TRAIL

thesis series

Saskia de Craen

The X-factor

A longitudinal study of calibration in young novice drivers

Propositions pertaining to the dissertation

"The X-factor: A longitudinal study of calibration in young novice drivers"

Saskia de Craen, March 16, 2010

I. Although calibration seems to be an important factor for safe driving, and experienced drivers are better 'calibrated' than novice drivers; calibration *cannot* explain the decrease in crash risk of young novice drivers during the first two years of their driving career.

II. Overall, drivers do *not* believe they are better drivers than the average driver. At best, young novice drivers can be said to believe they are better drivers than their actual performance suggests.

III. Professional driving examiners are rather successful in predicting drivers' (self-reported) crash liability, based on just a half-an-hour drive.

IV. The difference between practicing driving and driving experience is reflected by the way a driver relates to situations and learns from these.

V. "Ignorance more frequently begets confidence than does knowledge" (Darwin, 1871: p. 3).

VI. Standardizing driving tests to ensure a higher test reliability inevitably lowers predictive validity.

VII. Hazard or risk perception can never be seen as independent from drivers' beliefs about their own driving skills.

VIII. Paid maternity leave is just a small social investment compared to a female employee's contribution to the economy.

IX. In most societies the following still applies: "Whatever women do they must do twice as well as men to be thought half as good. Luckily this is not difficult" (Charlotte Whitton, Canada Month, June 1963).

X. The outward appearance (and lay-out) of a thesis is often underestimated. Could you imagine presenting your new born baby in a potato tray?

These propositions are considered to be opposable and defendable and as such have been approved by the supervisors prof. dr. K.A. Brookhuis, prof. dr. H. Elffers and dr. M.P. Hagenzieker.

Stellingen behorende bij het proefschrift

"The X-factor: A longitudinal study of calibration in young novice drivers"

Saskia de Craen, 16 maart 2010

I. Hoewel kalibratie een belangrijke voorwaarde lijkt te zijn voor veilig rijden, en ervaren automobilisten beter 'gekalibreerd' zijn dan onervaren automobilisten, kan kalibratie *niet* de afname in ongevalsrisico van jonge onervaren automobilisten in de eerste twee jaar van hun rijcarrière verklaren.

II. Over het algemeen vinden automobilisten zichzelf *geen* betere bestuurder dan de gemiddelde automobilist. Hooguit kan worden gesteld dat jonge onervaren automobilisten zichzelf betere bestuurders vinden dan hun werkelijke prestatie zou rechtvaardigen.

III. Professionele examinatoren zijn behoorlijk succesvol in het voorspellen van het (zelfgerapporteerde) ongevalsrisico van een automobilist, op basis van een rit van slechts een half uur.

IV. Het verschil tussen oefenen met autorijden en het opdoen van rijervaring zit hem in de manier waarop een bestuurder gebeurtenissen op zichzelf betrekt en hiervan leert.

V. "Onwetendheid ligt vaker ten grondslag aan zelfvertrouwen dan kennis" (Darwin, 1871: p.3).

VI. Het verhogen van de betrouwbaarheid van het rijexamen door middel van standaardisatie gaat ten koste van de validiteit van dit examen.

VII. Gevaar- of risicoherkenning kan nooit los gezien worden van hoe een automobilist denkt over zijn eigen rijvaardigheid.

VIII. Zwangerschapsverlof is een relatief kleine maatschappelijke investering, gezien de bijdrage die een vrouwelijke werknemer levert aan de economie.

IX. Voor de meeste samenlevingen geldt nog steeds dat "een vrouw twee keer zo goed moet zijn als een man, om half zo goed gevonden te worden. Gelukkig is dit niet moeilijk." (Charlotte Whitton, Canada Month, juni 1963)

X. Vreemd genoeg wordt het belang van het uiterlijk (en de opmaak) van een proefschrift vaak onderschat. Je legt je pasgeboren baby tenslotte ook niet in een aardappelbak.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren prof. dr. K.A. Brookhuis en prof. dr. H. Elffers en copromotor dr. M.P. Hagenzieker.

The X-factor

A longitudinal study of calibration
in young novice drivers

Saskia de Craen

The X-factor

A longitudinal study of calibration in young novice drivers

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 16 maart 2010 om 12:30 uur

door

Saskia DE CRAEN

Doctorandus in de Psychologie
geboren te Voorburg

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. K.A. Brookhuis
Prof. dr. H. Elffers

Copromotor: dr. M.P. Hagenzieker

Samenstelling Promotiecommissie:

Rector Magnificus	Voorzitter
Prof. dr. K.A. Brookhuis	Technische Universiteit Delft, promotor
Prof. dr. H. Elffers	Vrije Universiteit Amsterdam, promotor
Dr. M.P. Hagenzieker	Technische Universiteit Delft, copromotor
Prof. dr. R. Fuller	Trinity College Dublin
Prof. dr. E.M. Steg	Rijksuniversiteit Groningen
Prof. ir. F.C.M. Wegman	Technische Universiteit Delft
Drs. D.A.M. Twisk	Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam

Drs. D.A.M. Twisk heeft als begeleider in belangrijke mate aan de totstandkoming van het proefschrift bijgedragen.

Dit proefschrift is tot stand gekomen met steun van de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV. Hiernaast heeft het CBR een belangrijke bijdrage geleverd door het beschikbaar stellen van examinatoren en het geven van toegang tot de examencentra.

TRAIL Thesis Series T2010/2, the Netherlands TRAIL Research School
P.O. Box 5017
2600 GA Delft
The Netherlands
T: +31 (0) 15 278 6046
F: +31 (0) 15 278 4333
E: info@rsTRAIL.nl

ISBN 978-90-5584-127-1

Artwork on cover: Richard Bunschoten

Preface

Foremost, I would like to thank SWOV for the opportunity to work on this PhD study; and all colleagues at SWOV for their support, especially when the 'going got rough'. More specifically, I would like to thank: Niels Bos for the advice and interesting figures; Vincent Kars for his assistance with the website; Hansje Weijer for her help on the translations of the instruments; and the 'ladies from the library', Dennis and Ineke, for all the articles that had to be purchased and for the many trips to the SWOV basement for the somewhat older articles. I'm sure this thesis helped digitalising the SWOV literature database a bit further.

Sjoerd Houwing, thanks for all your support (and 'stophoest') during the most stressful periods of this endeavour. You are a true friend… uh, colleague. A special thanks for my colleague and friend Maura Houtenbos who was involved in every step of this project. From the selection of instruments, recruitment of participants, data analysis, to writing this thesis (you even made it to the stimulus material (see Appendix B – situation 2)). Besides being extremely helpful, you made this project a lot of fun.

I am very grateful for the cooperation with CBR (the Dutch Driving Test Organisation). It felt like nothing was too much trouble. I would especially like to thank Theo van Rijt and Patrice van Assendelft for opening the doors at CBR and the examiners for their warm welcome to their test locations. A special thanks to Cock Pleune, Ger(t) Roos, Hella Heeneman and Theo van der Drift, for their enthusiasm and dedication during the on-road driving

assessments. I learned a lot from talking to you, among other things how to become a better driver myself.

Divera Twisk, thank you for introducing me to the subject of young novice drivers. You encouraged me to make this project the best it could possibly be. Because of your critical view in the beginning of the project, you forced me to resolve many difficult issues early-on, which made the writing of this thesis a smooth(er) process.

Maura and Jacqueline, you played a very important role in the beginning of this project. During our "WOM" meetings we set the course for this project, and it is definitely due to your enthusiasm that 94% of all newly licensed drivers we contacted, agreed to participate in our study. So far, this is the highest response rate I have ever come across.

In this respect, I am also thankful for all the volunteers that joined us at the driving test locations to invite newly licensed drivers to participate in the study. And of course, thanks to all drivers (newly licensed and experienced) who completed all the questionnaires and driving diaries in this project with tireless enthusiasm and commitment.

I would like to thank my supervisors, Karel Brookhuis, Henk Elffers and Marjan Hagenzieker. This whole endeavour felt like a cooperation rather than a student-supervisor relationship. Your feedback has been very helpful, while you still gave me all the room for my own opinion. I especially appreciate how everyone worked so very hard to bring this thesis to completion in time to plan the defence before my maternity leave. I am deeply impressed by the way you helped me fulfil this wish.

I belief that it is impossible to succeed professionally (or in life for that matter) without a good solid 'social support system'. Fortunately, I am blessed with the best friends and family ever. Mariska, we have been friends for 20 years (as we are a bit ashamed to admit: "when did we turn so old?"). We have experienced ups-and-downs in both our lives; I hope that it will be up, up, up, from now on. Jolanda, I am "so" blessed to have you as a friend. We always got along great (to some people's surprise), but now our husbands and even our daughters are becoming BFF's. I hope we'll spent many more relaxed Sunday afternoons with both our families. To all friends from 'de Heische Tip': thanks for creating such a fun getaway from my professional and scientific life. Jolieke, Sjoerd and Maura, I'm looking forward to many more colleague evenings, and really hope that some day we'll see the Dutch entry for the Eurovision Song contest win! Finally,

Coranne, Kavita and Olga; although I see you way too little, you must know that I really have a blast every minute I spend with you guys.

To my dearest family: We are not the largest family, but easily make up for our limited size by how closely knit we are. Jeannette, Rob, Robbert, Renske and Irene thanks for all your love and support.

Oma (Granny): Mijn nieuwsgierigheid en het respect voor kennis heb ik zeker van u. Het is jammer dat u in uw leven niet de kansen heeft gekregen die ik in deze tijd heb. Het zou interessant zijn geweest, om te zien tot waar u het zou hebben gebracht.

Mum, I must have inherited my determination from you (if not by nature, then definitely by nurture). You are that kind of mother who would do and give everything, just to see your girls succeed in life. I hope we both made you proud; a major part of our success is on your conto.

I would like to end with a few words for the most important people in my life. Richard, we have been best friends since we were teenagers. You have always given me the feeling that I was worthwhile. And even during this difficult last year, you were so strong; you supported and comforted me more than the other way around. You are my true hero!

Daantje, you have helped me more with this thesis than you can ever imagine. Since you are in my life, I know what really matters. But more important: you are just great fun! Finally, to the little one in my belly: In a way we did this together; you shared in all the anxiety and stress. I wish you a less stressful life, at least until middle school.

I hope that the four of us (or whatever 'magic number' we will be) will have a lot of fun for a very long, long time.

Saskia de Craen, January 2010

Table of contents

1. Introduction

1.1. Background

Young, novice drivers have a higher crash rate than drivers from all other age categories (see Figure 1.1). In the Netherlands, a young novice driver (18-24 years old) has a four times greater chance of being involved in a crash than older, more experienced drivers (30-59 years old; SWOV, 2008). Crash rates are highest in the first months after licensing and drop substantially over the first two years of driving, with the most pronounced decline during the first six months or during the first 5000 kilometres of driving (OECD - ECMT, 2006).

There are basically two factors associated with the high crash risk: young age and lack of experience. The high crash risk for *young* drivers may be related to the fact that the human brain is still developing during adolescence (Paus et al., 1999; Sowell, Thompson, Holmes, Jernigan & Toga, 1999). Especially executive functions such as planning, impulse control, reasoning and the integration of information, which are relevant for safe driving, have not yet developed fully by the age of 18 (OECD - ECMT, 2006). In addition, specific subgroups of young drivers are even more at risk due to life-style factors, such as intentional risk taking, sensation seeking and peer pressure, often associated with young age (see Arnett, 2002).

Although young age is an important factor, crash studies suggest that the decrease in risk is more strongly related to gaining *experience* than to biological maturation. All novice drivers, irrespective of age, show an exponentially decreasing crash risk in the first years of their driving career (Maycock, Lockwood & Lester, 1991; Vlakveld, 2005). Therefore, this thesis focuses on how experience reduces crash risk over time, and which relevant processes are involved.

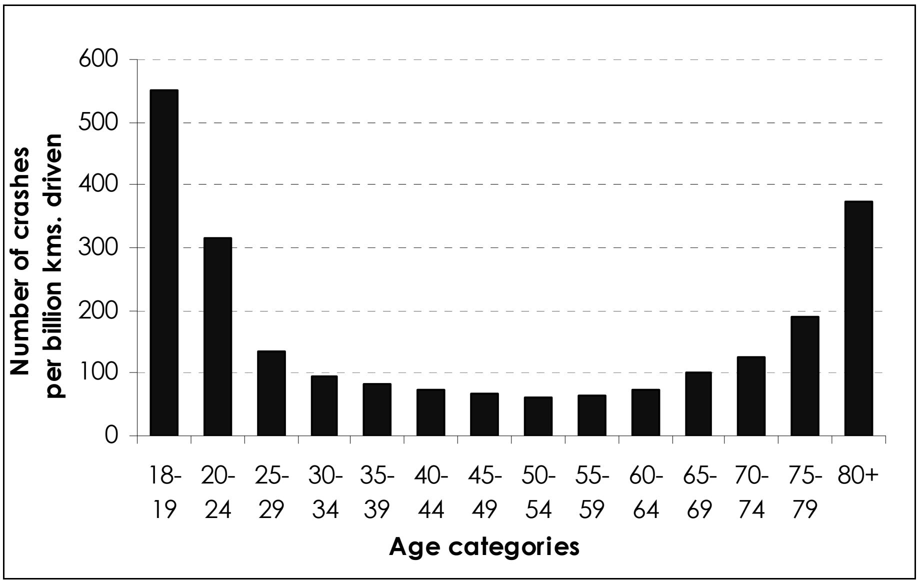

Figure 1.1. Number of crashes (fatal or with serious injuries) per billion motor-vehicle kilometres driven in the Netherlands in 1999-2007, for different age categories. Source: BRON (AVV); OVG (CBS until 2003); MON (AVV from 2004)[1]

Several authors argue that through practice, parts of the driving task (e.g. shifting gear) become automated (e.g. Groeger, 2000; Shinar, Meir & Ben-Shoham, 1998). One of the characteristics of automatic processing is that activities can be carried out without the need for active controlled processing or attention by the drivers (Shiffrin & Schneider, 1977). The more automated or routine a task becomes, the less mental capacity is required to perform the task (De Waard, 2002). This difference in mental workload may explain why driving a car is easier for more experienced drivers than for novice drivers (see Chapter 2 for a more detailed description of concepts such as mental workload and automated versus controlled processing).

[1] The estimated mileage is only available for the age group 18-24 year-old drivers as a whole, and not for 18-19 year-olds separately. There are indications that the 18-19 year-olds drive less than the 20-24 year-olds. Consequently, the crash risk for 18-19 year-olds presented in Figure 1.1 may be an underestimation of the actual crash risk.

Although mental workload has shown to have a significant contribution, an extra factor needs to be taken into account. The driving task is 'self-paced' (Taylor, 1964); that is, the driver can adjust the task demands (e.g. by reducing speed or increasing following distance), thus decreasing workload (Fuller, 2005). In theory, this strategy can help to overcome the limitations of novice drivers' performance; a novice driver can decrease the task demands to fit his[2] (deficient) level of automated driving.

However, studies have indicated that young novice drivers, as a group, do not use this strategy (Twisk, 1995). They tend to drive with (too) small safety margins (Engström, Gregersen, Hernetkoski, Keskinen & Nyberg, 2003) and are more likely, compared to other age groups, to engage in secondary behaviours (e.g. making a telephone call) while driving (Sayer, Devonshire & Flannagan, 2005). Calibration (see next section) may explain why young novice drivers do not adapt task demands sufficiently.

1.2. Calibration

A driver can decrease the task demands to fit his (poor) level of automated driving, for example by reducing speed or increasing headway. Whether this strategy can be applied successfully may be assumed to depend on three factors: a) the drivers' correct assessment of driving skills; b) the correct assessment of the complexity of the driving task; and c) the correct selection of behaviours that change task demands effectively. In a psychological context, the process of balancing task demands and capabilities has been referred to as *calibration* (Kuiken & Twisk, 2001; Mitsopoulos, Triggs & Regan, 2006).

Figure 1.2 introduces a model of the calibration process, inspired by Brown's (1989) model of subjective safety and Fuller's (2005) task-capability interface model. The model is not meant to provide an accurate description of reality, but rather to provide a simplified illustration of the elements of calibration and how they are presumably related to each other. Section 2.3 of this thesis will provide a more extensive description of calibration.

The separate elements in Figure 1.2 and processes similar to calibration have been described in the past. However, there is still not much agreement about what calibration is, how this affects traffic safety, and whether or how calibration develops over time (hence: "the X-factor"). For example, Brown (1989) and Gregersen (1995) describe the calibration process without actually

[2] In the remainder of the thesis, wherever 'he/his' is written, 'she/her' is also applicable

referring to it as *calibration*. Other studies mention calibration as a specific problem for young novice drivers (e.g. Triggs & Regan, 1998), without a clear description of the term. There are also studies which use a very narrow definition of calibration and describe only the self-assessment of skills (element A from Figure 1.2) as calibration (e.g. Harris & Drummond, 1998; Horrey, Lesch & Garabet, 2008). Finally, different authors use different terms to describe the same processes, for example: self-efficacy (Delhomme & Meyer, 2000; Sundström, 2008a), self-awareness (Mallon, 2006), self-regulation (Keating, 2007) and self-monitoring (Bailey, 2009) are used to describe what has been called "self-assessment of skills" in the current study.

The objective of this thesis is to investigate the high crash risk of young novice drivers, and more specifically, to explore whether a development in calibration skills could explain the substantial decrease in crash risk in the first years after licensing. To this end, the thesis will investigate whether empirical support for the concept of calibration can be found, starting from the model in Figure 1.2. In addition, it will analyse the development of calibration over time and with increasing experience.

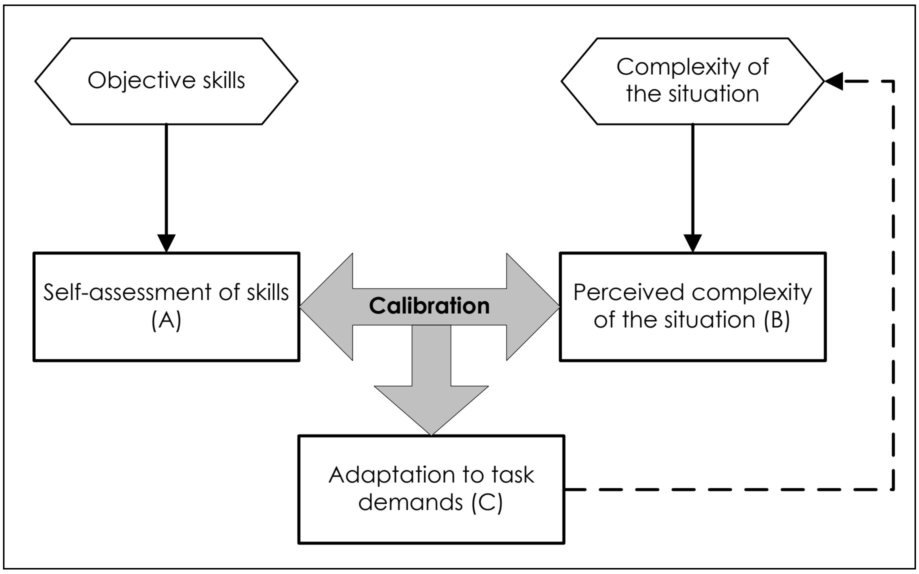

Figure 1.2. Model of the calibration process

Knowledge about (the development of) calibration of young novice drivers could help to improve driving education and driving tests. Attention in these fields has been focused on improvement of hazard perception; while the other elements of calibration (self-assessment of skills, and adaptation to task demands) are somewhat neglected. Or as Fuller (2008) stated: "Although there have been moves to address the issue of improving the perception of task demand [Perceived complexity] by the trainee driver, a similar response to the calibration problem on the perceived capability [Self-assessment of skills] side […] does not seem to have taken place to the same extent" (p. 340).

1.3. Research questions

This thesis aims to answer the following research questions:

1. *To what extent is poor calibration a contributing factor in the high crash risk of young novice drivers?*
   - *a. Do young novice drivers overestimate their skills more than experienced drivers?*
   - *b. Does an inadequate self-assessment of skills affect adaptation to task demands?*
   - *c. Is there a relationship between the elements of the calibration model and self-reported crashes?*

2. *How can calibration be measured?*

3. *How does calibration develop over time?*

It is expected that young novice drivers are worse at calibration than experienced drivers, and that they do not adapt to task demands sufficiently because they overestimate their driving skills and underestimate the complexity of the situation.

1.4. Method

In order to monitor the development of calibration, a group of young novice drivers was intensively followed from the moment of licensing over a period of two years. A study by Vlakveld (2005) indicated that, for the Dutch situation (in the years 1991-2001), the crash risk of novice drivers drops substantially during roughly the first four years of driving experience. A Canadian study (Mayhew, Simpson & Pak, 2003) indicated that a considerable drop already occurs during the first two years of independent driving. Considering these studies and practical considerations, a two-year period was chosen. Because of indications that the most distinctive drop in risk occurs in the first months of the driving career (Mayhew et al., 2003; Sagberg, 1998), the young novice drivers filled in the first questionnaire directly after they passed their driving exam.

To control for the fact that young novice drivers may change as a result of participating in this study, a small group of older, experienced drivers was also monitored during two years. The expectation was that this group will show no changes over time. As a second precautionary measure, the novice drivers were randomly assigned to two subgroups. The first group started completing questionnaires from the moment of licensing, while the second group of novice drivers started six months later.

During the two-year period, the participants completed questionnaires, kept a driving diary and participated in an on-road driving assessment.

The questionnaire contained items on self-assessment of skill and perceived risks in traffic. To monitor the internal processes of calibration, an instrument was developed and administered in the questionnaire to measure the outcome of the calibration process (the Adaptation Test; De Craen, Twisk, Hagenzieker, Elffers & Brookhuis, 2008).

In the driving diary, drivers reported on the trips that they made and the situations encountered in traffic. Drivers reported, for example, how much they had driven, if they had driven at night, with or without passengers, and if they had consumed any alcohol before driving.

To compare the reported experiences with actual driving performance, a subgroup participated in an on-road driving assessment. Driving skills were assessed on two occasions (in 2006 and 2007) in order to detect any changes over time as a result of driving experience. See Chapter 3 for a more extensive description of the design and instruments used in this study.

1.5. Outline

As can be seen in the graphical outline (Figure 1.3), the thesis consists of three parts. In the first part, Chapter 2 gives an overview of the literature describing the high crash risk of young novice drivers. In addition, Chapter 2 includes a theoretical framework for the calibration model with the following three elements: 1) self-assessment of skills, 2) perceived complexity of the situation, and 3) adaptation to task demands. Chapter 3 describes the methods that were used to study (the different elements of) calibration.

In the second part different elements of the calibration model are studied. Chapter 4 describes how *self-assessment of skills* can be measured best, and investigates whether novice drivers overestimate their driving skills more than experienced drivers. Chapter 6 studies whether inadequate self-assessment of skills is connected to insufficient *adaptation to task demands*.

Perceived complexity of the situation is always discussed in relation to the other elements of the calibration model, and is therefore not the main topic of a separate chapter. The Adaptation Test developed in the second part (Chapter 5) measures adaptation of speed to complexity of the situation, and can be used as an indication of calibration.

The third part of the thesis describes how calibration develops over two year's time, and describes the results of the driving diary (Chapter 7). The final chapter of this thesis (Chapter 8) will discuss the results of the preceding chapters in the context of other research findings and will draw some conclusions from the results of this study.

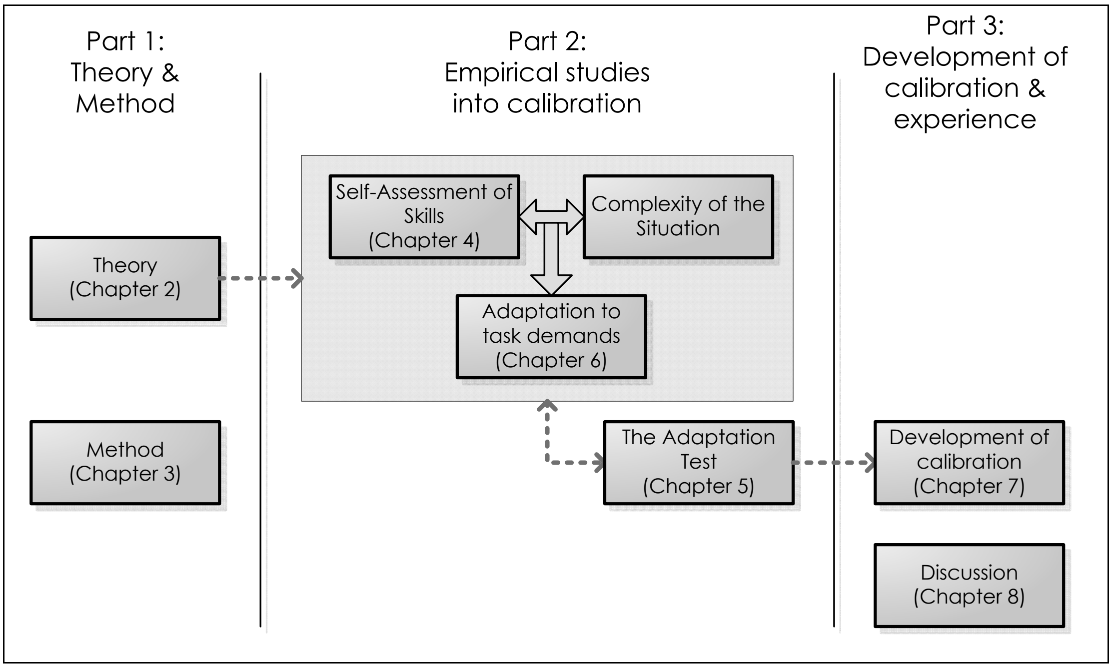

Figure 1.3. Graphical outline of the thesis

PART 1:

THEORY & METHOD

2. Theoretical background

This chapter gives an overview of the literature into the high crash risk of young novice drivers. In addition, it provides the theoretical background of the thesis. Section 2.1 describes the problem of young novice drivers; why are young novice drivers considered a risk group? It is argued that lack of experience is a larger contributor than young age.

Section 2.2 shows that the high crash risk of novice drivers, for a major part, can be attributed to the limited automation of driving subtasks, which leads to a higher mental workload for novice drivers compared to experienced drivers. The more routine a task becomes (automatic control), the less mental workload is required to perform it, and the driving task can be executed more efficiently and with less effort.

However, as is illustrated by the Motivational models in Section 2.3, the driving task is 'self-paced'. That is, drivers can make the driving task easier (or more demanding), for example by changing speed or headway. So, in theory, a novice driver can decrease the task demands to fit his (deficient) level of automated driving.

In Section 2.4, the 'self-paced' (motivational) models are incorporated into a new 'Calibration' model with three elements: 1) self-assessment of skills, 2) perceived complexity of the situation, and 3) adaptation to task demands.

2.1. The high crash risk of young novice drivers

Young, novice drivers have the highest risk compared to drivers from other age groups with respect to being involved in a traffic crash, in all motorised countries (see for example: Brorsson, Rydgren & Ifver, 1993; Engström et al., 2003; Gregersen & Bjurulf, 1996; Mayhew et al., 2003; Mayhew, Simpson, Singhal & Desmond, 2006; Murray, 2003; OECD - ECMT, 2006; Vlakveld, 2005; Williams, 2003).

There are typical characteristics of the crashes young, novice drivers are involved in (OECD - ECMT, 2006; Vlakveld, 2005). For example, crash records show that young, novice drivers are overrepresented in single-vehicle and loss-of-control crashes (Clarke, Ward, Bartle & Truman, 2006; Mayhew et al., 2003). High speed is a major factor in young novice drivers' crashes. Harrison, Triggs & Pronk (1999) found that speed related crashes were most common among young male drivers (almost 30% of all causation crashes) compared with young females (about 21%); in comparison, speeding was found to contribute to approximately 15% of older drivers' crashes.

Young novice drivers seem to have more problems during night hours. Gregersen and Nyberg (2002, as cited in OECD - ECMT, 2006) report Swedish data from 1994 to 2000 on time distribution of crashes, which showed that 32% of 18-19 year-old drivers' crashes occurred during darkness, while the corresponding share for other ages was 22%. In connection with this, fatigue is a common problem especially among young male night-time drivers (Vlakveld, 2005).

Alcohol and drugs seem to have more impact on young novice drivers. That is, in the Netherlands, young novice drivers do not drink-and-drive more often (AVV, 2007), but their crash risk is more greatly affected by alcohol, even at relatively low levels than those of older people (Preusser, 2002).

Finally, studies have shown that crash risk for young drivers is increased with the presence of teenage passengers (Brorsson et al., 1993; Preusser, Ferguson & Williams, 1998; Williams & Ferguson, 2002). However, this is only true for the presence of *teenage* passengers. When no differentiantion is made in the age of the passengers, crash risk seems to be positively influenced by the presence of passengers (Engström, Gregersen, Granström & Nyberg, 2008).

All factors related to the high crash risk of young novice drivers can be summarized in two elements; their young *age* and their lack of *experience*. In this regard, it is important to note that it has been difficult in the past to establish the relative contribution of young age and lack of experience (i.e. most novice drivers are also young drivers; Mayhew & Simpson, 1995; McCartt, Mayhew, Braitman, Ferguson & Simpson, 2009). And, as Groeger (2006) stresses, neither age nor inexperience are in themselves sufficient explanations for the high crash risk of young novice drivers, merely the factors associated with young age and inexperience can explain the high crash risk. The factors associated with young age and lack of experience are discussed in more detail in the next sections.

An intervening factor often mentioned in the young novice driver discussion is *gender*. It appears that young male drivers are even more at risk than their female counterparts (OECD - ECMT, 2006). Although this thesis focuses on the crash risk of *all* young novice drivers, gender will be discussed as a separate factor in Section 2.1.3.

2.1.1. Factors associated with young age

Figure 1.1 in Chapter 1 already showed that in the Netherlands the youngest age group has the highest crash risk. This level drops substantially for drivers over 25 years of age, and than increases again as the driver passes middle age.

In countries where driving is permitted from the age of 15 or 16, these age groups show an even higher crash risk (see OECD - ECMT, 2006). For example, Begg and Langley (2009) demonstrate in a review of several crash studies, that inexperienced and experienced *young* drivers have a high crash risk that decreases with age. Begg and Langley conclude that age, independent of experience, is a major determinant of risk. Waller, Elliott, Shope, Raghunathan & Little (2001) examined offences and crashes ('incidents') of 13,809 young adult drivers in Michigan. They found the highest risk for 16 year-old drivers, and a 5% reduction in total crash odds for each additional year of age at time of licensing.

At least part of the high risk of young drivers can be explained by biological factors, which apply to *all* young drivers. For example, neurobiological studies have shown that, at the age of 16, the human brain is still not fully matured (Paus et al., 1999; Sowell et al., 1999). Specifically those areas in the frontal lobe that deal with 'executive' functions like planning, impulse control, reasoning, and the integration of information, that are relevant if not crucial for adequate driving behaviour have not developed fully yet (OECD - ECMT, 2006).

These neurobiological studies are relatively recent, and studied brain activity during the performance of rather simple tasks. Moreover, these studies focussed on 16-year-olds, whereas in Europe most drivers are not licensed before the age of eighteen. Future neurobiological research will possibly reveal what these insights in brain development mean for the complex task of driving a car and for newly licensed drivers in Europe.

In addition to biological factors, specific *subgroups* of young drivers are considered to be even more at risk due to intentional risk taking (Ulleberg, 2002) or sensation seeking (see Jonah, 1997, for an overview). Arnett (2002) gives a good overview of life-style factors of young drivers, such as the power of friends (or peer pressure), the optimism bias, adolescent emotionality, growing importance of responsibility and freedom and risk in emerging adulthood.

