training was well understood. There were no indications that the young, novice drivers overestimated their skills, as a result of the training.

Questionnaire
The questionnaire contained items on risk awareness, self-assessment of skill and judgements of traffic situations on photo. The results from the questionnaire are somewhat unclear; some effects of the training were found, but not consistent and not always in the expected direction.

In line with expectations, the items concerning risk awareness confirmed that young drivers do not seem particularly concerned in general, and especially not about driving too fast. A least 60% of the respondents are not concerned about driving too fast. On the other hand, it turned out that young drivers are, overall, rather confident about their driving skills. At least 30% of the participants believe they are (very) strong in all skills, and in some skills more than 60% believe they are (very) strong.

It was expected that these opinions would improve as a result of the training day, in the sense that respondents would see more danger and be less confident about their driving skills. Detailed analyses showed no effect of training on these variables. Further research is needed to demonstrate that the questionnaire itself is sensitive enough to register changes as a result of a short-term intervention. The fact that there were significant gender differences in these issues, led to the conclusion that this part of the questionnaire possibly measures more stable attitudes or personality traits (which could not be changed with a one-day training course or within the period of a month).

On-road observation form
After the feedback drive, both the instructor and the participant filled out an on-road observation form. This form contained items on driving skill and assessment of complexity of the driving task. The young drivers' assessment of their own driving skills and task complexity did not change as a result of training. This implies, that the objective of the course to inform young drivers about their limited skills and the high complexity of the traffic situation did not result in a more cautious self-estimation. On the positive side, this result indicates that the training day and more in particular the track training did not lead to a higher estimation of skills and a lower estimation of the complexity of the driving task.

To study the accuracy of the driver's self image, their self-estimation scores were compared with the instructor's assessment of the young driver's

competencies. On 'vehicle control and general skills', instructors and participants did not differ in their assessment neither on the pre-test nor on the post-test. As expected on 'safe and defensive driving' in the pre-test, participants rated their performance higher than the instructor did. As the course was directed at improving self-assessment skills, it was expected accuracy to improve in the sense that their assessment would be more in line with that of the instructor after the training. This was not the case.

Generally, from the results from the on-road observation form, it can be concluded that while the instructors did see some improvement as a result of the training, the participants did not.

Driving Assessment
Task conditions between control group and experimental group differed systematically on the pre-test. Therefore, it cannot be excluded that the observed difference in task performance between control group and experimental group is a reflection of these test conditions rather than a significant difference between the two groups.

Within the experimental group, the performance of the participants of the two different training locations differed significantly. This, despite the fact that at both locations the participants had received exactly the same training (on paper). Where performance at location A (Lelystad) was improved by training, driving performance at location B (Rijssen) got even worse. Because the test conditions for the participants of the two locations were the same, this result is reliable.

The process evaluation indicated that despite their organisation's involvement in the NovEV project, the trainers from location B did not share the same opinion on the definitions of a 'useful' training. As a result, these trainers had to give a type of training they did not believe in. This could have (subconsciously) affected the way they gave the training, or the way the participants perceived the training. Research has shown (ADVANCED, 2002) that any education looses its strength if the educator is not absolutely convinced about what he/she is teaching. Moreover, that the effectiveness of the education is largely dependent on the person, the beliefs of the teacher, and his behaviour (Hale & Glendon,1987). For a more detailed discussion of the role of the 'teacher', see the ADVANCED report.

General conclusions

In the Dutch pilot, the recommendations of the ADVANCED report were closely followed with respect to the content of the course and the evaluation of its effects. However, as stated earlier, in practice these recommendations were not always followed in one of the two locations.

In this study, it has been demonstrated that, on the one hand, the second phase is recognized by the participants as a useful and necessary part of their driving career. On the other hand, the high refusal rate demonstrates that youngsters are not interested in participating on a voluntary basis. The effects of the course are limited, and can even be negative, if trainers are not fully equipped to give the course, indicating that a much greater effort is needed in training second phase trainers than has been the case in this project.

Contents

Foreword

This evaluation study was conducted within the framework of the European project *NovEV: Evaluation of post-licence training schemes for novice drivers* (Sanders & Keskinen, 2004), which was coordinated by CIECA. In this European project, the effect of a second phase driving course was evaluated in five countries.

The Dutch contribution to the NovEV project was based on two evaluations: a process evaluation, carried out by the Traffic Test company in the Netherlands, and an effect evaluation carried out by SWOV. The results of the effect evaluation are described in this report. The results of the process evaluation is reported in Vissers et al. (In press; in Dutch language) and in the European report (Sanders & Keskinen, 2004). It deals primarily with the implementation and organization of the course. The results are based on interviews with all partners involved: organizations, trainers, instructors and researchers. An abbreviated version of the process evaluation can be found in this report in *Chapter 3 (The process evaluation: implementing the training).*

The design, implementation and evaluation of the second phase course was carried out by a consortium of partners who all contributed in the financing, organization and expertise in the project (See *Appendix 1* for a list of partners). These partners came from many fields, like exam and training centres, driving schools, and research and governmental organizations. The general coordination was carried out by the ROVG (Regional Road Safety Council of Gelderland).

1. Introduction

The high accident risk of young/novice drivers has led to initiatives within Europe to find new methods for accident prevention. One such possibility is a new approach to driver training, in particular to post-license training. In the European project ADVANCED the basic principles of a successful advanced driver training are described. The ADVANCED project concluded that the primary objective of advanced driver training is to enhance and stimulate the development of higher order skills. These skills are related to hazard perception, self-assessment and situation awareness, and are known to develop relatively slowly in comparison to other driving related skills such as vehicle handling and the mastery of traffic situations.

Many studies indicate that there is a relation between hazard perception and self-assessment of skills. For example Brown (1989) claims that the perception of risk cannot be studied in isolation of both these elements. This balance between hazard perception and self-assessment has been called calibration. Calibration is seen to be an essential element in safe driving. At any moment in time, a driver needs to be actively engaged in assessing what the driving task requires in terms of actions or the avoidance of actions, and the potential difficulties involved (Kuiken & Twisk, 2001).

In brief, there are indications that young drivers underestimate the risk of an accident in a variety of hazardous situations At the same time there seems to be a problem with the assessment or evaluation of one's one driving skills. For example, the young driver underestimates what is needed to cope with a dangerous situation; they overestimate their own driving skill (Deery, 1999). McKenna et al. (1991) concluded that this overestimation of driving skills is caused by a 'positive self' rather than a 'negative other' bias. This could be caused by the fact that the young driver has encountered only a limited number of critical traffic situations, which may provide a false sense of mastery and safety. Therefore, a second phase driver training could be used to eliminate or reduce this false sense of mastery and safety. In other words, improve calibration skills.

There is evidence to suggest that especially the sub-standard levels of higher order skills in novice drivers are one of the main causes of their increased crash rate (Deery, 1999; Willems & Cuyvers, 2004; Engström et al., 2003). In the basic (pre-licence) training phase, drivers are trained with respect to vehicle control and the mastery of traffic situations. What is essential in this phase is the faultless and automatic application of such driving routines. Of course, issues such as hazard perception and risk awareness are addressed, but as driving experience is still very limited at this stage, the effect is probably relatively small. After completing his/her basic driving course, and passing the exam, the novice driver gains experience, but also is exposed to new risks. After six months of independent driving, he has driven in 'unfamiliar' situations, encountered new traffic situations, has started to develop his own driving style and to regard car driving as a means to an end (e.g. to go to a party to have fun) rather than as a meaningful activity in itself.

These new developments in the novice driver's career calls for a second phase in driver training. The objective of the second phase is to address these experiences and to contribute to the prevention of the associated risks. Furthermore, it is essential that any training should avoid overconfidence to develop. Research findings suggest that advanced multi-phased training for novice drivers that focuses on vehicle skills like skid control and emergency manoeuvring skills is counterproductive (Glad, 1988; Gregersen, 1996). Therefore, any second phase training needs to ensure, that such overconfidence does not result from training. This creates a dilemma for the second phase training. On the one hand, in order to be effective it is important for participants to be highly motivated and to find the courses attractive and stimulating. On the other hand the type of training that is most attractive (namely vehicle handling skills like skidding) should be excluded from the course.

To conclude, the expected effectiveness of second phase of driver training depends on:
- the adequacy of the training module to stimulate and to enhance higher order skills;
- the timing of the second phase in the total learning process;
- the extent to which the chance of overconfidence developing is minimized;
- its attractiveness for the target group.

The question that also needs to be answered is the relationship between the quality of the basic driving course and effectiveness of the second phase. Insight in this relationship is currently missing. On the one hand, it can be reasoned that a poorly educated driver does not benefit from advanced driving courses because of his poor driving routines. On the other hand, it seems likely that poorly educated drivers benefit the most, because for them there is still a lot to learn.

1.1. The content of the NovEV Dutch second phase training

In the Netherlands, driver training consists only of a basic (pre-license) driving phase. The content and training methods are not standardized, and it is left to the driving instructor to decide on how and what to teach. Driving standards are ensured by the content, reliability, and validity of the compulsory driving test. The government sets these standards.

The impact of the basic driving phase is limited and for this reason the Netherlands is participating in the NovEV project in which experts in the field have designed, implemented and evaluated a second-phase course.

The Dutch training course had the following objective: *"To enhance self-assessment skills, risk and safety awareness by feedback and training/coaching with respect to an individual's driving behaviour, personal style and decision-making characteristics".*

The structure and content of the training course followed closely the best practice recommendations of ADVANCED (p. 134-138), and consisted of the following elements:
- an on-road assessment drive (first feedback drive);
- training on a track;

- a group discussion;
- an on-road evaluation drive (second feedback drive; about one month later).

The training took place at two locations:
1. the ANWB (Dutch Automobile Club) track training site situated near Lelystad; in this report it is frequently referred to as 'Lelystad';
2. VVCR (post-licence driver training centre) track training site situated near Rijssen, referred to in this report as 'Rijssen'.

In both locations, the course structure and content was the same, and a detailed outline was described in the blueprint. However, differences did occur in the actual implementation and execution of the course. More detail on this can be found later in this report.

The exact content of the training elements depended on the 'specific needs' of a driver. To assess the specific needs of a driver, instruments were developed that were used to identify their particular weaknesses. These instruments were used for diagnostic purposes by the trainer and to provide feedback to the learner driver.

1.2. Research questions

In this evaluation study the following questions will be addressed:
- Which changes (in knowledge, attitudes, intended behaviour, and driving behaviour) can be observed that can be attributed to the second phase driving course?
- Does the training have an effect on those young drivers that are most at risk?
- How attractive is the training for young, novice drivers?

Originally it was the intention to address the following questions as well. However, for several reasons, it was not possible to answer these questions in this report:
- What is the relationship between the quality of the basic training and the effectiveness of the second phase?
  Because of the number of participants dropping out of the programme, there were not enough participants to reach sufficient power in the statistical analysis for a distinction between different qualities of the basic training (see *§ 2.3.4*: *Estimated power of the design*).
- What is the time span of the effect, and has the training led to new behaviour or insight that still develop after training?
  It would have been very interesting to see if the positive changes as a result of the training are still visible after a longer period, or if new developments occur after the training. However, it was not possible within the scope of this project to follow the participants for a longer time, (for example with a second post-test after a year).

1.3. Evaluation

To assess the effectiveness of the second phase training, the participants were divided into two groups. The first group (experimental group) would follow every part of the training. In other words, they would attend both feedback drives, the track training and the group discussion. The second group (control group) would only attend the feedback drives. The difference

between both groups would then reveal the effect of the track training and group discussion.

In order to measure the differences between the experimental and control groups, several instruments were developed:

- questionnaires concerning driving skills, self-assessment and risk awareness;
- diaries: semi-structured questionnaires in which driving events were reported by the participants;
- on-road observation form: an assessment tool to describe the driving performance of a driver;
- driving assessment: assessment by the instructor, based on the feedback drives, of the quality of driving in three fields: vehicle control, driving skills, and calibration skills[1];
- satisfaction questionnaire: this questionnaire contains questions on how satisfied participants were about the different components of the training day and the feedback drives.

The evaluation design and measurement instruments conform to the quality criteria for evaluation research. (ADVANCED, 2002; p. 139-150).

[1] Calibration is defined as the balance between self-assessment of skill and risk awareness. A central element in this balance is the skills a driver actually has, versus the skills the driver *thinks* he has.

2. Method

2.1. Participants/subjects

The aim was to have about 300 young novice drivers to participate in the project. Addresses of newly licensed drivers, between the age of 18 and 25, were obtained from the Central Licensing Bureau. Those that received their license in the summer of 2003 received a brochure and an invitation by mail to participate in a 'challenging' (and free of charge) safe driving course. The participant's travel expenses were refunded and, to stimulate participation, participants could win a holiday for two or free car insurance for one year by entering a lottery. As too few participants accepted the written invitation (about 10%), a new group of novice drivers were invited by telephone.

The young drivers who did not want to participate in the project (approximately 140) were asked to answer a few questions so they could be compared with the people who did want to participate, thereby checking for a selection bias. One of the questions asked concerned the reason why they did not want to participate in the project. *Figure 2.1* shows the percentages for each reason. (The percentages do not add up to a hundred, because it was possible to give more than one reason for non-participation).

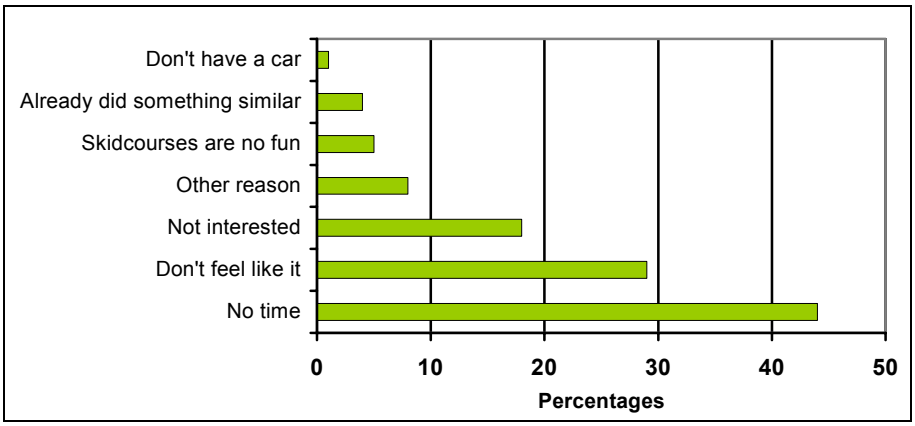

Figure 2.1. *Reasons for not participating in the project.*

The most common reasons for not participating in the project were lack of time and lack of interest.

After the round of phone calls, 376 young novice drivers agreed to participate in the project. Unfortunately, during the course of the project, many of the participants dropped out. The participants were expected to come to a location twice, the first time for a feedback drive and for some participants a training (experimental group), the second time only to attend a feedback drive. *Table 2.1* shows the attendance on both feedback drives.

		Post training feedback drive		Total
		Absent	Present	
Pre training feedback drive	Absent	198	15	213
Pre training feedback drive	Present	36	127	163
Total		234	142	376

Table 2.1. *Attendance in the feedback drives.*

Of the 376 young drivers that initially agreed to participate in this project, only 127 completed all parts of the project. Most of the participants dropped out before the first feedback drive, the first time they had to come to a location. One explanation is that participants became aware of what the contents of the project exactly were. We suspect that when the participants signed up for the project during the round of phone calls, most of them thought it was a one-day skid course. When they found out that they were expected to fill in questionnaires and had to come to a location twice, they dropped out. The extreme weather conditions – there was a traffic warning issued not to go outside unless you really had to, on some training days – and the distance to the training locations (Lelystad and Rijssen – a more than one hour drive for most of the participants) could also explain the high number of absentees.

Overall, this means that out of the 500 participants that were contacted to participate, about 340 were not interested enough to participate. About 140 refused immediately, and about 200 changed their mind later on. This indicates that a second phase training is not something that young, novice drivers would attend voluntarily. Of course in this project they had to do much more than just a one day training; they had to fill in a number of forms, and come to a location twice. Moreover, there was a strict timeframe when they had to attend the training; it was not even possible to attend one week later.

2.1.1. *Selection bias*

Because of the high percentage of dropouts, before and after the start of the project, it is important to realize that such a dropout can be selective, thereby causing a selection bias. If, for example, relatively more woman than men drop out or relatively more experienced drivers, the sample would not be representative anymore. To study the selectiveness, all novice drivers (whether they were willing to participate or not) were asked questions about their age, training and driving experience.

In order to estimate the selectiveness of the dropouts, five groups were compared (see *Table 2.2*). The first group in this table consists of young drivers who did not want to participate in the project when they were asked by phone (n=138). The second group agreed to participate, but dropped out before the first day, or did not show up on the first day. The third group consists of those who were present the first day, but dropped out before the second day (n=36). And the fourth group consists of those who could not make it on the first day (the pre-training feedback drive) but were motivated to come the second day (n=15). None of these groups were used in the main

analyses, but their scores were used illustratively in most of the figures. For example the group who did the first feedback drive, but was absent from the second, was used for their scores on the first feedback drive. The last group in *Table 2.2* consists of the respondents who finished all parts of the project (n=127). This group of 127 subjects was used in the main analyses.

		Not in the study				In the study
		Did not agree to participate after phone call (N=138)	Absent – Absent (N=198)	Present – Absent (N=36)	Absent – Present (N=15)	Present – Present (N=127)
Gender	% Male	50%	48%	56%	40%	61%
Age	Mean	21	20	20	21	20
	Standard	2	2	1	2	1
Number of months drivers' licence	Mean	9	9	8	7	8
	Standard	2	2	2	2	2
Hours of training for drivers' licence	Mean	42	42	39	46	39
	Standard	18	16	17	27	17

Table 2.2. *Selective drop-out: group comparison.*

As can be seen in the table, there is not much difference between the group of respondents that did not want to participate after the phone call, the respondents that did not finish all parts of the project and respondents who did participate in all parts. The only noticeable difference is that in the final sample of young drivers 60% is male. The results led to the conclusion that there is no problem with self-selection and selective drop-out.

It should be remembered that the respondents in the study and those who dropped out somewhere along the way, could only be compared for a couple of variables. These variables indicate that there is not much difference between the groups. Naturally, these groups do differ in at least one aspect, namely for one reason or another some completed the project and others did not.

2.2. **Training programme**

2.2.1. *Pre-test and post-test feedback drive*

The objective of the feedback drive was to present the driver with feedback about his driving performance. It consisted of a drive on different public roads. It was different from instruction drives, as the instructor confronted the driver with his 'expert' observations in order to make the participant 'think' and reflect. So he did not tell the participant what to do, but encouraged him to draw his own conclusions. The instructors were examiners form the Dutch driver testing centre and driver instructors with extra qualifications in the field of driver training and coaching.

In the pre-test feedback drive, the participant and instructor were accompanied by a second participant who rode along as a passenger. The drive was followed by a discussion between instructor, passenger and driver. As input for the discussion, the 'on-road observation form' for the participant-driver was completed by the participant-driver himself, by the passenger, and by the instructor. In the post-test feedback drive, no second participant was present.