Because it is difficult to reduce overall crash risk when focussing on a subgroup of drivers (e.g. Begg, Langley & Williams, 1999), life-style factors are not within the scope of this thesis.

2.1.2. Lack of experience

In addition to young age, the high crash risk of novice drivers has also been attributed to their inexperience. Crash rates drop substantially over the first two to four years of driving independently, with the most pronounced decline during the first six months or during the first 5000 kilometres of driving (OECD - ECMT, 2006).

According to McCartt et al. (2009) researchers have typically failed to partial-out the relative effects of age and driving experience when examining the driving skill of novice drivers. Young drivers are by definition inexperienced drivers. Although it is possible to find novice, but older drivers, this is usually an exceptional group. There are reasons why this group waited so long to get their driver's license. So if this group shows a particular behavioural pattern it cannot necessarily be traced back directly to their lack of experience but also to the special status of this group.

In Figure 2.1, from Vlakveld (2004; 2005), an attempt is made to establish the separate effect of young age and driving experience in the Netherlands. This figure was inspired by Maycock, Lockwood & Lester (1991) who used a similar figure to show the separate contributions of inexperience and young age for automobile crash rates in the UK.

Figure 2.1 shows self-reported crash risk, per 100,000 kilometres, for four different age groups from the start of their driving career and the years following. Figure 2.1 shows that for novice drivers under the age of 30, there is an enormous drop in crash risk in the first years of driving. The tops of each crash risk figure are connected to visualize the age effect. When comparing this age-effect to the separate crash risk curves for the different age groups, it is apparent that the decrease in crash risk due to experience is more profound than the age effect.

Figure 2.1 also illustrates the personal differences between drivers who obtain their drivers licence at different ages. Although drivers who pass the driving test at 18 start with a high crash risk, when they reach middle-age they have the smallest crash risk of all drivers. The group of drivers who obtain their license at 30-40 years of age, never reach a crash risk as low as drivers who were younger when they passed the driving test.

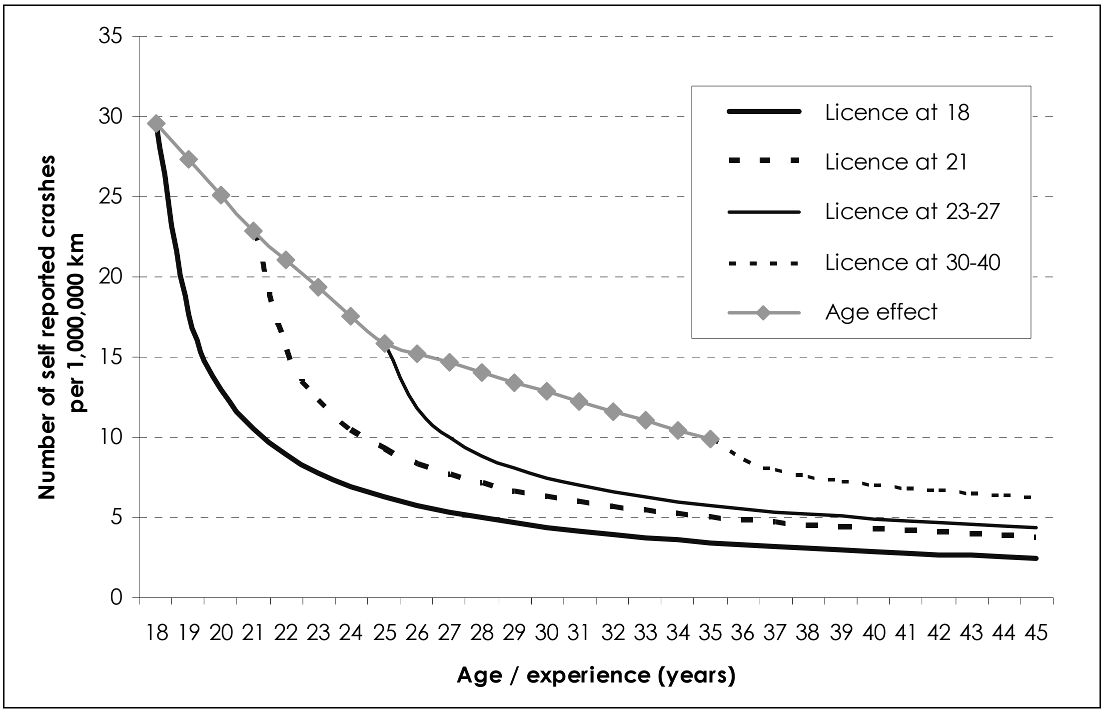

Figure 2.1. Decrease in crash risk for 18 year old novice drivers compared to other age groups. Source: PROV data 1990-2001; Vlakveld (2005)

Forsyth, Maycock and Sexton (1995) attempted to quantify the specific contribution of young age and inexperience on crash risk. Based on their survey of drivers at the end of their first, second, and third year of licensure, they report a 35-40% decrease in crash risk due to experience, for 17 year-old-drivers in the first year of driving. The reduction in crash risk due to age (i.e. from 17 to 18 years) was found to be 9%.

McCartt et al. (2009) reviewed eleven recent (1990 or newer) studies that tried to separate the effect of inexperience (length of licensure) and young age on crash risk (among which the studies by Vlakveld (2004), Maycock et al. (1991) and Forsyth et al. (1995)). The review excluded studies in which the age and experience factors were confounded, or where only the effect of age or only the effect of experience was studied. Based on the selected studies, the authors conclude that *both* inexperience and young age have an important, independent effect on crash risk. The age factor is primarily visible in the younger age groups (16-17 year old drivers), and there is strong evidence for a steep learning curve due to experience among drivers all ages.

To sum up, both young age and inexperience play an independent and important role in the high crash risk of young, novice drivers. Although it is difficult to separate the two factors, lack of driving experience seems to be a larger factor than young age. Therefore this thesis will focus on the development of driving experience.

2.1.3. Gender

Many crash risk studies report that young *male* drivers are much more likely to be involved in a serious crash than young *female* drivers (e.g. OECD - ECMT, 2006). This difference can partly be explained by the fact that males simply drive more than females do. Their exposure and therefore absolute number of crashes is higher. However, as illustrated by Figure 2.2, even when corrected for exposure, young male drivers have a much higher crash risk than their female counterparts (in the Netherlands).

Forsyth et al. (1995) have suggested that this is partly caused by the type of trips male and female drivers make. Females use their driver's licence mostly to get from one place to the next, while males spend more time driving simply for the sake of driving. Young male drivers typically drive more during leisure time, at night and with friends.

A Swedish crash study (Monárrez-Espino, Hasselberg & Laflamme, 2006) found differences in the types of crashes in which male and female drivers (aged 18-29 years) are involved. Male crash rate, in the first year as a licensed car driver, was five times higher than for females, but only for single vehicle crashes. There were no differences between males and females in crashes in which another motor vehicle was involved.

Research has also shown that male drivers are more frequently involved in risky driving, such as speeding and drinking and driving than female drivers. This was the result of a cohort-questionnaire study by Begg and Langley (2001), who also found that by the age of 26 many male drivers had 'matured out' of this behaviour. For female drivers aged 21 and 26 the risky driving and thrill-seeking was relatively low.

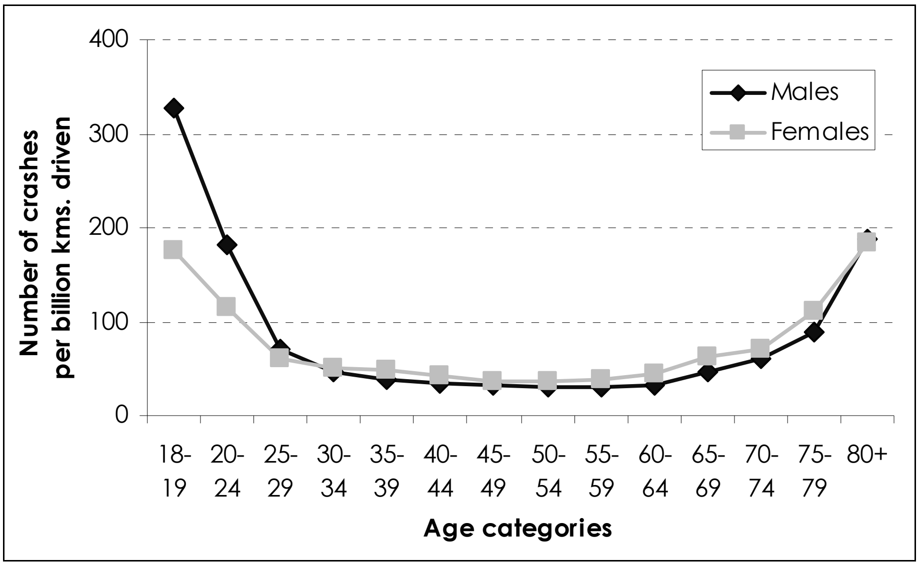

Figure 2.2. Number of crashes (fatal or with serious injuries) per billion motor-vehicle kilometers driven in the Netherlands in 1999-2007 for different age categories. Source: BRON (AVV); OVG (CBS until 2003); MON (AVV from 2004)[3]

In some studies, no difference between young male and young female drivers was found, when the data were corrected for exposure. For example, a study from Western Australia (Ryan, Legge & Rosman, 1998) suggested that, when taking mileage into account, there was no difference in crash risk for male and female drivers aged 17-24. Females tended to have fewer crashes than males, but also had less exposure in terms of kilometres driven per day.

Kweon and Kockelman (2003) analysed US crash records, and also found no substantial differences between the general crash rates for male and female drivers in the same age cohort when the data were adjusted for exposure.

[3] The estimated mileage is only available for the age group 18-24 year-old drivers as a whole, and not for 18-19 year-olds separately. There are indications that the 18-19 year-olds drive less than the 20-24 year-olds. Consequently, the crash risk for 18-19 year-olds presented in Figure 2.2 maybe an underestimation of the actual crash risk. This has no effect on the difference between males and females in this age group.

There are also differences between studies in the crash risk *development* of young male and female drivers. Crash records from the US (NHTSA, 2002), showed an increase in overall crash rate for young drivers (aged 15-20) between 1992 and 2002. For young males, driver fatalities rose by 15 percent, compared with a 42 percent increase for young female drivers. Skaar and Williams (2005) conclude, from the NHTSA crash records of 2002 to 2004, that adolescent and young adult females have become a critical cohort in the study of unsafe driving behaviour.

Twisk and Stacey (2007) found a different pattern in fatal crash risk development of young males and females in the Netherlands, Sweden and Great Britain. They compared the crash risk of young drivers to the crash risk of same-sex experienced drivers, over the period 1994 to 2001. They found that this *relative* risk of young male drivers is increasing, while the *relative* risk of young female drivers remains the same. The authors argue that young females seem to profit from traffic safety measures, whereas young males do not (compared to their experienced counterparts).

Figure 2.3 shows that, in the Netherlands, the crash risk (fatal or with serious injuries) for young male drivers is slightly decreasing since 1994, while the risk for female drivers remains the same. When considering that the overall crash risk in the Netherlands has decreased in previous years, it seems that the relative crash risk of young female drivers in the Netherlands is actually increasing. So, in contrast with the conclusions from Twisk and Stacey (2007) on development of fatal crashes, it seems that with respect to serious crashes in the Netherlands, young males seem to profit from traffic safety measures, whereas young females do not.

In conclusion, studies on the crash risk of young male and female drivers differ considerably. It is not quite clear if and why there is such a difference between young male and female drivers, or even if this difference is increasing or decreasing. In any case, there is no doubt that in the Netherlands young male drivers are (still) far more at risk to be involved in a crash than young female drivers. It seems that the same issues apply for young males and young females (the curve in Figure 2.2 is similar for both sexes), but that the magnitude of the problem is larger for young male drivers.

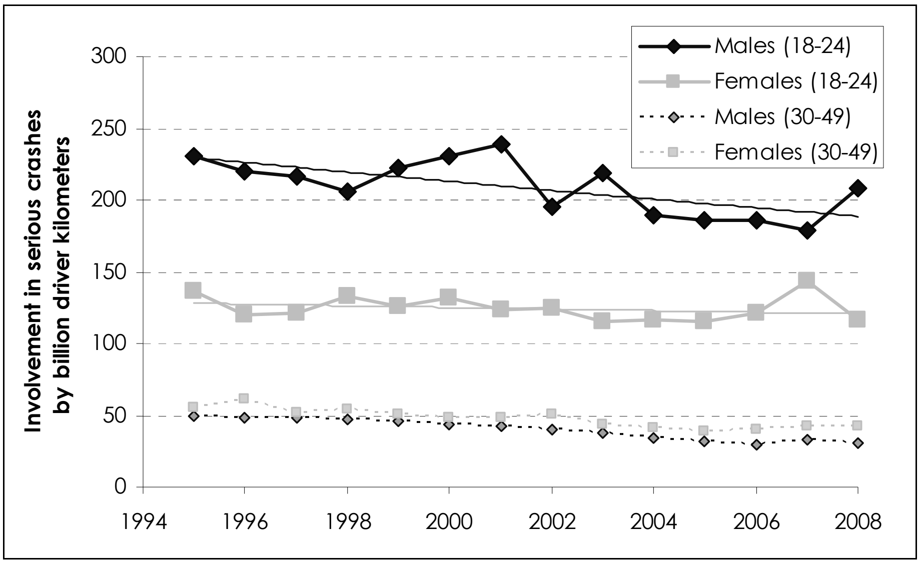

Figure 2.3. Number of crashes (fatal or with serious injuries) per billion motor-vehicle kilometers driven in the Netherlands from 1995-2008, for males and females aged 18-24 years and aged 30-49 years. Source: BRON (AVV); OVG (CBS until 2003); MON (AVV from 2004)

2.1.4. Conclusions: high risk of young novice drivers

The high crash risk of young novice drivers can be summarized in two elements, their young age and lack of experience. Although the two are both important and highly correlated, there are indications that lack of experience is a larger factor than young age. Therefore this thesis will focus on the role of experience in the decrease in crash risk. The following sections will discuss several models and theories that can explain the differences between experts and novices. Because the crash risk of young male drivers differs from the crash risk of young female drivers (in the Netherlands), this thesis will also focus on the factor gender.

2.2. Automation of driving subtasks

This section shows that the high risk of novice drivers has often been attributed to the limited automation of driving subtasks (Engström et al., 2003; Fuller, 2002a; Groeger, 2000; Rasmussen, 1986), which for many driving situations leads to a higher mental workload for novice drivers compared to experienced drivers (De Waard, 2002; Detweiler & Schneider, 1991; Patten, Kircher, Östlund, Nilsson & Svenson, 2006).

However, the driving task is 'self-paced' (Taylor, 1964); a driver can make every driving situation less demanding, for example, by reducing speed, increasing headway, or avoiding unnecessary distraction. Therefore, the next section (2.3) will introduce "motivational models" of driving behaviour, with a central role for the 'self-pacing' aspect of the driving task.

2.2.1. Automated processing versus controlled processing

The difference between automated and controlled processing can be understood with Norman's (1981) Activation-Trigger-Schema theory. According to this theory, every task performed by humans is represented by hierarchical, ordered *schemas*. For example: a visit to my grandmother ("Parent Scheme") contains a number of "Child Schemas" for dressing, leaving the house, driving my car, etc. Driving the car, in itself, contains the schemas: starting the car, accelerate, obeying traffic rules, navigating, etc. According to Norman each schema is 'triggered' for activation. For example, accelerate only happens after the car is started, not before. The completion of a task by using schemas is dependent on triggers provided by the situation, motivation, the presence of other competing schemas and strength of a schema as a result of frequent successful use.

The term *script* was introduced as a particular type of schema that describes the kind of knowledge that people can abstract from a common, frequently occurring event (Searleman & Herrmann, 1994). Scripts are not composed of memories for any one particular event, instead, they contain generic knowledge or memory about what usually happens. The benefit of having a script is that it allows a person to fill in missing details.

Shiffrin and Schneider (1977) differentiated two modes of information-processing. In the first, *automatic processing*, schemas are triggered automatically, without the necessity of active control or attention by the subject. Because many schemas can be effective at the same time this type of control is highly efficient. However, there is more room for error or slips as they are called; for example the activation of the wrong schema. For these situations Shiffrin and Schneider describe *controlled processing*. This mode is highly depended on feedback, when something goes wrong the system intervenes. The downside of this system is that it is a relatively slow system and requires much effort.

In order to activate the correct schema or script, a driver has to know what to expect in a certain situation. As a driver gains experience he develops expectancies on how traffic situations may evolve, which in turn increases anticipation (Van Elslande & Faucher-Alberton, 1997).

However, improved anticipation due to experience does not always have a positive effect on traffic safety (Houtenbos, 2008). Unjustified expectancy can have a major negative impact on traffic safety. Especially with looked-but-failed-to-see-errors, where car drivers looked in the direction of the other road user but did not see (or perceive) him, it appears that experienced road users are more likely to miss road users due to unjustified expectations (Herslund & Jørgensen, 2003).

But overall, experienced drivers utilize their expectation of the traffic environment to anticipate what is about to happen next. This gives them an advantage over novice drivers because it provides them with more time and space to decide and respond to the situation (Van der Hulst, 1999).

In the (holistic) theory of Situation Awareness (Endsley, 1995), the difference between novice and experienced drivers seems to boil down to a difference in (automated versus controlled) information processing and expectancy. Situation Awareness (SA) is defined as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1988, p. 789, as cited in Endsley, 2000).

Endsley (1995) describes efficient information processing, the deployment of attention, and high levels of automaticity as conditions for efficient SA. Gugerty and Tirre (2000) also found, in three experiments using a low-fidelity driving simulator, that SA ability is correlated (among others) with working memory ability. Finally, Bolstad and Hess (2000) conclude that experts perform better than novices, with respect to SA, because of: 1) their ability to draw on readily accessible knowledge structures to organize incoming information and formulate responses; 2) their use of operations with a high degree of automaticity; and 3) the fact that extensive practice often results in elimination of computational steps that may slow processing.

To conclude, with practise more and more (correct) schemas can be triggered with automatic processing, and driving a car becomes less effortful (Groeger, 2006). More specifically, a novice driver uses controlled processing to activate the correct schemas for shifting gear, for example. Much of his attention is focussed on this technical aspect of driving the car. As this driver gains more experience, the activation of correct schemas becomes a more and more automatic process. In other words, shifting gear does not require much active control or attention by the driver any more.

2.2.2. Mental workload

The human capacity for information processing is limited. Mental workload can be defined as the proportion of mental capacity actually required to perform a particular task (O'Donnell & Eggemeier, 1986), in other words, the amount of active control or attention that is needed for shifting gear. In all tasks, mental workload is determined by the interaction between the state or capability of the task performer (the driver) and the task (shifting gear) itself (De Waard, 2002). The more routine a task becomes (automatic processing), the less mental workload is required to perform it.

Mental workload imposed by a task can be measured objectively using a secondary task (Martens & Hoedemaeker, 2001; O'Donnell & Eggemeier, 1986; Wickens & Hollands, 2000). With this technique a participant performs a primary task (the driving task) and a secondary task. This secondary task can be a reaction time task, mental arithmetic or a memory search task. Presuming that the primary task performance stays level, the performance on the secondary task is assumed to be indicative of residual mental resources or capacity not utilized in the primary task.

The concept of mental workload has been used to describe some of the differences between novice and experienced drivers. In a field study, Shinar et al. (1998) compared novice and experienced drivers' performance in detecting road signs when driving cars with manual or automatic gears. The results showed that manual gear shifting significantly impaired sign detection performance of novice drivers. No such difference was found among experienced drivers. The authors concluded that gear shifting is a task that becomes automated over time.

Patten et al. (2006) used the secondary task method to explore the relationship between mental workload and driver experience. The main results showed a large and statistically significant difference in mental workload levels between experienced (professional) and inexperienced (regular) drivers. The authors conclude that (professional) drivers with better training and more experience are able to automate the driving task more effectively than their less experienced counterparts.

So, in general, experienced drivers endure less mental workload while driving, with the exception of elderly drivers (aged over 65), for whom driving leads to a greater mental workload compared to younger drivers (Cantin, Lavallière, Simoneau & Teasdale, 2009).

There are indications that novice drivers perform worse on two widely used tasks for research in traffic safety, namely visual search and hazard perception, because they have less mental capacity available for the execution of these tasks.

With respect to visual search, already in 1972, Mourant and Rockwell measured eye movements in traffic and discovered that novice drivers tend to look closer to the front of the car and less often in the rear-view mirror. Although Mourant and Rockwell based their results on the scanning behaviour of 6 novice and 4 experienced drivers, this result was replicated by Falkmer & Gregersen (2001) with 15 novice and 20 experienced drivers.

In addition, novices tend to fixate longer than experienced drivers, especially in dangerous situations (Chapman & Underwood, 1998); and use the same scanning pattern for all road types, where experienced drivers select visual strategies according to the complexity of the roadway (Crundall & Underwood, 1998).

Finally, the horizontal width of novice drivers' search patterns is less than that of experienced drivers (Underwood, Chapman, Bowden & Crundall, 2002).

In addition to visual search strategies, a lot of studies on the differences between experienced and novice drivers have focused on hazard perception skills (i.e. the ability to detect and respond to hazards). Several studies have found that experienced (and expert) drivers are better and faster in detecting hazards (Brown, 1997; McKenna & Crick, 1994; McKenna & Horswill, 1999; Whelan, Senserrick, Groeger, Triggs & Hosking, 2004). Some studies report a relationship between hazard perception skills and crash risk, in general and especially for inexperienced drivers (Congdon, 1999; ACER, 1999, as cited in Drummond, 2000; Pelz & Krupat, 1974).

Some authors argue that the deficiency in visual search and hazard perception of young novice drivers can be traced back to the lack of mental capacity for these activities (see OECD - ECMT, 2006). According to these authors, experienced drivers simply have spare capacity to look at more objects in the visual field.

The results of the previously mentioned fieldstudy by Shinar et al. (1998) supports this theory. Manual gear shifting significantly impaired sign detection performance of novice drivers, but did not impair the performance of experienced drivers. In other words, when novice drivers have to use more mental capacity on gear shifting, there is less capacity for visual search.

In addition, results of a study by McKenna and Farrand (1999) could indicate that novice drivers simply do not have spare mental capacity for hazard perception. In their study, McKenna and Farrand tested hazard perception skills of novice and experienced drivers, with and without a secondary task. They found that the secondary task interfered with hazard perception for both groups and that experienced drivers performed even worse than novice drivers on hazard perception when performing the secondary task. The authors concluded from these results that hazard perception itself is a task that cannot be automated; if it was, experienced drivers would not suffer from a secondary task this much. But you can also interpret these results that when experienced drivers have limited spare capacity (similar to novice drivers) they perform just as badly as novice drivers on a hazard perception test.

On the other hand, Underwood et al. (2002) concluded that the difference between novice and experienced drivers in visual search can probably *not* be attributed only to a lower level of automaticity and high mental workload for novice drivers. In their study novice and experienced drivers watched video-recordings taken from a car while their eye movements were recorded. With this procedure, the novice drivers did not have to control the vehicle, so all mental capacity was free to be used on scanning the road. The authors still found that novice drivers did not scan as much as experienced drivers.

It is most likely that both factors play a role. On the one hand visual search and hazard perception are skills that drivers learn and improve as they gain more experience. On the other hand, there is more mental capacity available for these activities as other driving tasks (shifting gear) can be executed more automatically.

In conclusion, with practice driving (sub)tasks become automated, leading to a decrease in mental workload. This makes the driving task easier for experienced drivers compared to novice drivers, and may even, in addition, be responsible for the fact that experienced drivers perform better at visual search strategies and hazard perception.

2.2.3. Hierarchical control models

Several authors have used a "hierarchical control model" (Ranney, 1994) to describe task performance in general or the driving task in particular. All models distinguish several levels of control, ranging from more automatic control (low level) to a higher level of control which needs more attention or mental capacity, and describe how experienced and novice drivers differ in the level at which they perform their tasks such as driving.

Anderson (1982) describes two stages in the acquisition of (cognitive) skills, these are: “the declarative stage” and “the procedural stage”. In the *declarative stage* the learner (driver) is first introduced to the new skill; performance is relatively unstable, as possible strategies are tested and rejected. Verbal mediation is often observed, because the learner needs to rehearse the facts in their working memory. In the case of distraction, task performance deteriorates considerably. After enough practice and experience, the learner reaches *the procedural stage*, at which verbal mediation does not exist, and the task performance is highly consistent and requires almost no effort. *Knowledge compilation* is the process of the skills transferring from the declarative stage to the procedural stage.

Michon (1985) distinguishes three levels of cognitive control of driving: Strategic (planning) level[4], Tactical (manoeuvring) level, and Operational (control) level. The *strategic level* includes the general planning of a trip, including the trip goals and route choice, plus an evaluation of the costs and risks involved. Within the boundaries set by the choices on the strategic level, at the *tactical level*, a driver performs manoeuvres such as obstacle avoidance, gap acceptance, turning and overtaking. The *operational level* is the lowest level in Michon’s hierarchy, and deals with vehicle handling such as the control of the vehicle on the road, steering, and shifting gears.

Finally, Rasmussen (1986) distinguishes task performance on three levels; a skill based level, a rule-based level and a knowledge-based level. *Skill-based performance* is routine, usually without feedback and corresponds with Shiffrin and Schneider’s (1977) automatic processing. *Rule-based performance* is guided by a general rule, which is applicable in several situations. Usually the specific rule is fairly simple (e.g. giving traffic from the right, right of way), but the situation in which the rule is applied is unique. Therefore, rule-based performance is partly automatic (execution of the rule) and partly conscious (applying the rule to the specific situation). For novel situations, *knowledge-based performance* is used. Reference must be made to a broader representation of knowledge or basic principles to formulate an appropriate solution for the new problem.

---

[4] In his article Michon refers to this level as the Strategic*al* level

Traffic crashes are often viewed as the result of human error (Carsten, 2002). In his Generic Error Modelling System (GEMS), Reason (1990) used Rasmussen's skill-rule-knowledge classification to describe how different types of errors are typically connected to different hierarchical control levels (see Table 2.1). For example, slips (execution of the wrong script) and lapses (omission to execute a script) are associated with an automated action. This type of error may be linked to a reduction in the efficiency of task performance (mental workload) or to an allocation of resources to other tasks (attention) (Carbonell & Martín-del-Río, 2002).

Table 2.1. Relationship between Reason's error types and Rasmussen's performance levels (Reason, 1990)

Performance level	Error type
Skill-based level	Slips and lapses
Rule based level	Rule based mistakes
Knowledge based level	Knowledge based mistakes

More recently, however, there has been a shift in (traffic) safety research from "human error" to "error of the system" (Dekker, 2005; Wegman & Aarts, 2006). Within this view human error is seen as the result of a failing system, rather than the cause for the system to fail. It is the challenge to change the system to prevent errors or at least minimize the consequences of an error.

Nevertheless, the original classification of human errors can be used to describe the differences between novice and experienced drivers. Because different types of errors correspond to task performance on different levels, novice drivers should, in theory, make different types of errors than experienced drivers. Research with the Driver Behaviour Questionnaire (DBQ - Reason, Manstead, Stradling, Baxter & Campbell, 1990), which was developed to measure the different error types of Reason's classification, shows different results. For example, Parker, Reason, Manstead and Stradling (1995) found that neither age nor mileage could predict self-reported errors.

On the other hand, Ozkan, Lajunen and Summala (2006), also administered the DBQ and did find that young drivers reported more mistakes especially for turning, steering, or overtaking. In addition, Åberg and Rimmö (1998), who added new error items to the DBQ, found that there are some slips and lapses that increase with age and experience of the drivers. Especially the factor of inattention errors is seen as an indication that driving errors may be a result of automation. So, there are indications that novice drivers make different types of errors because they perform (some of) the driving tasks on a different hierarchical level.

In conclusion, the classifications into different levels of the driving task provide an explanation for the importance of experience in traffic. On the one hand, experienced drivers have more schemas that can be executed on the lowest (skill-based) level (Van Leyden Sr., 1993). Therefore the driving task can be executed more efficiently and with less effort.

On the other hand, on a knowledge-based level, knowledge about traffic grows with experience so that experts have acces to a relatively extensive knowledge-base compared to novices. This can explain why novice drivers are more likely to produce a higher proportion of wrong 'solutions' when faced with a novel situation than experienced drivers (Fuller, 2002a).

This hierarchical approach has been used to structure the contents and goals of driver education. The "Goals for Driver Education" or GDE framework (Hatakka, Keskinen, Gregersen, Glad & Hernetkoski, 2002; Siegrist, 1999), revealed that currently driver education and testing in Europe is primarily focused on lower levels of the hierarchy, while the benefits of being an experienced driver lies mostly in task performance on a higher level.

2.2.4. Conclusions: automation of driving subtasks

The high risk of novice drivers can, for a major part, be attributed to the limited automation of driving subtasks, which leads to a higher mental workload for novice drivers compared to experienced drivers. The more routine a task becomes (automatic controlled processing), the less mental workload is required to perform it. Therefore the driving task can be executed more efficiently and with less effort.

This higher mental workload can even explain, according to some authors, why novice drivers perform worse at visual search, hazard perception, and Situation Awareness. Because the human capacity for information-processing is limited and novice drivers have not automated most parts of the driving task yet, there is less mental capacity for efficient visual search, hazard perception or Situation Awareness.

Finally, on a higher level of task performance, knowledge about traffic grows with experience. Therefore novice drivers are more likely than experienced drivers to produce a wrong 'solution' when faced with a novel situation.

However, limited automation of driving subtasks cannot completely explain the high crash risk of inexperienced drivers. This was also indicated by a study by McKnight and McKnight (2003) who conducted a crash analysis of more than 2000 crashes involving 16-19-year-old drivers. Although they found that the great majority of non-fatal crashes resulted from errors in attention, visual search and hazard perception, none of these factors contributed enough to explain the sharp decline in crash risk in the first years of independent driving.

Another factor needs to be taken into account, which is that driving is a 'self-paced' task (Taylor, 1964). A driver can adjust the task demands, for example by changing speed or headway (Van der Hulst, 1999), and avoiding unnecessary distraction (e.g. Summala, 2002). Thus, in theory, novice drivers can decrease task demands to fit their (deficient) level of automated driving. The next section describes several "Motivational models" of driving, which have incorporated the 'self-pacing' aspect of the driving task.

2.3. Motivational models of driving

This section will introduce some motivational models (see Ranney, 1994, for the classification of different types of models of driving). Motivational models incorporate the self-pacing aspect of the driving task, i.e. a driver can make the driving task easier, for example, by reducing speed or increasing headway. After a general description, the motivational aspects of human error (i.e. deliberate violations) are described, and finally two specific models (Brown's model of subjective safety and Fuller's task-capability interface model) are discussed in more detail.

2.3.1. General description

Wilde's (1982) Risk homeostasis theory, is based on the assumption that every driver has a relatively stable level of accepted subjective risk. A driver will change behaviour (e.g. drive faster or slow down) in order to maintain this constant risk level. The Risk homeostasis theory can explain why certain traffic safety improvements did not have the anticipated effect; because drivers also change their behaviour in an undesirable direction to maintain the same level of risk (Wilde, 1988).

Näätänen and Summala (1974; 1976) introduced the concept of a subjective risk-monitor. According to their Zero-risk theory (see also Summala & Näätänen, 1988), drivers do not constantly experience risk and are therefore not constantly adjusting their behaviour. Instead, only when perceived risk exceeds a certain threshold, a driver starts to feel uncomfortable and will compensate by changing his behaviour (fear avoidance). As a result of experience, this fear threshold shifts and a driver maintains too short safety margins and performs objectively dangerous manoeuvres.

Finally, in the risk-avoidance model, Fuller (1984; 1988) describes the driving task as constantly avoiding threat. However, the response to perceived threat is influenced by the rewards and punishments associated with each kind of alternative response. In other words, a driver may opt to experience some risk/threat if the reward for this behaviour is large enough. In contrast with Wilde and Näätänen and Summala, Fuller argues that drivers are not capable to monitor the probability of a rare event such as a crash. Instead, the subjective probability should not refer to the probability of of a crash, but to the likelihood of some *potential* aversive stimulus or threat.

2.3.2. Violations

In Section 2.2.3, Reason's (1990) model for human error was introduced. In addition to this error classification, Reason also describes *violations* which have a motivational and contextual cause. Violations are defined as "deliberate – but not necessarily reprehensible – deviations from those practices deemed necessary […] to maintain the safe operation of a potentially hazardous system" (Reason, 1990, p. 195).

Research with the DBQ, which measures self-reported errors and violations in traffic, shows that males and young drivers report more violations than females and more experienced drivers (Mesken, Lajunen & Summala, 2002; Ozkan et al., 2006; Parker, Reason et al., 1995). In addition, research with the DBQ has indicated that the tendency to commit driving violations is correlated with crash liability (Parker, West, Stradling & Manstead, 1995; Underwood, Chapman, Wright & Crundall, 1997).

2.3.3. Brown's 'model of subjective safety'

Brown (1989) describes driving as a self-paced task, in which drivers meet personal criteria of safety by attempting to match perceived hazards in traffic to their perceived abilities to cope with those hazards (see also Brown & Groeger, 1988). In Brown's Model of subjective safety (Figure 2.4) there is a central role for the balancing ("Comparator") of perceived hazards and the drivers coping ability. In order to do so, a driver has to make an *assessment* of his abilities and identification of road and traffic hazards.

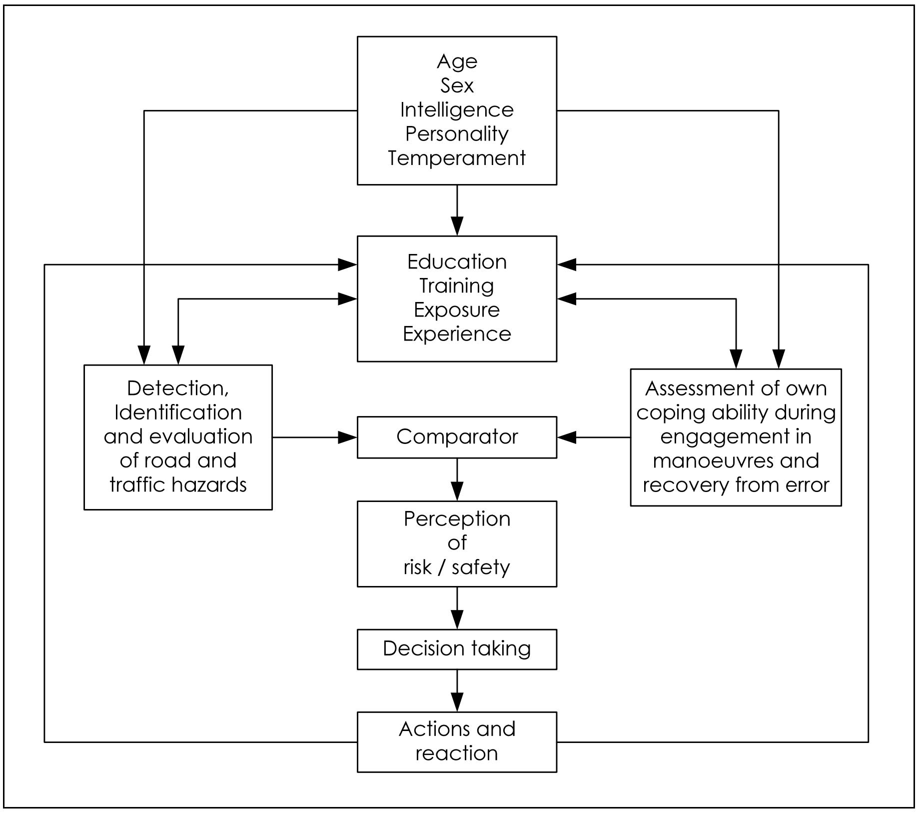

Figure 2.4. Brown's (1989) "Model of Subjective Safety"

In the description of his model, Brown focuses on the tactical aspects of the driving task (i.e. real-time interactions with the road and traffic). So, the result of the equation ("Actions and reaction") could be the change of speed or safety margins (vehicle control). This process described by Brown is therefore situation specific.

2.3.4. Fuller's task-capability interface model

Fuller introduced a model that is not situation specific. The initial task capability interface model (see Figure 2.5) was described in Fuller (2000), later versions can be found in Fuller (2002b; 2005). The most important aspect of this model is that the difficulty of a task is the product of the interface between the demands of the driving task and the capability of the drivers. Where capability exceeds demand, the task is easy; where capability equals demand the driver is operating at the limits of his/her capability and the task is very difficult. Where demand exceeds capability, the task is by definition too difficult which could result in a collision.

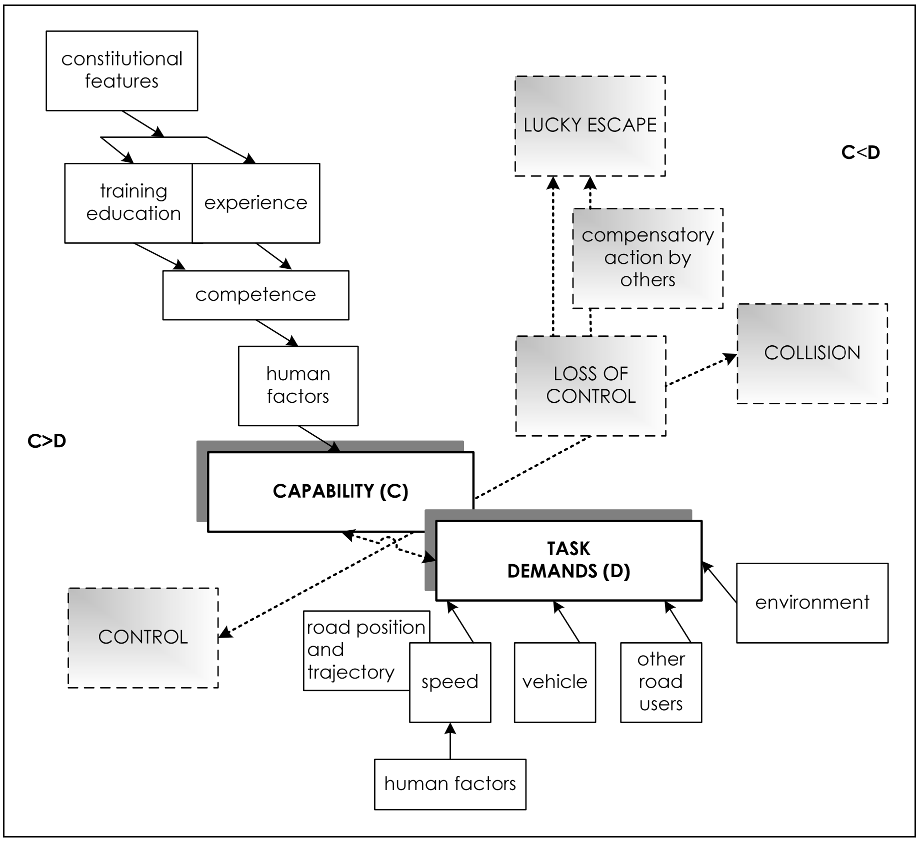

Figure 2.5. The task-capability interface model (Fuller, 2005)

Fuller includes the hierarchical nature of the driving task in his model. A driver can influence the task demands on a 'high level', for example by selecting a particular route (avoiding a difficult junction); or on a 'low level', for example changing speed (Fuller, 2005). Fuller's task-capability interface model is therefore not situation specific. Recently, Fuller (2008) has adapted his model into the task-difficulty homeostasis model. This new model focuses more on explaining drivers' choice of speed, in other words, only emphasises adaptation on a 'low level'.

2.3.5. Conclusions Motivational models of driving

According to the Motivational models a driver can make the driving task easier, for example, by reducing speed or increasing headway. Two Motivational models were described in more detail, Brown's (1989) 'model for subjective safety', and Fuller's (2005) task-capability interface model.

An asset of Fuller's model is that he incorporates the hierarchical nature of the driving task. A driver can make the driving task easier on different levels (a high level: choosing a safe route, or on a low level: decreasing speed). However, the aspect that is somewhat neglected in

Fuller's model is that it is crucial for a driver to make the correct *assessment* of the task demands or his capabilities. Otherwise he will never see the need to choose a safe route, or safe driving speed. Although later versions of Fuller's model do incorporate the assessment aspect (see e.g. Fuller, 2007), these later versions are more focussed on speed choice and trajectory (i.e. adaptation only on a low level of the driving task) than the task-capability interface model from 2005.

The aspect of *assessing* traffic hazards and coping abilities is a prominent element in Brown's model. It seems that a combination of Brown's and Fuller's model could be a good model to describe safe driving and help in explaining the high crash risk of young novice drivers.

To sum up, the fact that young drivers' performance improves with practice may be due to increased automation in combination with improved *adaptation to task demands* (i.e. making the driving task easier). In order to do so, a driver will have to make a correct *assessment of his skills* and the *situation's complexity*.

The balancing of task demands and skills has been referred to as calibration (Fuller, 2008; Kuiken & Twisk, 2001; Mitsopoulos et al., 2006). The next section introduces a new model for the calibration process, which combines elements of Brown's (1989) model of subjective safety and Fuller's (2005) task-capability interface model and incorporates the following three elements: self-assessment of skills, perceived complexity of the traffic situation and adaptation to task demands.

2.4. Calibration

2.4.1. What is calibration?

Figure 2.6 shows a new model of the calibration process, which was inspired by Brown's (1989) model of subjective safety and Fuller's (2005) task-capability interface model. The model is not meant to provide an accurate description of reality, but rather to give a simplified illustration of the elements of calibration and how they are related to each other. In words the calibration model describes that:

> *For safe driving, a driver needs to assess his own driving skills (A), weigh them against his perception of the complexity of the situation (B), and, as a final step, use the result of this balancing to adapt to task demands (C).*

In a psychological context, the process of balancing the assessment of skills and task demands has been called *calibration* (Deery, 1999; Kuiken & Twisk, 2001; Mitsopoulos et al., 2006) and is assumed to be related to the high crash risk of young novice drivers (Brown & Groeger, 1988; Brown, Groeger & Biehl, 1987; Gregersen, 1995; Mayhew & Simpson, 1995). In terms of the elements of the calibration model, young novice drivers are assumed to overestimate their skills, underestimate the complexity of the situation and, as a result they adapt insufficiently to task demands. However, clear empirical evidence for the relationship between calibration and the high crash risk of young novice drivers is missing.

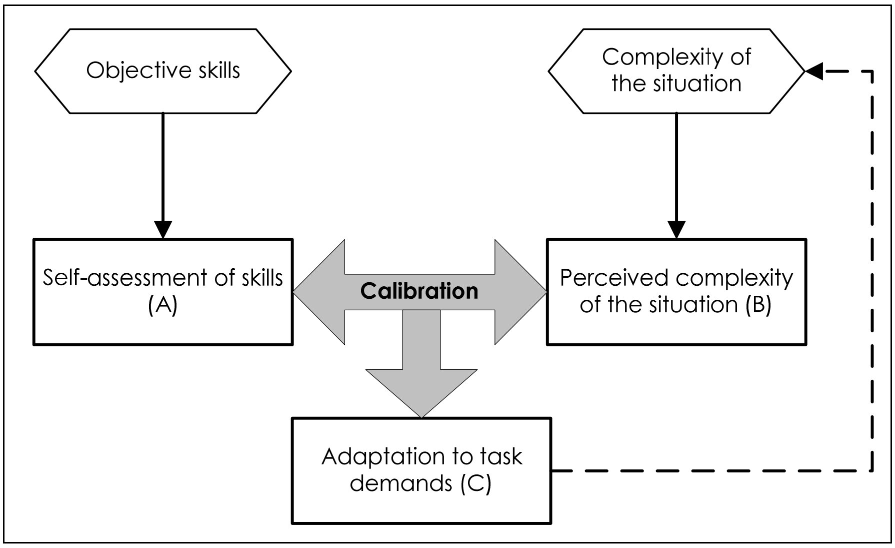

Figure 2.6. Model of the calibration process

The dotted line in Figure 2.6 indicates that when a driver adapts to the task demands of the specific situation, he changes the (perceived) complexity of this situation. This loop illustrates that calibration is an ongoing process. However, in some cases the complexity of the situation cannot be changed, for example, with extreme weather conditions, in these cases it may be best to decide not to drive at all (which in fact is rigorous adaptation to task demands).

A crucial element in the calibration model is that the balancing and adaptation can take place on different hierarchical levels of the driving task. For example on a high level, a driver can choose not to drive in the dark, because he believes this is too difficult, considering his skill level. On a lower level a driver can choose to reduce speed when he encounters a difficult situation. Similar to what was described in the hierarchical control models, adaptation on a low level can be executed with (almost) automatic control;

whereas adaptation on a higher level needs more attention or mental capacity.

In this section the three basic concepts of the model for calibration, self-assessment of skills, perceived complexity and adaptation to task demands will be discussed.

2.4.2. Self-assessment of skills

Self-assessment of skills can go either way, underestimation or overestimation of driving skills. Generally most attention is spent on overestimation of driving skills, being the most dangerous form of false self-assessment, and is believed to be related to the high crash risk of young, novice drivers (Gregersen, 1996; Mathews & Moran, 1986).

Traditionally, overestimation of skill, also called: the optimism bias (Deery, 1999; Svenson, 1981; Weinstein & Lyon, 1999) or self-enhancement bias (Brown, 1986; Walton, 1999), has been investigated by simply asking drivers to compare their skills with the 'average driver' in a questionnaire. Results show that, in general, drivers rate themselves to be better drivers than the average driver (Delhomme, 1996; McCormick, Walkey & Green, 1986; McKenna, Stanier & Lewis, 1991). More specifically, novice drivers are assumed to have an even poorer perception of their actual ability, they overestimate their ability more than experienced drivers (OECD - ECMT, 2006). In some studies (Nyberg & Gregersen, 2007; Tronsmoen, 2008), male drivers were found to overestimate their driving skills more than female drivers.

In a questionnaire study, McKenna et al. (1991) concluded that this overestimation is a result of 'positive self' rather than 'negative other'. That is, drivers view themselves better than average because they are 'good' drivers rather than that other drivers are 'bad'. In another questionnaire study, McKenna (1993) studied the perception whether overestimation was a result of imagined control (i.e. belief in one's ability) or high optimism (i.e. 'luck'). McKenna concluded that there is "clear evidence in favour of the illusion of control with no evidence in favour of unrealistic optimism" (p. 39).

However, not all studies into overestimation of skills reach the same conclusion. For example, DeJoy (1992), Nyberg and Gregersen (2007), and Tronsmoen (2008) all concluded, based on a questionnaire study, that male drivers overestimated their driving skills more than female drivers. That is, young male drivers assessed their own driving ability on a higher level than the young females did. In contrast, Mynttinen, Sundström, Koivukoski et al. (2009) and Mynttinen, Sundström, Vissers et al. (2009) recently compared drivers' self-assessments with their performance on the driving test, and concluded that males are *not* overconfident to a greater extent than females.

There is also disagreement as to whether novice drivers are more overconfident than experienced drivers. Mayhew and Simpson (1995) provide an extensive overview of studies into the assessment of skill. They have found studies indicating that young drivers are especially overconfident. However, there were also studies indicating that young drivers do not differ from older drivers in self-assessment, and studies that have shown young drivers are overconfident, but not in all driving situations. Finally, there were studies showing that young drivers express less overconfidence than older drivers.

More recently, Waylen, Horswill, Alexander and McKenna (2004) found that expert police drivers overestimate their skill to the same degree as novices, when they were asked to compare their skills to the 'average' and 'peer' driver. However, it must be said that the 'novice' drivers in this study had their drivers licence for 9.7 years on average and their mean age was 28.2 years. Matsuura (2005) asked male drivers to assess their driving compared to the average driver, and found that overconfidence existed in 'safety-oriented driving', but not in 'Skilful driving'. Horswill, Waylen and Tofield (2004) found that the overestimation bias (drivers rating themselves superior to their peers *and* the average driver) was greater for hazard perception skills than for either vehicle control skills or driving skill in general. Renge (1998) found that when subjects were asked to rate the confidence that they could cope with a specific situation (on video), more experienced subjects, especially driving instructors, were more confident in safe driving than less experienced subjects, and male subjects were more confident that females.

And finally, in a major questionnaire study that was conducted in the UK (Grayson & Elliott, 2004), one of the questions concerned self-reported confidence in driving ability. At four points in time (2 weeks, 6 months, 12 months and 24 months after passing the driving test) respondents indicate whether they felt 'very confident', 'fairly confident', 'not very confident', or 'not at all confident' in their driving ability. The results indicate that immediately after passing their exam drivers felt very confident in their driving skills. In the first year of driving, participants developed a more realistic view of the demands of traffic and their own skill. Six months and twelve months after passing the driving test, their confidence level is decreased by a statistically significant amount. After a year of driving experience the confidence level starts to increase again, but it never reaches the high level that was measured right after testing.

Most of the disagreement in the studies can be explained by the way the drivers were asked to assess their skills (see also Sundström, 2008b, for a review on the measurement of self-assessment). Sometimes drivers had to compare themselves with 'the average driver', sometimes they had to compare themselves with 'someone from their peers'. The respondents are not only asked to make an assessment of their own driving skill, but they also have to make an assessment of the driving skill of an 'average driver' or a 'peer'. But who is the 'average driver'? As Groeger (2000) points out, ‘average’ may be a negative rather than neutral descriptor. Traffic would be very unsafe if the average driver is only a mediocre driver.

Besides this theoretical problem, the interpretation of results is also very different. If a certain group expresses more confidence in their own skills, it is easily concluded that this group is over-confident with respect to these skills. But what if their skill level is in fact much higher than the comparison group? Some authors go even further, Mathews and Moran (1986), show results indicating that a group of young, novice drivers express about the same confidence in comparison with the average driver as an older, more experienced group. However, the authors reason, as the group of young, novice drivers have a higher crash risk, their expressed confidence is further from the truth that the confidence of the experienced group.

The solution to these difficulties with interpretation can easily be resolved, by letting the drivers make an assessment of their own driving skills and compare this with their actual driving skills. In domains other than driving, self-assessment of skills has been related to a more objective measure of the skill, rather than group average. Novices have been found to overestimate their skills in chess, people with more knowledge about chess assess their ability to remember chess material more accurately (Chi, 1978); physics, experts were more accurate at judging the difficulty of a problem than novices (Chi, Glaser & Rees, 1982); and tennis, novices are less likely than experts to successfully gauge whether specific play attempts were successful (McPherson & Thomas, 1989). More recently, Kruger and Dunning (1999) found that across four studies, participants scoring in the bottom quartile on tests of humour, grammar, and logic reasoning grossly overestimated their test performance and ability.

In some studies concerning driving skills, self-assessment was also compared to an independent measure. For example Matsuura (2005), compared the drivers' ratings of their own skill with observed driving skill. He found that young groups were more overconfident than older groups. But when the groups were divided into experience level, experienced drivers were more overconfident than inexperienced drivers.

Gregersen (1996) conducted an experiment in which novice drivers' ratings of their skill were compared to observed skill. The participants in this study were asked to estimate how many trials, out of five trials of braking and avoidance on a skid track, they believed they could manage correctly at 70 km/h. The estimation process was followed by the second task, in which the drivers were told to drive and the actual numbers of failures and successes were counted. In this study there was no distinction between experienced and novice drivers, but the novice drivers were divided into two groups, one which received a training to improve 'skills', the second group received a training focused on improving 'insight'. The results showed that this last group was better at assessing how many times they could successfully perform the brake/avoidance manoeuvre on the skid track. Gregersen therefore concludes that training novice drivers to improve skills produces more false overestimation than training insight. Gregersen did not compare drivers with different experience levels. However, because driving education and the driving test is still very skill oriented (in the Netherlands) his study could indicate that drivers who just passed their driving test overestimate their skills.

Delhomme and Meyer (2000) also compared drivers' expectation on how many cones they would knock over on a track with their actual performance on this track. The subjects in this study were all young male drivers divided into two groups (less and more experienced drivers) on the basis of a median split of the kilometres they had driven since they obtained their driving license. Each driver performed two experimental trials: one took place in normal visibility and the other in reduced visibility. Visibility was reduced by means of a stiff translucent plastic helmet. This was designed to make the driving task more difficult, emulating the situation of driving in fog. The results showed that the less experienced drivers had higher levels of expected performance on the task, even when visibility was reduced, indicating that these less experienced drivers were more overconfident about their driving skills.

Finally, Horrey, Lesch and Garabet (2008; 2009) conducted a series of experiments in which they compared drivers' expectation about their performance while driving around a course in an instrumented vehicle, with their actual performance. During the ride, participants were asked to complete a series of tasks on a hand-held or hands-free cell phone. The results showed that drivers were generally not able to assess the magnitude of the distraction effects. In one specific manoeuvre, in which the driver had to stop the car in time, drivers who reported the smallest estimates of distraction, showed the largest performance deficits (Horrey et al., 2008). More specifically, Horrey et al. found that young male drivers who thought they performed better in the stopping task actually performed worse than others. No such difference between estimated and actual performance loss was found for the older male drivers or the female drivers.

To conclude, studies on the subject of self-assessment of driving skills do not all reach the same conclusion. Some studies conclude that young novice drivers are especially overconfident, whereas other studies report no difference between young novice and older experienced drivers. This can (partly) be explained by different approaches used; a) comparison of a drivers' assessment with the group average, or b) comparison of the assessment of each driver to some more independent measure of skills. One of the objectives of this thesis is therefore to study if novice drivers indeed overestimate their driving skills more than experienced drivers (which is one of the predictions of the calibration model; Figure 2.6).

2.4.3. Perceived complexity

Hazard perception versus risk perception

Generally, a distinction can be made between hazard perception and risk perception. *Hazard perception* is seen as the ability to detect hazards. This skill is usually measured with reaction times needed to observe a potentially hazardous situation. *Risk perception* is generally described as a more subjective appraisal of risk.

Some study results also indicate that there is a difference between hazard perception and risk perception. For example, Sagberg and Bjørnskau (2006) used a video-based hazard perception test, to test for differences between three groups of drivers, having held a licence for 1, 5, and 9 months, and a group of drivers who had held their licences for several years. Sagberg and Bjørnskau found little evidence for improved hazard perception to explain the rapid decrease in young, novice driver crash risk over the first several months of driving. However, Renge (1998) also used a video-based hazard perception test, and found that hazard perception improved with experience, when he compared performance of novice, experienced drivers and driving instructors. The difference between these studies maybe caused by the fact that Sagberg and Bjørnskau measured reaction times (i.e. hazard perception), while the participants in Renge's study had to point out potentially dangerous situations without a time constraint (i.e. risk perception).

This is consistent with results of a study by Anders, Heustegge, Skotte, Müsseler and Debus (2006), in which they divided reaction time into two sub-processes. Novices tend to be slightly faster in detecting possible dangers than experts, when using an eye-tracking device (i.e. hazard perception). On the other hand, experts are faster in processing a potentially dangerous object; they have a faster reaction time measured from the time they spot the danger and react to it (i.e. risk perception). Although this difference was not statistically significant in the tested sample, it appeared consistently in high and medium dangerous situations.

With respect to the calibration model both hazard perception and risk perception apply. On a low hierarchical level, the perceived complexity can be seen as hazard perception, for example an intersection with unclear view is regarded as a potential hazard. On a higher hierarchical level, the perceived complexity in the calibration model responds more to risk perception, for example, the perception of the risks involved with driving in the dark.

Relationship between perception of complexity and self-assessment of skills

It must be stressed that in the calibration model *perceived complexity* includes the appraisal of danger or risk, and therefore cannot be separated from the *self-assessment of skills*. That is, the appraisal of the danger or risk of a situation is always related to how drivers view themselves and their skills. For example: "Yes, driving under the influence of alcohol is very dangerous. But, I am such a good driver, and I am used to drinking a lot, so a few drinks would not affect my driving".

This has also been the case for traditional hazard perception tests. The result of a hazard perception test is a combination of the perception of the hazard, but also how a driver believes he can cope with this hazard. Where hazard perception tests mostly measure the perceived complexity of a situation, this thesis will focus more on the self-assessment of skills aspect. In other words, the same problem will be studied, but from a different point-of-view.

The connection between perceived complexity and self-assessment was studied by Farrand and McKenna (2001), who showed traffic situations to their respondents. They found a significant inverse relationship between two risk perception ratings: a) "How would you estimate the risk in this situation" and b) "Compared to the average driver, how capable do you think you are of dealing safely with the risks in this situation". This result indicates that a high estimate of ability to deal with the risks was associated with a judgement that the scenario was less risky.

There is also evidence that both self-assessment of skills *and* perception of complexity play a role in safe driving. According to McKenna and Horswill (1999), it is a common finding that females rate hazards as more risky than males (Soliday & Allen, 1972; as cited in McKenna & Horswill, 1999). However, it has also been found that there is no difference between men and women in the speed with which they detect hazards (if they are matched for experience; McKenna, Waylen & Burkes, 1998). A mechanism such as calibration could explain these results. Hazard perception skills are the same for male and female drivers, but, because females are less confident about their driving skills they attribute the hazard as a greater risk. In other words, calibration describes more than hazard perception alone.

2.4.4. Adaptation to task demands

An important aspect of the calibration model displayed in Figure 2.6 is that the balancing of self-assessment and perceived complexity results in adaptation to task demands. It is assumed that, for example, overestimation of skills is a negative factor, because it results in less or insufficient adaptation to task demands. A driver who thinks of himself to be a very skilled driver is less likely to reduce his speed when it starts to rain.

A study by Quimby and Watts (1981) suggest that there is a relationship between risk perception and adaptation of speed to the situation. Their analysis showed a significant correlation between the subjects' median risk ratings on a test drive and their safety index (calculated from stopping distance and the forward visibility at locations during the test drive). This correlation indicates that those drivers whose speeds resulted in the greatest risk taking tended to consider the danger or risk to be low.

In contrast, Horswill et al. (2004) found little evidence for a relationship between self-assessment and behavioural adaptation. They did find that drivers rated themselves superior to their peers *and* the average driver. But this did not affect three risk taking intentions (photographic speed, questionnaire speed, photographic animation of car-following behaviour), whether compared individually or on aggregate.

There are indications that young, novice drivers, as a group, do not compensate for their (deficient) level of automated driving by decreasing task demands (Twisk, 1995). They tend to drive with too small safety margins (Engström et al., 2003) and are more likely, compared to other age groups, to engage in secondary behaviours (e.g. making a telephone call) while driving (Sayer et al., 2005).

2.5. Conclusions

This chapter described the crash risk of young novice drivers and the importance of lack of experience as an explanation compared to drivers' young age. There are many indications that with practice parts of the driving task develop (quickly) towards routine, i.e. driving (sub)tasks become automated, leading to a decrease in mental workload. However, lack of automation does not explain the high crash risk completely. Since the driving task is 'self-paced', the driver can adjust the task demands (e.g. by reducing speed or increasing following distance), thus decreasing workload. In theory, a novice driver can use this strategy to overcome the limitations of his performance; a novice driver can decrease the task demands to fit his (deficient) level of automated driving.

This self-pacing aspect of the driving task is described by motivational models. Two of these motivational models (Brown's (1989) model of subjective safety and Fuller's (2005) task-capability interface model) were incorporated into the calibration model, which describes that:

> *For safe driving, a driver needs to assess his own driving skills (A), weigh them against his perception of the complexity of the situation (B), and, as a final step, use the result of this balancing to adapt to task demands (C).*

As Deery (1999) suggested in his literature review, further research into calibration is needed: "… research has revealed that people's perception of risk in traffic hazards is predictive of their driving record. In this research, however, it is generally not possible to separate the relative contributions of young novice drivers' assessment of danger posed by the hazards and their self-assessed ability to deal with those hazards effectively" (p. 230).

Deery further reasons: "Although it is often said that experienced drivers are well-calibrated and that young novice drivers, particularly young males, are miscalibrated or overconfident, there is little direct scientific evidence to support this contention" (p. 232).

This thesis will empirically, based on the model in Figure 2.6, study the existence of calibration and its separate elements. In addition, it will analyse the development of calibration over time and/or with experience.

3. General method

The current chapter describes the methods that were used to study (the different elements of) the calibration model defined in Chapter 2. The results of Chapter 4 to 7 are based on the methods described here.

This chapter provides a description of the participants selected for the study (Section 3.1) and explains the use of a longitudinal design for the study, including the threats of such a design and drop-out rates (Section 3.2).

Furthermore, the chapter describes the instruments that were used: the questionnaire (Section 0), driving diary (Section 3.4) and on-road driving assessment (Section 3.5). Section (3.6) focuses on general issues concerning the validity and reliability of a driving assessment to measure 'ability to drive safely'. Section 3.6 also describes a small scale experiment to study whether the assessment of observers is influenced by the appearance of a driver.

Finally, the chapter describes the analyses that were conducted in the study and how missing data were handled (Section 3.7).

3.1. Participants

3.1.1. Selection

Two groups of drivers, novice and experienced drivers, were selected for this study. In collaboration with the CBR, the Dutch Driving Test Organisation, all drivers under the age of 25 who had passed their driving test during the last two weeks in September 2005, in the cities of Rotterdam, Rijswijk or Stellendam, were invited to participate in this project. Invitations were made in person, immediately after passing their driving test. During these two weeks 553 drivers were approached of whom 509 agreed to participate. This resulted in a response rate of 92% for the novice drivers.

The RDW, the Dutch Vehicle Technology and Information Centre, provided the names and contact information of 999 experienced drivers who held their licence for more than 10 years, and were not older than 50. These drivers were approached by mail with the request to participate in the project. Of the drivers that responded positively to an invitation to participate in this project only 179 participants were included in the project, for financial reasons. As a result, the participant rate for experienced drivers was 18%.

3.1.2. Participants from rural and urban area

The environment where a young novice driver takes the driving test and where he will usually drive after passing the test could possibly affect the way driving experience develops. For example, young novice drivers from a rural area will perhaps encounter less other road users, or find themselves in other traffic situations than those from urban areas (e.g. an intersection with a pedestrian crossing and a light rail). Also, youngsters in a rural area are likely to drive longer trips, for example to school or to a club.

The novice drivers were therefore selected from different driving test locations, ranging from a very urban area (Rotterdam), to a very rural area (Stellendam), and a driving test location with both types of areas (Rijswijk). The experienced drivers selected by the RDW resided in the same regions.

The purpose of the selection of participants in these different areas was to have a realistic, heterogeneous sample; not to study the differences between drivers from rural or urban areas. Analysis of the first two questionnaires showed no differences between drivers from a rural area and drivers from an urban area, therefore the novice drivers selected from the different test locations were analysed as one group.

3.1.3. Background characteristics

Table 3.1 shows the characteristics of the experienced and novice participating drivers at the start of the study. The mean age, for example, was calculated using the age of the participants in September 2005.