The locations of the feedback drives were rather different for the experimental group than for the control group. The feedback drive for the experimental group had to take place in the vicinity of the training location, which is located more than one hour's drive from their hometown. Probably for most participants, the environment was unfamiliar. The control group, however, performed their feedback drives in an area close to their own town. This was frequently the area in which the participant had taken driving lessons and/or their exam.

During the second feedback drive, all participants (both experimental and control group) drove in an exam-area close to their town. In contrast to the initial feedback drive, individual participants were alone with the trainer. The process evaluation interviews demonstrated that, in the instructors' opinions, their assessment of the participant's driving performance was partly dependent on the area in which the feedback drives took place. In their view, it is likely that the feedback drives in familiar conditions were more positively assessed than drives that took place in more unfamiliar conditions.

2.2.2. *Track training*

The trainers who gave the track training, were employees at the track-site and experienced trainers in voluntary, post-licence driving courses. The track training consisted of the following exercises:

- ABS and non-ABS braking exercises: 30 and 50km/h
  Goals: understanding the differences between ABS and non-ABS, experiencing the sensation of ABS braking, understanding the effect of speed on braking distances.
- Demonstration at 50 and 60km/h and showing the effect on braking distance.
- Driving on to the verge
  Goal: to experience the sensation of going on to the verge and semi-loss of control.
- Aquaplaning
  Goal: to show inability to steer when aquaplaning. Participants were inside the car when the instructor demonstrated the exercise.
- Driving around bends
  Goal: to show how small increases in speed can cause the vehicle to slide when driving around bends.
- Parallel braking exercises
  Goal: to show how easy it is to cause a pile-up unless proper safety margins are respected.

2.2.3. *Group discussion*

The group discussion was based on video sketches. The video depicted typical situations (incidents rather than accidents) involving young drivers (men and women). It highlighted issues such as distractions: music, mobile phones, peer pressure, passers-by; multi-tasking, pressure from other drivers, tailgating (insufficient safety margins), vehicle loading, etc. The moderator encouraged the youngsters to reflect on the events. The objective was to stimulate recognition of potentially hazardous situations in rather 'normal' social situations. The intention was to use the video sketches as a basis for further, more spontaneous discussion amongst the participants, led by the trainer.

2.3. **Evaluation design and timetable**

2.3.1. *Evaluation design*

The effect of the track training and group discussion was studied using a before-and-after design with a control group. Participants were randomly assigned to the control or the experimental (treatment) group. The *control group* (n=28) participated in both feedback drives. In addition to the feedback drive, the *experimental group* (n=99) also participated in the track training and the group discussion.

Table 2.3 shows a comparison between the experimental group and control group on four variables. Two other groups are also described. These are the 'no intervention' group, which consists of participants who were originally assigned to the experimental or control group but did not show up for the training day. Therefore they received no intervention at all, neither feedback drive, track training nor group discussion. The fourth group are those participants who did not show up for the second feedback drive. There is only information available on their performance before the training.

The comparison shows that the experimental group and control group do not differ on these four variables.

		Experimental group (N=99)	Control group (N=28)	No intervention (N=15)	Drop out (N=36)
Gender	% Male	61%	61%	40%	56%
Age	Mean	20	20	21	20
Number of months drivers' licence	Mean	8	9	8	9
Hours of training for drivers' licence	Mean	38	41	46	40

Table 2.3. *Comparisons between experimental group and control group.*

For research purposes, the most favourable design would be 'double-blind'. This would exclude shifts in results due to expectations of the participants and the assessor (the instructor). In a double-blind test design, neither the participant *nor* the assessor would know whether the participant is a member of the control or the experimental group and whether the feedback drive is

before or after the training. In this study, the instructors were very much aware if the feedback drive concerned the first or second drive, for the simple fact that it was also their first or second series of feedback drives. The instructors were, however, not aware of the assignment to groups. It was not possible, of course, to prevent the participants and instructors from talking about their experiences in the programme so far.

2.3.2. *Data collection methods*

The following instruments were used in the evaluation:

– Questionnaires (see *Appendix 2*): about driving skills, self-assessment and risk awareness. This questionnaire was based on previous work by Hatakka (1998). In this questionnaire, risk awareness was also measured using photographs of 'normal' traffic situations.

– Diaries: semi-structured questionnaires in which driving events were reported by the participants. The results of these diaries exceed the scope of this report, and will not be presented here.

– On-road observation form: an assessment tool to describe the driving performance of a driver. The driver himself, the passenger, and the driving instructor completed these forms after the feedback drive. This instrument was not only used for research purposes. By comparing the three assessments, the forms were also used as input for the discussion after the feedback drive.

– Driving assessment (see *Appendix 4*): assessment by the instructor of the quality of driving in three fields: vehicle control, driving skills, and calibration skills. In contrast with the on-road observation form, the instructor filled out the driving assessment in private. And the results were not discussed with the participants.

– Satisfaction questionnaire: this questionnaire contains questions on how satisfied participants were about the different components of the training day and the feedback drives. In addition, they were asked about what they thought they had learned from the training.

2.3.2.1. Website

For the administration of the instruments in the pre-test and post-test period (questionnaire and diary) a website was used. This website was only accessible for participants. This allowed for a day-to-day overview of those who responded. The participants that did not respond were encouraged by e-mails and phone calls to do so.

2.3.3. *Timetable and data collection*

	December 2003	January 2004	February 2004
	Pre-test One month before training day	*Training day*	*Post-test* One month after training day
Experimental group	Questionnaire Diary	*Pre-test feedback drive* - On-road observation form - Driving assessment *Track exercises* *Group discussion*	*Post-test feedback drive* - On-road observation form - Driving assessment - Satisfaction questionnaire
Control group	Questionnaire Diary	*Feed back drive* - On-road observation form - Driving assessment	*Feed back drive* - On-road observation form - Driving assessment - Satisfaction questionnaire

Table 2.4*. Timetable.*

2.3.4. *Estimated power of the design*

The original design was to conduct the study with a sample of 300 young, inexperienced drivers. A distinction was made in the research design (*Table 2.5*) between drivers who attended a regular driving education and drivers who attended a 'best practice' education to obtain their drivers licence (RIS = Rijopleiding in Stappen = Stepwise Driver Training). The programme of such a 'best practice' education is very structured and a candidate can only pass from one module to another if all the training objectives of the former module are fully met. As a didactic principle candidates first have to learn so called handling scripts (this is the traffic situation, I want to do this (e.g. turn to the left) so I must first do this (e.g. look in the mirror) and then do that). What is also different compared with the regular driver training is that a four hour track training is included. This is not a short skid course. The intention is to let the candidate feel how easy it is to lose control and that is better to avoid certain situations than to rely on your skills.

	Experimental group (training course)	Control (no course)
Regular driving education	100	50
'Best practice' education	100	50

Table 2.5*. Intended research design.*

Due to many respondents dropping out before and during the project, only 127 respondents finished the last part of the project. Before conducting any analysis, a power estimation was conducted to assess if there was a reasonable chance of finding any effects with these numbers of respondents. The question whether the remaining group of respondents was

still representative, and not affected by a selection bias, has already been addressed in *§ 2.1.1*.

	Group		Total
	Experimental	Control	
Regular driving education	60	15	75
'Best practice' education	39	12	51
Unknown	0	1	1
Total	99	28	127

Table 2.6. *Actual research design.*

The question is how this modification affects the power of the experiment. In short, the power of a statistical test is the chance of finding a significant difference, if one is there. The power is dependent on three factors (Stevens, 1996):

1. the significance level (α) set by the experimenter;
2. sample size (n);
3. effect size (d): how much of a difference the treatments make, or the extent to which the groups differ in the population on the dependent variable(s).

For the estimation of the power in this experiment, the assumption was made that there is a moderate effect of the training on the dependent variables. The significance level (α) was set to .05. The sample size that was used for the power estimation was the smallest group comparison needed (the number of respondents with a 'best practice' education). This results in the following power estimations for a two-sided T-test, F-test, and Chi-square tests (Cohen, 1988).

	T-test (α =.05; d=.50)	F-test (α =.05; f=.25; df=1)	Chi-square tests (α =.05; W=.30; df=1)
Original design (n=150; n=75)	.70	.86	.95
Actual design (n= 50; n=25)	.41	.42	.56
Actual design without 'best practice' vs. regular (n=127; n=63)	.79	.80	.92

Table 2.7. *Power estimations.*

As can be seen in the table, the modification of the design does result in a decrease in power. Stevens (1996) argues that a study with a power of .70 or .80 is a good investment of money and resources. Therefore a comparison between the drivers with a 'best practice' and a regular education does not seem feasible. Without such a comparison, the power estimations are quite promising.

A problem may also arise when other subdivisions are made in the groups. When for example the gender of the drivers is inserted as a factor, the groups are divided in half. Not only the loss of subjects in this experiment has decreased the power of the study; also the introduction of extra (sub) groups (e.g. gender) in the design may lead to loss of power. In other words, we have too few subjects in each (sub) group, so the chance of finding a statistically significant difference is greatly reduced, even if such a difference actually exists.

	T-test (α =.05; d=.50)	F-test (α =.05; f=.25; df=1)	Chi-square tests (α =.05; W=.30; df=1)
Original design (n=150; n=75)	.70	.86	.95
Actual design (n= 50; n=25)	.41	.42	.56
Actual design without 'best practice' vs. regular (n=127; n=63)	.79	.80	.92

Table 2.8. *Power estimations.*

As can be seen in the table, the modification of the design does result in a decrease in power. Stevens (1996) argues that a study with a power of .70 or .80 is a good investment of money and resources. Therefore a comparison between the drivers with a 'best practice' and a regular education does not seem feasible. Without such a comparison, the power estimations are quite promising.

A problem may also arise when other subdivisions are made in the groups. When for example the gender of the drivers is inserted as a factor, the groups are divided in half. Not only the loss of subjects in this experiment has decreased the power of the study; also the introduction of extra (sub) groups (e.g. gender) in the design may lead to loss of power. In other words, we have too few subjects in each (sub) group, so the chance of finding a statistically significant difference is greatly reduced, even if such a difference actually exists.

3. The process evaluation: implementing the training

The process evaluation was carried out to be able to document how successfully the blueprint was implemented and what lessons could be learned from the experiences of relevant actors.
On behalf of the process evaluation three discussion meetings were arranged:

- a discussion with the driving instructors and the examiners who were involved in the feedback drives (feedback drive instructors);
- a discussion with the track trainers;
- a discussion with the researchers.

In addition to these three meetings, any other relevant information on the process was also collected. This primarily concerns subjects discussed in the meetings of the 'second phase driver training' workgroup.

The following three sections (*§ 3.1*, *§ 3.2* and *§ 3.3*) contain the results of these discussions.

3.1. Results of the discussion with the feedback drive instructors

3.1.1. *General observations concerning the training programme*

The feedback drive instructors were very enthusiastic about the Dutch second phase training initiative. According to their experience, immediately after the driving exam, errors creep into the driving behaviour of young novice drivers. In the eyes of the feedback drive instructors, second phase training can be an effective way of correcting these errors in their driving style. They had the impression that the participants were open to critical remarks on their driving style and that they were willing to improve their driving behaviour. It is for this reason that the instructors have doubts about the representativeness of the group that took part in the experiment. In their view, the group of young novice drivers that really are a problem for road safety probably didn't participate. In their opinion, this means that the second phase training programme will have to be mandatory in order to be effective.

3.1.2. *Strong and weak points of the feedback drives*

Strong points:

- Most participants showed their normal driving behaviour. The feedback drive is not seen as a driving test and participants are not afraid to make mistakes.
- Participants were open to critical remarks and the instructors have the impression that participants are willing to make use of the advice they receive.
- The participants were very sensitive to the remarks of fellow participants. Their comments have, in general, more impact than those of the feedback drive instructor.
- Working with driver profiles (based on a questionnaire) was a good basis for discussing the strong and weak points in one's driving style.

– The overall driving performance of participants was good - to - reasonable.
– Women achieved better results in the feedback drives than men. This is different from the situation at the driving test. This may be due to the fact that the feedback drives measure different aspects of driving style than in the normal driving test.

Points that can be improved:
– Both participants and feedback drive instructors were not always fully informed about the goals and content of the project. Thus, participants that had been allocated to the control group didn't know they would not get road safety training on the track. This resulted in dissatisfied responses from a lot of participants. However, at the end of the feedback drive, these participants had positive opinions about the usefulness and attractiveness of the feedback drive.
– The time for discussion after the feedback drive is too short to be able to discuss all experiences. This is partly due to the fact that so much paperwork has to be done (filling out profile scores and on-road observation forms).

3.1.3. *Circumstances during the feedback drives*

When performing the feedback drives during the pre-test, the instructors were confronted with severe winter weather conditions: intensive snowfall and snow-covered roads. For many participants this was their first experience with such extreme weather conditions. According to the feedback drive instructors, this led the participants to drive extremely carefully by driving very slowly and by keeping larger safety margins than usual. According to the instructors, this influenced the way the driving behaviour of the participants was assessed. Because weather conditions were quite normal during the feedback drives during the post-test, it is more likely that participants displayed their normal driving behaviour. The differences in weather conditions between pre-test en post-test situation imply that it is difficult to compare the results of the feedback drives between pre- and post test and between experimental and control group.

3.1.4. *Planning and organization of the feedback drives*

Feedback drive instructors are rather critical about the way the feedback drives were planned. Instructors found that they were not always informed in time about the timetable and modifications in the schedule were not transmitted in time. Instructors found it very frustrating when participants didn't show up. One of the instructors went so far as phoning participants the day before they had their feedback drives to check if they were informed about their appointments and if they could keep them. This worked very well and all these participants eventually appeared.

3.2. **Results of the discussion with the track trainers**

3.2.1. *General observations concerning the training programme*

The trainers of the training sites in Lelystad (ANWB) and Rijssen (VVCR) were also very enthusiastic about the initiative to set up and evaluate the experiment with the second phase training programme. However, during the

discussion it appears that in the case of the ANWB trainers there was some uncertainty about the basis assumptions of the second phase training programme, especially when it comes to the contents and the working method of the track training.

For the ANWB trainers, working with this young age group is relatively new. The trainers of the VVCR already have a long tradition in working with young novice drivers. The VVCR was also involved in the EU-project ADVANCED, and, in the framework of the *Young Drivers Project*, the VVCR has already been carrying out training programmes for young novice drivers for some years. For this reason they already are more familiar with the basic principles of the second phase training programme.

During the meeting with the track trainers the discussion focussed partly on the benefits of skill-oriented track training. What emerged was that particularly the ANWB trainers seem to have a rock-solid faith in the usefulness of their skill-oriented training. Especially when training conditions are difficult (which was the case in the Dutch experiment due to the wintry conditions) and trainers do not have much or any experience in performing the training programme and in working with novice drivers (which applies to the ANWB trainers), trainers tend to fall back on their normal working methods. This perhaps also explains why ANWB trainers have difficulty in accepting the new principles of the second phase training programme, although initially they supported these assumptions and agreed upon the structure of the training programme. Considering the initial enthusiasm for the second phase programme, and the fact that the ADVANCED project philosophy of track training was never questioned during the working group meetings, the project management would only have been to predict this situation occurring in practice by conducting a full rehearsal.

What we can learn from the Dutch experience is:

- An introduction to the training programme of three hours is not enough to teach the trainers the skills and motivation necessary for an effective execution of the course, especially if trainers do not have experience in working with the target group of young novice drivers.
- It is necessary to have a more profound discussion about the assumptions of the training programme. Having trainers state that they support the training programme is not enough.
- Trainers with little or no experience in working with young novice drivers need time to build up experience with the training programme in practice.

3.2.2. *Track training*

The track trainers in general agreed upon the structure of the track training and upon the exercises that are part of the track training. In the case of the ANWB trainers, this seems to contradict the fact that they questioned the basic assumptions of the second phase training programme. Perhaps this has to do with the fact that, in essence, they support the exercises (because they are also part of their own skill-oriented training programme), but they do not yet understand the different methods involved through which exercises support the principles of the second phase.

A general comment from the trainers was that, in some cases, they would have liked to have more time available, so that participants could experience

the effects of the manoeuvres more frequently, e.g. the influence of speed on the braking distance.

Due to the poor weather conditions most of the training sessions in Lelystad could not be performed according to the blueprint. Because of the exposure of participants to snow and rain, it was not possible to have a discussion with the whole group after each exercise on the track. In Rijssen, participants could stand in a bad weather shelter, so discussions were carried out according to plan.

In the case of the ANWB trainers, there was some misunderstanding concerning the degree to which they could give *instruction*. This led to the conclusion that in Lelystad there had been insufficient communication about the objectives of the exercises.

Another point of concern is that participants need to have the possibility to experience the exercises outside the car, when standing on the side of the track. In the braking exercises, for instance, the impact of speed is sometimes felt more outside than inside the car. In Rijssen the training group was always split in two: one group driving and one group observing. Due to the bad weather conditions in Lelystad, this was not the case.

3.2.3. *Group discussion*

The group discussions in general went well. The video sketches are a good means to initiate the discussion. As the group discussions took place at the end of the training day, some training groups were already really tired. The track trainers said that in these cases it was difficult to fill up the time available for discussion and therefore the group discussions were sometimes concluded earlier. We have to take into account that some track trainers didn't have much or any experience in carrying out group discussions with young people. It is likely that a more experienced and better trained course leader would have no problem motivating the group, despite it being the end of an already long day and the participants were a little tired.

Some trainers hadn't entirely understood the procedure for the group discussion. They thought it was important to show and discuss each of the video sketches. In those cases, the trainers had to break off the conversation and ultimately there was little to no interaction between participants.

3.2.4. *Planning and organization of the training days*

The training days sometimes progressed in a rather chaotic manner. Because of the bad weather conditions, participants arrived too late or not at all. And participants who had agreed to use their own cars were ultimately afraid to use them. In these cases, the training centres had to provide one. According to the trainers, the participants were badly informed about the contents of the training programme. A lot of participants expected they would get a skill-oriented track training. This led to problems, especially in the case of Lelystad. In Lelystad, the regular training programme of the ANWB was taking place next to the NovEV training. Quite a few participants were disappointed that they could not have the regular (more spectacular?) ANWB track training. The ANWB trainers said it was difficult to keep these

participants interested in the second phase training programme. In Rijssen, 'regular' training also took place simultaneously, but this was not mentioned as a problem.

In the opinion of the trainers, most of the organizational problems could have been prevented if there had been someone co-ordinating all training activities on each training day.

It would also have been helpful if all participants had done the training in hired cars. A ‘hired car’ or ‘rented car’ is a car that is provided by the project organisation. In the blueprint for the Dutch second phase training programme an important principle was that participants should do the training in their own car (or the car they are using most, in most cases being the car of one of the parents). This applies to the feedback drive as well as for the track training. Some of the participants couldn’t come to the training with their own car, so a rented car was arranged for them (a car from the training institute or a car from a driving instructor). This complicated planning and organisation quite a lot. In addition, due to the bad weather conditions participants that came to the training or the driving audit in their own car ultimately refused to drive in their own car or weren’t allowed by their parents (because it was their car). In these cases, considerable improvisation was necessary to provide a car for the participants. If we had worked with hired cars for everyone from the outset, none of these planning or organisational problems would have occurred.

3.3. **Results of the discussion with the researchers**

3.3.1. *Recruitment of the participants*

Looking back at the recruitment of the participants by telephone, it is possible that the participants had been given a too positive picture of the project and the training programme. In the first stage of recruitment, all young drivers were told that they could participate in a (spectacular) skid training day. In most cases it was not mentioned that the participants had to come back a month after the training for a second feedback drive and that they had to fill out several instruments (questionnaires and diaries). This may have caused a lot of participants to quit the project when they found out they were not able to do a skid training (but a safety training instead), they had to be present on two days and they had to fill out a questionnaire and keep a diary twice.

It probably would have been useful to ask participants to confirm their participation, and all the details of the project in writing, once they had committed themselves during the initial telephone recruitment.

When it comes to incentives for participation, a fee of 25 euros and a raffle with the chance of winning one of two travel vouchers or free car insurance for a year, the conclusion must be that these incentives were not attractive enough for young people. Perhaps a more personal approach, in which young drivers are recruited by their driving instructor or by their examiner, would have been more effective.

3.3.2. *Website, transfer of information, and data collection*

The website had two main goals: transfer of information to participants and collection of data by Internet. As far as data collection is concerned, the website proved to be of great value. Most young people have access to the Internet and filling out questionnaires and diaries using the website was very efficient.

Transfer of information through the Internet was less successful. Sometimes relevant information was available too late (for instance, the description of the route to the training centres). The forum function of the website (encouraging discussion groups online) also didn't work as planned. There was not enough time to provide new information on a regular basis and to stimulate the participants to discuss issues with each other on the forum site.

4. Results: questionnaire

The effect of the training was measured with several instruments. The results from each of these instruments will be discussed in the following chapter. This chapter shows the results from the questionnaire which the participants filled out before and after the training. And the Feedback form in which the participants expressed their satisfaction with the training. In *Chapter 5*, the Driving assessment form will be discussed. This is a form, which was filled in by the instructors (privately) after the feedback drives before and after the training. Finally in *Chapter 6* we will discuss the direct calibration scores that were given by the instructors, but also by the participants themselves, about their performance in the feedback drive.

The questionnaire (see *Appendix 2*) was filled out by the participants approximately one month before and one month after the training. The participants were invited by email to visit the website and fill in the questionnaire. The participants who did not have access to the internet received the questionnaire by mail.

This questionnaire focused on self-assessment of driving skills and risk, and safety awareness. The questionnaire consisted of four parts:

1. general questions (age, gender, how often do you drive?, etc.);
2. items focusing on self-assessment: weak and strong skills;
3. items focusing on risk and safety awareness: the degree of difficulty and complexity of the driving and traffic task;
4. judgement of traffic situations.

4.1. Satisfaction questionnaire

After the final part of the project, namely the second feedback drive (one month after the training day), the young drivers all filled out a questionnaire in which they could indicate how satisfied they were with the training and the project. The most important purpose of this questionnaire was to find out what the young drivers (thought they had) learned during the training.

One of the questions was how 'Fun' and 'Useful' the participants thought the different parts of the project had been. *Table 4.1* shows the percentages of respondent who 'highly agreed' and 'agreed' with the statements that the different parts were 'Fun' and 'Useful'.

Table 4.1 shows that the respondents appreciated both feedback drives (before and after the training) the most. The young drivers were the least satisfied with the Group Discussion.

		Males (n=82)	Females (n=58)	Overall
Pre training feedback drive	Fun?	94	92	93
	Useful?	85	86	85
Training on track	Fun?	92	92	92
	Useful?	85	87	86
Group Discussion	Fun?	55	65	59
	Useful?	67	65	66
Post training feedback drive	Fun?	90	93	91
	Useful?	91	97	93

Table 4.1*. Percentage of respondents 'highly agree' or 'agree'.*

Because the training took place at different locations and therefore with different instructors, *Table 4.2* shows the percentages (highly) agree with 'Fun' and 'Useful' at the different locations.

		Experimental		Control (n=34)	Overall
		Lelystad (n=34)	Rijssen (n=62)		
Pre training feedback drive	Fun?	94	94	90	93
	Useful?	88	84	83	85
Training on track	Fun?	82	98		92
	Useful?	79	90		86
Group Discussion	Fun?	47	67		59
	Useful?	44	79		66
Post training feedback drive	Fun?	94	86	97	91
	Useful?	94	94	91	93

Table 4.2*. Percentage of respondents 'highly agree' or 'agree'.*

A remarkable result from *Table 4.2* is that the control group seemed to appreciate both feedback drives almost the same as the experimental group. Usually, this is a problem for an experimental-control group design. The control group is often far less motivated, because they received a comprised version of what was promised. In this study, there does not seem to be a problem with the motivation of the control group.

Table 4.2 shows that, overall, participants in Rijssen were more content with the training and discussion. There was not much difference in the assessment of both feedback drives. Participants in Lelystad found both feedback drives slightly more fun and useful, but this could be because they were not so content with the training on the track and the group discussion, therefore appreciating the feedback drives even more. For both locations, the group discussion was seen as the least attractive module in terms of 'fun' and 'usefulness'. However, when the two locations are analysed separately it

becomes clear that there are differences on all modules, and that the modules of the training day (group discussion and track training) are rated higher in Rijssen than in Lelystad. This indicates that the Rijssen group was more content than the Lelystad group. This phenomenon is the strongest in the group discussion. Here we saw a 35-percentage point difference on the 'usefulness' of the group discussion. It would appear that, according to the participants, in Rijssen the group discussion succeeded in getting a 'message across', in contrast to the group discussion in Lelystad.

It was also important for participants to receive the right message. In order to verify if this was the case, the questionnaire contained several statements. *Figure 4.1* shows the percentages of '(highly) agree' and 'disagree' with several statements. The percentages in the figure represent the answers of both control and experimental groups, because there was no difference whatsoever between those groups in what they thought they had learned from the training. Promisingly, but not surprisingly, most respondents disagreed with the statements that they should look for more challenges and should drive in a more sporty manner. What is also important is that they do not think they drive better than the average motorist. Less than 10% of the participants agreed with this statement. It would have been a very undesirable side effect of the training, if the young drivers thought that they improved that much. However, only 40% of the participants believe they perform less well than the average driver. This must mean that more than half of the participants believe they are at least as good a driver as the average driver, with a lot more experience.

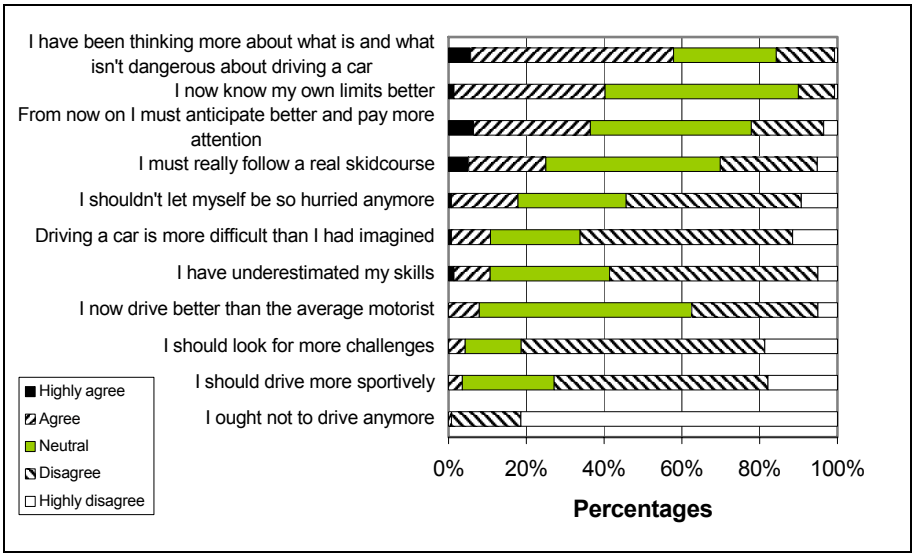

Figure 4.1. *Percentages agree - disagree on what was learned during training.*

4.2. Self-assessment and risk awareness

Figure 4.2 shows the participants' opinions on several statements concerning risk awareness. A distinction has been made between risks caused by the young, novice driver himself and risks caused by other road users. The higher the percentage in *Figure 4.2*, the less young drivers are concerned with that item, indicating low risk awareness. So, on the left are

subjects that concern the young drivers very much, and on the right are subjects that the young drivers are not concerned so much about. An alarming result is the fact that the young drivers do not seem to be concerned about driving too fast. A least 60% of the respondents are not concerned about driving too fast.

For some of the risk factors, the respondents had to indicate the perceived risk of the same situation twice. Once the situation where the young drivers were the cause (internal risks) and once when other road users caused the risk (external risks). There seems to be a slight difference in estimation of internal and external risks. For example, young drivers are more concerned when other drivers take risks in traffic than when they take risks themselves. A Principal Component Analysis was conducted to investigate if the internal and external caused risks were two different constructs. The results indicated this was not the case. Young novice drivers generally assess risks as less or more dangerous, irrespective of the cause of that risk.

The percentages indicating risk awareness are plotted for several groups, experimental-control; before and after the training. A third group is plotted in the figure, the 'no intervention' group (before and after training). This group consists of respondents who did not show up for the first feedback drive or training day, but did fill in their questionnaire twice (pre- and post-'training'). This is a special group because there is information in the before and after situations, but they did not have any intervention whatsoever. This means that any development within this group can be attributed to the passing of time.

The hypothesis is that the experimental - after group, which should have been affected by the training, scores a lot lower on most of the risk awareness items. However, there does not seem to be a pattern of differences between groups in the figure, for risks caused by the driver himself or by other road users.

A Repeated Measures Analysis was conducted on the questionnaire items concerning risk awareness. This analysis compared the scores before and after the training (within components) of the experimental group with the control group (between components). No effect of the training was found.

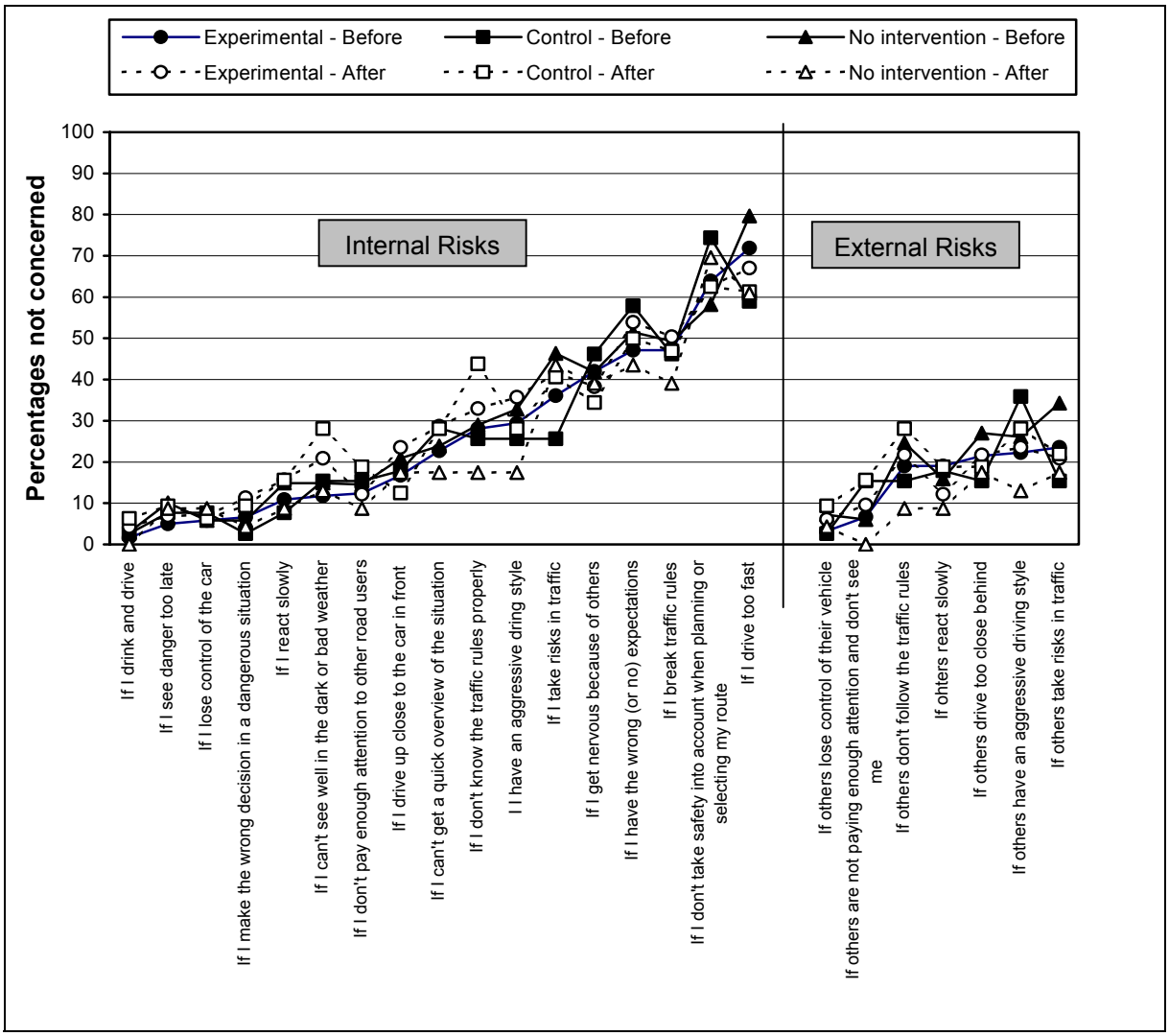

Figure 4.2. *Risk Awareness (grouped into internal and external risk factors) – percentage not concerned about issue.*

Figure 4.3 shows the scores on the self-assessment of driving skill items. A distinction was made between skills involving 'safe driving' (e.g. obeying traffic rules) and 'demanding driving' (e.g. reacting accurate to emergency situations).The higher the percentage, the more participants think they are (very) strong in that aspect, indicating a possible problem of overestimation of skills. So, on the left are driving skills where the participants think they are not so strong; on the right, are aspects of driving where the young drivers think they are (very) strong.

There seem to be some 'outliers' in the figure, especially in the 'No intervention - After' group. This is a result (gimmick) of the way the data was presented, and does not actually indicate extreme answers. The items were ordered on the basis of the percentages of the 'Experimental - Before' group, because this was the largest group. The outliers merely indicate that those groups have a different ordering of aspects they think they are (very) strong in. They do not, overall, have a bigger problem with overestimation of skill.

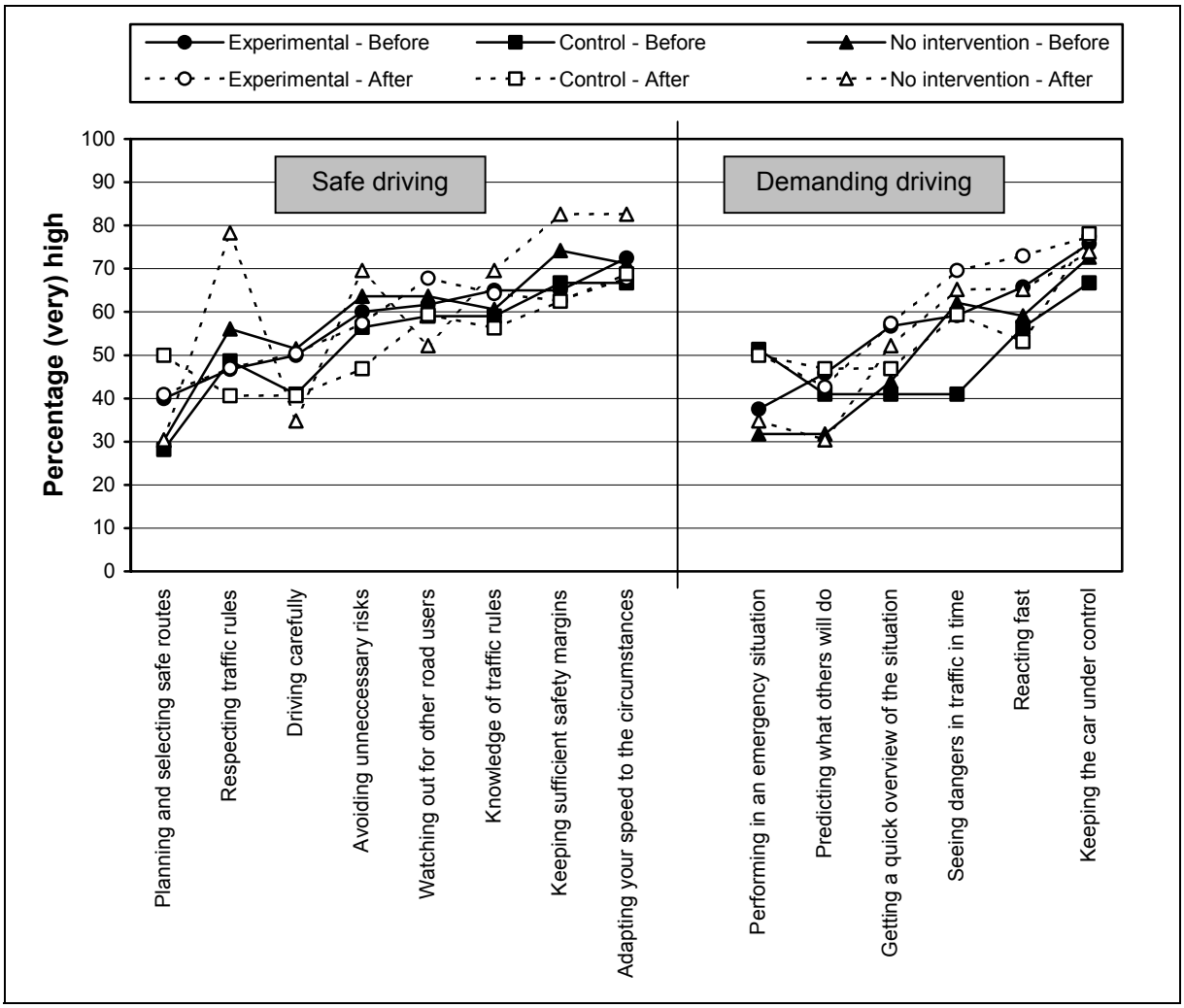

Figure 4.3. *Self-assessment of driving skill (grouped into 'safe driving' skills and 'demanding driving' skills) – Percentage (very) strong*

Figure 4.3 shows that young drivers are, overall, rather confident about their driving skills. At least 30% of the participants believe they are (very) strong in all skills, and in some skills even more than 60% believe they are (very) strong. There is no difference in skills concerning 'safe driving' and risks concerning 'demanding driving'. A Principal Component Analysis also indicated that young novice drivers generally do not make a difference between 'safe driving' and 'demanding driving' when assessing their skills.

A Repeated Measures Analysis was conducted on the questionnaire items concerning self-assessment of driving skill. This analysis compared the scores before and after the training (within components) of the experimental group with the control group (between components). No effect of the training was found.