Table 3.1. Characteristics of the experienced and novice drivers at the start of the study

	Experienced (n = 179)		Novice (n = 509)		
	Mean	SD	Mean	SD	Significance
Age	40.6	5.6	20.0	1.9	$F_{1,680}$ = 5237.29; p < .001
Years of licensed driving experience	20.3	5.7	0.0	0.0	$F_{1,680}$ = 6510.17; p < .001
Driving education & exam					
No. lessons	31.0	15.5	35.0	13.3	$F_{1,602}$ = 9.84; p < .05
No. hours	31.7	17.4	40.3	17.0	$F_{1,597}$ = 30.37; p < .001
No. exams	2.0	1.2	1.9	0.9	n.s.
	Count	**%**	**Count**	**%**	**Significance**
Gender					
Males	87	49%	264	52%	
Females	91	51%	245	48%	n.s.
Total	178	100%	509	100%	
Highest school education*					
VMBO/MAVO	73	42%	200	45%	
HAVO	31	18%	102	22%	
VWO	27	16%	100	21%	$\chi^2_{(3,N=609)}$ = 32.22; p < .001
Else	42	24%	34	13%	
Total	173	100%	436	100%	
Driving education					
Regular	171	99%	386	89%	
Driver training in steps (RIS)	-	-	11	3%	$\chi^2_{(2,N=609)}$ = 16.97; p < .001
Crash course	2	1%	39	9%	
Total	173	100%	436	100%	

Note. Some information, for example school and driving education, was not available for all participants due to drop-outs.
*These education levels roughly correspond to the following: VMBO/MAVO = Vocational Education / Lower Secondary Education; HAVO = Higher Secondary Education; VWO = Pre-university Education.

As can be seen in Table 3.1, the experienced drivers were, on average, 20 years older than the novice drivers and had a driver's licence for 20 years. The experienced drivers reported obtaining less driving lessons before obtaining their driver's licences. At least some differences between novice and experienced drivers, for example in school and driving education, are most likely caused by the generation difference. For example, twenty years ago there was no 'driver training in steps (RIS)'.

3.2. Design

3.2.1. A longitudinal study

In order to monitor the development of calibration, the drivers in our study were followed intensively for two years. A study by Vlakveld (2005) indicated that, for the Dutch situation (in the years 1991-2001), the crash risk of novice drivers drops substantially after roughly four years of driving experience. A Canadian study (Mayhew et al., 2003) indicated that a considerable drop already takes place during the first two years of driving. Considering these studies and practical considerations, the two-year period was chosen. Because of indications that the most distinctive drop in risk occurs in the first months of the driving career (Mayhew et al., 2003; Sagberg, 1998), the novice drivers completed the first questionnaire a few weeks after they passed their driving exam.

During this two-year period, participants completed six so-called *sessions;* one session every four months. A *session* consisted of filling out one questionnaire and keeping a weekly driving diary for three weeks.

The questionnaire contained items on self-assessment of skills, perceived risks in traffic and items from the Driver Behaviour Questionnaire (DBQ). In addition, a new instrument was developed (the Adaptation Test) which measures adaptation to task demands as a function of perceived complexity of the situation. See Section 0 for a description of the questionnaire.

In the weekly driving diary, participants reported the trips that were made and the situations that were encountered in traffic in the preceding week. The participants reported, for example, how much they had driven, if they had driven at night, with or without passengers, and if they had consumed any alcohol before driving, etcetera. See Section 0 for a description of the driving diary.

Each session was completed online, using a website that was designed for this study. When a specific questionnaire or diary was supposed to be completed, participants received an email with a hyperlink to the specific questionnaire or diary.

To relate the reported experiences in the questionnaire and driving diary to actual driving performance, a selection of novice and experienced drivers participated in an on-road driving assessment (in 2006 and 2007).

With this longitudinal design, it is possible that drivers who, on a regular basis, answer questions about traffic behaviour and traffic safety will become more aware of the risks of driving than drivers who perhaps never even think about traffic safety (see e.g. Falk, 2010). Therefore an attempt was made to rule out a 'Hawthorne effect', that is, measuring certain behaviour may change that behaviour (Bouchet, Guillemin & Briançon, 1996; Murray, Swan, Kiryluk & Clarke, 1988). The novice drivers were randomly divided into two groups. The first group consisted of 330 novice drivers (Novice I) and started filling out questionnaires and diaries directly in October 2005. The second group, consisting of the remaining 179 novice drivers (Novice II), was selected for participation in October 2005, but was not invited to complete the sessions until May 2006. See Table 3.2 for the complete design of the study.

Table 3.2. Design – number of participants completing the sessions and participating in the driving assessment

	First year of the study				
	Selection Sept '05	Session 1 Oct '05	Session 2 Jan '06	Driving assessment April '06	Session 3 May '06
Group					
Experienced	179	167	165	47	159
Novice I	330	283	272	46	257
Novice II	179	--	--	37	132
Total	688	450	437	130	548

	Second year of the study			
	Session 4 Sept '06	Session 5 Jan '07	Driving assessment April '07	Session 6 May '07
Group				
Experienced	158	157	35	157
Novice I	251	249	53	251
Novice II	127	117	24	120
Total	536	523	112	528*

* The number of participants in Session 6 increased compared to Session 5 because participants were 'allowed' to miss one or two sessions

3.2.2. Incentive

In order to keep the drivers motivated to continue participating in the study, several strategies were applied. Participants were rewarded €20 for each completed session, of which half was available directly and half was saved in a virtual piggybank until the end of the study. In practise, this meant that, towards the end of the study, the amount of money saved grew larger and larger; so it became more and more rewarding for participants to remain in the study and finish the final session.

In addition to this official reward, during the two years, the participants received several free 'gadgets' (e.g. a beach ball in the summer and shawl in the winter) to thank them for their participation in the study. The idea behind these gifts is that participants start to feel obligated to complete the sessions for which they have already been rewarded. Research has indicated that this 'guilt' mechanism is effective with even the smallest incentive (see e.g. Dillman, 2000).

A final strategy to keep the participants in the study was to have as much personal contact as possible. The participants received a personalised e-mail to announce each scheduled questionnaire or diary. If they did not respond to the (reminder) e-mail or complete the requested questionnaire or diary, participants were contacted by telephone. During periods between sessions, participants received newsletters about the project and the 'gadgets' described above.

3.2.3. Drop-out

Despite our efforts, participants dropped-out of the study for various reasons. Table 3.3 shows the number of active participants per session. Active participants completed an entire session consisting of a questionnaire and three weekly driving diaries. It is possible that some participants completed the questionnaire but failed to complete all diaries. These participants are considered drop-outs in Table 3.3, although their responses to the questionnaire are still used as data in the separate chapters of this thesis. This explains the differences between the number of active participants in Table 3.3, and the number of participants in the designs of several chapters.

Table 3.3 shows a difference in drop-out rate between the novice and experienced drivers (24% and 31% versus 12% respectively). This is presumably caused by the way the participants were approached to participate in the study. The experienced drivers received an invitation in the mail. If they were interested in the study, they had to fill in a response card and mail this back to us. In other words, they had to actively sign up for the study and were probably highly motivated and committed participants.

The novice drivers, on the other hand, were approached in the driving test centres. They did not have to do anything active to sign up for the study other than give their personal information. In addition, they had just passed their driving tests, so most of them were euphoric and probably could not oversee the implications of participation in a two-year study.

Table 3.3. Number of active participants that completed each session and the cumulative drop-out (in percentages)

Session	Selection	1 Oct '05	2 Jan '06	3 May '06	4 Sept '06	5 Jan '07	6 May '07
Experienced							
Active	179	167	165	159	158	157	158
Cum. % drop-out		7%	8%	11%	12%	12%	12%
Novice I							
Active	330	283	272	257	251	249	251
Cum. % drop-out		14%	18%	22%	24%	25%	24%
Novice II							
Active	179	178	177	132	127	117	123
Cum. % drop-out		1%	1%	26%	29%	35%	31%
Total							
Active	688	628	614	548	536	523	532
Cum. % drop-out		9%	11%	20%	22%	24%	23%

Table 3.3 furthermore shows the highest drop-out rate in the Novice II group (31%). These were participants that signed up for the study six months before they actually started their first session. In these six months we did not contact them at all. It is possible that these participants felt less obligated to participate in the study because of this delay.

Finally, Table 3.3 shows that the drop-out rate in Session 6 is less than the drop-out rate in Session 5. Because participants were 'allowed' to miss one or two sessions the number of active participants can also increase from one session to the next. After Session 4 and 5 drop-outs were stimulated to fill in Session 6 anyway and receive their rewards that had been saved from previous sessions.

3.2.4. Background characteristics of the drop-out

As can be seen in Table 3.4 there are no significant differences within the group of *experienced* drivers, between the participants that dropped-out of the study and those who participated until the final session. Although relatively more males dropped out of the study, the total numbers were too small to be able to detect a significant difference.

Table 3.4. Difference between drop-outs and active participants within the group of experienced drivers

	Active (*n* = 158)		Drop-out* (*n* = 21)		
	Mean	SD	Mean	SD	Significance
Age	40.6	5.5		39.9	n.s.
Years of licensed driving experience	20.4	5.7		19.5	n.s.
Driving education & exam					
No. lessons	30.8	15.2	33.7	18.1	n.s.
No. hours	31.7	17.6	32.2	15.9	n.s.
No. exams	1.9	1.1	2.4	1.8	n.s.
	Count	%	Count	%	Significance
Gender					
Males	74	47%	13	65%	
Females	84	53%		35%	n.s.
Total	158	100%	20	100%	
Highest school education**					
VMBO/MAVO	67	42%		63%	No test possible due to small numbers
HAVO	29	18%		15%	
VWO	22	14%		16%	
Else	40	25%		7%	
Total	158	100%	15	100%	
Driving education					
Regular	157	99%	14	93%	No test possible due to small numbers
Driver training in steps (RIS)	0	0%	0	0%	
Crash course	1	1%	1	7%	
Total	158	100%	15	100%	

*Some information, for example school and driving education, was not available for all drop-outs.
**These education levels roughly correspond to the following: VMBO/MAVO = Vocational Education / Lower Secondary Education; HAVO = Higher Secondary Education; VWO = Pre-university Education.

In contrast, Table 3.5 shows that *novice drivers,* who dropped out of the study, were significantly more often male, lower educated, and needed less driving education to obtain their drivers' licence.

Table 3.5. Difference between drop-outs and active participants within the group of novice drivers (Novice I + Novice II)

	Active (n = 374)		Drop-out* (n = 135)		
	Mean	SD	Mean	SD	Significance
Age	19.9	1.8		20.3	n.s.
Driving education & exam					
No. lessons	35.6	13.5	31.2	11.7	$F_{1,433}$ = 5.81; p < .05
No. hours	41.2	17.6	35.2	12.3	$F_{1,429}$ = 6.54; p < .05
No. exams	1.9	0.9		0.9	n.s.
	Count	%	Count	%	Significance
Gender					
Males	179	48%	85	63%	$\chi^2_{(1,N=509)}$ = 9.06;
Females	195	52%	50	37%	p < .01
Total	374	100%	135	100%	
Highest school education**					
VMBO/MAVO	161	43%	39	63%	
HAVO	93	25%	9	15%	$\chi^2_{(3,N=436)}$ = 8.63;
VWO	90	24%	10	16%	p < .05
Else	30	8%	4	7%	
Total	374	100%	62	100%	
Driving education					
Regular	333	89%	53	85%	
Driver training in steps (RIS)	8	2%	3	5%	n.s.
Crash course	33	9%	6	10%	
Total	374	100%	77	100%	

*Some information, for example school and driving education, was not available for all drop-outs.
**These education levels roughly correspond to the following: VMBO/MAVO = Vocational Education / Lower Secondary Education; HAVO = Higher Secondary Education; VWO = Pre-university Education.

It is difficult to assess the effect of this selective drop-out on the results of the study; especially in comparison with other questionnaire studies. The initial response rate for the novice drivers was extremely high; 92% for a longitudinal study, where the average response rate for a *single* mail survey is 74% (Dillman, 2000). In comparison, Harrison (2004) conducted a two-year longitudinal study among learner drivers. In his study only 35% of the drivers, who had already expressed interest in the study, actually signed up and 30% finished the study.

It can be argued that the drop-out among novice drivers in the current study are people who usually would not sign up for a study if invited by mail (i.e. the otherwise unknown non-response group). In our study the characteristics of this otherwise non-response group are known and are known to differ from (92% of) the whole population. As a result, it must be kept in mind that the sample in our study is not completely representative for the whole population, as is the case for all questionnaire studies.

3.3. Questionnaire

The questionnaire was administered every four months, a total of six times, during two years. To announce a new questionnaire, participants received an email with a direct link to the online questionnaire. Completing the questionnaire took about 30 minutes.

The questionnaire included general questions about driving behaviour; for example, whether the participants owned a car, whether they had borrowed a car from friends or parents, etcetera. In addition, the questionnaire included specific questions about two elements of the calibration model (Figure 1.2): self-assessment of skills, and adaptation to task demands.

3.3.1. Self-assessment of skills

The Driver Confidence Questionnaire (see Appendix A) was designed for this study to measure self-assessment of skills. There has been much debate in the past on how to ask drivers about their skills. The traditional approach to measure self-assessment of skills is to ask drivers to compare their skills with the skills of 'the average driver' (Delhomme, 1996; Mathews & Moran, 1986; McCormick et al., 1986; McKenna et al., 1991).

However, as Groeger (2000) points out, asking drivers to compare themselves with the average driver could be misleading, because 'average' may be a negative rather than neutral descriptor (i.e. comparable to 'mediocre') at least for some people. Therefore, in the current study, we asked drivers to compare themselves with 'the average driver', but also to make a comparison with a 'peer driver'. This is comparable to the method applied by Svenson (1981), who asked his participants to compare themselves with the other participants (university students) in the room. Also similar to the study by Svenson (1981), participants did not only assess their *skills*, but also gave an estimation of their *crash risk*, compared to the average and peer driver.

Finally, the Driver Confidence Questionnaire included some neutral questions (i.e. without comparison to the average or peer driver) on how confident the participants are as a driver, and how much danger they perceive in traffic. These neutral questions were included because, if you ask drivers to make a comparison either with the average or peer driver, it is difficult to separate if the assessment of their own skills was incorrect, or if they simply have wrong ideas about other drivers.

A five point scale (rather than a finer scale) was used for these questions. This scale was chosen because the most important distinction for analysis in this study is between drivers who view themselves as (much) better, similar or (much) worse drivers compared with the average or peer driver.

3.3.2. Adaptation to task demands

An important aspect of the calibration model is that the balancing and adaptation can take place on different hierarchical levels of the driving task (see e.g. Fuller & Santos, 2002; Rasmussen, 1986). This thesis considers adaptation to task demands on the two highest levels of Michon's (1985) driving task hierarchy. First, on the *strategic* level, the driver could choose to avoid a difficult junction or choose (not) to drive in the dark. Second, on the *manoeuvring* level, a driver reduces speed or increases headway, when encountering more complex situations.

The influence of inadequate calibration on reported violating behaviour is also studied, which can be viewed as both strategic, as it involves a strategic choice (not) to violate traffic rules, as well as manoeuvring, as the specific situation has a large influence on driver behaviour (Reason, 1990). The lowest level of the driving task hierarchy (i.e., the control or vehicle handling level) which is often performed automatically was not considered here, as adaptation to task demands is assumed to involve a considered assessment of the situation.

Strategic level

The questionnaire contained two questions concerning adaptation to task demands on the strategic level. Drivers were asked: a) whether, in the preceding four months, they had cancelled an intended driving trip due to adverse conditions (e.g., darkness or weather conditions); and b) whether, in the preceding four months, they had avoided a difficult (i.e. busy or complex) intersection.

Manoeuvring level – the Adaptation Test

In order to measure how drivers adapt to task demands, as a function of the complexity of the situation, the Adaptation Test was developed. Chapter 5 describes how the test was developed and evaluated as an instrument for this study.

The Adaptation Test consisted of 18 traffic scenes presented in two (almost) identical photographs (see Appendix B). The photographs differed in one single detail, thereby increasing the *complexity of the situation* (right-hand side of the calibration model, see Figure 1.2). The pictures were presented randomly and participants were kept unaware of the varying level of complexity.

The respondents were asked to assess at what speed they would normally drive in the depicted situations. It was stressed that respondents were not asked to determine the legal speed limit. Neither in the instructions nor in the pictures was explicit reference made to the legal speed limit. The situations were selected in such a manner that the 'extra' element would increase the complexity of the situation, without legally obliging the driver to lower his speed.

Although absolute speeds are relevant, speed assessment on the basis of photographs is unreliable (see e.g. Groeger, 2000, for information on the assessment of driving speeds). Therefore, only the direction of the difference between reported speed in the simple and complex version was used to test for differences between groups. A response was considered 'correct' if the reported speed was lower in the complex situation than in the corresponding simple situation. A higher reported speed in the complex situation was regarded to be equally unwanted as no speed difference at all. For both cases such a response was considered 'incorrect'.

As it was important that participants would not notice the mechanism behind the Adaptation Test and consciously compare pairs of photos, the photos were randomly distributed throughout the web based questionnaire, without the possibility of navigating back to previous photos. The photos were distributed under two conditions: 1) in half of the situations the participants viewed the complex situation first and in the other half they viewed the simple situation first; and 2) a simple and complex version of the same traffic situation never immediately succeeded each other. A pilot study, using a similar photo judging task, revealed that participants were not aware of the differences between the photos when they were presented in random order following the abovementioned conditions. The participants assessed six new traffic situations (i.e. 12 photos) in three subsequent sessions (see Table 3.6).

Table 3.6. Distribution of the 18 situations of the Adaptation Test over the 6 sessions in the study

Session		Situation 1 – 6	Situation 7 -12	Situation 13-18
1.	October 2005	x		
2.	January 2006		x	
3.	May 2006			x
4.	September 2006	x		
5.	January 2007		x	
6.	May 2007			x

With the use of photographs, task complexity could be easily manipulated, which is fairly impossible in a more natural environment (e.g. a real driving task). In addition, the easy implementation of photographs on a website makes this test very cost-effective and applicable in a longitudinal study with a large number of participants.

Self reported violations – the DBQ

The questionnaire contained an abbreviated Dutch version of the Driver Behaviour Questionnaire (DBQ, Verschuur, 2003b; Verschuur & Hurts, 2008); see Appendix C.

The original 50 item Driver Behaviour Questionnaire was a self-report questionnaire of unsafe driving behaviour experienced by drivers over the past year (Reason et al., 1990). Behaviours in the DBQ are classified as errors or violations, consistent with Reason's (1990) Generic Error Modelling System (GEMS). From the original DBQ, Parker et al. (1995), selected 24 items that explained the most variance on the three factors identified in the original DBQ: Errors, Lapses and Violations. A Dutch translation of these 24 items by Verschuur (2003b; Verschuur & Hurts, 2008) was used in the current study.

Several studies indicated a relationship between violations and crashes (Parker, West et al., 1995; Sullman, Meadows & Pajo, 2002; Underwood et al., 1997); and also between errors and crashes, although less frequent (Mesken et al., 2002; Verschuur, 2003b).

3.3.3. Self-reported crashes

In five out of six questionnaires, drivers were asked to report whether they had been in a crash in the preceding four months. In the first questionnaire, this question was omitted as the novice drivers had no unaccompanied driving experience at this point. A crash was defined as an incident where there was at least material damage or personal injury. Only three crashes with personal injury were reported, compared to 142 crashes with material damage only. Due to the small number of crashes with personal injury, no distinction will be made in the severity of the crash in this study.

3.4. Driving diary

The driving diary was developed to investigate the relationship between experiences in traffic and the development of calibration skills. For example: does a novice driver with a high mileage develop calibration skills faster than drivers who hardly ever drive? Or, does the way a novice driver reflects on incidents encountered in traffic, have an effect on the development of calibration skills?

In order to answer these questions, the participants were asked to fill in three diaries per session, each reporting over a period of one week. At the end of each week the participants reported in the diaries how much they had driven, if they had driven mostly during weekdays or rather in the weekend, or if they, for example, had consumed alcohol before driving (see Appendix D for the specific questions in the driving diary).

In addition, participants described one specific incident that they had encountered that preceding week in more detail. They described what had happened, who they considered at fault, and if they had learned anything from the incident.

When a driving diary was scheduled, participants received an email with a direct link to the online diary. The driving diary would take about 5-10 minutes to complete if participants had driven that week. If not, it was also possible to report not driving at all that particular week.

3.5. On-road driving assessment

To compare the reported experiences from the questionnaires and diaries with actual driving performance, a subgroup of novice and experienced drivers participated in an on-road driving assessment. Driving skills were assessed on two occasions (in 2006 and 2007) in order to detect any changes in time as a result of driving experience.

This section discusses the procedure and participants of the on-road driving assessment in this longitudinal study. The next section (3.6) focuses on general issues concerning the validity and reliability of a driving assessment to measure 'ability to drive safely'. Section 3.6 also describes a small scale experiment to study whether the assessment of observers is influenced by the appearance of a driver.

3.5.1. Procedure

The driving assessment consisted of a half hour drive on different road types. The drives departed from two Licensing Centres in the cities of Rotterdam and Rijswijk. In both cities, participants drove a standardised route, which was selected in such a way that the routes in Rijswijk and Rotterdam were as similar as possible. On a few occasions an alternative route was driven because of circumstances like traffic density or a lorry blocking the way. In order to make the drive similar to an ordinary trip, the participants were instructed to follow sign posts on parts of the route (e.g., "Use the sign posts to find your way to the station"). For the remaining part, traditional "turn left / turn right" instructions were used to give directions to the participant.

The driving assessments were conducted by four professional driving license examiners. It would have been preferred to have each drive assessed by more than one examiner, similar to the behaviour-observation method Wiener Fahrprobe (Risser, 1985; Turetschek & Risser, 2008). With this method one observer registers behaviour in a standardised way (the Coding Observer), while the second concentrates on all kind of deviating behaviour and communication aspects of the drive (the Free Observer). Unfortunately, it was too costly and time consuming to have all drives observed by two examiners, so each participant was assessed by one examiner in the role of the Free Observer.

The examiners rated the drivers on their ability to drive safely, on a scale from 0 to 10; 5.5 being the pass-fail criterion had it been a real driving test. In addition, the examiners provided a separate assessment of the drivers' speed and headway behaviour.

Previous to the actual drives, each examiner drove both routes in Rotterdam and Rijswijk three times, with different participants. The purpose of these 'pilot-drives' was to familiarize the examiners with the routes, and to standardize the assessments of the examiners.

The driving examiners were aware of the objective of the study. As this could bias their assessment (e.g. young drivers are 'most likely novice drivers'), the examiners were explicitly informed that the group consisted of different types of drivers, some older who rarely drove, and some younger who drove on a regular basis (e.g. as a professional courier). In addition, the participants were instructed not to mention anything about their prior driving experiences to the examiner.

3.5.2. Participants

It would have been too costly and time consuming to have *all* participants' driving skills assessed by professional examiners. Therefore, a random sample was drawn, with about 40 participants from each group (Novice I, Novice II and Experienced drivers), from the total pool of participants in the study. In 2006, a total of 130 drivers participated in an on-road driving assessment; 83 novice drivers (age: M = 20 and SD = 2; 52% male) and 47 experienced drivers (age: M = 41 and SD = 5; 57% male).

The aim was to include the drivers of this first driving assessment in the second driving assessment as well. However, not all drivers were willing to participate a second time. Therefore, in 2007 a new group of drivers was invited to participate in a driving assessment. A total of 112 drivers participated in the second driving assessment; 77 novice drivers (age: M = 21 and SD = 2; 49% male), and 35 experienced drivers (age: M = 42 and SD = 6; 51% male).

As can be seen in Table 3.7, 34 novice drivers and 20 experienced drivers participated in both on-road driving assessments; 49 novice and 27 experienced drivers participated only in the first assessment; and 43 novice and 15 experienced drivers participated only in the last assessment.

Table 3.7. Number of participants in the driving assessments in 2006 and 2007

	2006	2007
Novice I		
Both assessments	20	20
2006 only	26	--
2007 only	--	33
Total	46	53
Novice II		
Both assessments	14	14
2006 only	23	--
2007 only	--	10
Total	37	24
Experienced		
Both assessments	20	20
2006 only	27	--
2007 only	--	15
Total	47	35
Total	130	112

3.6. Validity and reliability of the on-road driving assessment

As with any method, it is essential that the measures obtained are reliable and valid. A valid method measures what it is supposed to measure. In the case of the driving assessment, it is supposed to measure 'safe driving'. A reliable method measures consistently, irrespective of the circumstances. This means that the same driver should pass the driving test with any examiner (inter-rater reliability), on any day and at any location.

3.6.1. Previous studies on validity and reliability

The driving assessment relies on the quality of the assessment of the examiners, for which doubts have been expressed with respect to the reliability and validity (see Senserrick & Haworth, 2005 for a review). For example, Maycock (2002) found that women have more difficulty passing the driving test than men. Similarly, Crinson and Grayson (2005) reported that of the 17-year-olds, only 56% of the females versus 60% of the males pass the driving test in the UK. Nevertheless, after passing the test, female drivers are involved in fewer crashes than male drivers. This suggests that the driving test is not very successful in distinguishing safe drivers from unsafe drivers (i.e. lack of validity).

However, Nyberg, Gregersen and Wiklund (2007) did not find the same gender effect in Sweden. Both males and females had the same percentages of passing the driving test (M: 66% F: 67%). Even after controlling for age there was no difference between males and females. Moreover, Nyberg et al. also did not find an other often reported result (OECD - ECMT, 2006) that females need more driving education before taking the test.

Baughan (2005) argues that drivers act differently during an assessment with an observer, than in everyday traffic, that is, without an observer. A study by Quimby, Maycock, Palmer and Grayson (1999) on speeding behaviour, however, indicates that the assessment of an expert or examiner is a good reflection of actual driving style. In their study, natural driving behaviour was captured with a speed camera before participants were invited for an on-road driving assessment. Participants were unaware that they were being observed. As it turned out, there was a strong correlation between the examiners assessment of the speeding behaviour and the behaviour that was monitored in the natural setting.

Kay, Bundy, Clemson and Jolly (2008) studied the validity and reliability of an on-road driving assessment with healthy older drivers and older drivers with a vision deficit. The driving assessment seemed to measure one single construct, namely driving errors, which is indicative of safe driving, according to the authors.

The most important feature of a driving assessment or the official driving test is that it predicts crash liability. This is very difficult to determine in the case of the driving test, as only drivers who pass the driving test can be included in a crash analysis determining the quality of the examiner's decision. Drivers who do not pass the test are not allowed to drive (without a licensed instructor), and are consequently not exposed to the same traffic risks as drivers who pass the driving test. There are indications, however, that there is some predictive validity in the assessment of an expert.

For example, Grayson et al. (2003) found a relationship between expert ratings on a driving assessment and crash liability. That is, drivers that were assessed as safer drivers were found to be involved in fewer crashes than drivers with a less positive assessment. Unfortunately, there were not enough participants (N=96) to achieve significance, but the authors conclude that "… all relationships were 'sensible' statistically, and were in accord with earlier work" (p. 23).

More convincingly, Maycock and Forsyth (1997) studied the relationship between minor errors on the UK driving test (L-test) and number of crashes. In the UK, a candidate driver is scored for minor and major errors during the driving test. A major error results in directly failing the driving test, whereas, it is possible to pass the test with a few minor errors. Maycock and Forsyth found an association between these minor errors committed on the driving test and the number of crashes in the first three years of driving. The more errors that were committed during the test, the more crash involved the driver turned out to be.

Because of the studies that have found relationships between the expert assessments and actual driver behaviour or even crash risk, the use of a driving assessment was considered an appropriate tool to measure 'ability to drive safely' in the current study. Nevertheless, literature does suggest that a driving assessment has some reliability and validity issues. Therefore some precautions were taken to control for bias related to validity and reliability in this study.

3.6.2. Controlling for bias in the current study

For validation purposes, the driver's self-reported speeding behaviour from the questionnaire was compared with the examiner's opinion about the speeding behaviour of the participant; and a significant correlation was found (Pearson r = .18; $p < .05$). Of course, this does not say anything about 'actual' driving behaviour (i.e. normal behaviour in traffic without an observer in the car), which was not available in the current study. However, the correlation does suggest that participants' behaviour was not affected much by the presence of an observer.

In the current study, it was not feasible to calculate inter-rater reliability. This would have entailed more than one examiner assessing the same participant. Unfortunately, the availability of driving examiners did not allow for this.

In conclusion, there were not many ways to control (or check) for bias in the driving assessment. In addition, there were concerns about the fact that the examiners were not blinded for the experience level (or at least the age) of the participants. Therefore, a small scale experiment was conducted to further examine the validity of driving assessments.

3.6.3. Small scale experiment: the effect of driver appearance on the assessment of driving skills[5]

Introduction

Each driving assessment was conducted by one of four professional examiners. These examiners were aware of the objective of the study (i.e. to study differences between experienced and novice drivers). This posed a problem because it was possible that the examiners would assess young drivers ('most likely novice drivers') differently than the older drivers ('most likely to be experienced').

In spite of the specific instructions that participants could have very different levels of experience (see Section 3.5.1) there were still some concerns that the appearance of participants would influence the assessment

[5] These results were based on:

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., & Brookhuis, K.A. (to appear). Novice drivers' development 0.5 and 1.5 years after licensing; results of an on-road driving assessment. In: *Proceedings of the 4th International Conference on Traffic & Transport Psychology*.

Van der Zwan (2009). *De invloed van leeftijd op de beoordeling van rijvaardigheid*. (Afstudeerscriptie 0248347; in Dutch). Amsterdam: Universiteit van Amsterdam.

of the examiner. As can be seen in Figure 3.1, the individual examiners assessed the novice and experienced drivers differently. There is much agreement about the performance of the novice drivers as a group, but there is much less agreement in the assessments of the experienced drivers. Especially Examiner 1 and Examiner 3 seem to differ in their opinion about the experienced drivers. ANOVA showed that the difference between examiners was significant ($F_{3,228} = 3.79$, $p < .05$, $\eta^2 = .05$) [6]. Bonferroni Post Hoc tests indicated a significant difference between Examiner 1 and 2. Examiner 3 and 4 assessed considerably less participants than examiners 1 and 2, which could explain why there was no significant difference between these examiners and examiner 1.

Because all examiners assessed different participants it cannot be ruled out that the differences between assessments were correct reflections of the performance of the drivers. Therefore a small scale experiment was designed to study how observers are influenced by the appearance of drivers.

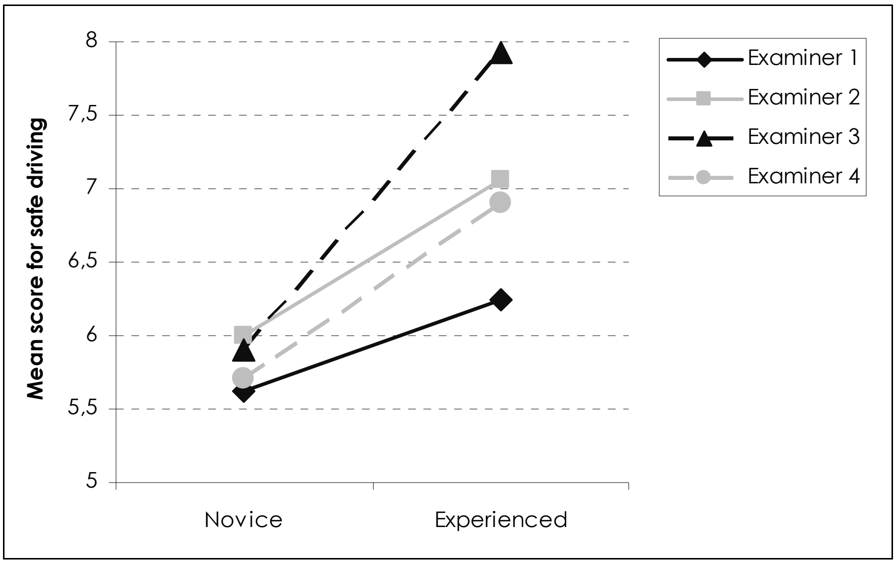

Figure 3.1. Assessments of the individual examiners on the novice and experienced drivers in 2006 and 2007

Method

With the help of the Dutch Driving Test Organisation (CBR), two video recordings of a driving trip were constructed; one trip showing a good

[6] In the case that one participant was assessed twice (in 2006 and 2007) by the same examiner, only the first assessment (from 2006) was used for analysis.

performance and one with a poor performance with respect to safe driving (e.g. respectively larger or smaller headway). These trips were recorded using a 'head-mounted' camera, showing the trips from the perspective of the driver. By doing so, the actual person driving the car was invisible to the person viewing the video.

Two videos were also constructed with a young and older driver getting into the drivers' seat of a car and placing the camera on their head. These drivers were respectively 23 and 59 years of age. Using different combinations of these videos it was possible to vary the appearance of the driver (young or older) independently from actual driving performance (good or poor; see Table 3.8).

Table 3.8. Combinations of driving performance and appearance of the driver

	Appearance of the driver	
	Younger	Older
Driving performance		
Good	Video 1	Video 2
Poor	Video 3	Video 4

Participants

The original intention was to have these videos assessed by professional CBR examiners. However, because it was not possible for the CBR to participate in this experiment, the videos were assessed by students from the University of Amsterdam instead. A total of 55 students participated in the study (82% females, mean age = 21 years), of whom 73% had their drivers' licence.

The students were randomly assigned to one of four videos (see Table 3.8). They were asked to view the driving trip from the perspective of a driving test examiner; and to rate the drivers on their 'ability to drive safely'. The students were instructed to use a scale from 0 to 10; with 5.5 being the pass-fail criterion.

Results

Because it was important to know if the participants in the study in fact perceived the appearance of the drivers, they were asked to estimate the age of the driver in the video. Paired sample t-tests indicated that the older driver (true age = 59; assessed age M = 43, SE = .79) was perceived as significantly older than the younger driver (true age = 23; assessed age M = 23, SE = .38; $t(53) = 25.94$, $p < .001$).

Figure 3.2 shows that the good performance drives were indeed assessed as safer than the poor performance drives. Two-way between subjects ANOVA indicated that this difference was significant ($F_{1,51} = 11.91$, $p < .001$).

In addition, Figure 3.2 shows that there was a tendency to rate older drivers as somewhat more safe than younger ones. Please recall that the *drives* with the older and younger driver were exactly the same. Still, the drive of the younger driver is rated lower than the drives of the older driver, especially for the poor performance drive. This difference was, however, not significant.

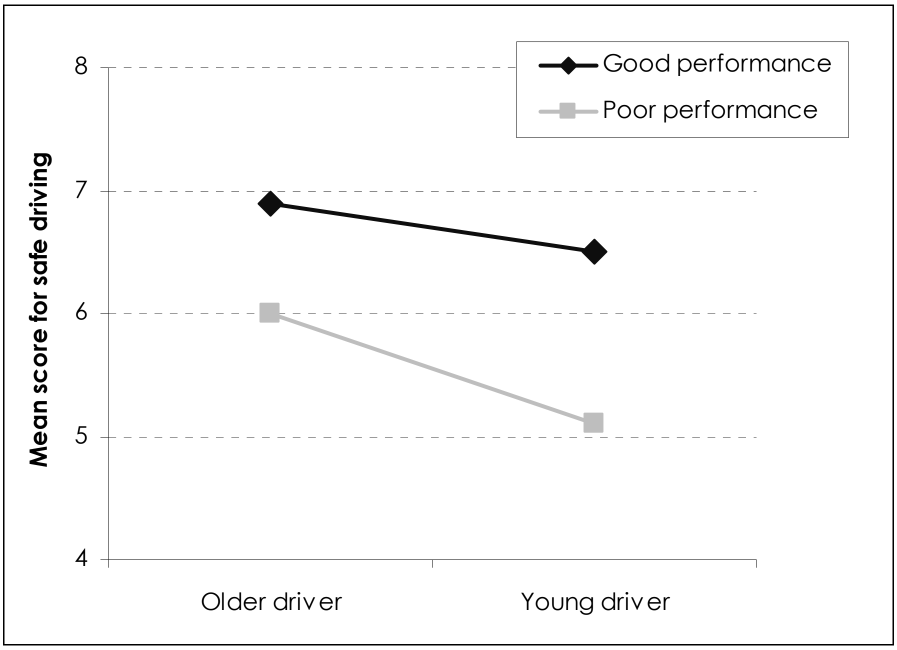

Figure 3.2. Mean scores for 'ability to drive safely' for older and young driver on the good and bad performance drive

Conclusions

Although no significant bias of appearance of the driver was found, this does not exclude the existence of one. Especially, since a small trend towards an age bias seems present in Figure 3.2.

It is unfortunate that the CBR examiners were unable to assess the videos, since it is likely that they are less influenced by the appearance of the driver, because they are trained to give objective ratings of a driver's skills.

In any case, previous research and the trend in our own experiment indicate that the results of the driving assessment should be interpreted with care. A direct comparison between novice and experienced drivers on aspects measured in the driving assessment is unfavorable, since it is possible that other influences, such as the appearance of the driver, influenced the assessment. Within the same group of drivers (i.e. within the novice drivers) the risk of a bias is less likely.

3.7. Data analysis

The categorical variables in this study were analysed using a Chi-square test ($\alpha = .05$). For interval data, ANOVA was used to test for significance between groups. The longitudinal data were analysed using *F*-test Repeated Measures ANOVA.

Besides significance of the results, the effect size (Partial èta squared, $\eta 2$) was also considered with $\eta 2 \approx .01$ as a small, $\eta 2 \approx .06$ as a medium, and $\eta 2 \approx .14$ as a large effect size (Cohen, 1988).

3.7.1. Assumptions of the *F*-test by Repeated Measures ANOVA

One of the assumptions of the *F*-test is that each population of scores is *normally distributed*. Visual inspection of the data showed that this was not the case for each variable. However, the *F*-test is robust against violations from the normality assumption. For sample sizes larger than 30 the type I error (α) is nearly the same as when normality can be assumed (Jaccard & Becker, 2002). Therefore, Repeated Measures ANOVA was considered to be an appropriate tool.

A second assumption of the *F*-test Repeated Measures ANOVA, of which violation has more serious consequences, is the *sphericity assumption*. Whether this assumption holds can usually be tested with the Mauchly test. However, studies have suggested that this test is not always reliable (Jaccard & Becker, 2002; Stevens, 1996). Because for some variables the sphericity-assumption was definitely violated, the decision was made to assume that for *all* analysis the sphericity-assumption was violated. As a result, the degrees of freedom in the *F*-test had to be adjusted. Two of the most common encountered adjustment factors are the Huynh-Feldt epsilon and the Greenhouse-Geisser epsilon (Jaccard & Becker, 2002). Because the latter is the more conservative adjustment (Stevens, 1996), this statistic was chosen to test for significance.

3.7.2. Missing data in the longitudinal analysis

Sources for missing data

Like most longitudinal analysis techniques the *F*-test Repeated Measures ANOVA needs complete series of data. And, also like most longitudinal studies, the current data had missing data due to different causes. Bijleveld et al. (1999) distinguish three sources of missing data. The first, *missing subjects*, are respondents who refused to participate at all in the project. This source of missing data has consequences for the generalisability of the results. However there are no consequences for the analysis of the data, as, these subjects produced no data to analyse.

A second source of missing data according to Bijleveld et al. are the *missing occasions.* These occur when for some occasions measurements are available for a certain subject, but not for all. In general, these subjects would belong to the so-called drop-out group, who stop participating somewhere throughout the study. In the current study, there is an extra cause for missing occasions. The Novice II group started with the study after the first two sessions. So, for these respondents the first two sessions are missing occasions by design.

A third source of missing data described by Bijleveld et al. concerns the *missing value* problem. This occurs when participants are available for a certain session, but do not fill in every question. This is especially an issue with the diary study. The participants were asked to fill in the diary even if they had not driven a car in a particular week. Particularly novice drivers did sometimes not drive for a whole week, resulting in many missing values. With *F*-test Repeated Measures, one missing value at a certain instance causes not only that instance to be useless for that subject, but also the entire series (two years) for that subject to be compromised.

Missing data in the questionnaire

With respect to the variables from the questionnaire, respondents with two or more missing answers (out of six sessions) were considered missing *subjects*. Nine participants missed one *occasion*, these missing values were replaced with the mean value of the preceding and following entry. With the exception for the first session, missing data in this session were replaced with information in the second session; and in the final session, a missing value was replaced with the entry in session 5. After this correction, there were 158 experienced drivers and 252 novice drivers with complete series of data

Missing data in the driving diary

The three sources of missing data combined show that out of 330 novice drivers who started the study, only 62 drivers (19%) drove in all 18 weeks *and* reported this in the diaries. Of the 179 experienced drivers, 84 (47%) drove in all 18 weeks and reported this in the diaries. In theory, this would mean that the Repeated Measures ANOVA can only be done for 19% of the novice and 47% of the experienced drivers. In addition, none of the Novice II participants can be included in these analyses, because they all have missing data for the first 6 diaries.

Before the data from the driving diary were analysed, several steps were taken in order to deal with the missing data. In the first step, every participant with more than two missing sessions (six diaries) was considered either a *missing subject* or a *missing occasion* (drop-out from the study). A total of 92 (=13%) participants did not fill in any diary (i.e. missing subjects); and 51 (=7%) participants missed more than two sessions (six diaries). The latter participants (*missing occasions)* were considered to be drop-outs and were removed from the analyses. See Table 3.9 for the number of missing subjects and occasions in each group.

Table 3.9. Missing subjects and occasions in the study

	Novice I		Novice II		Experienced		Total	
	n	%	n	%	n	%	n	%
Missing subjects	40	12%	42	24%	10	6%	92	13%
Drop-out (Missing occasions)	31	9%	14	8%	10	6%	55	8%
Active participants	259	79%	123	69%	159	89%	541	79%
Total	330	100%	179	100%	179	100%	688	100%

In the second step, the information in the three diaries of each session was averaged. This reduces the amount of missing data due to participants reporting not driving in one week (*missing values*), but with valid entries in the other two weeks of that particular session. Table 3.10 shows the number and percentage of participants with at least one missing value with and without averaging of the sessions. For the Novice I group, for example, the table shows that this step reduces the number of participants with incomplete data series from 74% to 30%.

Table 3.10. Number of participants with at least one missing value on the variable 'driving on a weekend' within the 18 diaries and within the (averaged) 6 sessions

	Novice I		Novice II		Experienced		Total	
	n	%	n	%	n	%	n	%
Missing values within:								
Diaries	192	74%	78	63%	75	47%	345	74%
Sessions	77	30%	32	26%	10	6%	119	22%
Active participants	259		123		159		541	

The third step in dealing with missing values was done for each variable separately. In this step, missing information from one session was replaced with the mean value of the preceding and following entry. With the exception for the first session, missing data in this session was replaced with information in the second session; and the final session, a missing value was replaced with the entry in session 5. If more than one succeeding value was missing, no data were replaced. For the Novice II group, who did not complete the first two sessions, only missing values on the last 4 sessions were replaced. Appendix E shows how many entries were replaced with an estimate, for each of the variables of the diary.

Novice II group

The Novice II group started with the study after the first two sessions. So, by definition the Novice II group had incomplete series of data. For both the questionnaire and the diary variables, the responses from the Novice I group were compared to the responses from the Novice II group on the last four sessions. Repeated Measures ANOVA indicated that there were no differences between these groups of novice drivers. Therefore, only the data of the Novice I group, with complete series of data, were used to test for development over all the variables.

Conclusions missing data

The steps described in this section to deal with the problem of missing data were executed only to enable testing for significance on the longitudinal data. Where possible, for example when comparing novice and experienced drivers on a single data point, all available data were used.

3.8. Summary of the methods used

To monitor the development of calibration, about 500 novice drivers were followed intensively for two years. About 180 experienced drivers, with at least 10 years of driving experience, served as a control group. Of all the novice drivers invited for this study, no less than 92% agreed to participate. For the experienced drivers, the response rate was 18%. In total, 23% of the participants who started the study, dropped-out during the course of two years; 12% of the experienced drivers versus 27% of the novice drivers.

During the two years the participants completed six sessions, one session every four months. Each session consisted of a questionnaire and three driving diaries, reporting over a period of one week.

The questionnaire included items on self-assessment of skill and perceived risks in traffic (see Chapter 4). To measure how drivers adapt to task demands as a function of the complexity of the situation, the Adaptation Test was developed (see Chapter 5). Finally, the questionnaire included items to measure adaptation on a strategic level, and an abbreviated Dutch version of the Driver Behaviour Questionnaire (DBQ; see Chapter 6).

The driving diary was developed to investigate the relationship between experiences in traffic and the development of calibration skills. In this diary participants reported, for example, how much they had driven, if they had driven mostly during weekdays or in the weekend, or if they had consumed alcohol before driving (see Chapter 7).

A subgroup of drivers participated in an on-road driving assessment on two occasions; in 2006 and 2007 (see e.g. Chapter 4).

PART 2:

EMPIRICAL STUDIES

INTO CALIBRATION

4. Self-assessment of skills[7]

Chapter 2 introduced three elements of the Calibration model: 1) self-assessment of skills, 2) perceived complexity of the situation, and 3) adaptation to task demands.

The current chapter describes how *self-assessment of skills* can be measured. More specifically, it studies whether novice drivers overestimate their driving skills more than experienced drivers. Two approaches for measuring self-assessment of driving skills are used. In the first (more traditional) approach the drivers' assessment of their own skills is compared to the group average. Results of this approach suggest that novice drivers are modest about their skills. That is, experienced drivers are more optimistic about their driving skills and risks in traffic than novice drivers.

In the second approach, the drivers' assessment is compared to a more independent measure of driving skills (the assessment of an examiner). The results of this approach indicate that novice drivers overestimate their driving skills more than experienced drivers. Especially novice drivers with a

[7] This chapter is based on:
De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., & Brookhuis, K.A. (2007a). Do young novice drivers overestimate their driving skills? In: *Proceedings of the Young Researchers Seminar*. Brno: CDV.

low score for 'safe driving' on the driving assessment are more confident than their performance would suggest. The experienced drivers with the lowest scores on the driving assessment are less confident about their driving skills.

It is concluded that a comparison with an independent measure of driving skills is necessary to measure over-estimation (or under-estimation) of skills.

In Chapter 5, a method is developed to measure *adaptation to task demands* as a function of *perceived complexity of the situation*.

4.1. Introduction

In Chapter 2 several approaches to measure self-assessment of driving skills were described. In the traditional approach drivers were asked to compare their skills with the skills of the 'average' driver. Results showed that drivers generally rate themselves to be better drivers than the average driver (Delhomme, 1996; Mathews & Moran, 1986; McCormick et al., 1986; McKenna et al., 1991; Svenson, 1981).

But it is not clear if this overestimation of skills is a larger problem for young novice drivers compared to more experienced drivers, as study results differ. The review of Mayhew and Simpson (1995) showed that there are: 1) studies indicating that young drivers are especially overconfident; 2) studies indicating that young drivers do not differ from older drivers in self-assessment; 3) studies indicating that young drivers are overconfident, but not in all driving situations; and 4) studies showing that young drivers express less overconfidence than older drivers.

In addition, the traditional approach of simply asking drivers to compare themselves with the average driver has been criticised from a theoretical point of view. Groeger (2000) points out that asking drivers to compare themselves with the average driver could be misleading, because 'average' may be a negative rather than neutral descriptor (i.e. comparable to 'mediocre') at least for some people. Traffic would be very unsafe if the average driver is only a mediocre driver.

This theoretical issue can be resolved, by letting the drivers make an assessment of their own driving skills and compare this with their actual driving skills. In some studies self-assessment was compared to a more objective measure. For example Sundström (2008a) compared novice drivers' performance on the theory and practical driving test with drivers' assessment on how they would perform on these tests. Sundström found only a weak relationship between perceived and actual performance on these tests; performance on the theory test had the strongest relationship with perceived performance on this test.

Horrey et al. (2008) compared drivers' own expectation about their performance while driving around a course in an instrumented vehicle, with their actual performance. During the drive, participants were asked to complete a series of tasks on a hand-held or hands-free cell phone. The results showed that drivers were generally not able to assess the magnitude of the distraction effects. In one specific manoeuvre, in which the driver had to stop the car in time, drivers who reported the smallest estimates of distraction, even showed the largest performance deficits.

Although these studies indicate that drivers, in general, are not very capable in assessing their performance, no comparison has been made between novice and more experienced drivers. That is, it is still inconclusive if novice drivers overestimate their skills *more* than experienced drivers do. Only one study was found in which novice drivers were compared to experienced drivers. Matsuura (2005) compared drivers' ratings of their own skill with in-car observations by driving instructors. He found that young groups were more overconfident than older groups, but experienced groups were more overconfident than novice groups. This study included only male drivers.

The objective of this chapter is to study whether novice drivers indeed overestimate their driving skills more than experienced drivers, which could indicate that novice drivers have lacking calibration skills. Both approaches for measuring self-assessment of driving skills will be compared. In the first (more traditional) approach the drivers assess their own skills relative to the 'average' or peer driver. In the second approach, the drivers' assessment is compared to a more independent measure (an assessment of an examiner). Of course a driving assessment does not provide a genuine 'objective' measure of a driver's skills, as the examiner gives a subjective opinion. But in this context the assessment is considered independent, as the measure is not influenced by the driver's opinion and attitudes about what safe driving is.

4.2. Method

4.2.1. Design

During the two years of the study the participants filled in six questionnaires and on two occasions a sub-sample participated in the on-road driving assessment (see Chapter 3 -- General method). For the purpose of the present chapter the responses to only one questionnaire and one driving assessment was required. The participants received feedback about their performance on the driving assessment, which could influence their self-assesment of skills. Therefore, the responses from questionnaire 2 (January 2006; before the driving assessment) were used for measures of self-assessed skills. For the Novice II group, who did not participate in the first two sessions, the responses from questionnaire 3 (May 2006) were used (see Table 4.1).

For the comparison of the self-assessments with the driving assessment (April 2006), this meant that: half of the novice drivers completed the questionnaire first and then participated in an on-road driving assessment, and half of the novice drivers first participated in the driving assessment and then completed the questionnaire. All experienced drivers first completed the questionnaire and then participated in the driving assessment.

Table 4.1. Design – number of participants completing the Questionnaires and/or participating in the driving assessment

	Session 2 January 2006	Driving assessment April 2006	Session 3 May 2006
Group			
Novice I	279	46	--
Novice II	--	37	142
Experienced	166	47	--

4.2.2. Participants

The total sample in this chapter consisted of 587 participants; 166 experienced drivers and 421 novice drivers. At the start of the study (September 2005) the experienced drivers had a mean age of 41 (*SD* = 5.6 years), an average of 20 years driving experience (*SD* = 5.7 years), and 49% of the experienced drivers were male. The novice drivers had only two weeks of post-license driving experience when they completed the first questionnaire; they had a mean age of 20 (*SD* = 1.8 years), 51% were male.

The sub-sample of participants in the driving assessment consisted of 130 drivers; 47 experienced and 83 novice drivers. The experienced drivers had a mean age of 41 (*SD* = 5.1 years), and 57% were male. The novice drivers had a mean age of 20 (*SD* = 1.7 years), and 52% were male.

4.2.3. Instruments

The Driver Confidence Questionnaire used in this chapter (see Appendix A), included questions on how confident the participants feel as a driver, and how much danger they perceive in traffic. The participants were also asked to compare themselves, on a five-point scale, with the average and peer driver, on their driving skill, ability to cope with hazards, and their risks to get involved in a crash. A five point scale (instead of a finer scale) was used for these questions because the most important distinction for analysis in the current study is between drivers who view themselves as (much) better, similar or worse drivers compared with the average driver. These measures from the questionnaire were used as the self-assessment of skills. See Chapter 3, General method, for a more detailed description of the Driver Confidence Questionnaire; and Appendix A for the specific questions in the questionnaire.

The ratings for 'ability to drive safely' obtained from the driving assessments (see Chapter 3) were used as a more independent measure of driving skills. The driving assessment consisted of half an hour driving on different road types. The driving assessments were conducted by three professional examiners. The examiners were asked to rate the drivers on the ability to drive safely, on a scale from 0 to 10; with 5.5 being the pass-fail criterion in a real driving test.

4.2.4. Data analysis

Respondents answered the questions about confidence and danger in traffic on a five-point scale. Some categories were combined afterwards, because of extreme low frequencies. These categorical variables were analysed using a Chi-square test. For interval data, ANOVA was used to test for significance.

4.3. Results

No significant differences were found between the novice drivers who completed the questionnaire before the driving assessment (Novice I) and drivers who did this afterwards (Novice II). These drivers were analysed as one group of novice drivers for the remaining analyses.

4.3.1. Perceived confidence and danger

There was no difference between experienced and novice drivers in the amount of danger they perceived in traffic. However, a difference was found in how much confidence drivers had in their own driving skills (see Figure 4.1). Experienced drivers had more confidence in their own driving skill than novice drivers ($\chi^2_{(3,N=587)} = 27.89$; $p < .001$)[8].

In addition, a gender effect was found in the amount of danger drivers perceive in traffic and how much confidence drivers have in their own driving skill. Male drivers perceived less danger ($\chi^2_{(2,N=587)} = 14.20$; $p < .01$)[9] and had more confidence ($\chi^2_{(3,N=587)} = 36.62$; $p < .001$) than their female counterparts.

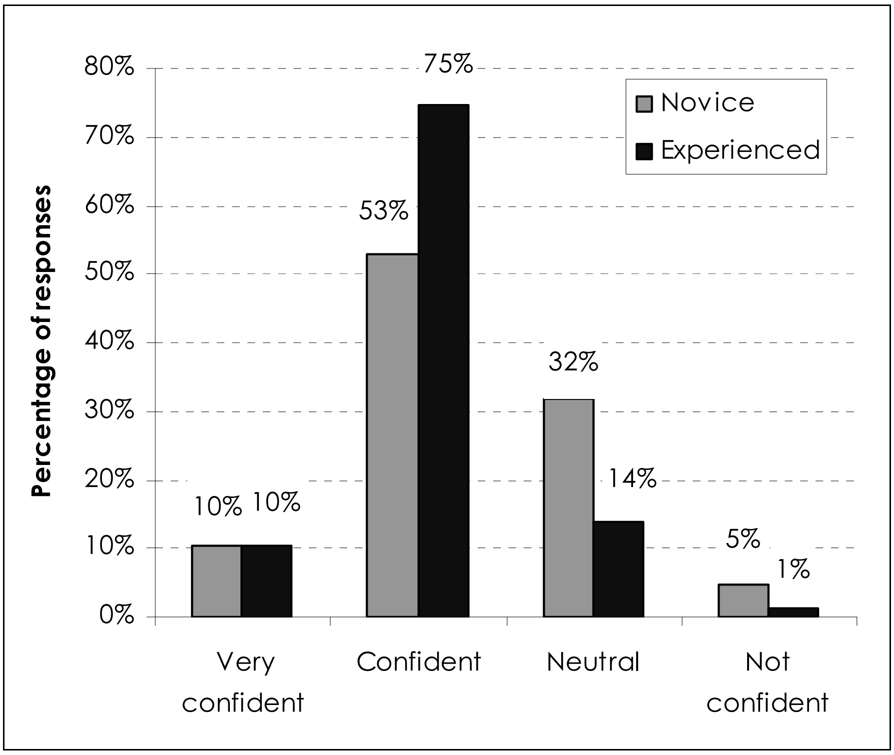

Figure 4.1. Confidence in own driving skill

[8] Due to small numbers, the categories 'Somewhat insecure' and 'Very insecure' were combined into the new category 'Not confident'.

[9] Due to small numbers the categories 'Traffic is very dangerous' and Traffic is quite dangerous' were combined; and the categories 'Traffic is very safe' and 'Traffic is quite safe' were combined.

4.3.2. Comparison 'average' driver and peers

The participants were asked to compare themselves with the average driver and with their peers. Table 4.2 shows the percentages of positive answers; for example, the percentage of drivers believing to be '(much) better drivers'.

The difference between novice and experienced drivers was tested using Chi-square analysis. The right-hand column of Table 4.2 shows the significance level of these tests. There is a difference when drivers compare themselves with the average driver or with their peers. When comparing with the average driver, significantly more experienced drivers see themselves as '(much) better drivers', having '(much) less risk in traffic', and being '(much) better at coping with hazards'. However, when comparing to peer drivers, experienced drivers are less positive and novice drivers are more positive, resulting in not much difference between novice and experienced drivers. Only for risks in traffic, novice drivers are significantly more optimistic than experienced drivers.

Table 4.2. Comparison of novice and experienced drivers to 'the average driver' and peers (analysed with Chi-square analysis)

	Novice (N = 421)		Experienced (N = 166)		
	n	%	n	%	Significance
I am a (much) better driver when I compare myself with…					
the average driver	135	32 %	85	51 %	.000
peer drivers	206	49 %	68	41 %	n.s.
I have (much) less risk in traffic when I compare myself with…					
the average driver	124	30 %	73	44 %	.001
peer drivers	186	45 %	55	33 %	.009
I am (much) better in coping with hazards in traffic when I compare myself with…					
the average driver	154	37 %	94	57 %	.000
peer drivers	186	44 %	72	43 %	n.s.

4.3.3. Comparison with expert's opinion

The self-assessments of skills and risks were also compared to a more independent measure of skills (i.e. the examiner's assessment of 'ability to drive safely'). Table 4.3 shows the statistics of this general mark for safe driving. ANOVA showed that this general mark was significantly different for experienced and novice drivers ($F_{1,126}$ = 29.34; p < .001). The effect size

($\eta 2$ = .16) indicates that this is a large effect. There was no difference between male and female drivers in the driving assessment.

Table 4.3. Mean score and pass rate

	Novice drivers	Experienced drivers
N	83	47
Mean score	5.8	6.8
Standard deviation	1.2	0.9
Pass rate	70%	94%

In order to make the measures of confidence and the examiners opinion comparable the data were converted into percentiles. That is, the total group of 130 drivers was ordered on the basis of these variables and was assigned a rank on each variable. For example, percentile 5% on driving skill indicates that the driver belongs to the worst 5% of drivers. Figure 4.2 shows a plot of these constructed percentiles. The dotted line indicates the driver's percentile on the marks provided by the examiner. The solid line represents the percentile on the confidence scale ("How much confidence do you have in your driving skill"). The higher the line, the higher the examiners mark (dotted line), or the higher one's confidence (solid line).

The whole sample was divided into four groups, each with approximately 25% of all drivers in the sample, displayed on the x-axis of Figure 4.2. All drivers scoring less than 5.5 on the driving assessment form one group. These are drivers who would have failed the driving assessment if it were an actual driving test. The second group is formed by drivers who scored between 5.5 and 6 on the driving assessment. The third group scored between 6 and 7. And the final group performed best on the driving assessment, with a score higher than 7.

Figure 4.2 shows that novice drivers with the lowest scores on the driving assessment (scoring less than 5.5) have an average confidence. The whiskers, indicating the 95% confidence interval, show that this is a significant difference between confidence and driving test result. The bottom figure shows that for experienced drivers there was no such difference. However, the experienced drivers performing best on the driving assessment also have an average confidence; they underestimate their driving skills. The larger standard deviations in Figure 4.2, indicated by the whiskers, are caused by the small numbers of novice drivers scoring higher than a 7, and the small numbers of experienced drivers, scoring lower than 5.5.

In general, the figure shows that for all drivers, there is little correspondence between the performance on the driving assessment and the driver's confidence. Whether a driver 'failed' the driving assessment or was assessed as a (very) good driver, he reported about the same level of confidence in the questionnaire.

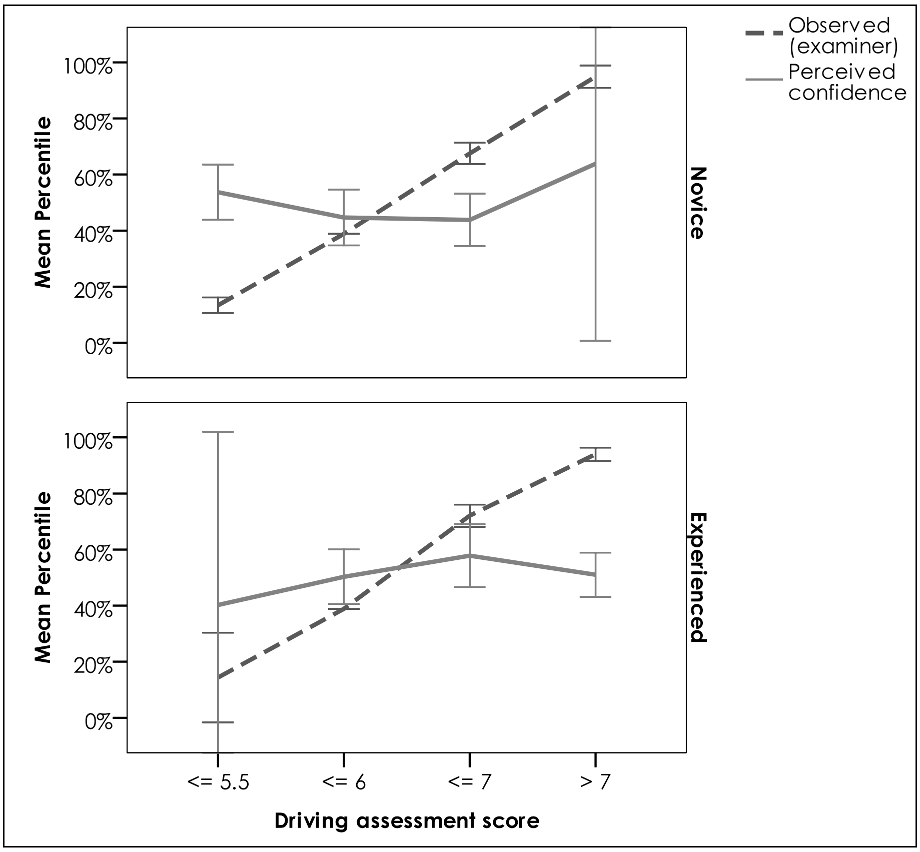

Figure 4.2. Observed driving skill versus perceived confidence

The variables from Table 4.2 were also compared to the examiner's opinion. Cronbach's alpha of .72 for the comparison with the average driver and .81 for the comparison with peers, suggested that the three variables could be added up to form one new variable. The left hand side of Figure 4.3 shows the comparison of this compound variable with the examiner's opinion. The higher the solid line, the more drivers believe they are better drivers, better in coping with hazards and are less at risk in traffic. The figure is quite similar to Figure 4.2 with a significant more optimistic view of the novice drivers with the lowest scores on the driving assessment; and a significant more moderate view of experienced drivers performing best on the driving test.