4.2.1. *Gender differences*

Further investigation was conducted to see if there were effects of the training that were not so obvious. For example the Repeated Measures Analysis was conducted for Females and Males separately. A Multivariate

Variance Analysis (MANOVA) indicated that there is a significant main effect of gender on risk awareness and self-assessment (*Table 4.3* and *Table 4.4* show the Pillai's Trace statistics). Males score worse on risk awareness, in the sense that they see less risks. Males also score worse on self-assessment, in the sense that they are more confident than females.

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Intercept	Pillai's Trace	.999	12985.660[a]	24.000	367.00	.000	.999
Gender	Pillai's Trace	.115	1.995[a]	24.000	367.00	.004	.115
a) Exact statistic							

Table 4.3. *MANOVA – risk awareness; design: intercept & gender.*

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Intercept	Pillai's Trace	.984	1641.133[a]	14.000	380.00	.000	.984
Gender	Pillai's Trace	.320	12.754[a]	14.000	380.00	.000	.320
a) Exact statistic							

Table 4.4*. MANOVA – self-assessment of skill; design: intercept & gender.*

Within these groups of females and males, no effects of the training were found. That is, there was no effect of the training found for either the males or females. In the power estimation (*§ 2.3.4*), it was concluded that there would not be enough participants to include an extra factor (gender) and still have a reasonable chance at significant results. It is, however, unlikely that this was the reason for not having any significant results in this analysis, because the effect size was also quite small.

The fact that an effect of gender on the several statements was found, indicates that the questionnaire does measure something. It is possible that the questionnaire measures more stable personality traits, such as those that can be expected to differ between males and females, (in other words, not something that can be changed with a one-day training course).

4.2.2. *Driver education*

The analyses were also repeated for the participants with a regular driver education and for participants with a 'best practice' driver education. No effects of the training were found within these groups. In *§ 2.3.4 (Estimated power of the design)* the conclusion was reached that the chance of finding significant effects of the training with the distinction between regular and 'best practice' education would be very slim. So this result was not very surprising.

However, the power should not be a problem for finding the main effects of driver education. But there was no main effect of the driver education found.

The participants with a regular driver education filled in the questionnaire in the same way as the participants with a 'best practice' education.

4.3. Calibration

Calibration is defined as the balance between self-assessment of skill and risk awareness. Those who see little or no danger in traffic, but are at the same time highly confident of their own skills form a risk group in terms of calibration. In order to analyse if the training had any effect on calibration, as was measured by the questionnaire, the group of participants was split in half, twice. First the group was split in half on the basis of their assessment of their driving skill. Participants who had a lower score than the median formed one group; they perceive themselves with the weakest skills. Participants with scores above the median formed another group; they perceive themselves with the strongest skills.
Secondly, the group was split in half, in the same manner, on the basis of their perception of risks in traffic. This resulted in *Table 4.5*, for the experimental group.

The risk group, in terms of calibration, is highlighted. These are the participants who express the most confidence about their own driving skill and, at the same time, see little danger in traffic. The expectation for the training is that the number of participants in this group is smaller after the training. But there is another group which indicates bad calibration (also highlighted). This is the group of drivers who are very insecure of their own driving skill and, at the same time, see much danger in traffic. This is not a risk group, in itself, but it is important that this group does not get bigger as a result of the training. It is not the intention of the training to create young drivers who are too insecure of their driving skills and who see too much danger in traffic.

Unfortunately, as can be seen in *Table 4.5*, there was no shift in the experimental group from 'bad' calibration to 'good' calibration. There is even some increase in the number of participants who see no danger in traffic and assess themselves to be very skilful. This is however not enough for a significant effect of the training. Also in the control group (*Table 4.6*) there was no distinct shift in calibration groups.

Pre training		Self-assessment of skill				Post training		Self-assessment of skill		
		Weak	Strong	Total				Weak	Strong	Total
Risk awareness	No danger	28	17	45		Risk awareness	No danger	26	20	46
	Much danger	20	32	52			Much danger	21	30	51
	Total	48	49	97			Total	47	50	97

Table 4.5. *Calibration development – experimental group.*

Pre training		Self-assessment of skill			Post training		Self-assessment of skill		
		Weak	Strong	Total			Weak	Strong	Total
Risk awareness	No danger	7	2	9	Risk awareness	No danger	10	4	14
	Much danger	10	9	19		Much danger	6	8	14
	Total	17	11	28		Total	16	12	28

Table 4.6. *Calibration development – control group.*

Table 4.7 shows the effect of the training on calibration for the individual participant. The rows in the table denote the 'calibration-group' of the participant before the training and the columns denote the 'calibration-group' after the training. This means for example that of the 60 participants who were in the 'good-calibration' groups before the training, 40 remained in that group, 12 ended up in the 'insecure-group' and 8 in the 'risk-group'.

Table 4.7 indicates that there is some change between groups before and after the training, but there is no specific pattern. Some participants improve, whereas others end up in a worse group than before the training. The same effect can be seen in the control group.

Before training	After training							
	Experimental group				Control group			
	Good	Insecure	Risk	Total	Good	Insecure	Risk	Total
Good	40	12	8	60	12	1	3	16
Insecure	11	7	2	20	5	5	0	10
Risk	5	2	10	17	1	0	1	2
Total	56	21	20	97	18	6	4	28

Table 4.7. *Shifts in calibration.*

4.4. Situation questions

Besides the questions concerning 'Risk awareness' and 'Self-Assessment of skills', the respondents were asked to judge traffic situations on photos. The respondents were asked to estimate how fast they would drive through the situation displayed on the photo (see *Appendix 3* for the two situations that were used). Both situations were shown in pairs, randomly scattered throughout the questionnaire. In one of the photos, the situation was slightly more complicated (because of the presence of a cyclist). The objective of these questions was to measure if youngsters take the difficulty of the situation into account when judging how to handle the situation. In this case, it was not important that they responded with low speeds, as such, but it was important that they adapted their speed to the complexity of the situation.

These situations were tested in a pilot study (n=10), by means of the website. During a discussion afterwards, it turned out that none of the young drivers in the pilot study noticed the small differences in the photos. There

were even some complaints that the website did not work correctly because the same photos seemed to be shown several times. However, it turned out that the young drivers subconsciously did alter their speed, dependent on the complexity of the situation.

Figure 4.4 shows the percentage of 'good' responses for the different groups concerning the first situation. A 'good' response means that the person reported a lower speed when the situation was more complex. A 'bad' answer was considered to be the case when there was no difference in speed between the two situations, or when the speed was even higher in the more difficult situation. The figure shows the responses of (1) the Experimental group, which attended a feedback drive, training on a track, and a group discussion; (2) the control group, which only attended the feedback drive; but also (3) an extra group – no intervention – these are participants who did not show up for the first day, but did complete the questionnaire before the second feedback drive. The 'no intervention' group did not get 'anything' between the before-training and after-training moments.

As can be seen in the figure, there were no effects found in the training on the judgement of the situation.

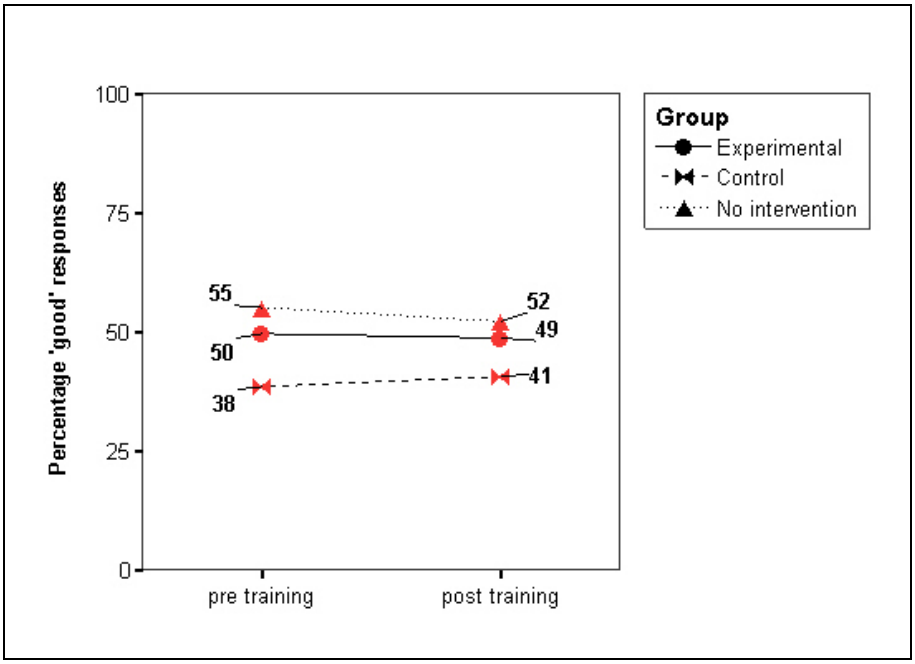

Figure 4.4. *Percentage of 'good' responses in situation 1.*

Appendix 3 also shows a second situation that was displayed on photo, after which the participants were asked to estimate the speed they would drive. *Figure 4.5* shows the percentage of good responses by the participants from the experimental and control groups, before and after the training. A 'good' answer means that the respondent reported a lower speed when the situation was more complex.

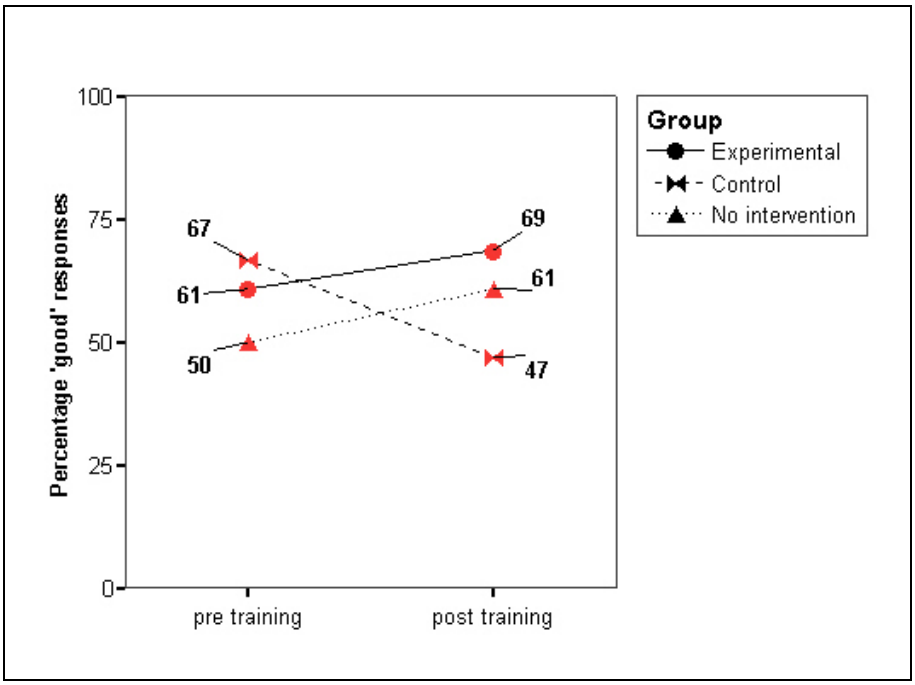

Figure 4.5. *Percentage of 'good' responses in situation 2.*

Figure 4.5 shows a (slight) improvement in the experimental group after the training. The control group diminishes substantially at the same time, which can not be explained so easily. Another puzzle is the improvement in the 'no intervention' group, which is about the same as in the experimental group.

Table 4.8 shows the exact number of participants with a 'right' or 'wrong' response to the situations for experimental versus control group, before and after the training.

		Experiment	Control	No intervention	Total
Pre training	Right response	73	26	33	132
	Wrong response	47	13	33	93
	Total	120	39	66	225
Post training	Right response	79	15	14	108
	Wrong response	36	17	9	62
	Total	115	32	23	170

Table 4.8. *Frequencies of right and wrong responses to situation 2.*

Using Log linear analysis several models were tested for significance to find the factors which could explain the data (*Table 4.9*). The factors involved were [A] right or wrong response, [B] before or after the training, and [C] Experimental, control or no intervention group. A model without the training-group interaction did not fit the data (significant deviation). A slightly more complicated model, including the training interaction, did fit the data sufficiently. In order to test if the interaction factor was the cause of the model fitting the data, the difference between the chi-square and degrees of

freedom between the models was calculated. This resulted in a significant contribution of the training – group interaction in the fit of the model.

Model		Chi-square	Df	Significance
Without training interaction	[AB][AC]	18.456	4	< 0.0001
With training interaction	[AB][AC][BC]	4.828	2	n.s.
Training significance	[BC]	13.628	2	< 0.01

Table 4.9. *Log linear analysis situation 2.*

Of course, these analyses still do not explain why the improvement that was found in the experimental group also took place in the group that did not receive any form of intervention, nor even a feedback drive. Nor does it explain why the control group, which should have remained at the same level, showed such deterioration. Both effects could have been caused by the small number of participants in the control and no intervention group. The log linear analysis was repeated for the control and no intervention scores added up (because both groups represented 'no training'). There was still a significant effect of the training which can be attributed to the training.

4.5. Conclusions questionnaire

The results from the questionnaire are somewhat unclear: some effects of the training were found, but they were not consistent, and were not always in the expected direction.

An alarming result from the questions concerning risk awareness is the fact that the young drivers do not seem particularly concerned in general, and especially not about driving too fast. A least 60% of the respondents do not believe that it is dangerous to drive too fast. On the other hand, it turned out that young drivers are, overall, rather confident about their driving skills. At least 30% of the participants believe they are (very) strong in all skills, and in some skills more than 60% believe they are (very) strong. Unfortunately, these opinions did not improve after the training. The fact that there were significant gender differences in these issues, led to the conclusion that this part of the questionnaire probably measured more stable attitudes or personality traits (which could not be changed with a one-day training course or within the period of a month).

There were some significant effects in the situation judgements, but they were somewhat contradictory. There was only an effect in one out of two situations. However, the control group, which should theoretically remain at the same level, scored lower on the post-test. Moreover, there is no explanation why the group without any intervention improved in the same way the experimental group did.

5. Results: driving assessment

For every on-road feedback drive, an assessment form was completed by the instructor (*Appendix 4*). The participants were judged on twenty skills (items), according to the scale: insufficient (1 point), sufficient (2 points), or good (3 points). The twenty skills can be divided into three groups, Vehicle control, Driving skills, and Calibration skills. Three new variables were formed by adding up all the scores in each group. So, for example, for vehicle control, the total score of a participant could range from 3 (insufficient on all three items) to 9 (good on all three items). The internal consistency, Cronbach's alpha, is depicted after each scale.

Vehicle control (Range: 3 - 9; Cronbach's alpha: 0.67)

1. Preparation for drive / end of drive
2. Vehicle handling
3. Vehicle control

Driving skills (Range: 8 - 24; Cronbach's alpha 0.71)

13. Driving on straight and bendy roads
14. Behaviour at junctions
15. Behaviour when turning
16. Entering and exiting traffic
17. Overtaking and passing
18. Being overtaken
19. Changing lanes and lateral positioning
20. Driving on different surfaces

Calibration skills (Range: 6 - 18; Cronbach's alpha: 0.84)

7. Defensive behaviour: anticipation
8. Defensive behaviour: effective observation skills
9. Defensive behaviour: safety cushion and safety margins
10. Risk awareness, danger recognition, and traffic insight
11. Adapted and decisive driving: speed
12. Adapted and decisive driving: decisive handling

The Cronbach's alpha's show there is substantial internal consistency *within* the three constructs, with the exception of the Cronbach's alpha for 'Vehicle control', which is somewhat small. Analysis showed there also is a significant correlation (ranging from .55 to .65) *between* these three aspects of the driving assessment ($p<.000$). Participants, who score high on for example driving skill, also score high on vehicle control.

Because the training course was aimed at improving calibration, the largest effect was expected in the calibration score. Note that this calibration (measured with the feedback drive) is not exactly the same as calibration measured by the questionnaire; *§ 4.3*).
Repeated Measures Analysis on each skill was used to test this hypothesis.

5.1. Vehicle control

The total score on vehicle control for the test and control group, before and after the training, is shown in *Figure 5.1*. The vertical axis depicts the mean

value of the total score on vehicle control. The range of this total score was from 3 (insufficient on all three items) to 9 (good on all three items). Because the instructors rarely used the 'insufficient' category, the range in *Figure 5.1* has been modified from 6 (sufficient on all three items) to 9 (good on all three items). The horizontal axis depicts the value on the first and second feedback drive.

Two extra groups are displayed. The 'No intervention group' indicates the participants who could not come to the first feedback drive, but were still invited to attend the second. This group did not receive any form of training or even feedback. The 'drop out group' consists of respondents who dropped out of the project after the first feedback drive, so we do have information on their performance during the first feedback drive. There is only information on their driving skill at one moment. The whiskers shown in *Figure 5.1* display the 95% confident interval.

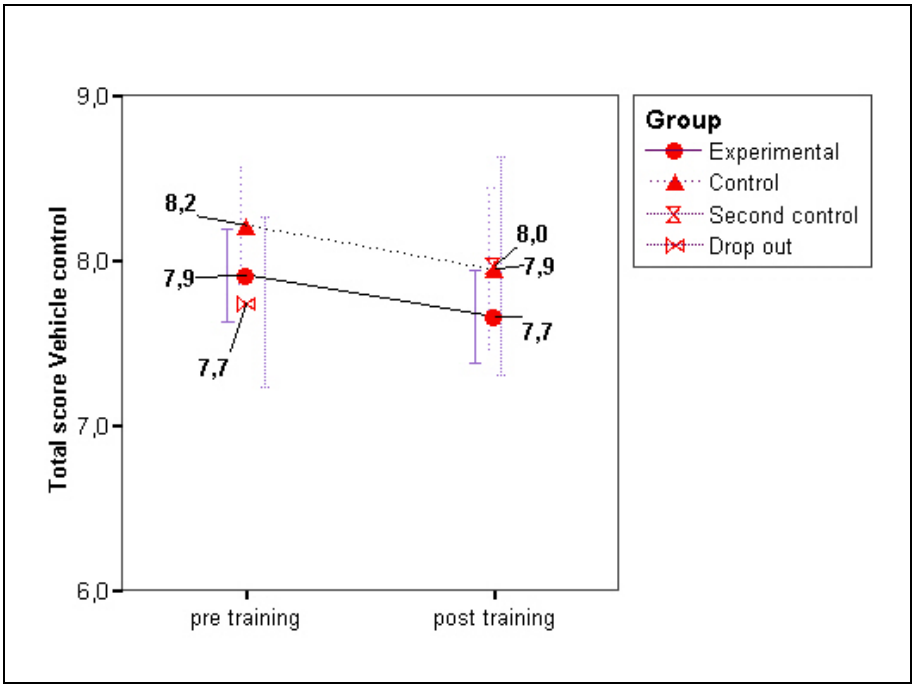

Figure 5.1. *Vehicle control (total score from driving assessment).*

As can be seen, both groups (control and experimental) perform slightly worse on driving skill after the second feedback training. What is remarkable is that there seems to be a consistent difference between experimental and control group. Although respondents were assigned randomly to the conditions, the control group performed (slightly) better on the first drive.

A Repeated Measures Analysis was conducted to discover if the differences between before and after training (and differences between control and experimental group) were significant. Although all participants of either one or two feedback drives are shown, the analysis was conducted only on the young drivers participating on both days. The respondents who did not show up for the first (no intervention group) or second feedback drive (drop out) were deleted from the analysis. No significant effects were found. Also, the effect sizes of these factors were negligible.