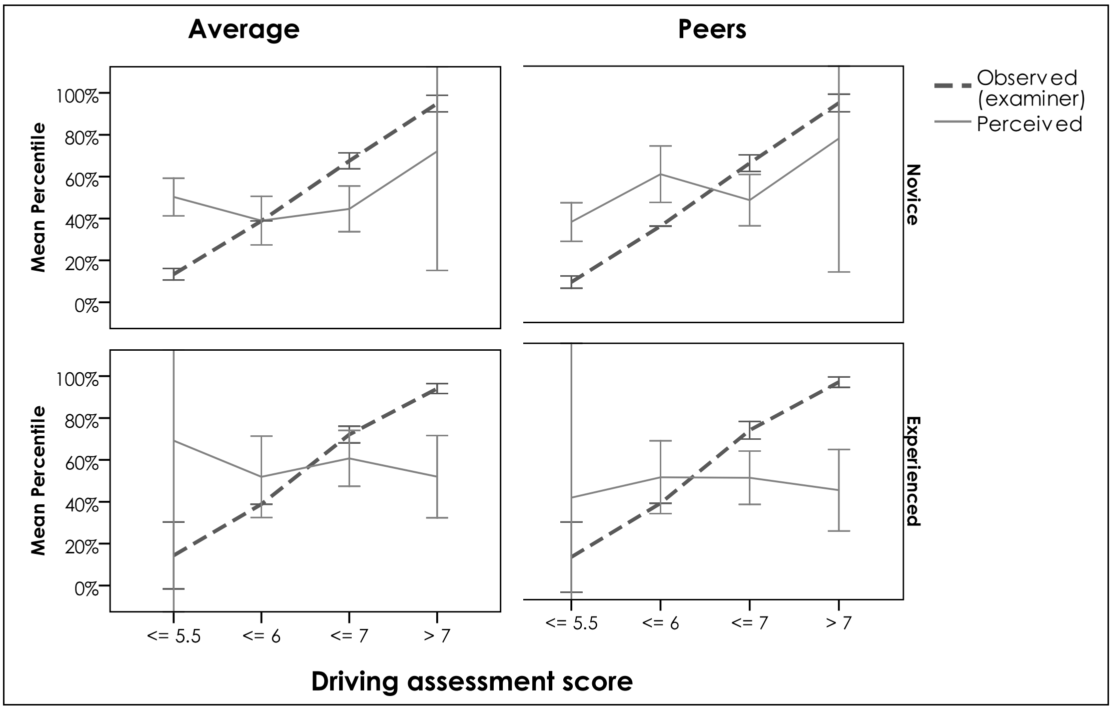

Figure 4.3. Observed driving skill versus perceived comparison with the average driver (left) and comparison to peers (right)

The right hand side of Figure 4.3 shows that when novice drivers compare themselves with their peers the figure resembles more the examiner's opinion. Nevertheless, the group that just barely 'passed' the driving test (score between 5.5 and 6) is still significantly more optimistic about their driving skills and risks in traffic. Again, all figures show that confidence level appears to be relatively independent from performance on the driving assessment.

4.4. Discussion

The objective of this chapter was to study whether novice drivers overestimate their driving skills more than experienced drivers. Two approaches were used to answer this question, and results differ for each approach. When drivers are asked to compare themselves with 'the average driver' on confidence and perceived danger in traffic, it seems that novice drivers are modest. That is, experienced drivers are more optimistic about their driving skill and risks in traffic than novice drivers. Compared with the average driver significantly more experienced drivers see themselves as '(much) better drivers', having '(much) less risk in traffic', and being '(much) better in coping with hazards'. This pattern changes when drivers compare themselves to 'peer drivers'. In this case novice drivers become more optimistic and experienced drivers are more pessimistic; and novice drivers are significantly more optimistic than experienced drivers about the risks in traffic.

The general finding reported in the literature, that the majority of drivers rate themselves to be better drivers than the average or peer driver (Delhomme, 1996; Horswill et al., 2004; McCormick et al., 1986; McKenna et al., 1991; Svenson, 1981), was not replicated in this study. Partly, this can be explained by the structure of the questionnaire. In our study, participants assessed their skills on a five point scale, whereas, for example, Svenson (1981) and McCormick et al. (1986) used respectively ten and seven point scales.

In addition, our method differs from the abovementioned studies in that we used a website to collect our data, instead of participants being in the same room with the researcher and other participants. More specifically, Svenson (1981) explicitly asked his participants to compare themselves with the other participants in the room. Groeger and Brown (1989) replicated Svenson's results, but explained most of the overconfidence as an artefact; "people seek not to respond accurately, but to appear in a good light with respect to their peers".

On the other hand, McKenna, Stanier and Lewis (1991) used a mail system to collect their data anonymously, comparable to the method in our study, and still found that drivers overestimate their skills in different driving scenarios. In conclusion, our deviating results can only partly be explained by the different methods used.

In the second approach, self-assessment of skills was compared with a more independent measure of these skills, rather than the group averages. The results indicate that novice drivers who performed worse on the driving assessment (scoring less than 5.5) have an average confidence. Experienced drivers with the lowest scores on the driving assessment were less confident about their driving skills. For experienced drivers, a difference was found between self-assessment and performance in the group who performed best on the driving assessment. These drivers were less optimistic relative to their performance on the driving assessment.

The general lack of correspondence between perceived driving skills and actual ability (see e.g. Mynttinen, Sundström, Koivukoski et al., 2009; Sundström, 2008a) that was mentioned in the introduction was replicated by the current study. In addition, the current study showed that novice drivers are worse in assessing their driving skills than more experienced drivers.

When comparing the results from the two approaches we can conclude that novice drivers are not as optimistic about their driving skills as had been thought in the past. They seem to recognize that they are not as skilled (yet) as the average driver. However, when comparing their self-assessment with their actual behaviour there are indications that they overestimate their driving skills. Because of this result and theoretical considerations (see Section 4.1), for the measurement of self-assessment of skills, it is necessary to compare the drivers' assessment to an independent measure of driving skills.

5. The development of the Adaptation Test[10]

Chapter 4 showed that *self-assessment of skills* can be measured by comparing the drivers' assessment of their driving skills to a more independent assessment of these skills.

This chapter introduces a method (the Adaptation Test) which measures *adaptation to task demands* as a function of *perceived complexity of the situation.* The Adaptation Test consists of 18 traffic scenes presented in two (almost) identical photographs, which differ in one single detail, increasing the situation's complexity. The direction of the difference in reported speed between the two pictures is used as an indication of drivers' inclination to adapt their speed to the complexity of the traffic situation.

---

[10] This chapter was based on:

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., & Brookhuis, K.A. (2008). The development of a method to measure speed adaptation to traffic complexity: Identifying novice, unsafe, and overconfident drivers. *Accident Analysis & Prevention, 40*(4), 1524-1530.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., & Brookhuis, K.A. (2009). The adaptation test: The development of a method to measure speed adaptation to traffic complexity. In: *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Big Sky Montana, USA.

The results presented in this chapter show that novice drivers performed worse on the Adaptation Test (i.e. less often reported a lower speed in the more complex situation) than experienced drivers. In addition, unsafe drivers and overconfident drivers, as identified in the on-road driving assessment, performed worse on the Adaptation Test. However, the expected correlation between performance on the Adaptation Test and self reported crashes was not found.

Because of its consistent ability to differentiate between groups of drivers (experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers), it is concluded that the Adaptation Test is effective at measuring adaptation of driving speed to the complexity of the situation.

The Adaptation Test is one of the instruments that are used in Chapter 6 to study how inadequate self-assessment of skills affects adaptation to task demands; in Chapter 7 the Adaptation Test is used to monitor the development of calibration over a two year period.

5.1. Introduction

The fact that young novice drivers' performance improves with practice may be due to increased automation in combination with improved self-assessment of skills, improved perception of the situation's complexity, and as a result, adequate adaptation to task demands (see Chapter 2 – Theoretical background and Figure 5.1). To measure the effect of perceived complexity on adaptation to task demands a new method was developed. This chapter reports the results of the development of this method: The Adaptation Test.

The Adaptation Test consists of 18 traffic scenes presented in two (almost) identical photographs (see Appendix B). The photographs differ in one single detail, thereby increasing the *complexity of the situation* (right-hand side of Figure 5.1). As the pictures were presented randomly and participants could not return to previous pictures, participants were kept unaware of the varying level of complexity. With the use of photographs, task complexity can be manipulated, which is fairly impossible in a more natural environment. In addition, the easy implementation of photographs on a website makes this test very cost-effective and applicable in a longitudinal study with many participants.

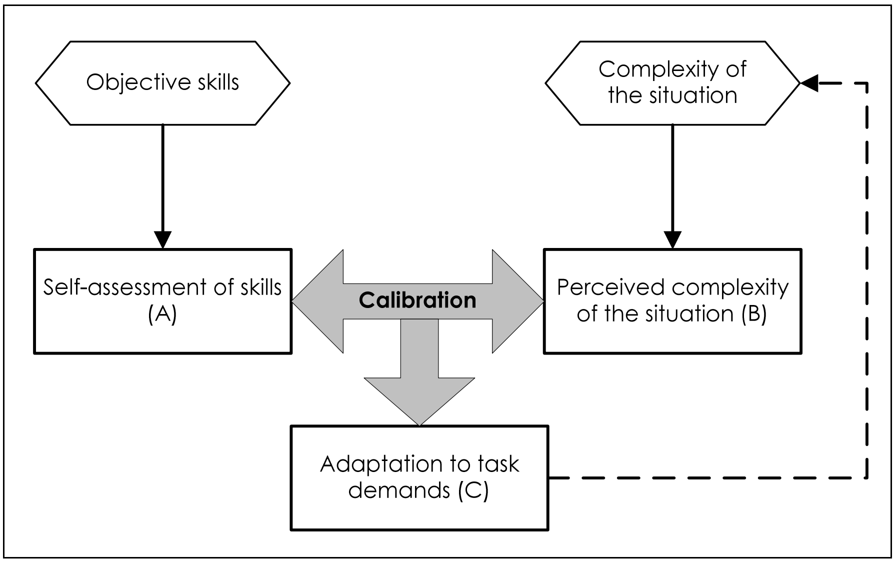

Figure 5.1. Model of the calibration process

To study the extent to which complexity affects *adaptation to task demands* (bottom Figure 5.1) we used driving speed, because reducing speed directly reduces the demands of the driving task (e.g. Fuller, 2008; Quimby & Watts, 1981). The respondents were asked to assess at what speed they *would* drive in the depicted situations. Neither in the instructions nor in the pictures was explicit reference made to the legal speed limit.

The situations were selected in such a manner that the 'extra' element would increase the complexity of the situation, without legally obliging the driver to lower his speed. In Figure 5.2, for instance, a driver is allowed to drive at the same speed, and has right of way, in both situations (see Appendix B for all situations of the Adaptation Test).

Although absolute speeds are relevant, speed assessment on the basis of photographs is unreliable (see e.g. Groeger, 2000, for information on the assessment of driving speeds). Therefore, only the direction of the difference between reported speed in the simple and complex version was used to test for differences between groups. A response was considered 'correct' if the reported speed was lower in the complex situation than in the corresponding simple situation. A higher speed in the complex situation was regarded to be equally unwanted as no speed changes; for both cases such a response was considered 'incorrect'.

Figure 5.2. An example of a simple (left) and complex (right) situation in the Adaptation Test

A basic concept in the development of the test was not to measure a driver's hazard perception skill, but the adaptation of his speed as a result of the specific situation and his beliefs about his own driving skills. For example, a driver may spot the danger (or hazard), but not decrease speed because he believes to be capable of handling the situation. Because the objective of the Adaptation Test was not to respond as quickly as possible to potential hazards, there was no time constraint for participants to respond (as is often the case in (classic) hazard perception tests (see e.g. Groeger, 2000)).

As it was important that participants did not become aware of the mechanism behind the Adaptation Test and consciously compare pairs of photos, the photos were randomly distributed throughout the web based questionnaire, without the possibility to scroll back to previous photos. A pilot study, using a similar photo judging task, revealed that participants are not aware of the differences between the photographs when they are presented in random order.

5.1.1. Hypotheses

The objective of this chapter is to evaluate the Adaptation Test. Poor calibration may explain the high crash rate of young, novice drivers, so it is assumed that novice and/or unsafe drivers will incorrectly assess their driving skills more often, will underestimate the demands of the driving task, and will consequently show less adaptation to task demands. In addition, a relationship between performance on the Adaptation Test and (self reported) crashes is expected. The Adaptation Test should meet these basic criteria:

1. In general, drivers adapt their speed to the complexity of the situation (report lower driving speeds in the more complex traffic situations);
2. Inexperienced, overconfident, or unsafe drivers perform worse on the Adaptation Test than more experienced, well-calibrated and safe drivers;
3. There is a relationship between performance on the Adaptation Test and (self reported) crashes.

5.2. Method

5.2.1. Design

The results reported in this chapter are based on the first three sessions of the longitudinal study (October 2005, January 2006 and May 2006), and the results of the first driving assessment (April 2006). With the exception of the self-reported crashes, which are all crashes reported during the complete two-year period.

Chapter 3 (General method) already mentioned the potential problem with our longitudinal design, in which novice and experienced drivers fill in questionnaires during two years. With this design, it is possible that drivers who, on a regular basis, fill in questionnaires about traffic behaviour and traffic safety will become more aware of the risks of driving than drivers who perhaps never even think about traffic safety. More specific for the Adaptation Test, it is possible that participants figure out that each situation is depicted in pairs, and try to find the differences. That is, they 'learn' how to successfully take the Adaptation Test.

In order to rule out these and other learning effects (see also Bouchet et al., 1996) as alternative explanations for our results, the novice drivers were randomly divided into two groups. A total of 297 novice drivers (Novice I) started filling in questionnaires and diaries in October 2005, the remaining 137 novice drivers (Novice II) were selected for participation in October 2005, but did not start filling in the questionnaires until May 2006 (see Table 5.1).

Table 5.1. Design – number of participants completing the Adaptation Test and/or participating in the driving assessment

	Session 1	Session 2	Driving assessment	Session 3
	(sit. 1 – 6) Oct. 2005	(sit. 7 – 12) Jan. 2006	April 2006	(sit. 13 – 18) May 2006
Group				
Novice I	297	279	46	265
Novice II	--	--	37	137
Experienced	173	166	47	163
Total	466	445	130	569

5.2.2. Participants

The sample of participants in the present chapter consisted of 607 drivers; 173 experienced drivers and 434 novice drivers. At the start of the study, the experienced drivers had a mean age of 41 (SD = 5.6 years), an average of 20 years driving experience (*SD* = 5.7 years), and 49% of the experienced drivers were male. The novice drivers had only two weeks of driving experience since passing their driving test when they filled in the first questionnaire; they had a mean age of 20 (SD = 1.8 years), 52% being male.

The sub-sample of participants in the driving assessment consisted of 130 drivers. The 47 experienced drivers had a mean age of 41 (*SD* = 5.1 years), and 57% of the experienced drivers were male. The novice drivers had only two weeks of post-test driving experience when they filled in the first questionnaire; they had a mean age of 20 (*SD* = 1.7 years), 52% being male.

See Chapter 3 (General Method) for more information about the participants in the study.

5.2.3. Instruments

The Adaptation Test

The participants assessed six new traffic situations in three consecutive sessions (see Table 5.1). The photographs of traffic situations were randomly distributed throughout the questionnaire, without the possibility to scroll back, under two conditions: 1) in half of the situations the participants viewed the complex situation first and in the other half they viewed the simple situation first; and 2) a simple and complex version of the same traffic situation never immediately succeeded each other. For the evaluation of the Adaptation Test, in this chapter, the eighteen traffic situations collected in the first three questionnaires were analysed as one.

On-road driving assessment

The results of the on-road driving assessment were used as an independent measure of driving skills (as opposed to the drivers' own assessment of their driving skills). In this driving assessment, the examiners rated drivers on 'the ability to drive safely', on a scale from 0 to 10; with 5.5 being the pass-fail mark in a real driving test. See Chapter 3 (General method) for more information on the on-road driving assessment.

5.2.4. Data analysis

The direction of the difference between reported speed in the simple and complex version was used for between group analyses. One point was assigned when the reported speed was lower in the complex situation than in the corresponding simple situation. A higher speed in the complex situation was regarded to be equally unwanted as was no speed change, and no points were assigned. So, a participant could have a maximum score of 18 in the case that all situations were assessed correctly. When for one of the photos no speed was reported, or the speed was unlikely to be correct (e.g. 555 km/h), the score was coded as 'missing' (in .3 % of all entries).

Because not all participants had 18 values (i.e. the Novice II group assessed only six situations), the percentage of correct responses was used for analysis and not the total sum of scores. This adaptation score (i.e. percentage of correct responses) was only calculated for participants with at least six valid scores and was used to test for differences between groups.

Analysis of variance of the adaptation score was used to test for significant differences between groups. For the evaluation of the situations, Repeated Measures ANOVA with reported speed on the simple and complex situation (i.e. not the adaptation score) as the within-subjects-factor was used.

5.3. Results

No significant differences were found between the Novice I and Novice II groups concerning the questionnaire completed in May 2006, indicating that there is no evidence of a learning effect in this study. Thus, the Novice I and Novice II group were considered as one group of novice drivers for the following analyses. The adaptation score was less sensitive for the Novice II group compared to the Novice I group (i.e. based on 6 situations instead of 18). As a result, the significance tests became more conservative (i.e. smaller chance of finding a significant result when there is, in fact, a difference).

5.3.1. Evaluation of the situations

Table 5.2 shows the mean and standard deviation of the reported speed for the eighteen situations in the Adaptation Test. Repeated Measures ANOVA , with reported speed on the simple and complex situation as within-subjects-factor, was carried out for all situations and for each of the eighteen situations separately. The results showed a main effect of complexity; the mean reported speed was generally lower in the complex situations than in the simple situations ($F_{1,603}$ = 655.10; p < .001), with a large effect size ($\eta_P^2 = 0.52$).

Table 5.2. Mean and standard deviation of the reported speed (km/h) in the eighteen situations of the Adaptation Test

		Simple		Complex	
Situation	Legal limit	Mean	SD	Mean	SD
1	50	41.9	8.1	37.6	9.2
2	50	38.2	8.5	36.0	9.8
3	50	38.9	10.6	31.1	12.3
4	50	32.2	8.5	30.5	9.4
5	50	43.6	8.3	41.5	8.8
6	120	102.1	14.9	92.6	18.0
7	50	46.5	7.4	44.3	8.3
8	80	82.5	12.3	65.7	13.4
9	30	31.5	8.4	29.0	7.7
10	50	47.7	8.7	46.3	10.3
11	50	40.0	9.0	28.6	11.4
12	80	74.4	16.1	74.2	14.9
13	80	66.4	13.6	61.1	14.0
14	50	31.8	10.0	32.9	9.7
15	50	46.1	8.2	44.5	8.3
16	50	44.7	9.8	35.6	11.9
17	50	36.1	9.1	34.6	8.9
18	50	48.2	8.9	42.1	10.7

Repeated Measures ANOVA for each situation separately showed that the effect of complexity on adaptation was significant (p < .01) in sixteen out of eighteen situations. In situation 12, depicting a rural road (speed limit 80 km/h; see Appendix B), no significant difference was found between the complex and simple situation whereas in situation 14, depicting an urban situation (speed limit 50 km/h; see Appendix B), a significant difference was found in the unexpected direction.

These two situations were not excluded from the analyses, because no definite explanation could be given as to why these two situations produced a different result than expected. Moreover, excluding the situations from the Adaptation Test would impair the continuity of the longitudinal study. The two situations cause the Adaptation Test to be less sensitive, and therefore making the statistical tests more conservative.

5.3.2. Experience

Figure 5.3 shows the adaptation score for the experienced and novice driver groups, for both male and female drivers. The solid lines indicate the median, the boxes the interquartile range, and the whiskers indicate the minimum and maximum values.

The analysis of variance shows that there is a significant main effect of both experience ($F_{1,603}$ = 26.77; p < .001) and gender ($F_{1,603}$ = 8.42; p < .01) on the adaptation score. Experienced drivers performed better (45% correct responses) than novice drivers (36% correct responses). Males performed better than females (respectively 43% and 38% correct responses). The effect size is considered small for both experience level (η_P^2 = .04) and gender (η_P^2 = .01). No significant interaction was found between experience level and gender.

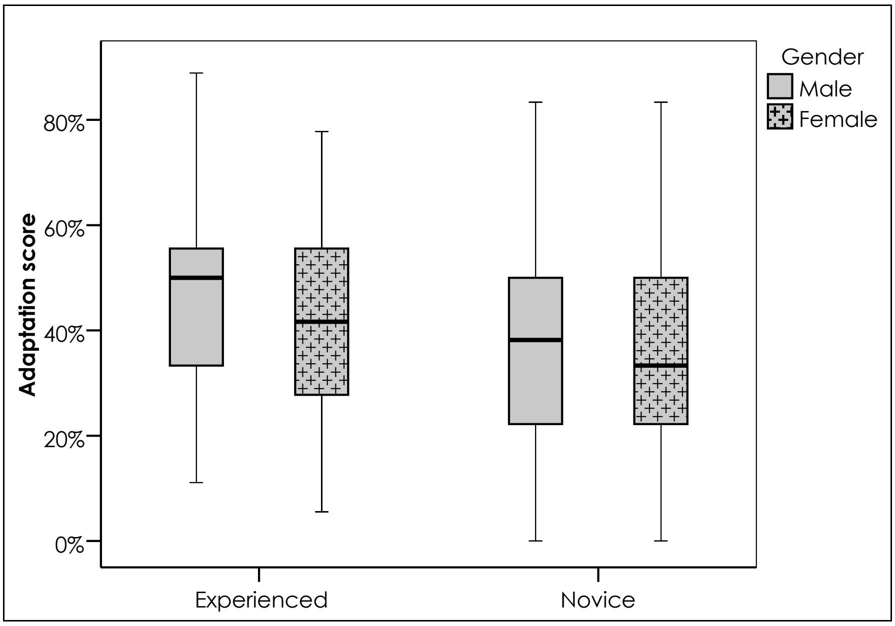

Figure 5.3. Effect of experience level and gender on adaptation score; higher percentage reflects better adaptation

5.3.3. Driving skills

Pearson correlation coefficient between the examiners' 'ability to drive safely' score and the adaptation score from the photograph assessments indicated a moderate but significant correlation ($r_{N=130}$ = .31; p < .01). Drivers who were rated higher by the examiners performed significantly better on the Adaptation Test.

Based on the scores provided by the examiners, two types of drivers were distinguished, safe and unsafe drivers. Drivers who scored less than 5.5 were considered to be unsafe drivers, and would have failed if the driving assessment was an actual driving test. Note that the distinction between 'safe' and 'unsafe' drivers is based on the assessment; there is no information whether these drivers actually portray safe or unsafe driving behaviour in every day life. Using this criterion, 28 drivers were considered unsafe drivers and 102 drivers were considered safe drivers.

The unsafe drivers responded correctly in 28% of the situations. The safe drivers gave correct responses in 42% of the situations. Figure 5.4 shows the median and (interquartile) range of the adaptation score for safe and unsafe drivers. The analysis of variance showed that there was a moderate (η_p^2 = .09) and significant difference ($F_{1,128}$ = 12.45; p < .01) between safe and unsafe drivers concerning the way they assessed the depicted traffic situations. No interaction effect between gender and experience was found.

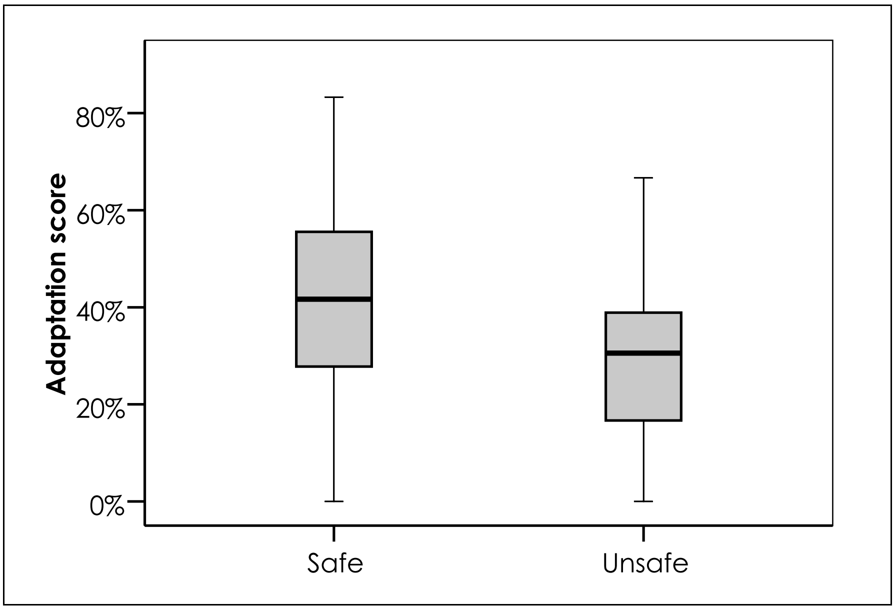

Figure 5.4. Effect of safe driving on adaptation score

5.3.4. Self-assessment of skills

In the questionnaire, the participants judged, on a five-point scale, how confident they were about their own driving skills (see Chapter 3 – General method). Based on these scores[11], participants were divided into two groups: a) high confidence drivers (drivers who said they were (very) confident); and b) low confidence drivers (drivers who responded with neutral confidence or no confidence at all).

Table 5.3 shows the cross tabulation of the drivers' own confidence concerning their driving skills and the categorization into safe and unsafe drivers based on the examiners' safe driving scores. This cross tabulation results in three 'calibration' groups: a) 83 (75 + 8) well-calibrated drivers (share the examiner's opinion about their driving skills), b) 20 overconfident drivers (were confident about their driving skills, but failed the driving assessment) and c) 27 insecure drivers (were less confident about their driving skills, but passed the driving assessment).

Table 5.3. Calibration groups – based on the first questionnaire

	On road driving assessment		Total
	Passed	Failed	
Confidence			
High	75 (58%)	20 (15%)	95 (73%)
Low	27 (21%)	8 (6%)	35 (27%)
Total	102 (79%)	28 (21%)	130 (100%)

Chi-square analysis indicated that the group of experienced drivers consisted of significantly more well-calibrated drivers (79%) than the group of novice drivers (55%) ($\chi^2_{(1,N=130)} = 7.06$; $p < .01$). Within the group of novice drivers 21% were overconfident, compared to 6% of the experienced drivers, while 24% were insecure, compared to 15% of the experienced drivers. There was no significant difference between males and females in the three calibration groups. There was some overrepresentation of females in the insecure group, but not in the overconfident group. These calibration groups were used in a further analysis of the assessment of situations depicted on the photographs.

Figure 5.5 shows median and (interquartile) range of the adaptation score for the three calibration groups. A moderate (η_p^2 = .09) and significant difference

[11] The scores of the first questionnaire that a participant filled in were used. This means that for the Experienced drivers and Novice I group the responses of the first questionnaire (October 2005) were used, and for the Novice II group the responses of the third questionnaire (May 2006) were used.

between the calibration groups ($F_{2,127}$ = 6.46; $p < .01$) was found. The well-calibrated drivers scored an average of 43% correct responses, the insecure group scored 38%, whereas the overconfident group of drivers scored worst with only 26% correct responses.

Post Hoc Bonferroni tests show that the overconfident drivers significantly differed from the well-calibrated drivers and the insecure group ($p < .01$). No significant interaction effect was found between calibration groups and gender.

The calibration groups were constructed using the scores on the driving assessment and confidence level. So in theory it is possible that the difference in well-calibrated versus insecure and overconfident drivers can be explained completely by the performance on the driving assessment, and that confidence level does not influence this effect. In this case there would not be an interaction effect between driving assessment score and confidence level. Analysis of variance showed a main effect of passing or failing the driving assessment on adaptation score ($F_{1,126}$ = 6.36; $p < .01$) but no main effect of confidence on adaptation score. A small (η_p^2 = .02) interaction effect was found, however, due to low power (30%), only 8 drivers failed the assessment *and* had predicted this, this effect was not significant.

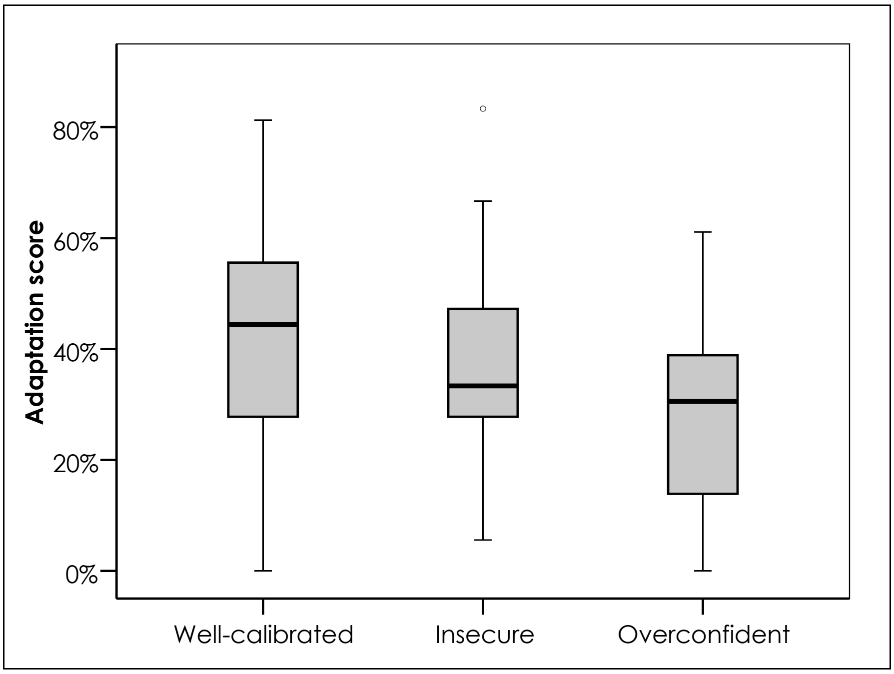

Figure 5.5. Effect of self-assessment on adaptation score

5.3.5. Relationship with self-reported crashes

During the two years of the study 121 participants reported a total of 145 crashes. A crash was defined as an incident that resulted in at least material damage or personal injury. Table 5.4 shows that there was no difference between males and females in the number of reported crashes. Novice drivers reported significantly more crashes than experienced drivers ($\chi^2_{(1,N=585)} = 19.17$; $p < .001$). Finally, no difference in performance on the Adaptation Test was found between drivers who had a crash and who had not reported a crash. Both groups had an average of 40% correct responses to the situations.

Table 5.4. Total number of self-reported crashes in two years

		No crash		One or more crashes	
Subgroups	N	n	%	n	%
Gender					
Males	292	233	80 %	59	20 %
Females	293	231	79 %	62	21 %
Experience					
Experienced	166	151	91 %	15	9 %
Novice	419	313	75 %	106	25 %
Total	585	464	79 %	121	21 %
% correct responses on Adaptation Test		40 %		40 %	

5.4. Discussion

The objective of this chapter was to evaluate a new methodology, using traffic situations depicted on photographs, to measure adaptation of driving speed (the Adaptation Test). The results of the reported mean speed show that drivers reported a lower speed for the complex situation in sixteen out of eighteen situations. It was expected that novice and/or unsafe drivers incorrectly assess their driving skills and the demands of the driving task more frequently and consequently adapt their speed less often to the situation than experienced and/or safe drivers. For the analysis of the differences between these groups the adaptation score was calculated. If a participant reported a lower speed in the more complex situation, he gained one point, while no 'negative' points were assigned. Analysis of these adaptation scores indicate that novice drivers are worse at adapting speed to the situation than experienced drivers.