There were some indications that the training that took place in Lelystad differed somewhat from the training in Rijssen, for example, in appreciation by the respondents. Therefore, the results were also analysed for Lelystad and Rijssen as a factor (see *Figure 5.2*). The Repeated Measures Analysis revealed no significant effects.

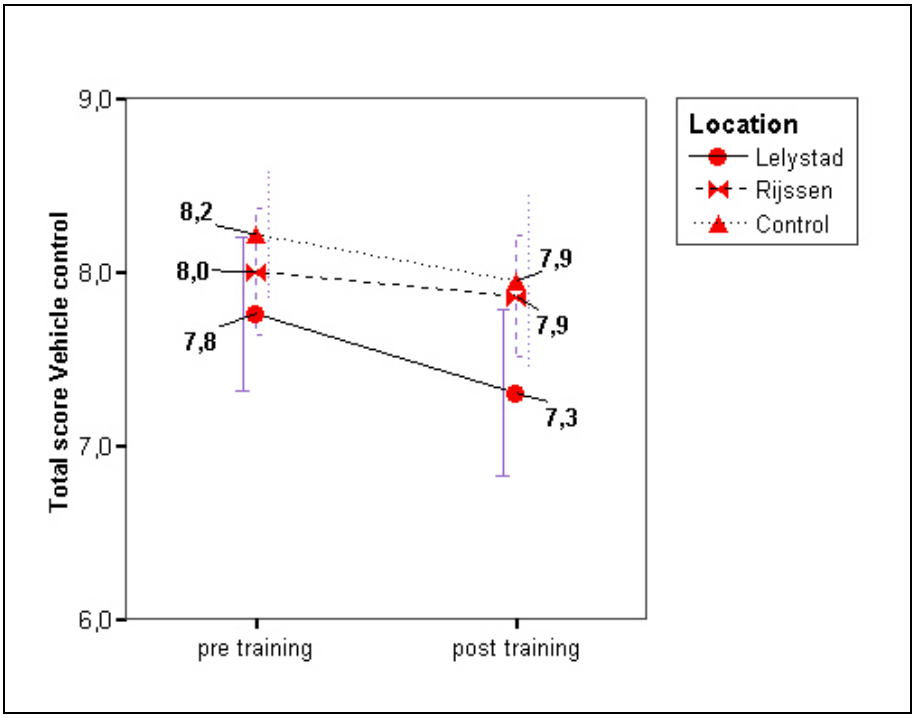

Figure 5.2. *Vehicle control (Lelystad and Rijssen separate).*

These results (no significant effects of the training) were just as expected. Because the training was not designed to affect vehicle control performance, it is not surprising that no effect was found.

5.2. Driving skill

Figure 5.3 shows the total score on driving skill, before and after the training The vertical axis depicts the mean value of the total score on driving skill. The range of the total score was from 8 (insufficient on all eight items) to 24 (good on all eight items). Because the instructors rarely used the 'insufficient' category, the range has been modified from 16 (sufficient on all eight items) to 24 (good on all eight items). The horizontal axis depicts the value on the first and second feedback drive.

No effect of the training is expected, because the training was not aimed at improving driving skill. Of course, it is not a negative result if there are some effects on driving skills, as long as they are in the desired direction.

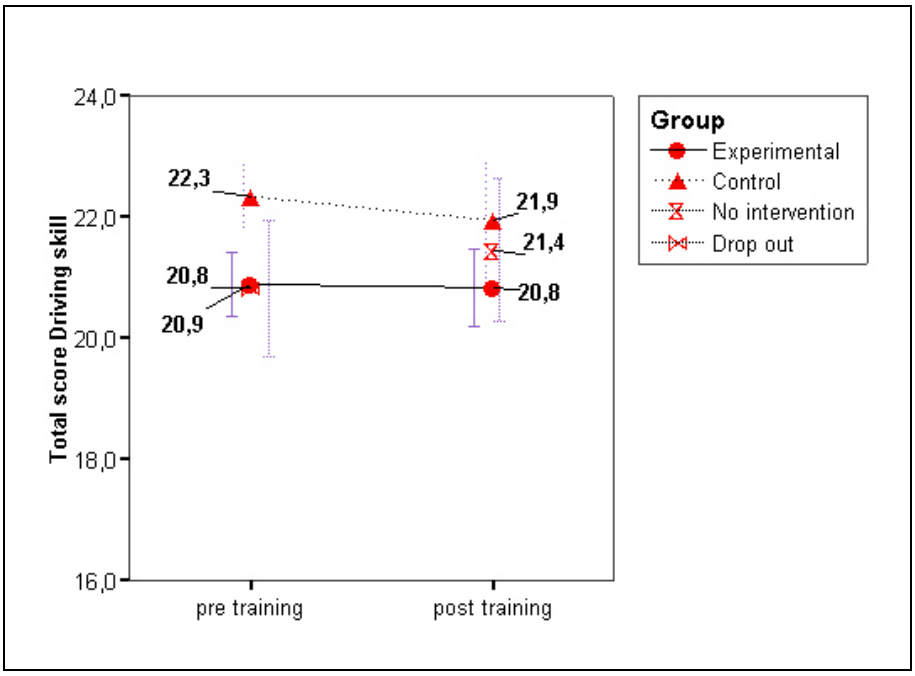

Figure 5.3. *Driving skill (total score from driving assessment).*

Again these results were analysed with Repeated Measures Analysis, and a significant effect of 'Group' was found between the subjects in the analysis (*Table 5.1*). In other words, there was a significant difference in performance (before training) between experimental and control group. This is remarkable, because respondents were assigned randomly to the conditions, so there should not be a difference between control group and experimental group during the first feedback drive. An explanation for this phenomenon could be the following: the instructors in the experimental conditions had to come to an unfamiliar location. They indicated themselves that it had been a complicating factor to carry out the feedback drives in unknown surroundings. The instructor who performed the feedback drives for the control group did this in a familiar location. There was also a difference between the locations themselves. The experimental locations were rather remote. The control locations, on the other hand, were located in the middle of a city.

This difference between the control group and experimental group has some consequences for the interpretation of the results. But this will be addressed later, because in driving skill there were no effects found as a result of the training (*Table 5.2*).

Source	Type III Sum of squares	df	Mean square	F	Sig.	Partial Eta^2
Intercept	1266.515	1	1266.515	10534.650	.000	.988
Group	1.122	1	1.122	9.335	.003	.070
Error	14.908	124	.120			

Table 5.1. *Tests of between-subjects effects - driving skill.*

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.000	.031[a]	1.000	124.000	.860	.000
Training * group	Pillai's Trace	.001	.151[a]	1.000	124.000	.699	.001
a) Exact statistic							

Table 5.2. *Multivariate tests - driving skill; design: intercept & group within subjects design: training.*

Figure 5.4 shows the experimental group split up in both locations, Lelystad and Rijssen.

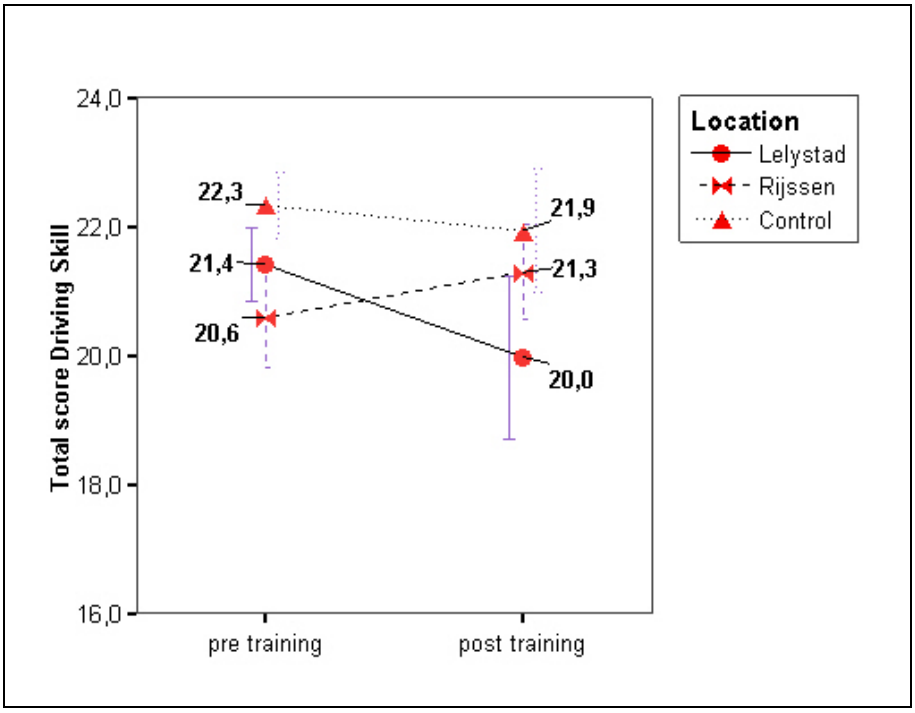

Figure 5.4. *Driving skill (Lelystad and Rijssen separate).*

This figure shows a clear difference between Rijssen and Lelystad. The performance of the young drivers who participated in the training in Lelystad shows roughly the same decline as the control group. On the other hand, the Rijssen participants improved their performance in a statistically significant manner. Repeated Measures Analysis indicated that the effect of the training on the different locations was significant (*Table 5.3*). The Partial Eta squared indicates the effect size of the training. An Eta squared of .055 can be seen as a moderate effect size (Stevens, 1996).

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.003	.385[a]	1.000	123.000	.536	.003
Training * location	Pillai's Trace	.055	3.604[a]	2.000	123.000	.030	.055
a) Exact statistic							

Table 5.3. *Multivariate tests – driving skill (Lelystad and Rijsen separate); design: intercept & location within subjects design: training.*

5.3. Calibration

Because the training was specifically aimed at improving calibration, the expectation is that the calibration score from the driving assessment was affected the most. *Figure 5.5* shows this total score on calibration skill, before and after the training. The vertical axis depicts the mean value of the total score on calibration skill. The range of this total score was from 6 (insufficient on all six items) to 18 (good on all six items). Because the instructors rarely used the 'insufficient' category, the range in *Figure 5.5* has been modified from 12 (sufficient on all six items) to 18 (good on all six items). The horizontal axis depicts the value on the first and second feedback drive.

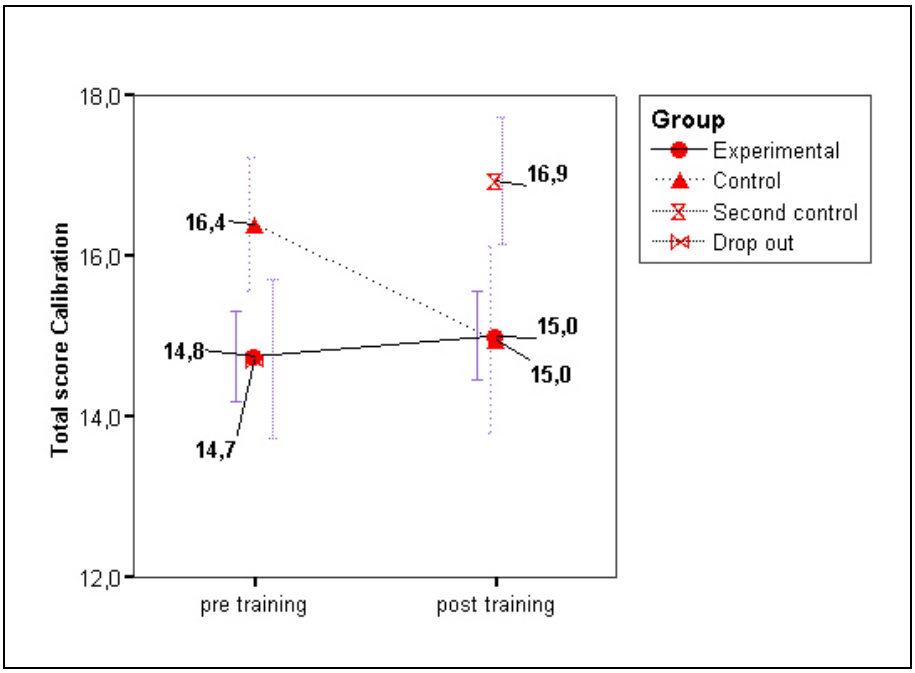

Figure 5.5. *Calibration score (total score from driving assessment).*

This figure shows an interaction effect between Group and Training, which turned out to be significant (*Table 5.4*). However, the experimental group shows almost no improvement. The interaction is actually caused by the deterioration in the control group, after they had been assessed remarkably highly during the first feedback drive. We already saw the same significant difference between the experimental and control group in driving skill. This

leads to the tentative conclusion that the control group and experimental group were assessed differently during the first feedback drive. The interaction effect in calibration score can be explained fully by this phenomenon. The interaction effect can be reduced to a different assessment during the first feedback drive. At the time of the second feedback drive, the difference was removed and the groups performed identically.

Also noticeable in *Figure 5.5* is the outstanding performance of the 'No intervention' group. These are participants who could not come to the first feedback drive, but were still invited for the second. In the post-training situation they drove for the very first time. It's inexplicable how it is possible that this group without any training, or even a feedback drive, performed so well. A cautious guess is that these young drivers were more impressed and therefore more motivated than the rest of the group. Either that, or they were judged more leniently by the instructors, because it was their first time.

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.018	2.234[a]	1.000	124.000	.138	.018
Training * group	Pillai's Trace	.034	4.309[a]	1.000	124.000	.040	.034
a) Exact statistic							

Table 5.4. *Multivariate tests – calibration sore; design: intercept & group within subjects design: training.*

Figure 5.6 shows the performance of Lelystad and Rijssen.

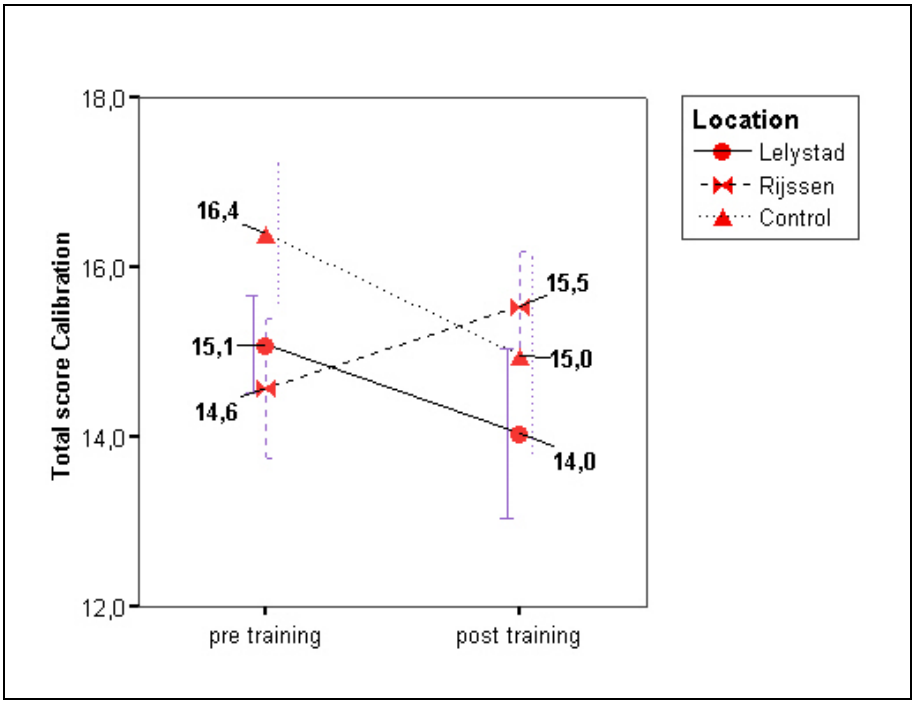

Figure 5.6. *Calibration score (Lelystad and Rijssen separate).*

Again there is a difference in performance between Rijssen and Lelystad. The young drivers who went to Lelystad perform roughly the same as the control group who did not receive a training and group discussion. The participants from Rijssen show a totally different development: they improve from the first to the second feedback drive, in a statistically significant manner. What is important in this comparison is the fact that there was no difference between the performance of Lelystad and Rijssen during the first feedback drive. The effect of the training (*Table 5.5* indicates that this effect is significant) can be attributed completely to the training.

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.017	2.187[a]	1.000	123.000	.142	.017
Training * location	Pillai's Trace	.083	5.559[a]	2.000	123.000	.005	.083
a) Exact statistic							

Table 5.5*. Multivariate tests – calibration score (Lelystad and Rijssen separate); design: intercept & location within subjects design: training.*

5.4. Conclusions driving assessment

The results on driving assessment show that there is a significant effect of the training on driving and calibration skill. There is a difference between experimental and control group. Because there is a significant difference between these groups before the training, this result is difficult to interpret.

However, there is also a significant difference between the performance in both locations, Lelystad and Rijssen. Because the performance in Lelystad and Rijssen was the same before the training, there are no problems with interpreting this result. The participants from Rijssen perform significantly better than the participants form Lelystad, as a result of the training.

In all measures (vehicle control, driving skill, and calibration skill) there seems to be a difference between experimental and control group (significant for driving skill and calibration skill), before training. This could be a problem, for example in *Figure 5.5*, where the effect of the training could be explained by the differences between control group and experimental group during the first driving assessment. The respondents were assigned to the conditions randomly, so the groups should be the same. Another explanation for this effect could be that the young drivers in the experimental group were assessed by different instructors than the drivers in the control group, or more likely, the different type of location of the experimental and control group. Because of the organization of the training day, the experimental feedback drive was in the vicinity of the training site (so not in a familiar environment for participants), which was a rather secluded area, while the control group feedback drive was in the vicinity of the participant's hometown and in a more urban area. When comparing the performance of Lelystad and Rijssen, these analytical problems are no longer an issue. The performance of the participants in Lelystad or Rijssen are exactly the same during the pre-training drive, and only differ on the post-training drive.

An effect that is apparent from all measures is that only the respondents who received their training in Rijssen perform better in the second feedback drive than in the first (with the exception of vehicle control were no group improves their performance). The respondents in the control group and from Lelystad perform even worse than before. How is this possible? The respondents in the control group should, in theory, remain at the same level. One of the explanations could be the fact that during the first feedback drive there was extremely bad weather, snowstorms, slippery roads, etc. This group of young drivers had never experienced these conditions. It is possible that they were so in awe of the extreme conditions that they all accepted very large safety margins, therefore scoring better on the feedback drive with regard to safe driving. During the second feedback drive the circumstances were normal and it is possible that they were assessed worse. But, if this is the case, all scores during the first feedback drive should have been lower. Unfortunately we can only speculate, and will never know what the scores would have been if the weather had been normal.

6. Results: on-road observation form

The on-road observation form consisted of three parts and was to be completed by the instructor as well as the participant after the feedback drive. During the first feedback drive, another participant who was in the car also filled in the observation form. The opinion of this extra passenger has, however, not been analysed because there was no passenger present during the second feedback drive.

The first three items on the observation form (part 1), concerning assessment of driving skill, were to be rated on a scale from 'weak' (1) to 'strong' (5). The fourth item (part 2), concerning perception of the driving task, was to be rated on a scale from 'difficult' (1) to '(too) easy' (5). The final item (part 3) concerned the so-called 'driver profile'. The instructor or the participant could choose one out of five profiles that felt most appropriate. These profiles included:

1. very insecure driver who sees danger everywhere and allows for too large safety margins;
2. calm driver, very aware of the dangers of driving;
3. smooth driver, but sees the dangers of driving and the limits of his own abilities;
4. sharp driver who leaves himself limited room for manoeuvre, but has an eye for dangerous situations;
5. overconfident driver who leaves insufficient (safety) margins.

As previously mentioned, the training took place at two locations: Lelystad and Rijssen. In the previous section it has been shown that the two locations of the experimental group did not behave in the same fashion. Therefore, in this section, all analyses are performed comparing 'locations' (Lelystad, Rijssen, and Control) rather than comparing 'groups' (Experimental and Control).

In the following section, the results of several analyses are presented per item on the form. Furthermore, the results are divided into three subsections. The first subsection describes the scores provided by the instructors and compares the scores given before and after training. The second subsection describes the scores given by the participants and also compares before and after training scores. The third subsection compares the scores provided by the instructors to scores provided by the participants and considers the relationship between these scores before and after training.

6.1. Scores provided by the instructors

To analyse the effect of training on the scores provided by the instructors (for participants in each location) Repeated Measures Analyses were performed. The results for each item on the on-road observation form are presented and discussed below.

6.1.1. *Vehicle control and general skills*

The results, which can be found in *Table 6.1*, show a significant difference in scores before and after training over the different locations. Also, an Eta squared of .094 could be considered to be moderate to large (Stevens, 1996), which is an indication of the effect size of training.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.022	2.698[a]	1.000	118.000	.103	.022
Training * location	Pillai's Trace	.094	6.126[a]	2.000	118.000	.003	.094
a) Exact statistic							

Table 6.1. *Multivariate tests - vehicle control and general skills; design: intercept & location within subjects design: training.*

Figure 6.1 shows the scores on vehicle control and general skills, before and after the training. The mean score is depicted on the vertical axis, where (1) indicates 'weak' skills and (5) indicates 'strong' skills. The whiskers shown in *Figure 6.1* illustrate the 95% confidence intervals.

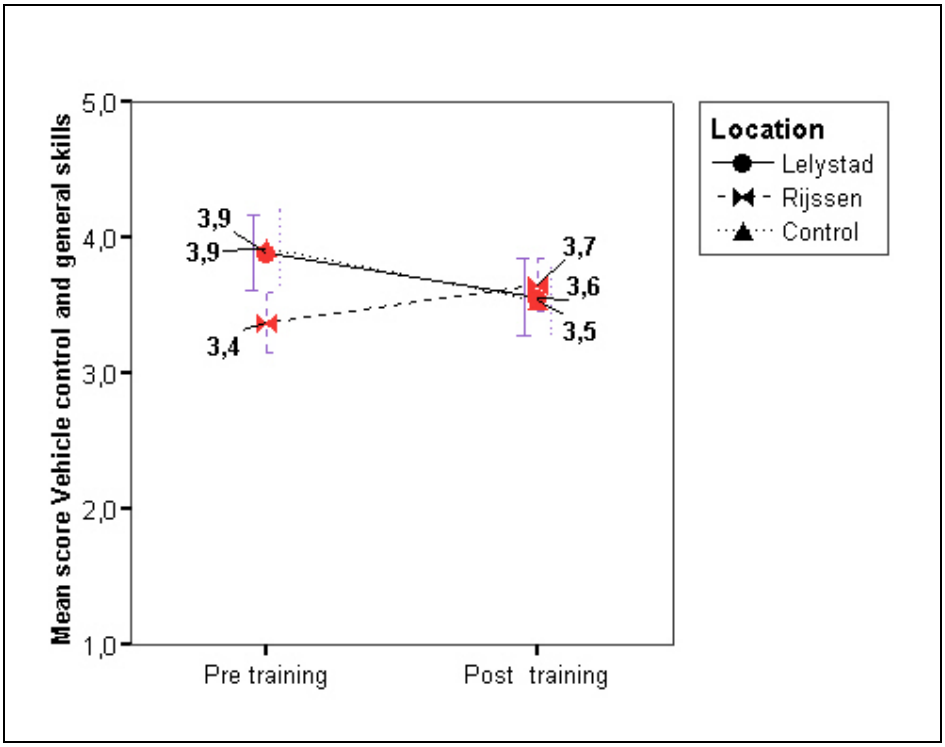

Figure 6.1. *Vehicle control and general skills.*

Figure 6.1 displays a difference in particular between Rijssen versus Lelystad and the control group. It shows that participants in Rijssen received a lower score from the instructor before training than participants in Lelystad and the control group. This is actually contrary to expectation, as all groups are expected to start at a comparable level. Analyses in the previous

sections already discovered a difference between the experimental groups and control group. However, in those analyses it was the control group that scored differently on the first feedback drive. In the current analysis it is the Rijssen location that shows a deviation.

It also shows that, according to the instructor, only participants in Rijssen were stronger in controlling the vehicle after training than before. Participants in Lelystad and the control group were considered to be weaker in controlling the vehicle after training than before.

6.1.2. *Safe and defensive driving*

Figure 6.2 shows the scores on safe and defensive driving, before and after the training. The mean score is depicted on the vertical axis, where (1) indicates 'weak' skills and (5) indicates 'strong' skills.

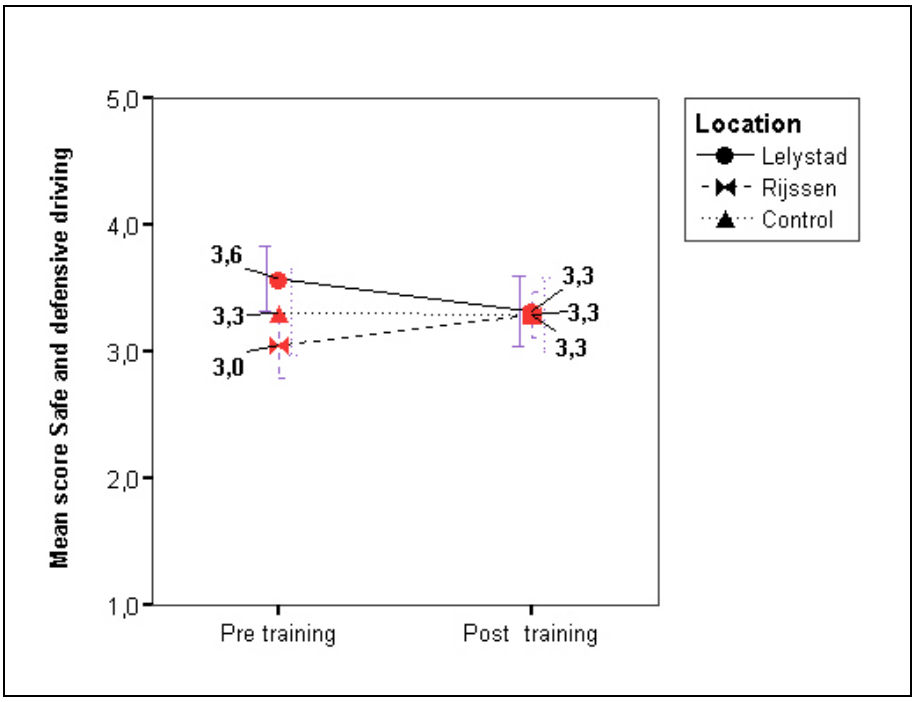

Figure 6.2. *Safe and defensive driving.*

The results for the 'safe and defensive driving' item, show no significant differences in scores before and after training, either dependent or independent of location. As also found for the first item on the on-road observation form, *Figure 6.2* shows that, according to the instructors, the participants in Rijssen were weaker before training in safe and defensive driving than participants in Lelystad or the control group. Another similarity with the first item is that only participants in Rijssen received higher scores after training than before training.

6.1.3. *Anticipation*

The results, for the 'anticipation' item, which can be found in *Table 6.2*, show a significant difference in scores before and after training over the different

locations. Eta squared amounted to .208 and could be considered to be rather large, which is an indication of the effect size of training.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.009	1.133[a]	1.000	119.000	.289	.009
Training * location	Pillai's Trace	.208	15.653[a]	2.000	119.000	.000	.208
a) Exact statistic							

Table 6.2. *Multivariate tests – anticipation; design: intercept & location within subjects design: training.*

Figure 6.3 shows the scores on anticipation, before and after the training. The mean score is depicted on the vertical axis, where (1) indicates 'weak' skills and (5) indicates 'strong' skills.

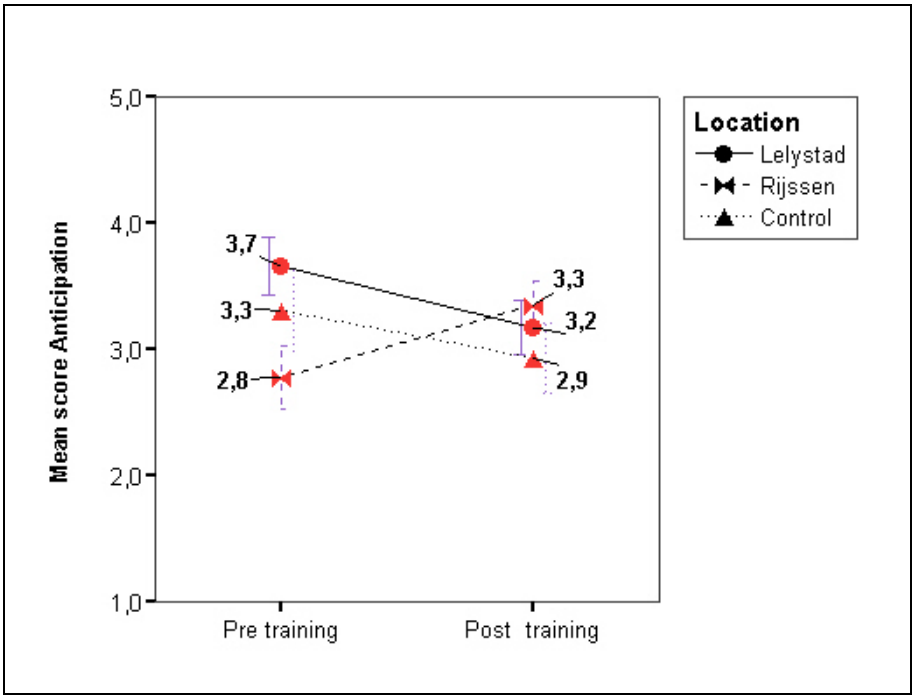

Figure 6.3. *Anticipation.*

Similar to the results for the previous two items on the on-road observation form, *Figure 6.3* shows that, according to the instructors, the participants in Rijssen were weaker before training in anticipation than participants in Lelystad or the control group. Also, an opposite effect is found for the participants in Rijssen versus the participants in Lelystad and the control group. Participants in Rijssen are considered by the instructor to be stronger in anticipation after training than before training. Participants in Lelystad and the control group are given lower scores for anticipation after training than before training.

6.1.4. *Estimation of the complexity of the driving task*

The results, which can be found in *Table 6.3*, show a significant difference in scores before and after training independent of the location. Eta squared amounted to .066 and could be considered to be moderate, which is an indication of the effect size of training.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Training	Pillai's Trace	.066	8.095[a]	1.000	114.000	.005	.066
Training * location	Pillai's Trace	.026	1.515[a]	2.000	114.000	.224	.026
a) Exact statistic							

Table 6.3. *Multivariate tests – estimation of the complexity of the driving task; design: intercept & location within subjects design: training.*

Figure 6.4 shows the scores on estimation of the complexity of the driving task, before and after the training. The mean score is depicted on the vertical axis, where (1) indicates that the participant perceives the driving task as difficult, and (5) indicates that the participant perceives the driving task as (too) easy. This means that a high score is unfavourable. (In the previous figures, a high score represented a favourable score).

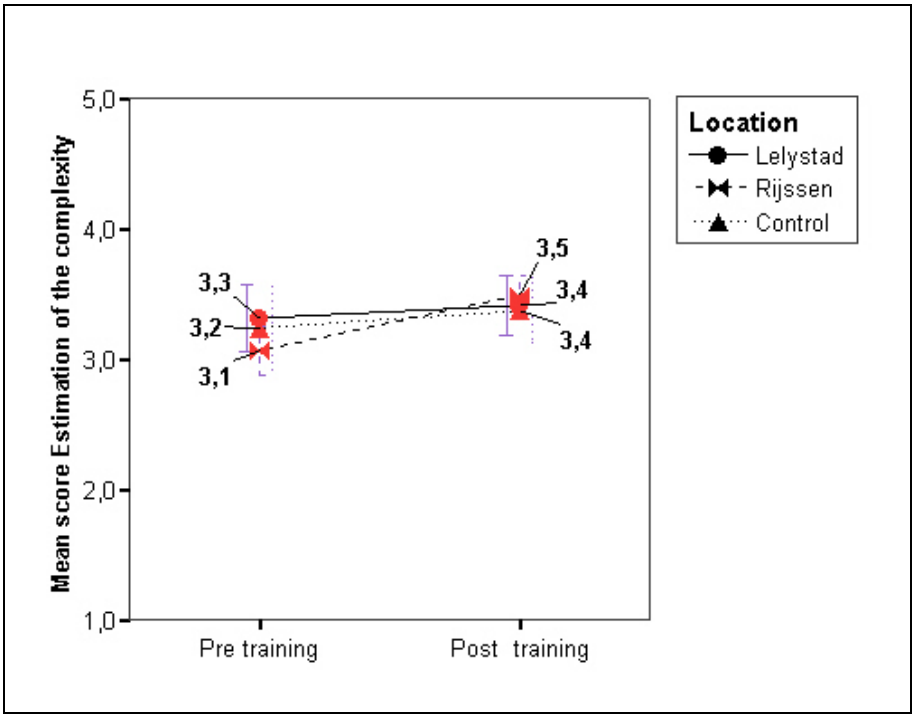

Figure 6.4. *Estimation of the complexity of the driving task.*

This item on the on-road observation form shows a somewhat different result compared to the first three items. *Figure 6.4* displays that the instructor rated the participants from all locations to estimate the complexity of the driving

task as harder before the training compared to after the training. Or, in other words, according to the instructor, during the second feedback drive, the participants underestimated the driving task more than during the first feedback drive,. Because this effect is found for all locations, this implies an effect of the feedback drive, and / or maybe even an effect of the instructors (interpreting this item differently the second time). Most probably this effect is caused by the extreme weather conditions during the first feedback drive. On some training days, a traffic warning was even issued not to go outside unless absolutely necessary. So it is unlikely, with such extreme conditions, that the participants underestimated the driving task.

6.1.5. *Profile scores*

The scale belonging to the item 'driver profile ' on the on-road observation form has a somewhat different scale than the previous four items. Namely, this item does not vary from 'weak' to 'strong' or from 'difficult' to 'easy'. In this case, one could assert that the scale varies from 'undesirable' to 'desirable' (the central profile) and back to 'undesirable' again. As this does not constitute an interval scale, it did not seem appropriate to conduct a Repeated Measures Analysis. Thus, frequencies and percentages were computed to detect differences (see *Table 6.4* and *Figure 6.5*).

The results show that independent of location or effect of training, the central 'optimal' profile is chosen most. The most salient effect of training seems to occur around profile 3 and 4, which is, however, largely due to the smaller number of participants in Lelystad (45 versus 23) and Rijssen (73 versus 47) after training.

	Lelystad				Rijssen				Control			
	Pre training		Post training		Pre training		Post training		Pre training		Post training	
Profile 1	0	0%	0	0%	2	3%	2	4%	1	4%	0	0%
Profile 2	6	13%	5	22%	16	22%	13	28%	5	21%	8	22%
Profile 3	24	53%	13	57%	29	40%	24	50%	11	46%	23	62%
Profile 4	12	27%	5	22%	23	32%	6	13%	5	21%	5	14%
Profile 5	3	7%	0	0%	3	4%	2	4%	2	8%	1	3%

Table 6.4*. Frequency and percentage for each profile before versus after training.*

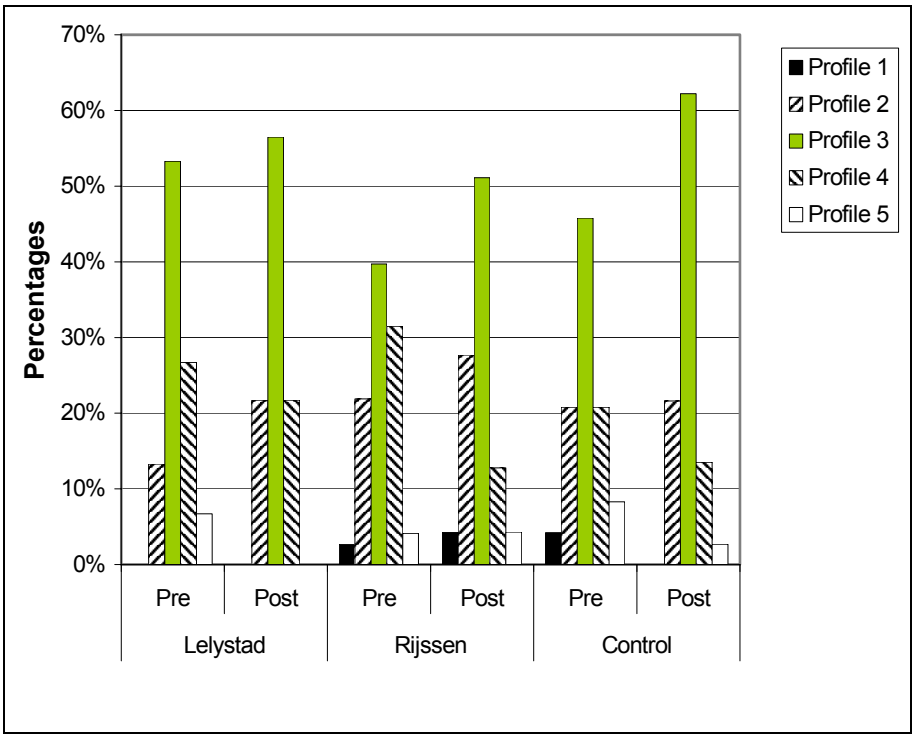

Figure 6.5. *Percentage each profile is chosen by an instructor.*

6.1.6. *Conclusions*

The results presented in the above section, seem to lead to the conclusion that the training did have an effect on the scores given by the instructors. The effect of training is mainly found in the 'Rijssen' location and less in the 'Lelystad' location. However, to account for this effect is rather difficult, as a strange phenomenon seems to have occurred in Rijssen before the training. For all items, Rijssen starts out with the lowest scores compared to the other two locations before training. As mentioned earlier, this is contrary to expectation, as all locations should behave about the same before a manipulation is administered. In previous sections, a difference between the two experimental locations and the control group was also noted, for which there actually was a satisfactory explanation. But at this moment in time, no evident explanation comes to mind that can account for the deviation of Rijssen, although it seems that this effect should not be attributed to differences between the participants in the Rijssen location. A more plausible explanation would be that the instructors in Rijssen interpreted the items on the form in a different way than instructors in the other locations or than after training.