A sub-sample of 130 drivers (83 novice drivers and 47 experienced drivers) participated in an on-road driving assessment. A significant positive correlation between performance on the on-road driving assessment and assessment of the traffic situations with the photographs was found, indicating that drivers with a high score for 'safe driving' adapt their speed more often to the situation. When distinguishing between safe and unsafe drivers, based on the performance on the driving assessment, unsafe drivers responded correctly (i.e. reported a lower speed in the more complex situation) to 28% of the situations, whereas, safe drivers responded correctly to 42% of the situations.

The performance on the on-road driving assessment was used to construct three self-assessment groups: well-calibrated drivers, overconfident drivers, and insecure drivers. The results on assessing traffic situations were as expected: well-calibrated drivers performed best, the insecure group performed somewhat worse, while the overconfident drivers performed significantly worse on the adjustment to traffic situations. This is consistent with previous research where overconfident drivers, rather than insecure drivers, were described as a problem group (Gregersen, 1996; Mathews & Moran, 1986).

Because the calibration groups were constructed using the scores from the driving assessment and confidence ratings, in theory, the results could have been caused by one of these factors (and not the combination). Because the power to test for interaction effects was too low (i.e. too few participants in aggregated groups) this could not be tested.

Ideally we would have expected a relationship between the Adaptation Test and self-reported crashes; with lower scores on the Adaptation Test correlating with more crashes. Although, as expected novice drivers reported more crashes than experienced driver, no correlation between performance on the Adaptation Test and self-reported crashes was found. A closer look at the reported crashes revealed that these mostly concern minor crashes (e.g. a fender bender on a parking place). It is possible that these crashes were caused by poor vehicle handling skills, while the Adaptation Test measures a higher-order-skill: adaptation of speed to the traffic situation. Nevertheless, this is only speculating. It is impossible to assess in hind sight whether the reported crash was either caused by poor vehicle handling skills, or for example that the driver lost control over the car because of a wrong speed choice.

Experienced drivers and safe drivers were expected to be better at assessing their skills, i.e. more aware of the limits of their skills. All three subgroups, experienced drivers, safe drivers, and drivers who correctly assessed their driving skills, indeed performed better at adapting speed to the situation.

However, there was some inconsistency as well. Males were significantly better at adapting speed to the situation than female drivers. This is an unexpected result, since (young) males have a higher crash risk than (young) female drivers (OECD - ECMT, 2006).

In addition, there was no significant difference between males and females in the calibration groups. So, although the difference between males and females on the adaptation score was small (η_P^2 = .01), there appears to be a gender effect that cannot be explained by the self-assessment of skills. Unfortunately the sample was too small to find significant interaction effects between self-assessment groups and gender.

With respect to the newly developed method, two points deserve attention. First of all, the test uses self-reported speed to measure adaptation to task demands which is not the same as actual speed behaviour. However, using reported speed enables complete control over the complexity level of the traffic situation. In the current study, we take into account that reported speed is not the same as actual speed choice; therefore, the between group analyses are based on the difference between the reported speed in the complex and simple traffic situation, rather than absolute speed.

Secondly, none of the groups performed very well on the test. Of all the subgroups, experienced drivers performed best on the test, but still only scored ‘correctly’ on 45% of the situations; in more than half of the situations, even experienced drivers do not report a decreased speed for the more complex situation. A closer look at the situations revealed that often the same speed is reported for both the simple and complex situation. Despite explicit instructions that the speed limit in the traffic situations should be ignored, reported speed often fluctuated around the speed limit. It seems that for some traffic situations the road design determines speed choice more than the specific circumstances do (i.e. complexity of the situation). In addition, it is also possible to deal with the increased complexity without changing speed (i.e. increasing vigilance, preparedness to respond or with dropping subsidiary tasks). This response is not measured with the Adaptation Test.

It can be argued that there was not enough difference between the simple and complex photograph. Increasing this difference could indeed lead to more reported differences between the complex and simple version of the traffic situation. However, it is not so obvious that this would improve the test in differentiating between novice and experienced drivers. Because the complexity would be more obvious, novice drivers would probably also be better at recognizing this difference. Moreover, the photographs were constructed with the restriction that in both photographs the priority situation and legal limit was the same. With this restriction, the differences between the simple and complex photographs had to be a subtle one. In any case, when a different speed was reported, this was almost always in the correct direction, i.e. a decreased speed for the more complex situation.

Overall, the results indicate that the Adaptation Test can be used to measure adaptation of driving speed to the situation. The test is not very sensitive in differentiating between individual drivers (i.e. even experienced drivers do not report a decreased speed for the complex situation in more than half of the situations), and no correlation was found with self-reported crashes.

However, because of its consistent ability to differentiate between groups of drivers (experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers), the Adaptation Test is considered a useful method for assessing calibration.

6. The effect of self-assessment of skills on adaptation to task demands[12]

This chapter describes if and to what extent inadequate *self-assessment of skills* (Chapter 4) affects *adaptation to task demands*. This is measured with the Adaptation Test, which was described and evaluated in Chapter 5; and with responses to the questionnaire introduced in Chapter 3.

The results show that overconfident drivers reported significantly more violating behaviour than well-calibrated and insecure drivers. Overconfident drivers also showed significantly less instances of adaptation to traffic complexity, as measured with the Adaptation Test. Finally, 50% of the overconfident drivers reported one or more crashes in the following year; which was significantly more than the well-calibrated and insecure drivers (respectively 18% and 10%).

[12] This chapter was based on:
De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., & Brookhuis, K.A. (2007b). Overestimation of skills affects driver's adaptation to task demands. In: *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Stevensen, Washington, USA.

The conclusion of this chapter is that overconfidence is related to inadequate adaptation to task demands. Both elements (overconfidence and inadequate adaptation) are used in the next chapter (Chapter 7) to describe how calibration develops over time.

6.1. Introduction

An important assumption of the calibration model (see Figure 6.1) is that inadequate self-assessment of skills results in an insufficient adaptation to the task demands. For example, a driver who thinks of himself to be a very skilled driver is less likely to reduce his speed when it starts to rain.

Available research shows that this assumption is not conclusive. Some findings support the assumption that inadequate self-assessment of skills results in an insufficient adaptation to the task demands. For example, a study by Quimby and Watts (1981) showed a significant correlation between the subjects' own median risk ratings on a test drive and their safety index (calculated from stopping distance and the forward visibility at locations during the test drive). This correlation indicates that those drivers whose speeds resulted in the greatest risk taking behaviour, tended to consider the danger or risk to be low.

On the other hand, Horswill, Waylen and Tofield (2004) found little evidence for a relationship between self-assessment of skills and behavioural adaptation. They did find that drivers rated themselves superior to their peers ánd the average driver. But this did not affect three risk taking intentions: 1) self reported speeding behaviour; 2) speed choices assessed with four photographs of different road situations; and 3) car following behaviour assessed with a photograph of a motorway scene.

This chapter studies whether inadequate self-assessment of skills (left-hand side of Figure 6.1) results in an insufficient adaptation to the task demands (bottom of Figure 6.1). An important aspect of the calibration model is that the balancing and adaptation can take place on different hierarchical levels of the driving task (e.g. Fuller & Santos, 2002; Rasmussen, 1986). This thesis distinguishes adaptation on the two highest levels of Michon's (1985) driving task hierarchy. First, on the *strategic* level, the driver might avoid a difficult junction or choose (not) to drive in the dark. Second, on the *manoeuvring* level, a driver might not sufficiently reduce speed or increase headway, when encountering more complex situations.

The influence of inadequate self-assessment on reported violating behaviour is also studied, which can be viewed as both strategic, as it involves a strategic choice (not) to violate traffic rules, and manoeuvring, as the specific situation has a large influence on driver behaviour (Reason, 1990). The lowest level of the driving task hierarchy (i.e., the control or vehicle handling level) was not considered here, as adaptation to task demands is assumed to involve a conscious assessment.

Finally, this chapter studies if (inadequate) self-assessment of skills has an effect on the number of self reported crashes in the following year of the study.

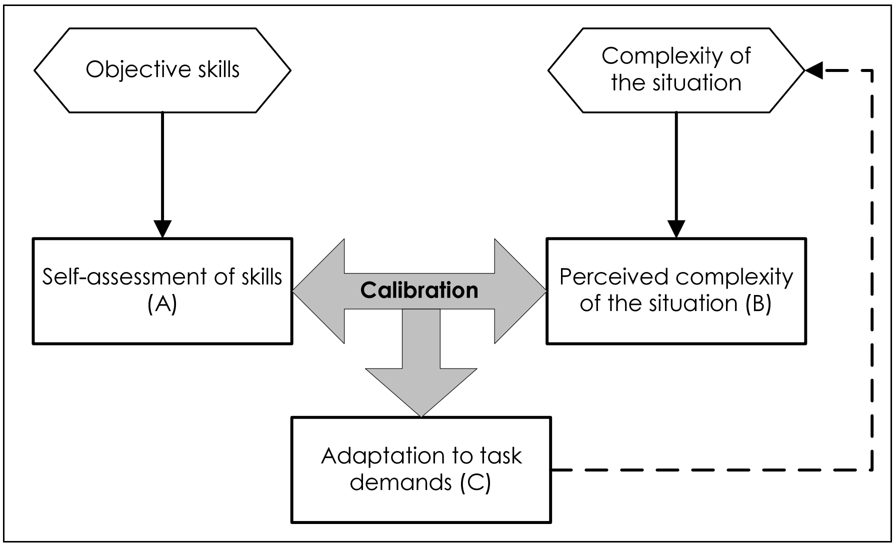

Figure 6.1. Model of the calibration process

6.2. Method

6.2.1. Design

During the two years of the study the participants filled in six questionnaires and on two occasions a sub-sample participated in the on-road driving assessment (see Chapter 3 -- General method). For the purpose of the present chapter the responses to only one questionnaire and one driving assessment was required. For the Novice I and the experienced drivers, the data from session 2 (January 2006) were used. The Novice II group did not participate in the first two sessions, therefore the data from session 3 (May 2006) were used (see Table 6.1).

For the comparison of the self-assessments with the driving assessment (April 2006), this meant that: half of the novice drivers completed the questionnaire first and then participated in an on-road driving assessment, and half of the novice drivers first participated in the driving assessment and then filled in the questionnaire. All experienced drivers first completed the questionnaire and then participated in the driving assessment.

There was one exception. The analysis of the Adaptation Test (see Chapter 5) was based on the assessment of the 18 traffic situations, gathered in the first three sessions.

Table 6.1. Design – number of participants completing the Questionnaires and/or participating in the driving assessment

	Session 2 January 2006	Driving assessment April 2006	Session 3 May 2006
Group			
Novice I	279	46	--
Novice II	--	37	137
Experienced	166	47	--
Total	445	130	137

6.2.2. Participants

The sample of participants in the present chapter consisted of the 130 drivers who participated in the first driving assessment. At the start of the study the 47 experienced drivers had a mean age of 41 (*SD* = 5.1 years), and 57% of the experienced drivers were male. The novice drivers had only two weeks of post-license driving experience when they filled in the first questionnaire; they had a mean age of 20 (*SD* = 1.7 years), 52% being male. See Chapter 3 (General Method) for more information about the participants in the study.

6.2.3. Instruments

On-road driving assessment

Examiners rated drivers on their ‘ability to drive safely’ on a scale from 0 to 10; 5.5 being the pass-fail criterion in a real driving test. As in a regular driving test, this general mark for ‘safe driving’ is a combination of many different skills on different levels of the driving task (e.g., to *manoeuvre* safely through traffic, a driver must possess a certain level of vehicle *control*). See Chapter 3 (General Method) for more information about the on-road driving assessment.

Questionnaire

The questionnaire used in this study (see Chapter 3 – General method), contained questions on how confident the participants are as a driver. A five-point-scale was used for these questions because the most important distinction for analysis in the current study is between drivers who report to be (very) confident, neutral or (very) insecure. The responses to the 'driver confidence' question compared with the examiners assessment for 'safe driving' was used to construct a measure for self-assessment of skills (see also Chapter 4 – Self-assessment of skills).

The questionnaire also contained questions regarding the adaptation to task demands. Relating to the strategic level, drivers were asked: a) whether, in the last four months, they had cancelled an intended driving trip due to adverse conditions (e.g., darkness or weather conditions); and b) whether, in the previous four months, they had avoided a difficult intersection.

To measure adaptation to task demands on the manoeuvring level the Adaptation Test was used (see Chapter 5 – The development of the Adaptation Test).

Finally, the questionnaire contained an abbreviated Dutch version (Verschuur, 2003a; Verschuur & Hurts, 2008) of the Driver Behaviour Questionnaire (DBQ, see Parker, Reason et al., 1995). See also Chapter 3 – General method and Appendix C. The 8 items of the DBQ indicating violating behaviour were used to measure the effect of inadequate self-assessment on violating behaviour, which relates to the strategic level as well as the manoeuvring level.

To study the relationship between self-assessment of skills and crash risk, the self reported crashes in the second year of the study were analysed. Only the second year was used to limit confounding of results. The experience of a crash could change a driver's self-assessment of skills. For example, it could make an overconfident driver less optimistic about his driving skills. Therefore, the self-assessment of a driver was based on the reports in the first year of the study (questionnaire 2 and 3) and the self reported crashes were from the second year (questionnaire 4, 5 and 6).

6.2.4. Data analysis

Self-assessment was computed by comparing the examiner's rating for 'safe driving' with the participants' rating of 'driver confidence'. Chi-square analysis was used to test for significant differences in frequencies. Univariate Analysis of Variance (ANOVA) was used to test for significant effects on the score from the Adaptation Test. The DBQ items were analysed using Multivariate Analysis of Variance (MANOVA) with Pillai's Trace criterion.

6.3. Results

6.3.1. Calibration groups

The variable 'driver confidence' was used to distinguish two driver groups: a) high confidence drivers (drivers who said to be (very) confident); and b) low confidence drivers (drivers who responded with neutral or not (at all) confident). Next, the examiner's 'safe driving' score was used to divide the drivers into two groups: a pass group, consisting of those who would have passed the driving test (scoring 5.5 or higher) and a fail group consisting of those who would have failed (scoring less than 5.5). Table 6.2 shows a cross tabulation of 'driver confidence' and the examiner's 'safe driving' score. The cross tabulation results in three calibration groups[13]: a) 88 (82 + 6) well-calibrated drivers (share the same opinion about their driving skills), b) 20 insecure drivers (are not confident about their driving skills, but passed the driving assessment) and c) 22 overconfident drivers (are confident about their driving skills, but failed the driving assessment).

The reported results in this chapter are fairly robust for the pass-fail criterion for 'safe driving'. That is, the results are similar when a pass-fail criterion of, for example, 6 (in stead of 5.5) is used.

Table 6.2. Calibration groups – based on the second and third questionnaire ([a] = well-calibrated drivers; [b] = insecure drivers and [c] = overconfident drivers)

	Safe driving		Total
	Passed	Failed	
Driver confidence			
Confident	82[a]	22[c]	104
Insecure	20[b]	6[a]	26
Total	102	28	130

Table 6.3 shows some characteristics of the three calibration groups. Chi-square analysis indicated that the group of experienced drivers consisted of significantly more well-calibrated drivers (85%) than the group of novice drivers (58%); ($\chi^2_{(1,N=130)} = 10.21$; $p < .01$).

[13] This classification into calibration groups differs somewhat from the classification in Chapter 5 (The development of the Adaptation Test). In Chapter 5 the "driver confidence" ratings from the first questionnaire were used to calculate the calibration groups; in the present chapter the ratings from the second and third questionnaire were used.

The difference between males and females in the three calibration groups was also significant ($\chi^2_{(2,N=130)}$ = 8.83; p < .05); 77% of the males were well-calibrated drivers; of the female drivers only 57% were well-calibrated. Female drivers were significantly overrepresented in the insecure group (25% of the female drivers compared to 7% of the male drivers; $\chi^2_{(1,N=130)}$ = 7.91; p < .01). No gender effect in the overconfident group was found.

Finally, Table 6.3 shows some difference between calibration groups in the number of hours that they reported driving in one week in the first year of the study. ANOVA, however, indicated that this difference was not significant.

Table 6.3. Characteristics of calibration groups

	n	Well-calibrated		Insecure		Over-confident	
Group							
Experienced	47	85%		11%		4%	
Novice	83	58%		18%		24%	
Gender							
Male	70	77%		7%		16%	
Female	60	57%		25%		18%	
		Mean	SD	Mean	SD	Mean	SD
Number of days driving in one week		4.1	2.0	4.1	1.7	4.3	2.0
Number of hours driving in one week		8.0	6.8	5.9	5.2	8.3	6.7

6.3.2. Adaptation to task demands

Strategic level

The questionnaire contained two questions concerning adaptation to task demands on the strategic level. Drivers were asked: a) whether, in the last four months, they had cancelled an intended driving trip due to adverse conditions (e.g., darkness or weather conditions); and b) whether, in the previous four months, they had avoided a difficult intersection. With respect to the first question, results show that overall 25% of the drivers reported 'not driving because of adverse conditions'; no difference was found between the well-calibrated group and the other two groups.

A larger proportion of drivers (44%) reported having avoided a difficult intersection. However, further analyses showed that drivers avoided such intersections because of traffic congestion, not because of traffic complexity. Consequently, there was no difference between calibration groups. Also no effect of 'driver confidence' or performance on the driving assessment ('safe driving') on these variables was found.

Manoeuvring level

ANOVA showed that there was a moderate (η^2 = .09) and significant difference between calibration groups ($F_{2,124}$ = 6.15; p < .01) in the responses to the Adaptation Test[14]. Post Hoc Bonferroni tests indicated that overconfident drivers adapted their speed to the complex situation significantly less often (26% correct responses; p < .01) than well-calibrated drivers (42% correct responses) and insecure drivers (36% correct responses). No significant interaction effect was found between calibration groups and gender.

Unsafe drivers, as determined in the driving assessment, responded correctly to the Adaptation Test in 28% of the situations. The safe drivers gave correct responses in 42% of the situations. ANOVA showed a moderate (η_P^2 = .09) and significant difference ($F_{1,128}$ = 12.45; p < .01) between safe and unsafe drivers concerning the way they assessed the depicted traffic situations. No effect of 'driver confidence' was found on the Adaptation Test.

Violations

The eight items of the DBQ indicating violating behaviour (see Table 6.4) were used to study whether well-calibrated drivers reported less dangerous decisions. MANOVA, with calibration group and gender as fixed factors, indicated that there was a large (η^2 = .13) and significant difference between calibration groups ($F_{16,236}$ = 2.11; p < .01) in their reported violating behaviour. The overconfident group generally reported more violating behaviour than the other groups.

Tests of between-subject effects specify which of the eight items contribute to the general multivariate effect. In this case, tests of between-subject effects indicate that, except for being angered by another driver (item 4), and having an aversion to a class of road users (item 7), all items contribute significantly to this difference between calibration groups[15].

[14] These results are slightly different from the results reported in Chapter 5 (The development of the Adaptation Test), due to the different classification into calibration groups.

[15] NB: the asterisks in Table 6.4 do *not* indicate these between-subject effects

Table 6.4. Mean score and Bonferroni Post Hoc tests on DBQ items

	Mean score		
	Well-calibrated	Insecure	Overconfident
Become impatient with a slow driver in the outer lane and overtake on the inside	2.0	2.1	2.8*
Drive especially close to the car in front as a signal to its driver to go faster or get out of the way	1.9	2.0	2.6*
Cross a junction knowing that the traffic lights have already turned against you	1.7	1.8	2.3*
Angered by another driver's behaviour, you give chase with the intention of giving him/her a piece of your mind	1.2	1.1	1.1
Disregard the speed limits late at night or early in the morning	2.8	3.6	3.5
Drive even though you realise that you may be over the Legal blood-alcohol limit	1.2	1.1	1.6*
Have an aversion to a particular class of road user, and indicate your hostility by whatever means you can	1.5	1.6	2.0
Get involved in unofficial 'races' with other drivers	1.4	1.2	1.9

NB: * indicates a significant difference (α = .05) with well-calibrated drivers

The results from the multivariate analysis indicate that there is a difference between calibration groups in reported violating behaviour. But it is also important to know which group differs from the other two groups. Bonferroni Post Hoc tests were used for this purpose. Table 6.4 shows the mean score for each group on the 8 DBQ items indicating how often they would indulge in that behaviour (0 = Never; 5 = Nearly all the time). Significant deviations (α = .05) from the well-calibrated group, based on Bonferroni Post Hoc tests, are indicated by an asterisk. Note that the between-subjects-effect in the previous section specified which *items* contributed to the significant multivariate effect. The Post Hoc tests specify which *group* contributed to the effect.

The asterisks in Table 6.4 indicate where the overconfident drivers differ significantly from the well-calibrated group. The lack of asterisks in the insecure column indicates that there was no significant difference between the insecure group and the well-calibrated drivers.

No main effect of gender was found on the DBQ-items. However, the interaction of gender and calibration group was significant ($F_{16,236}$ = 2.16; $p < .01$). Tests of between–subject effects showed that the interaction effect was exclusively caused by 'driving over the Legal blood-alcohol limit' (item

6). Only in the overconfident group, male drivers reported significantly more drink and drive behaviour than females. In the well-calibrated and insecure group, no effect of gender was found.

The items of the DBQ were further analysed for the effects of 'driver confidence' and performance on the driving assessment ('safe driving'). Confident drivers reported more violations than less confident drivers ($F_{8,119} = 2.47$; $p < .05$). However, tests of between–subject effects showed that only item 8: 'Getting involved in unofficial races with other drivers', contributed significantly to this multivariate effect. No effect of 'safe driving' was found.

6.3.3. Relationship with self-reported crashes

Of the 130 participants in the driving assessment, for which the calibration groups were constructed, a total of 29 drivers reported one or more crashes (with damage and or personal injuries) in the second year of the study. Two experienced drivers reported a crash, while 27 of the novice drivers reported a crash.

Table 6.5 shows differences between the calibration groups in the number of reported crashes in the second year of the study. Only 10% of the insecure drivers reported a crash and 18% of the well-calibrated drivers reported a crash. However, 50% of the overconfident drivers reported a crash. Chi-square analysis indicated that this difference between calibration groups was significant ($\chi^2_{(2,N=130)} = 12.35$; $p < .01$).

For comparison purposes, Table 6.5 also shows the number of self-reported crashes for drivers that either 'passed' or 'failed' (scoring lower than 5.5) the driving assessment. Chi-square analysis indicated that this difference is also significant ($\chi^2_{(1,N=130)} = 15.79$; $p < .001$). Of the drivers who passed the driving assessment only 15% reported one or more crashes, compared to the drivers who failed the driving assessment, of which 50% reported one or more crashes.

Finally, Table 6.5 shows the number of insecure and confident drivers (according to the questionnaire responses) that reported one or more crashes. This difference was not significant according to the Chi-square analysis.

Table 6.5. Number of self reported crashes in the second year of the study for each of the three calibration groups

	No crash		One or more crashes	
	n	%	n	%
Calibration group				
Well-calibrated	72	82%	16	18%
Insecure	18	90%	2	10%
Overconfident	11	50%	11	50%
Result driving assessment				
Passed	87	85%	15	15%
Failed	14	50%	14	50%
Driver confidence				
Insecure	21	81%	5	19%
Confident	80	77%	24	23%
Total	101	78%	29	22%

6.4. Discussion

One of the assumptions of the calibration model is that inadequate self-assessment, and specifically overestimation of skills, results in inadequate adaptation to task demands. To measure self-assessment of skills, drivers' self-reported confidence was compared with actual driving performance. The cross tabulations resulted in three groups: a) well-calibrated drivers, b) overconfident drivers and c) insecure drivers. Only half of the novice drivers were 'well calibrated' as opposed to 85% of the experienced drivers. As expected, a relatively large proportion of novices (24%) belong to the 'overconfident' group.

The main hypothesis of this chapter, that inadequate self-assessment of skills is associated with (reporting) dangerous behaviour (irrespective of experience level), was supported by the finding that overconfident drivers generally reported more violating behaviour. Our results also show that overconfident drivers reported less instances of adaptation of driving speed to the complexity of traffic situations as measured with the Adaptation Test (*manoeuvring level*). Although insecure drivers seemed to perform somewhat worse on the Adaptation Test, the percentage of correct responses was not significantly different from the well-calibrated drivers.

No evidence was found to suggest that inadequate self-assessment of skills affects adaptation on the *strategic level* of the driving task. This could have been a result of the choice of questions to measure adaptation on this level. Only two questions were used, of which one seemed to have measured something else than it was intended to measure (i.e. drivers considered congestion as a reason to avoid a difficult intersection, instead of complexity of the traffic situation). The second question (whether drivers had cancelled an intended driving trip due to adverse conditions, e.g., darkness or weather conditions) could have been influenced by the uncontrolled element of opportunity. Drivers may not have cancelled a trip simply because difficult circumstances did not arise.

Results also showed a relationship between self-assessment of skills and self reported crashes. Of the overconfident drivers, 50% reported a crash; this is significantly more than the well-calibrated and insecure drivers (respectively 18% and 10%).

This is in contrast with a recent study by Tronsmoen (2008) who found that the better respondents perceived their driving ability to be, the lower their level of risk was (also based on self-reported crashes). In this latter study, however, participants were asked to assess their driving skills ánd report their crash involvement in the same questionnaire. It is very likely that this diffused the causal relationship. If a driver had experienced a crash, it is likely that this negatively influenced his perception of his driving skills. In addition, people have a tendency to give coherent answers. That is, it is probably difficult for a driver to report that he believes he is a very skilled driver, after reporting the occurrence of one (or more) crashes. In this chapter we report on self-assessment scores obtained prior to the self reported crashes, which makes the interpretation of a causal relationship between self-assessment of skills and crash risk more plausible.

On the other hand, the construction of the calibration groups was based on a comparison of self-reported confidence ('driver confidence') and the examiners' 'safe driving' scores. It can be argued that reported differences between calibration groups are solely caused by 'safe driving', and that 'driver confidence' did not contribute to the reported differences. For example, all overconfident drivers - by definition - failed the driving assessment. Moreover, the relationship between the driving assessment and self reported crashes also indicated that 50% of the drivers who failed the driving assessment reported one or more crashes in the following year; and no relationship was found between self reported driver confidence and self reported crashes.

It is difficult to give a decisive answer about this alternative explanation, because the number of participants is too small to have a reasonable power for tests of interaction between confidence and examiners' opinion. There were, for example, only six well-calibrated drivers who also failed the driving assessment (but were correctly insecure about their driving skills). No interaction effects of reasonable size can be found with such small numbers.

However, with respect to adaptation to task demands, there are indications that neither ‘driver confidence’ nor ‘safe driving’ in itself can explain all the results. For example the violation items of the DBQ indicated a significant main effect for ‘driver confidence’ but no significant main effect for ‘safe driving’ was found. In contrast, for the Adaptation Test, a significant effect for ‘safe driving’ was found but not for ‘driver confidence’. These results suggest that the combination of ‘safe driving’ and ‘driver confidence’ (i.e. self-assessment of skills) explains more variation than both factors separately.

Prior research has indicated calibration as a relevant factor for safe driving, and has linked calibration to the high crash rate of young, novice drivers (e.g. Brown & Groeger, 1988; Gregersen, 1995; Mitsopoulos et al., 2006). The current chapter provides evidence for the behavioural consequences of ‘overconfidence’, and possibly a relationship with crash risk. At least on the manoeuvring level, overconfident drivers adapt their behaviour less to the traffic situation.

PART 3:

DEVELOPMENT OF CALIBRATION & EXPERIENCE

7. The development of calibration skills

This chapter describes how calibration develops over time. To do this, the development of the separate elements of the calibration model is described. *Self-assessment of skills* (Chapter 4) is compared from one year to the next and the development in *adaptation to task demands* (Chapter 6) is described on six occasions during two years. The scores on the Adaptation Test (Chapter 5) during the two years of the study are compared for the novice and experienced drivers.

In addition, this chapter reports the results of the driving diary that the participants kept during six periods over two years of the study. In this diary, they reported on how much they had driven, under what circumstances (e.g. in the dark) and if they had violated any traffic rules (e.g. driving after consumption of alcohol or speeding).

The results show that, although a difference was found between novice and experienced drivers in their calibration skills, as was already found in previous chapters, the expected *improvement* in calibration skills of the novice drivers was not found in the two years after receiving their driver's licence.

The results of the driving diary suggest that novice drivers differ from experienced drivers with respect to the type of trips they make. Novice drivers drive more at night, during weekends, during leisure time and with passengers. In addition, novice *male* drivers drive faster and without seatbelts more often than novice *female* drivers. There are no indications that these trip characteristics change much in the first two years after licensing.

7.1. Introduction

The previous chapters of this thesis suggested that the different components of the calibration model (see Figure 1.2) exist. The results in Chapter 4 showed that novice drivers seem to overestimate their driving skills more than experienced drivers, but only when their self-assessment is compared to their actual behaviour. Because of this result and theoretical considerations (see Section 4.1), it was concluded that, for the measurement of self-assessment of skills, it is necessary to compare the drivers' assessment to a more independent measure of driving skills.

Chapter 6 showed that inadequate self-assessment, and specifically overestimation of skills, is connected to inadequate adaptation to task demands. More specifically, overconfident drivers generally reported more violating behaviour as measured with the Driver Behaviour Questionnaire (DBQ). The results in Chapter 6 also showed that overconfident drivers reported less instances of adaptation of driving speed to the complexity of traffic situations as measured with the Adaptation Test (see Chapter 5). Although insecure drivers seemed to perform somewhat worse on the Adaptation Test, the percentage of correct responses was not significantly different from the well-calibrated drivers.

One of the main objectives of this research is to investigate if and how calibration develops over time. Therefore, in this chapter, the development of the different components of the calibration model, *self-assessment of skills* and *adaptation to task demands,* is described over the two years of the study. The expectation is that novice drivers will start at a lower level compared to experienced drivers, and that they will improve their calibration skills during the two years of the study.

Because in the calibration model *perceived complexity* cannot be separated from the self-assessment of skills (see Section 2.4.3), this element is always discussed in relation to the other elements of the calibration model. Consequently, there is no specific measure of perceived complexity that is described over the two years of the study. Perceived complexity is, however, incorporated in the Adaptation Test, which measures the effect of perceived complexity on adaptation to task demands.

This chapter also discusses the results of the driving diary, in which the participants reported on the kind of trips they made, to determine if the *type* of trips novice and experienced drivers make, can explain the differences in crash risk. Also, the driving diary was analysed to see if there were differences between the type of trips that novice males and novice females make. In the Netherlands, young *male* drivers run almost twice as much risk to be involved in a crash than young *female* drivers. Research has suggested that this is partly caused by the type of trips male and female drivers make (Forsyth et al., 1995).

7.2. Method

7.2.1. Participants

This chapter describes the results of different variables from different instruments of the study (questionnaire, driving assessment, etc.). For each variable or instrument a different selection of participants was used.

To study the development of *self-assessment of skills*, a comparison with the results of the driving assessment needed to be made. Therefore, only the drivers who participated in one or both on-road driving assessment(s) were used. The sample consisted of 130 drivers in 2006 and 112 drivers in 2007. A total of 34 novice drivers and 20 experienced drivers participated in both on-road driving assessments; 49 novice and 27 experienced drivers participated only in the first assessment; and 43 novice and 15 experienced drivers participated only in the last assessment (see also Section 3.5.2).