6.2. **Scores awarded by the participants**

To analyse the effect of training on the scores awarded by the participants in each location, Repeated Measures Analyses were performed. As none of the results were significant, only figures are shown.

Initially it was intended that the on-road observation form would be filled in, not only by the instructor, but also by the participant, following each

feedback drive. However, due to miscommunication, not all participants completed the form. Especially the last item on the form, the driver profile, was seldom filled in by the participant. Therefore, differences between driver profiles provided by the participant were not analysed.

6.2.1. *Items 1 to 4*

The results, which can be found in *Figure 6.6*, show no significant differences in scores before and after training.

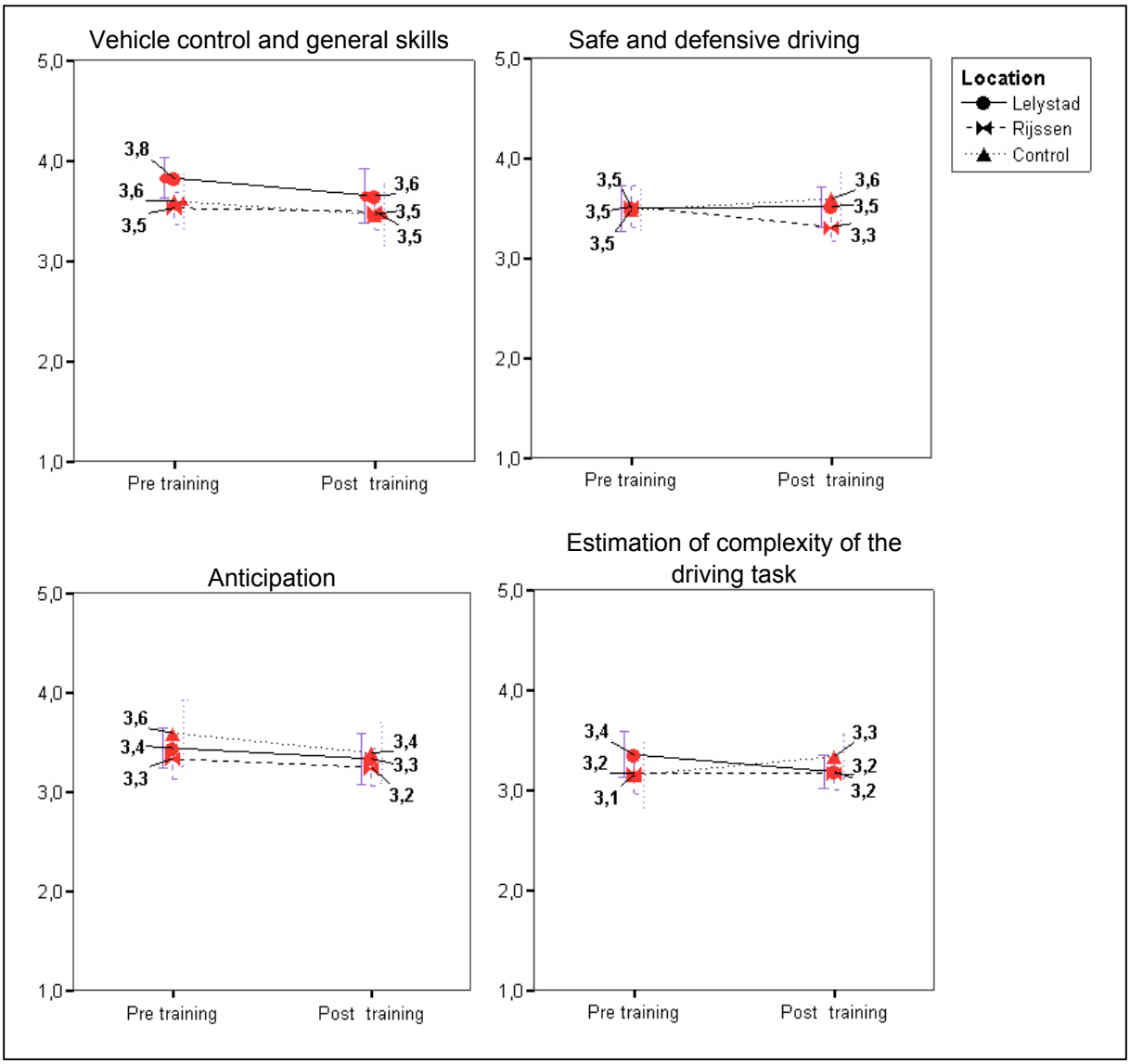

Figure 6.6. *Results for the participants - mean score*.

6.2.2. *Conclusions*

An important objective of the training was to improve calibration. This would be the case if the participants did not see any improvement in their skills, after the training, but perceived the complexity of the driving task as being more difficult.

Repeated Measures Analyses were performed to detect significant differences between scores provided by the participants before and after training. However, none of the items on the on-road observation form filled out by the participant showed significant differences between before and after training, although an effect of training was expected. It seems that , according to the young drivers themselves, the training did not have an effect on the assessment of skill or complexity of the driving task. This is, on one hand, positive because the participants did not think they improved their skills as a result of the training. On the other hand, there was no change in their perception of the complexity of the driving task.

6.3. **Comparing the instructors with the participants**

The first step in analysing the differences between the scores awarded by the instructors and scores given by the participants, was to calculate correlations between the scores. The Pearson correlation coefficients, although most of them significant, ranged between .13 and .43. Because of this weak but significant correlation, Repeated Measures Analyses were conducted to find more specific patterns in the comparison between instructors and participants. It is to be expected that before training, participants will overestimate or perhaps underestimate their skills more than after training. Therefore, two Repeated Measures Analyses were conducted per item. The first considering the scores awarded by both the instructor and the participant before training, and the second considering the scores given after training.

6.3.1. *Vehicle control and general skills*

Table 6.5 shows some results for both raters in all locations concerning the first item.

	Condition	Pre training			Post training		
		Mean	Standard Deviation	N	Mean	Standard Deviation	N
Instructor	Lelystad	3.80	.795	44	3.58	.830	33
	Rijssen	3.36	.901	70	3.67	.769	61
	Control	3.89	.667	36	3.48	.671	42
	Total	3.61	.850	150	3.59	.755	136
Participant	Lelystad	3.82	.540	44	3.67	.777	33
	Rijssen	3.60	.646	70	3.49	.698	61
	Control	3.61	.645	36	3.48	.707	42
	Total	3.67	.620	150	3.53	.719	136

Table 6.5*. Descriptive statistics - vehicle control and general skills.*

6.3.1.1. Pre training

The results show no significant differences between instructors and participants before training. This finding is actually not what is expected, but is less surprising than if scores had differed significantly in the unexpected direction, (if the participants had rated themselves to be less skilful than the instructors rated them). The results, which can be found in *Table 6.6*, also show an effect of location, which is again unexpected and even undesirable before training.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Rater	Pillai's Trace	.000	.003[a]	1.000	147,000	.958	.000
Rater * location	Pillai's Trace	.051	3.947[a]	2.000	147,000	.021	.051
a) Exact statistic							

Table 6.6. *Multivariate tests – vehicle control and general skills (pre training); design: intercept & location within subjects design: rater.*

6.3.1.2. Post training

The results again show no significant differences between instructors and participants. This is not very surprising as it would be expected that, after training, the difference between instructors and participants would not exist, or at least be less than before training. As in this case, there was no difference found before training, an effect of training in the expected direction should not be found.

6.3.2. *Safe and defensive driving*

6.3.2.1. Pre training

The results, contrary to the results for the first item, show a significant interaction effect between rater and location before training (*Table 6.7*). Eta squared for this effect could be called moderate (.048). This effect was not expected, as all participants should be rated about the same before training in each location.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Rater	Pillai's Trace	.024	3.712[a]	1.000	148.000	.056	.024
Rater * location	Pillai's Trace	.048	3.705[a]	2.000	148.000	.027	.048
a) Exact statistic							

Table 6.7. *Multivariate tests – safe and defensive driving (pre training); design: intercept & location within subjects design: rater.*

6.3.2.2. Post training

The results (*Table 6.8*) show a significant effect of rater independent of the location the participants were in. It would be expected that the difference in scores between instructor and participant would diminish after training. Therefore it seems hard to explain how the difference between raters can be significant after training, while it was not before training. *Table 6.9* shows that the difference in mean score between the raters is actually less after training compared to before training and is therefore not an unexpected result. However it should be noted that this effect cannot be attributed to the training, as this result is found independent of location. This result could therefore be interpreted as an effect of the feedback drive.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Rater	Pillai's Trace	.030	4.146[a]	1.000	132.000	.044	.030
Rater * location	Pillai's Trace	.012	.811[a]	2.000	132.000	.446	.012
a) Exact statistic							

Table 6.8*. Multivariate tests – safe and defensive driving (post training); design: intercept + location within subjects design: rater.*

Table 6.9 shows some descriptive statistics for both raters in all locations concerning safe and defensive driving.

	Condition	Pre training			Post training		
		Mean	Standard Deviation	N	Mean	Standard Deviation	N
Instructor	Lelystad	3.39	0.841	44	3.30	0.810	33
	Rijssen	3.07	0.976	71	3.28	0.710	61
	Control	3.33	0.793	36	3.29	0.716	41
	Total	3.23	0.903	151	3.29	0.732	135
Participant	Lelystad	3.48	0.628	44	3.52	0.566	33
	Rijssen	3.55	0.789	71	3.31	0.564	61
	Control	3.28	0.701	36	3.54	0.596	41
	Total	3.46	0.728	151	3.43	0.580	135

Table 6.9*. Descriptive statistics – safe and defensive driving.*

6.3.3. *Anticipation*

6.3.3.1. Pre training

The results, which can be found in *Table 6.10*, show a significant effect of rater independent of the location the participants were in. Eta squared equals .065 and could be considered moderate. It would be expected that before training, participants tend to score themselves as 'stronger' at anticipation compared to instructors. *Table 6.11* shows that this is indeed the case. Besides an effect of rater, an interaction effect between rater and

location is also found (*Table 6.10*), which seems to be an undesirable effect. However, *Table 6.11* shows that again, a strange phenomenon has occurred with the instructors in Rijssen before the training. Unfortunately, it is not yet clear what this phenomenon exactly is.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Rater	Pillai's Trace	.065	10.310[a]	1.000	148.000	.002	.065
Rater * location	Pillai's Trace	.044	3.435[a]	2.000	148.000	.035	.044
a) Exact statistic							

Table 6.10*. Multivariate tests – anticipation (pre training); design: intercept & location within subjects design: rater.*

	Condition	Pre training			Post training		
		Mean	Standard Deviation	N	Mean	Standard Deviation	N
Instructor	Lelystad	3.45	.791	44	3.15	.619	33
	Rijssen	2.89	1.008	71	3.34	.750	61
	Control	3.19	.786	36	3.00	.775	41
	Total	3.13	.926	151	3.19	.738	135
Participant	Lelystad	3.45	.589	44	3.33	.736	33
	Rijssen	3.42	.822	71	3.25	.745	61
	Control	3.53	.878	36	3.27	.807	41
	Total	3.46	.772	151	3.27	.757	135

Table 6.11*. Descriptive statistics – anticipation.*

Table 6.11 shows that before training, participants rate themselves stronger on anticipation than the instructors do, which is a result in accordance with expectation.

6.3.3.2. Post training

The results show that the significant difference between raters that was found before training is not found after training. Again, as the interaction between rater and location is not found significant, this effect can be attributed to the feedback drive.

6.3.4. *Estimation of complexity of the driving task*

6.3.4.1. Pre training

The results show no significant differences between instructors and participants before training. This finding is actually not what was expected, but less surprising than when scores would have differed significantly in the

unexpected direction. Also, the results show no effect of location, which is to be expected before training.

6.3.4.2. Post training

The results, which can be found in *Table 6.12*, show a significant difference between instructors and participants after training. Eta squared is .060 and can be considered moderate. However, this effect is not found in combination with location, and therefore this effect cannot be attributed to training, but perhaps to the feedback drive.

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta^2
Rater	Pillai's Trace	.060	8.094[a]	1.000	126.000	.005	.060
Rater * location	Pillai's Trace	.041	2.667[a]	2.000	126.000	.073	.041
a) Exact statistic							

Table 6.12*. Multivariate tests – estimation of complexity of the driving task (post training); design: intercept & location within subjects design: rater.*

Table 6.13 shows that, after training, participants consider the driving task to be easier than the instructor does, compared to before training, which is an unintended effect. Among other things, the training intended to make participants more aware of the complexity of the driving task. The results here, however, indicate that the opposite effect has been achieved.

	Condition	Pre training			Post training		
		Mean	Standard Deviation	N	Mean	Standard Deviation	N
Instructor	Lelystad	3.39	.722	44	3.41	.665	32
	Rijssen	3.13	.803	69	3.53	.570	57
	Control	3.24	.855	34	3.25	.630	40
	Total	3.23	.794	147	3.41	.620	129
Participant	Lelystad	3.34	.608	44	3.19	.471	32
	Rijssen	3.23	.789	69	3.18	.658	57
	Control	3.24	.781	34	3.25	.670	40
	Total	3.27	.734	147	3.20	.617	129

Table 6.13. *Descriptive statistics – estimation of complexity of the driving task.*

6.3.5. *Conclusions*

The results discussed above do not present strong evidence for an effect of the training, but should perhaps be interpreted as evidence for a positive effect of the feedback drive.

Overall, the hypothesis was that differences between instructors and participants would diminish as an effect of training. This is found in a few

instances, but the results in previous sub-sections show that this is mostly due to a change in the instructors instead of a change in the participants. Therefore, based on the results of the on-road observation form, it does not seem that the training had the intended effect on the participants.

Generally, from these results, it can be concluded that while the instructors did see some improvement as a result of the training, the participants did not.

7. Discussion and conclusions

7.1. The evaluation study: strengths and weaknesses

The evaluation was a true field study in the sense that not all conditions met the scientific criteria. Some 'complications' were apparent from the start (like the familiarity with the traffic situations during the feedback drive), others happened just 'by accident' (the trainers' opinions about the principles of the track exercises). Here, an overview is given of the strength and weaknesses of the study, as well as an analysis of the impact the 'weaknesses' have had on the reliability, generalizability, and the validity of the results.

7.1.1. *The selection of subjects/participants*

Any study that aims to document the effects of a voluntary course like this one, needs to be able to reject the claim that the participants in the study were a very small and select group: namely persons that are very interested in safety issues. The study was able to document non-response both at the start and during the study. There was a good overview of certain critical variables about the characteristics of the 68% that refused to participate in the study. A comparison between the refusers, dropouts, and participants on four variables (gender, age, driving experience and hours of training) did not show major differences between the groups, which led to the conclusion that, based on these variables, the participants could still be regarded as an a-select group. However, they did of course differ in the most important variable: namely to accept the invitation to sign up for the course and/or to finish it.

The participants were randomly assigned to the two research conditions (control and experimental group). Only in a few cases this was not possible, such as in the case of scheduling problems, or the (im)possibility of a participant to come to one of the training sites.

7.1.2. *The validity of the feedback drive*

In studies of this nature, the actual driving behaviour is seldom assessed. In this study the driver was assessed during his pre-test feedback drive and the post-test feedback drive. The value of these assessments can be questioned on three aspects:

1. As it was part of the course, the instructor influenced the candidate by commenting on his behaviour. To make the conditions between control and experimental group comparable, he did this for both the control and the experimental group. The consequence is that the 'control' group experienced some form of intervention as well.
2. Because of the organization of the training day, the pre-test feedback drive was in the vicinity of the training site (so in an unfamiliar environment for the participant) while the post-test feedback drive was in the vicinity of the participant's hometown. For the control group both feedback drives were in the vicinity of their hometown. Taking these circumstances into account, it is to be expected that in comparison to the control group, the experimental group would perform worse on the pre-test feedback drive. Moreover, as both groups took their post-test

feedback drive in a familiar driving environment, the experimental group should improve in comparison to the pre-test drive, and the control group should neither improve nor worsen.
3. The weather conditions during the drives differed to a large extent between participants and between pre- and post tests.

7.1.3. *The quality of the assessment*

When assessing the quality of driving (safety. suppleness. etc) the 'expert eye' is superior to any so-called objective measurement using instrumented cars or driving simulator performance. In this study, 'real life driving' has a high face validity in comparison to other methods. For this reason, the feedback drive and the expert assessment are strong features of this study. Despite this, there are three elements in this study that are a threat to the validity of the assessment:

1. On the pre-test feedback drive of the experimental group, the expert assessors were aware of the fact this was the pre-test drive of a person of the experimental group. Knowing that this group was in need of training, this knowledge may have influenced their judgment and expectation, leading to a rather negative assessment.
2. On the post-test feedback drive, the assessors were aware of the fact that these were post-test feedback rides. A strong point was that they did not know whether a participant belonged to the experimental or to the control group.
3. The pre- and post feedback drives were not assessed by the same instructor. It is unclear to what extent this may have led to differences between groups.

7.1.4. *The research design: the task of motivating the control group*

A problem in using a control group, is how to keep the control group interested in the study. A demotivated control group is not a good control group. In this study it was unavoidable to frustrate the control group, as all participants signed up for the 'challenging safety course'.
In order not to lose too many members from the control group at the outset of the study, we chose not to communicate to the control group that they were the control group. Instead, we spoke about participating in different courses. Although it led to confusion and demotivation on the training day itself, we expected that refusal rates would have been much higher had we mentioned the control group membership earlier in the study.

The question remains whether the control group was too frustrated with their 'training' to allow for a valid measurement. The satisfaction questionnaire, however, shows that 83% of the control group agreed or highly agreed with the statements that the pre-training feedback drive was 'useful', and that 90% agreed or highly agreed with the statement that the drive was 'fun'. In this respect, the control group's responses did not differ from the experimental group. Compared to the pre-training feedback scores, the scores of the control group on the post-training feedback driver were even higher. This leads us to conclude that the control group was highly motivated and, in this respect, did not significantly differ from the experimental group.

7.2. **The results of the study**

The following results are primarily based on the effect study.

7.2.1. *How attractive is the training for those who did not participate?*

Of 500 potential participants, 28% did not sign up for the course, 40% did not show up for the pre-test feedback drives, and 7% did not participate in the post-test feedback drive. About 25% finished all elements. From these figures it might be concluded that the majority of young drivers (68%) are not interested in such a course, even when a) all travel expenses are paid for, b) the course is free of charge, and c) incentives are used (lottery). The motives young drivers give for not signing up (no time, don't feel like it, not interested), also present indications in this direction.

7.2.2. *How attractive is the course for those who did participate?*

All the participants that completed the course filled out a questionnaire asking how useful and fun the different components had been. The results show that both control group as well as experimental group are similar on their judgment of the feedback drive. The experimental group also participated in a track training and a group discussion. The latter gets the lowest rating, while the track training gets the highest. However, the relative minor difference with the feedback drive opens up the discussion about the necessity of track training in these safety programmes. It is frequently mentioned that track training is needed in order to provide an attractive programme. The findings in this study lead to the conclusion that attractiveness is not dependent on the presence of track training after all.

Although exactly the same training programme was offered on two locations, the satisfaction scores differ between locations, both on the track-training programme and in the group discussion. The information from the process evaluation demonstrated that the trainers on the Lelystad track site were not happy with the assumptions underlying the programme. For instance, the programme emphasized that manoeuvring skills should not be trained, but that risks should be demonstrated in order to increase self-awareness about risks and to avoid over-confidence. The ANWB trainers, however, were convinced about the value of skill training. From participants we learned that these views were expressed during the course, possibly leading to confusion and dissatisfaction in the participants. This finding stresses again the importance of the personal views of the trainer for the success of the education programme, as has already been demonstrated in other educational fields, and has been documented by the ADVANCED project.

7.2.3. *The power: consequences of a smaller sample*

As only 33% of the participants actually completed the programme, the power of the study in order to be able to detect changes (given one is there) has been greatly reduced. From the power analysis it was concluded that the comparison between 'regularly' trained candidates and the candidates that had followed a 'best practice education' had to be dropped from the analyses because of too few participants in each cell.

Has the training resulted in a change in knowledge, attitudes, and behaviour with respect to the objectives of the course to increase 'risk awareness'?

7.2.4. *Effects on self-assessment, risk awareness and calibration*

The questionnaire measured the effect on risk awareness and self-assessment of skills. No statistical differences were found on the post-test between the control and the experimental group. To rule out the possibility that the questionnaire was not sensitive enough to find differences, the data were also analysed for gender differences. In line with expectations, differences showed up between men and women. Men were more confident and positive about their driving styles and saw less hazards. This, however, did not change after the training. This led to the assumption that the questionnaire measures (more stable) personality traits and attitudes which can not be altered by a one-day training course.

There were some significant effects in the situation judgements, but these were somewhat contradictory. There was only an effect in one out of two situations. However, the control group. which should theoretically remain at the same level, performed worse after the training. Moreover, there is no explanation why the group without any intervention improved in the same way the experimental group did.

7.2.5. *Does the training have an effect on those young drivers who are most at risk?*

The training's objective was to improve calibration. Calibration is defined as being the balance between self-assessment of skill and risk awareness. Those youngsters that are high on self-assessment (who think that they are extremely good drivers) and have a low risk awareness (regard dangerous behaviours as not dangerous at all) are particularly at risk. Another group that needs particular attention is the group highly insecure drivers with a high risk awareness. The training did not significantly affect either of these two risk groups, with respect to their self-perception and risk awareness.

7.2.6. *Effects on self-assessment (on-road observation form)*

The participants were asked after the on-road feedback drive about the quality of their own driving with respect to vehicle control, safe and defensive driving, and anticipation. The training did not have an effect on the assessment of skill or complexity of the driving task, according to the young drivers themselves. This is, on the one hand, positive because the participants did not think they improved their skills as a result of the training. On the other hand, there was no change in the perception of the complexity of the driving task.

It is possible that the young drivers already had a very accurate image of their quality as (car) drivers and about the complexity of the traffic situation. To study this the scores of the participant were compared to the scores the instructor had awarded them on the same four items;

- vehicle control and general skills;
- safe and defensive driving;
- anticipation;
- estimation of the complexity of the driving task.

On 'vehicle control and general skills', instructors and participants did not differ in their assessment neither on the pre-test nor on the post-test. On 'safe and defensive driving', in the pre-test. participants rated their performance higher than the instructor did. As the course was directed at improving self-assessment skills, one would have expected this to improve in the sense that their assessment would be more in line with that of the instructor. This was not the case.

7.2.7. *Driving assessment*

The driving performance of the participants was assessed by the driving instructor during the on-road feedback drive, before and after the training. Three aspects of driving were assessed:

– vehicle control;
– driving skill;
– calibration.

The results on the driving assessment show that there was a significant effect of the training on driving skill and calibration skill. This is visible in the difference between experimental and control group, but with some analytical complications and inexplicable patterns. The biggest problem is the difference between the experimental and control group during the first feedback drive. The positive effect of the experimental group and negative effect of the control group can be explained completely by the absence of this difference during the second feedback drive. The difference during the first feedback drive is probably caused by the fact that the participants in the experimental group had to drive in an unfamiliar environment, whereas the control group could drive nearby their homes; most of the times in the same place were they got their drivers licence. Although this undesirable phenomenon can be explained, it still interferes with the interpretation of the effects of the training.

There also seems to be a difference between the performance in the two locations, Lelystad and Rijssen. The performance of the participants on calibration skill increased after the training in Rijssen, but decreased after the training in Lelystad. This time the performance of the participants during the first feedback drive is exactly the same for both locations. Therefore the difference between Lelystad and Rijssen can be attributed completely to the effects of the training.

The process evaluation indicated that the trainers in Lelystad and Rijssen did not share the same opinion on the definition of a useful training. Without drawing any conclusions as to who is right in this matter, it is very possible that the trainers in Lelystad had to give a training which they basically did not agree with. This could have (subconsciously) affected the way they gave the training, or the way the participants perceived the training. Research has shown (ADVANCED, 2002) that any education, looses its strength if the educator is not absolutely convinced about what he/she is teaching. Moreover, that the effectiveness of the education is largely dependent on the person, the beliefs of the teacher, and his behaviour (Hale and Glendon, 1987). For a more detailed discussion of the role of the 'teacher' , see the ADVANCED report.

7.3. Conclusions

In the Dutch pilot, the recommendations of the ADVANCED report were closely followed with respect to the content of the course and the evaluation of its effects. However, in practice these recommendations were not always followed, as was the case in one of the two locations.

In this study, it has been demonstrated that, on the one hand, participants recognize the second phase as a useful and necessary part of their driving career. On the other hand, the high refusal rate demonstrates that youngsters are not interested in participating on a voluntary basis. The effects of the course are limited, and can even be negative, if trainers are not fully equipped to give the course, indicating that a much greater effort is needed in training second phase trainers than has been the case in this project.

In the course that was delivered by extra-qualified trainers, effects were found in the self-awareness calibration factor of the on-road feedback drive. Moreover, participants also valued these modules highly. On many other measures -including the questionnaire- the course did not result in any changes, which in itself leads to the question on how to interpret the results. Are the effects found real effects, and are the other instruments insensitive to measuring these effects? Or alternatively, are the statistically significant effects not to be found in real-life traffic? Answering this question is only possible when more evaluation studies of this type of training are carried out.

References

Brown, I.D. (1989). *How can we train safe driving?* Traffic Research Centre, University of Groningen, Haren.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale NJ.

Deery, H.A. (1999). *Hazard and risk perception among young novice drivers*. In: Journal of Safety Research, vol. 30, nr. 4, p. 225-236.

Engström, I., Gregersen, N.P., Hernetkoski, K., Keskinen, E. & Nyberg, A. (2003). *Young novice drivers, driver education and training; Literature review.* VTI Rapport 491A. Swedish National Road and Transport Research Institute VTI, Linköping.

Glad, A. (1988). *Phase 2 driver education; Effect on the accident risk.* Institute of Transport Economics TOI, Oslo.

Gregersen, N.P. (1996). *Young drivers' overestimation of their own skill; An experiment on the relation between training strategy and skill*. In: Accident Analysis and Prevention, vol. 28, nr. 2, p. 243-250.

Hale, A.R. & Glendon, A.I. (1987) *Individual behaviour in the control of danger*. Industrial Safety Series, vol. 2. Elsevier, Amsterdam.

Hatakka, M. (1998). *Novice drivers' risk- and self-evaluations; Use of questionnaires in traffic psychological research; Method development, general trends in four sample materials, and connections with behaviour*. University of Turku, Turku.

Kuiken, M.J. & Twisk, D.A.M. (2001). *Safe driving and the training of calibration; Literature review.* R-2001-29. SWOV Institute for Road Safety Research, Leidschendam.

McKenna, F.P., Stanier, R.A. & Lewis, C. (1991). *Factors underlying illusory self-assessment of driving skill in males and females*. In: Accident Analysis and Prevention, vol. 23, nr. 1, p. 45-52.

Sanders, N. (2002). *The EU ADVANCED project; Description and analysis of post-licence driver and rider training*. International Commision of Driver Testing Authorities CIECA, Rijswijk.

Sanders, N. & Keskinen, E. (2004). *EU NovEV project; Evaluation of post-licence training schemes for novice drivers.* International Commision of Driver Testing Authorities CIECA, Rijswijk.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Third edition. Lawrence Erlbaum Associates Publishers, Mahwah NJ.

Vissers, J.A.M.M., de Craen, S., Houtenbos, M., Twisk, D.A.M. & Kooistra, A. (2004). *Tweede fase opleidingsprogramma; Resultaten van de Nederlandse proef.* Regionaal Orgaan Verkeersveiligheid Gelderland ROVG, Arnhem.

Willems, B. & Cuyvers, R. (2004). *Ervaring en Ongevalbetrokkenheid*. RA-2004-30. Steunpunt Verkeersveiligheid bij Stijgende Mobiliteit, Diepenbeek.

Appendix 1 Organisations involved

- Gelderland Provincial Road Safety Board (project leader);
- ANWB (training Lelystad);
- BOVAG (involved in the project with regard to the role of driving schools in the training programme);
- Central Licensing Bureau. Eastern region (organization of the feedback drives);
- The Ministry of Transport (involved because of possible introduction of compulsory second phase training in the Netherlands);
- FAM (involved in the project with regard to the role of driving schools in the training programme);
- Novem (involved because of the integrated part of the 'New Driving' training programme);
- SWOV (realisation of the evaluation research);
- Traffic Test (organization and guidance of the training programme);
- VVCR (training Rijssen).

Appendix 2 Questionnaire

Questionnaire

Instructions:

This questionnaire contains two types of questions. There are questions about your driving style, things which you consider dangerous in traffic or that you are good at. But there are also some photos of traffic situations, of which you will be asked to comment on. While filling out the questions you may come across the same photo or one that is very similar to a previous one. Please try to judge each photo on its own merits and answer the question accordingly.

Personal details

First name:		
Last name:		
Gender:	Male / Female	(Circle the answer)
Date of birth	_ _ / _ _ / _ _ _ _	
Street:		Nr.
Zipcode		
City:		
Telephone number:	_ _ _ _ _ _ _ _ _ _	
Mobile number:	06 - _ _ _ _ _ _ _ _	
Date of receiving drivers licence:	_ _ / _ _ /_ _ _ _	
Driver education	Regular / 'Best practice'	(Circle the answer)
Did you receive an intensive course?	Yes / No	(Circle the answer)
Do you have a car at your disposal?	Yes / No	(Circle the answer)
Number of driving lessons:		
Length of driving lessons in minutes:		
What is the level of high school education, that you have finished or still taking?	VMBO / MAVO HAVO VWO Different	(Circle the answer)

Question 1

Imagine you are driving the above minivan. You want to turn right at the junction. The cyclist wants to go straight on.

We would like you to allocate a total of 10 points to the following two questions. The higher the points, the more frequent you would do it.

a. How often would you accelerate and carry on turning? _____ points

b. How often would you brake and let the cyclist go first? _____ points

Question 2

How often do you drive a car?

a. Practically every day
b. 2 or 3 times per week
c. on average once a week
d. a few times a month
e. less than once a month

(circle the correct answer)

Question 3

Roughly how many kms have you driven since passing your driving test?

a. 0 - 500 km
b. 501 – 2000 km
c. 2001 – 5000 km
d. 5001 – 10.000 km
e. more than 10.000 km

(circle the correct answer)

Question 4

Please indicate how fast you would drive in the above situation.

NB: the question is not how fast you are allowed to drive but how fast you would be driving if this photo was taken from your car (it may be faster or slower than the speed limit)

Answer _____ km/h

Question 5

Please indicate how fast you would drive in the above situation.

NB: the question is not how fast you are allowed to drive but how fast you would be driving if this photo was taken from your car (it may be faster or slower than the speed limit)

Answer _____ km/h

Question 6

You have now got your driving licence; but as you look back at your driver training, how well do you think it prepared you for the following things:

	badly	ok	well?
a. steering and manoeuvring the car	1☐	2☐	3☐
b. speed selection in a number of situations	1☐	2☐	3☐
c. recognising dangers and risks in traffic	1☐	2☐	3☐
d. driving in difficult circumstances (rain, darkness, etc)	1☐	2☐	3☐

Question 7

Which of the following statements about your driving style do you agree with?

a. I drive hard and sharp and like taking little risks
b. On the whole, I drive carefully. Driving can be dangerous.
c. None of the above; I am in between

(circel the correct answer)

Question 8

You are approaching a zebra crossing. A pedestrian arrives.

We would like you to allocate a total of 10 points to the following two questions. The higher the points, the more frequent you would do it.

a. How often would you accelerate and drive through? _____ points

b. How often would you brake to allow the pedestrian to cross? _____ points

Question 9

Imagine you are driving the above minivan. You want to turn right at the junction. The cyclist wants to go straight on.

We would like you to allocate a total of 10 points to the following two questions. The higher the points, the more frequent you would do it.

a. How often would you accelerate and go ahead? _____ points

b. How often would you brake and let the cyclist go first? _____ points

Question 10

Please indicate how dangerous you find the following subjects in traffic.

1 = not dangerous at all
2 = hardly dangerous
3 = not dangerous / not without risk
4 = quite dangerous
5 = very dangerous

a.	If I lose control of the car, I find that…	1☐	2☐	3☐	4☐	5☐
b.	If I drive too fast, I find that…	1☐	2☐	3☐	4☐	5☐
c.	If others react slowly, I find that…	1☐	2☐	3☐	4☐	5☐
d.	If I see danger too late in traffic, I find that	1☐	2☐	3☐	4☐	5☐
e.	If I don't pay enough attention to other road users, I find that…	1☐	2☐	3☐	4☐	5☐
f.	If others have an aggressive driving style, I find that…	1☐	2☐	3☐	4☐	5☐
g.	If I have the wrong (or no) expectations, I find that …	1☐	2☐	3☐	4☐	5☐
h.	If others lose control of their vehicle, I find that …	1☐	2☐	3☐	4☐	5☐
i.	If I make the wrong decision in a dangerous situation, I find that …	1☐	2☐	3☐	4☐	5☐
j.	If I don't know the traffic rules properly, I find that …	1☐	2☐	3☐	4☐	5☐
k.	If others don't follow the traffic rules properly, I find that…	1☐	2☐	3☐	4☐	5☐
l.	If others drive too close behind me, I find that…	1☐	2☐	3☐	4☐	5☐

Question 11

You are approaching a zebra crossing. A pedestrian arrives.

We would like you to allocate a total of 10 points to the following two questions. The higher the points, the more frequent you would do it.

a. How often would you accelerate and drive through? _____ points

b. How often would you brake to allow the pedestrian to cross? _____ points

Question 12

Please indicate how fast you would drive in the above situation.

NB: the question is not how fast you are allowed to drive but how fast you would be driving if this photo was taken from your car (it may be faster or slower than the speed limit)

Answer _____ km/h

Question 13

Please indicate how dangerous you find the following subjects in traffic.

1 = not dangerous at all
2 = hardly dangerous
3 = not dangerous / not without risk
4 = quite dangerous
5 = very dangerous

a. If I can't get a quick overview of the traffic situation, I find that…	1☐	2☐	3☐	4☐	5☐
b. If I break traffic rules, I find that…	1☐	2☐	3☐	4☐	5☐
c. If I can't see well in the dark or in bad weather, I find that…	1☐	2☐	3☐	4☐	5☐
d. If I react slowly, I find that…	1☐	2☐	3☐	4☐	5☐
e. If I don't take safety into account when planning or selecting my route or schedule, I find that…	1☐	2☐	3☐	4☐	5☐
f. If others take risks in traffic, I find that…	1☐	2☐	3☐	4☐	5☐
g. If I have an aggressive driving style, I find that	1☐	2☐	3☐	4☐	5☐
h. If I drink and drive, I find that…	1☐	2☐	3☐	4☐	5☐
i. If others are not paying enough attention and don't see me, I find that…	1☐	2☐	3☐	4☐	5☐
j. If I drive up close to the car in front, I find that…	1☐	2☐	3☐	4☐	5☐
k. If I get nervous because of others, I find that	1☐	2☐	3☐	4☐	5☐
l. If I take risks in traffic, I find that...	1☐	2☐	3☐	4☐	5☐

Question 14

Please indicate how fast you would drive in the above situation.

NB: the question is not how fast you are allowed to drive but how fast you would be driving if this photo was taken from your car (it may be faster or slower than the speed limit)

Answer _____ km/h

Question 15

You are approaching a zebra crossing. A pedestrian arrives.

We would like you to allocate a total of 10 points to the following two questions. The higher the points, the more frequent you would do it.

a. How often would you accelerate and drive through? _____ points

b. How often would you brake to allow the pedestrian to cross? _____ points

Question 16

Every driver has strengths and weaknesses as a driver. If you look at your own skills, what are your strengths and weaknesses?

	Very weak	Weak	Not weak not strong	Strong	Very strong
a. Keeping the car under control	1☐	2☐	3☐	4☐	5☐
b. Adapting your speed to the circumstances	1☐	2☐	3☐	4☐	5☐
c. Reacting fast	1☐	2☐	3☐	4☐	5☐
d. Seeing dangers in traffic in time	1☐	2☐	3☐	4☐	5☐
e. Getting a quick overview of the situation	1☐	2☐	3☐	4☐	5☐
f. Watching out for other road users	1☐	2☐	3☐	4☐	5☐
g. Predicting what others will do	1☐	2☐	3☐	4☐	5☐
h. Performing in an emergency situation	1☐	2☐	3☐	4☐	5☐
i. Knowledge of traffic rules	1☐	2☐	3☐	4☐	5☐
j Respecting traffic rules	1☐	2☐	3☐	4☐	5☐
k. Keeping sufficient safety margins	1☐	2☐	3☐	4☐	5☐
l. Avoiding unnecessary risks	1☐	2☐	3☐	4☐	5☐
m. Planning and selecting safe routes	1☐	2☐	3☐	4☐	5☐
n. Driving carefully	1☐	2☐	3☐	4☐	5☐

Appendix 3 Photo's of situations

Situation 1:

Situation 2:

Appendix 4 Driving assessment form

Fill-in date: .....
Name of participant: .....
Name of driving analyst: .....

Criteria	***Assessment***			***Comments***
Driving style	***G***	***V***	***O***	
1. Preparation for trip / end of trip				
2. Vehicle handling				
3. Vehicle control				
4. Energy-saving driving				
5. Driving independently				
6. Social traffic behaviour				
7. Defensive behaviour: anticipation				
8. Defensive behaviour: effective observation skills				
9. Defensive behaviour: safety cushion and safety margins				
10. Risk awareness, danger recognition and traffic insight				
11. Adapted and decisive driving: speed				
12. Adapted and decisive driving: decisive handling				
Situation-specific behaviour	***G***	***V***	***O***	
13. Driving on straight and bendy roads				
14. Behaviour at junctions				
15. Behaviour when turning				
16. Entering and exiting traffic				
17. Overtaking and passing				
18. Being overtaken				
19. Changing lanes and lateral positioning				
20. Driving on different surfaces				

Note: G= Good, V = Satisfactory and O= Unsatisfactory