For the remaining variables from the questionnaire all participants were used. These data were analysed using the *F*-test by Repeated Measures ANOVA, which can only analyse complete series of data. After correction for missing data there were 158 experienced drivers and 252 novice drivers with complete series of data for questionnaire variables (see also Section 3.6.2).

For the driving diary, there were much less participants with complete series of data, due to the fact that participants could report if they had not driven during the preceding week. Out of all 330 novice drivers, only 62 (19%) drove in all 18 weeks *and* reported this in the diaries. Of the 179 experienced drivers, 84 (47%) drove in all 18 weeks *and* reported this in the diaries. If we were to use only the participants who completed *all* diaries, the number of participants would be drastically reduced to only 19% of the novice and 47% of the experienced drivers. However, we attempted to reduce the number of incomplete series of data. One of the remedies was to average the entries of three weeks into one average for the whole session. So, if for example a report for one week was missing (because the participants did not drive) the reports of the other two weeks in the session were averaged to obtain a response for the whole session. See Section 3.6.2 for an extensive description of the correction for missing data.

By definition, the Novice II group, which started filling out questionnaires six months later than the rest of the participants, would not have a complete series of data for the two years of the study. Moreover, the amount of missing data for the Novice II group was too much to enable correction for missing data. Instead, for each of the variables that were analysed, the Novice I group was compared to the Novice II group on the last four sessions only. Repeated Measures ANOVA indicated that there were no differences between both groups of novice drivers. Therefore, only the data of the Novice I group was used to test for development in responses to the questionnaires and driving diaries.

7.2.2. Instruments

On-road driving assessment

Examiners rated drivers on their 'ability to drive safely' on a scale from 0 to 10; 5.5 being the pass-fail criterion in a real driving test. As in a regular driving test, this general mark for 'safe driving' is a combination of many different skills on different levels of the driving task (e.g., to *manoeuvre* safely through traffic, a driver must possess a certain level of vehicle *control*).

Questionnaire

The questionnaire used in this study contained questions on how confident participants are as a driver (the Driver Confidence Questionnaire; see Appendix A). A five point scale was used for these questions because the most important distinction for analysis in the current study is between drivers who view themselves as (much) better, similar or (much) worse drivers compared to the average driver. The responses to the 'driver confidence' question compared with the examiners assessment for 'safe driving' (mentioned above) was used to construct a measure for self-assessment of skills (see also Chapter 4 – Self-assessment of skills).

Besides the Driver Confidence Questionnaire, the questionnaire also contained questions regarding the adaptation to task demands on the strategic and manoeuvring level. Relating to the strategic level, drivers were asked: a) whether, in the last 4 months, they had cancelled an intended driving trip due to adverse conditions (e.g., darkness or weather conditions); and b) whether, in the previous 4 months, they had avoided a difficult intersection. The previous chapter already showed that this second question did not refer to adaptation to task demands in the sense we intended, because participants reported avoiding intersections because of traffic congestion, not because of traffic complexity. Therefore only the percentage of drivers cancelling an intended trip due to adverse conditions was used as an indication of adaptation on the strategic level.

To measure adaptation to task demands on the manoeuvring level the Adaptation Test was used (see Chapter 5 – The development of the Adaptation Test).

Finally, the questionnaire contained an abbreviated Dutch version (Verschuur, 2003b; Verschuur & Hurts, 2008) of the Driver Behaviour Questionnaire (DBQ, see Parker, Reason et al., 1995). See also Chapter 3 – Section 3.3.2 and Appendix C. The eight items of the DBQ indicating violating behaviour were used to measure the effect of inadequate self-assessment on violating behaviour, which relates to the strategic level as well as the manoeuvring level.

In the last five questionnaires, drivers were asked whether they had been in a crash as a driver in the preceding four months (this was not asked in the first questionnaire because the novice drivers had no licensed driving experience at this point). Only three crashes with personal injury were reported, compared to 142 crashes with material damage only. Because of the small number of crashes with personal injury, no distinction was made in this study in the severity of the crash.

Driving diary

For each session the participants completed one questionnaire and three weekly driving diaries. At the end of each week the participants reported how much they had driven, if they had been driving mostly during weekdays or in the weekend, or if they had, for example, consumed alcohol before driving. It was also possible to report not to have driven at all that week (see Appendix D for the specific questions in the driving diary).

Several of the variables in the driving diary were asked on an eight-point-scale. For example, the participants indicated on a scale from 0 to 7 if they had driven: 0 = “not at all during the weekend” or 7 = “all trips during the weekend”. Other variables were asked in a yes/no question (binary variables). For example, the participants were asked to indicate if they had: 0 = “no” or 1 = “yes”, driven after the consumption of some alcohol in the preceding week.

7.2.3. Data analysis

Categorical and Interval variables
The categorical data (e.g. from the Driver Confidence Questionnaire) were analysed using a Chi-square test. The interval variables were analysed using *F*-test by Repeated Measures ANOVA. Because the sphericity-assumption was violated for some of the variables the Greenhouse-Geisser epsilon, with adjusted degrees of freedom, was used to test for significance (see e.g. Jaccard & Becker, 2002; Stevens, 1996).

Binary variables
The binary variables were analysed using two strategies. First by aggregating the three weekly scores into one binary session score and, second, by adding all 3 weekly scores into one interval sum score for the session.

The first strategy included recalculating the yes/no responses from each of the three weeks of the diary into one new binary variable for each session. This new variable indicated if participants had driven after the consumption of some alcohol in the preceding *three* weeks instead of one (0 = “no” or 1 = “yes”).

The three-way contingency table with this new variable was then analysed using Chi-square analysis. Although this is a correct method to analyse frequencies, there is a drawback in the interpretation of the results. If the Chi-square test indicates that the frequencies in the table are significantly different from the expected frequencies it is difficult to discover which cell of the table causes this significance. For example, when analysing the three-way contingency table[16], it is almost impossible to determine whether there is a difference between novice or experienced drivers, or that the reports for the novice drivers are decreasing over time, or for example that the frequency in session three is deviant.

In the second strategy, the yes/no responses from the three weeks of the diary were aggregated into a new ‘interval’ variable. This variable could have four values: 0 = “never driven after consumption of alcohol”, 1 = “in one week driven after consumption of alcohol”, 2 = “in two weeks driven after

[16] The three-way contingency table with the levels:
Experience level (novice / experienced) x session (1 to 6) x drinking and driving (yes / no)

consumption of alcohol", and 3 = "in all three weeks driven after consumption of alcohol".

This new 'interval' variable was analysed with the *F*-test by Repeated Measures ANOVA. Strictly, this is not the correct method to analyse this data, since the new variable, with only four categories, is not really a measurement on the interval level. In addition, the variable is highly skewed towards the lower end of the scale. However, the advantage of this strategy is that it is possible to locate *where* the data are significantly different. For example, when comparing novice with experienced drivers, it is possible to indicate whether there is a difference between novice and experienced drivers, or that the reports for the novice drivers are decreasing over time, or for example that the frequency in session three is aberrant.

A combination of both strategies was used to test the originally binary variables in the driving diary. First, Chi-square analysis on the binary variable was used to test for significance, and second, *F*-test by Repeated Measures ANOVA on the new 'interval' variable was used to interpret the significance.

The data from the driving diaries have undergone several steps to deal with missing values (see also Section 3.6.2) or enable interpretation of results. In all cases, the variables for each week were aggregated into session reports, and only the responses of the Novice I group were tested. The figures in this chapter show the results of these aggregated variables. Appendix F shows a summary of the raw data from the driving diary, including the Novice II group.

7.3. Results

7.3.1. Self-assessment of skills

Chapter 4 introduced three questions from the Driver Confidence Questionnaire, in which drivers were asked to compare themselves with 'the average' and 'peer' driver (see Table 4.2). Of these variables, a development was only found in the variable where drivers stated that they were a (much) better driver than either their peers or the average driver (see Figure 7.1). Novice drivers see themselves as increasingly better drivers throughout the two years of the study, when they compare themselves to their peers

($\chi^2_{(5,N=2114)}$ = 17.43; $p < .01$) as well as when they compare themselves to the average driver ($\chi^2_{(5,N=2115)}$ = 15.12; $p < .05$)[17].

It is difficult to interpret this result because due to increased driving experience, novice drivers are probably actually becoming better drivers. So, have they become more overconfident over the two years of the study, or simply became better drivers and reported this correctly in the questionnaire? As was concluded in Chapter 4, a comparison with a more independent measure of driving skills is necessary to draw conclusions about self-assessment of skills.

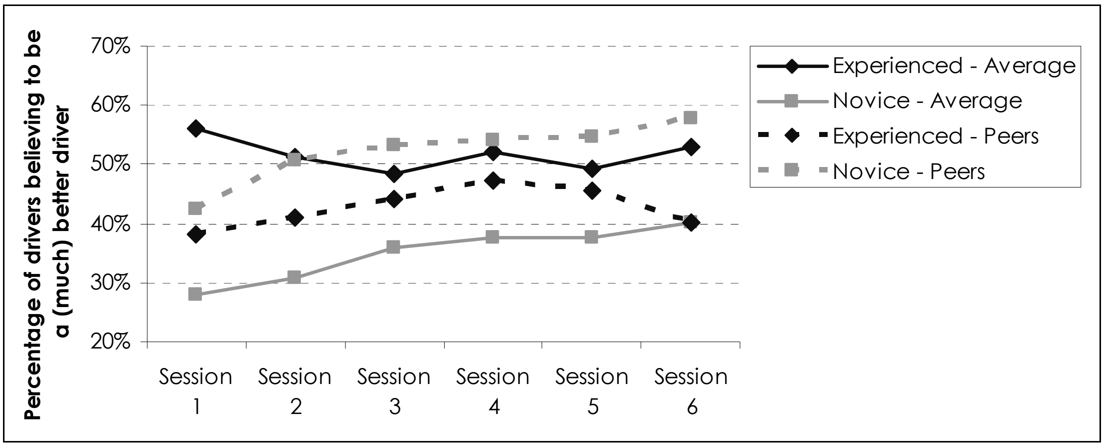

Figure 7.1. Percentage of novice and experience drivers believing to be a (much) better driver compared to their peers (dotted lines) and the average driver (solid line)

Figure 7.2 shows such a comparison between the observed skills (examiners opinion) with the drivers own confidence ratings. To make a comparison between the two variables possible, the variables were recoded into percentile scores (see Section 4.3.3. for a description of the procedure). The left hand side of Figure 7.2 shows the percentile scores of the confidence ratings and the examiners opinion during the driving assessment in 2006 and is exactly similar to Figure 4.2. The right hand side of Figure 7.2 shows the same comparison for the data from 2007.

[17] Because of the unequal distribution over categories this variable was considered to be an ordinal variable (rather than an interval variable) and was analysed using Chi-square analysis. In this analysis, the responses from each session were analyzed as if they stemmed from different respondents, causing the large N.

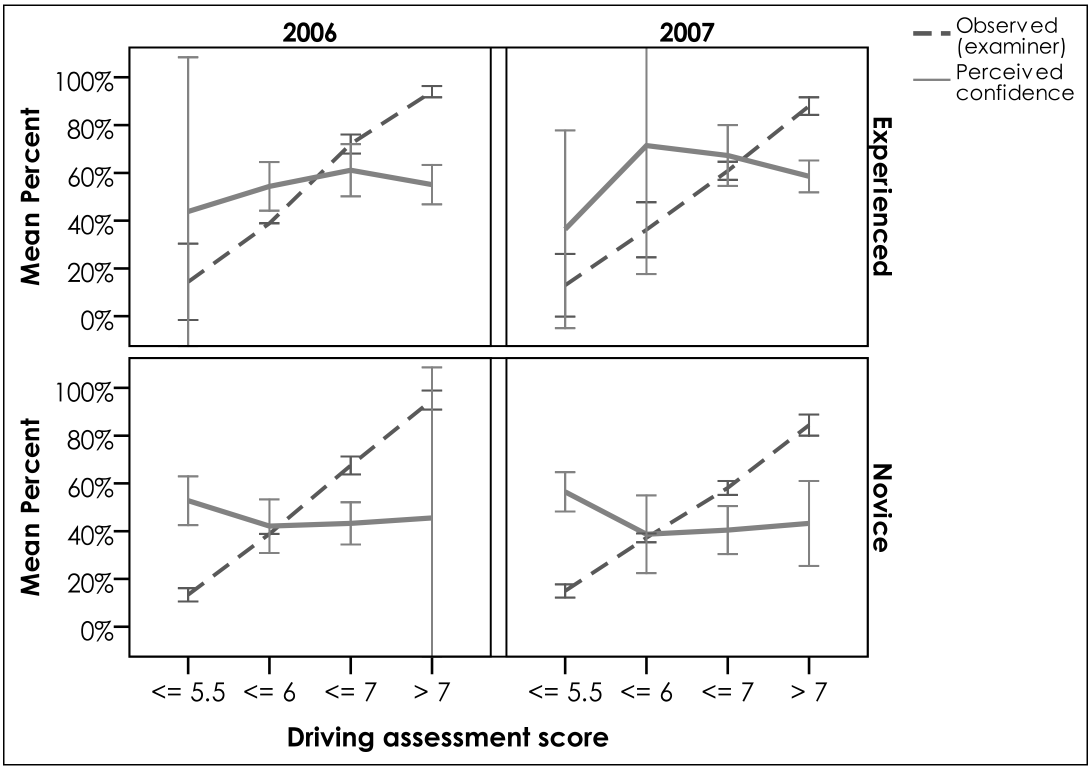

Figure 7.2. Mean percentiles of observed skills and perceived confidence (in 2006 and 2007)

Figure 7.2 shows that, in 2006, there was little correspondence between the performance on the driving assessment and a driver's confidence. Whether a driver failed the driving assessment or was assessed as a (very) good driver, he reported roughly the same amount of confidence in the questionnaire. In 2007 this picture did not change much. The smaller 95% confidence interval (indicated by the whiskers) for novice drivers scoring higher than 7 for safe driving only indicate that there are more drivers in this category in 2007 compared to 2006. However, the mean percentage of confidence level did not change from 2006 to 2007. There is still a significant difference, as indicated by the whiskers, between the ranking based on the *perceived* confidence in driving skills and *observed* skills by the examiner. Novice drivers scoring the lowest on the driving assessment seem to report the most confidence in their driving skills.

7.3.2. Adaptation to task demands

Strategic level

Figure 7.3 shows that, for the final five sessions, there is not much difference between experienced and novice drivers in cancelling an intended trip due to adverse conditions. The first session is an exception, however; this difference can be explained by the fact that the novice drivers only held their driver's licence for 2-4 weeks when they completed the first questionnaire. In comparison, the experienced drivers reported situations from the period of four months preceding the questionnaire, resulting in a relatively high percentage of situations in which they decided not to drive. So, novice drivers reported fewer instances of adverse conditions, simply because the reporting period was shorter.

The high percentage of drivers reporting cancelling a trip in Session 2 can (partly) be explained by bad weather conditions. These questionnaires were completed during January and February 2006. During this winter period there were extremely bad weather conditions in the Netherlands, possibly explaining the high percentage of cancelled trips. However, Session 5 does not show these high percentages of cancelled trips, and was also completed during the months January and February (but in 2007).

Still, even when considering the winter periods during Session 2 and 5, there seems to be an unexplained trend of fewer drivers reporting that they cancelled at least one intended driving trip due to adverse conditions.

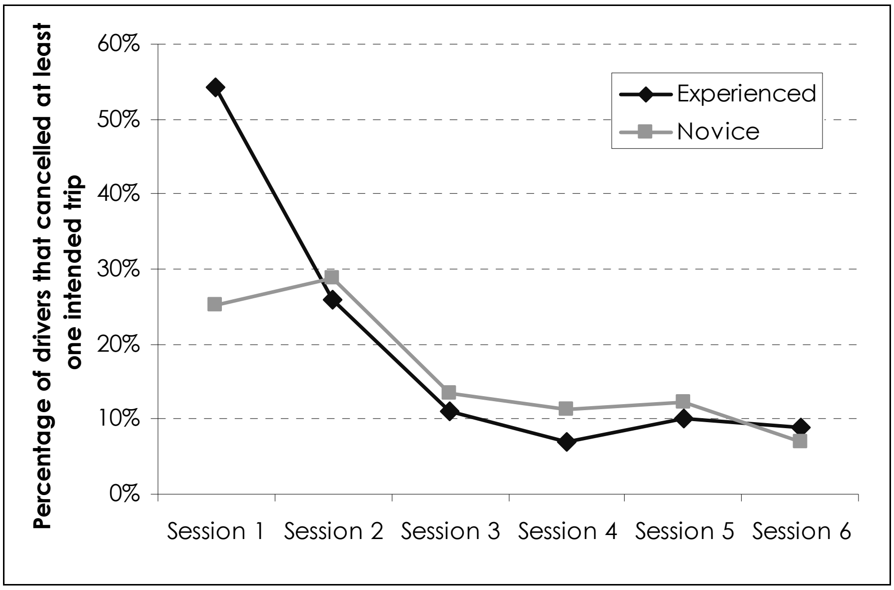

Figure 7.3. Percentage of drivers that cancelled at least one intended driving trip due to adverse conditions

Manoeuvring level – the Adaptation Test

As can be seen in Figure 7.4 experienced drivers performed better on the Adaptation Test during the whole study. Repeated Measures ANOVA, with the adaptation score on the six questionnaires as the within-subjects factor, showed a significant difference between experienced and novice drivers ($F_{1,408}$ = 11.40; p < .01; η_P^2 = .03), and a significant improvement over time ($F_{5,1924}$ = 24.40; p < .001; η_P^2 = .06). The Repeated Measures ANOVA showed no interaction effect between experience level and time. Thus, the expected effect, that experienced drivers remained at the same level while the novice drivers improved their performance on the Adaptation Test, was not found.

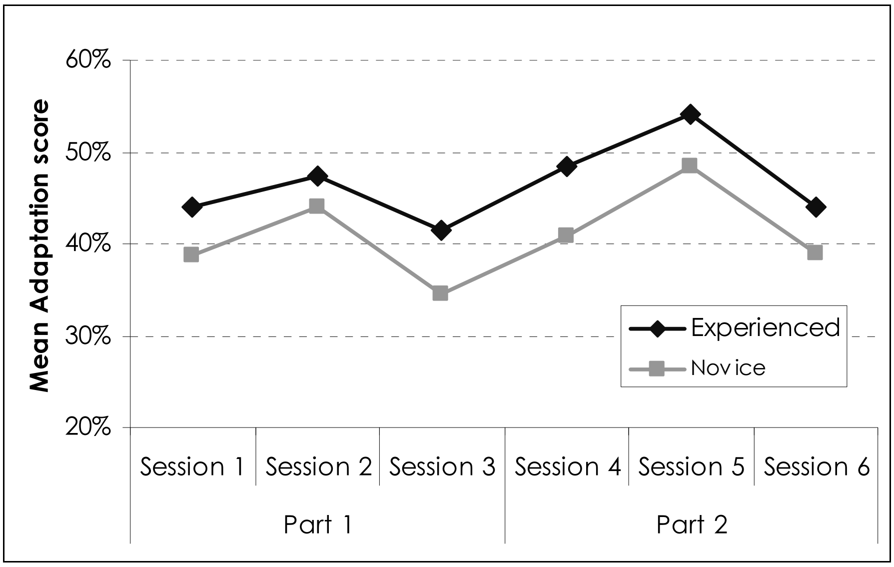

Figure 7.4. Mean adaptation score for experienced and novice drivers in the six sessions

The complete Adaptation Test consisted of 18 traffic situations each depicted on two photos. To avoid recognition by the participants, only six situations were assessed in Session 1 (Photoset 1). In the second session six new situations were assessed (Photoset 2), and finally in Session 3 again six situations were assessed (Photoset 3). The assessment of these photosets were than repeated in Session 4, 5 and 6. This pattern is visible in Figure 7.4. It seems that Photoset 2 (in Session 2 and 5) was easier to assess than Photoset 3 (in Session 3 and 6).

In order to take these differences between photosets into account, a separate analysis was conducted with experience level (novice or experienced) and photoset (1, 2 or 3) as the between-subjects factors and Part (1 or 2, for the moment of assessment) as the within-subjects factor. The results again showed a small, but significant effect over time ($F_{1,1345}$ = 28.56;

$p < .001$; $\eta_P^2 = .02$) and a significant main effect of both experience level ($F_{1,1345} = 31.48$; $p < .001$; $\eta_P^2 = .02$) and photoset ($F_{2,1345} = 24.92$; $p < .001$; $\eta_P^2 = .04$). Indeed, Photoset 2 was easier to assess than the other two photosets. However, no interaction effect between photoset and experience level was found, indicating that the situations were similarly difficult or easy to assess for novice and experienced drivers in our study. Also, no interaction was found between photosets and time, indicating that all three photosets measured a similar development over time.

Reported violations as measured with the DBQ

The eight items of the Driver Behaviour Questionnaire (DBQ) concerning violating behaviour (see Table 6.4 for the individual items) were used as an indication of the amount of dangerous decisions made by drivers. Figure 7.5 shows that, while novice drivers start at the same level as experienced drivers, the number of reported violations quickly increases during the first year of independent driving. The number of violations reported by the experienced drivers remains at the same level. Repeated Measures ANOVA indicated that there is a significant effect over time ($F_{5,1926} = 11.66$; $p < .001$; $\eta_P^2 = .03$), a significant difference between novice and experienced drivers ($F_{1,409} = 15.65$; $p < .001$; $\eta_P^2 = .04$), and a significant interaction effect ($F_{5,1926} = 11.90$; $p < .001$; $\eta_P^2 = .03$).

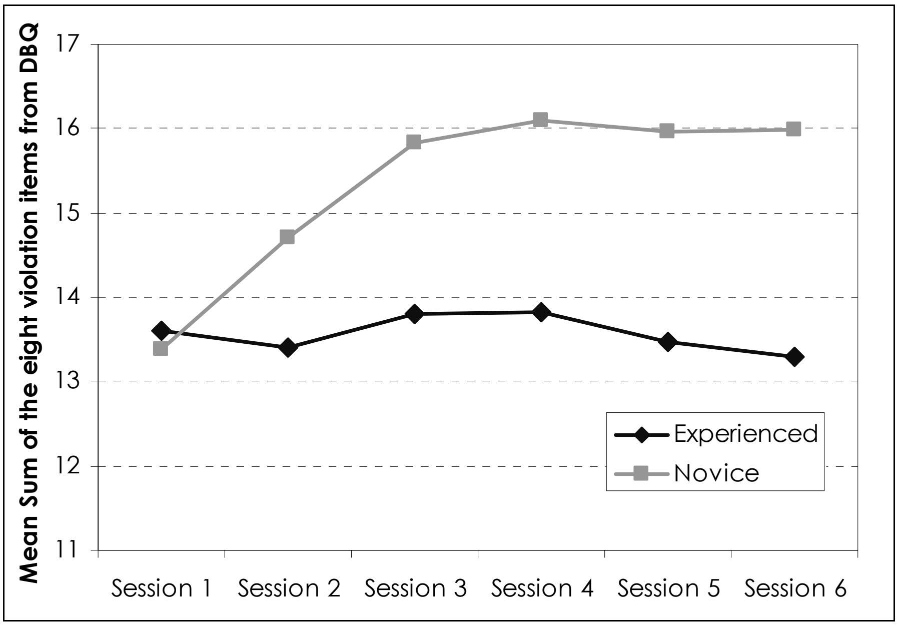

Figure 7.5. Mean sum of the eight violation items in the DBQ

7.3.3. Development in self-reported crashes

During the two years of the study 121 participants reported a total of 145 crashes (with personal injury or material damage). Figure 7.6 shows that novice drivers reported more crashes than experienced drivers. In two years, 9% of the experienced drivers reported a crash, compared to 25% of the novice drivers. This difference between novice and experienced drivers is significant ($\chi^2_{(1,N=585)} = 19.17$; $p < .001$). However, no clear development over time is present, possibly due to the relatively small number of crashes reported.

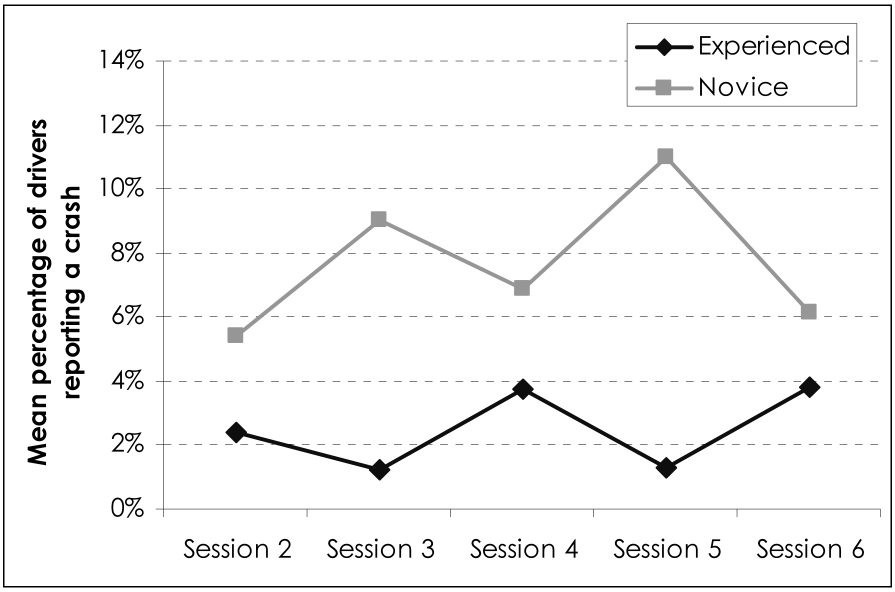

Figure 7.6. Percentage of participants reporting a crash (either with material damage and/or personal injury) in the previous four months (NB: there is no crash rate for the first questionnaire)

7.3.4. Differences between drivers with and without improvement on the Adaptation Test

The longitudinal analysis in the previous section did not provide any indications of development of calibration skills in young novice drivers relative to experienced drivers. In general, experienced drivers perform better than novice drivers (e.g. on the Adaptation Test), but the expected relative improvement of young novice drivers was not found.

In theory it is possible that only a part of the novice drivers improved their calibration skills, but that this development was not visible in the analysis of the whole group of participants. The characteristics of this group of improved drivers could be informative in learning how calibration develops. The current section describes explorative analyses in search of the drivers who did or did not improve their calibration skills in the two years of the study.

Selection of groups

The scores of the Adaptation Test were used to decide which of the young novice drivers developed either positively or negatively in calibration skills. For the Novice I group with valid values on all six occasions of the Adaptation Test (n = 260), the Adaptation Score from Part 1 (Session 1, 2 and 3) was subtracted from that of Part 2 of the study (Session 4, 5 and 6). From these scores, the 20% of novice drivers with the most improvement and the 20% of novice drivers with the least improvement (deterioration) were selected[18].

In addition, there was a Novice II group (n = 119) which started completing questionnaires (and therefore the Adaptation Test) from Session 3. For these drivers the Adaptation Score from Session 3 was compared to the Adaptation Score from Session 6 (both Photoset 3).

The total group of novice drivers was classified in one of the following groups: 81 novice drivers who improved on the Adaptation Test; 71 novice drivers[19] who deteriorated on the Adaptation Test; and 227 with the least change. Figure 7.7 shows the mean Adaptation Score on all 6 Sessions for the three groups.

[18] The 20% cut-off was chosen to select the smallest (most distinct) groups, with still enough participants for reliable comparisons.

[19] These groups are not exactly 20%-20%-40% of the total group, because all drivers with the same amount of improvement or deterioration were classified into the same group. Therefore the cut-off point was not exactly at 20% of the drivers, but rather an approximation of 20% of the drivers.

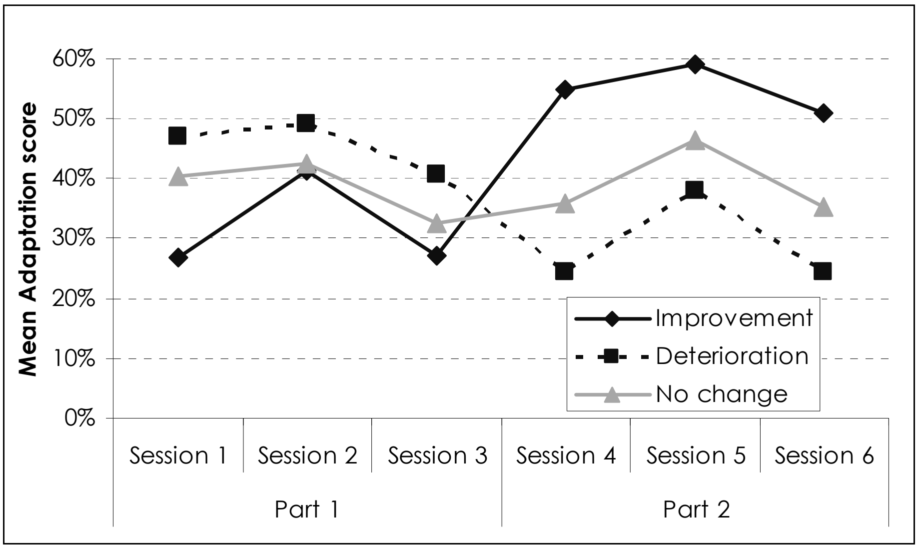

Figure 7.7. Mean adaptation score for the group who improved, deteriorate or did not change on the Adaptation Test

Characteristics of the groups

Table 7.1 shows that females are significantly overrepresented within the group of drivers who detoriated on the Adaptation Test. In addition, the table shows a significant difference in education level. The group of drivers detoriating on the Adaptation Test are significantly lower educated than the drivers who improved on the Adaptation Test.

No difference in age was found. This was not unexpected, since all drivers were young (between the age of 18 and 24) at the start of the study. No difference was found in driver education (number of lessons and total hours of lessons) or the number of times a driver had to take the driving test before passing it. Finally, there was no difference in how much the drivers drove (in number of days or hours of the week) between the groups who improved or deteriorated on the Adaptation Test.

Table 7.1. Characteristics of the novice drivers who improved the most, deteriorated the most or showed little or no change with respect to their Adaptation scores

	Improvement (*n* = 81)		Deterioration (*n* = 71)		No change (*n* = 227)		Significance*
	Mean	SD	Mean	SD	Mean	SD	
Age	19.9	1.7	19.7	1.5	20.0	1.9	n.s.
Driving education & exam							
No. lessons	35.9	12.5	36.5	13.9	34.7	13.5	n.s.
No. hours lesson	42.0	17.4	42.9	18.3	39.6	16.7	n.s.
No. exams	2.0	.9	1.9	.9	1.9	.9	n.s.
Driving experience							
Days per week	3.0	2.0	3.0	2.0	3.1	2.0	n.s.
Hours per week	5.8	8.6	5.9	8.9	4.8	5.0	n.s.
	Count	%	Count	%	Count	%	
Gender							
Males	44	54%	26	37%	114	50%	$\chi^2_{(1,N=152)}$
Females	37	46%	45	63%	113	50%	= 4.77
Total	81	100%	71	100%	227	100%	$p < .05$
Highest school education**							
VMBO/MAVO	31	38%	39	55%	96	42%	$\chi^2_{(3,N=152)}$
HAVO	14	17%	14	20%	64	28%	= 8.97
VWO	27	33%	17	24%	47	21%	$p < .05$
Else	9	11%	1	1%	20	9%	
Total	81	100%	71	100%	227	100%	
Driving education							
Regular	71	88%	66	93%	198	88%	
Driver training							No test
in steps (RIS)	1	1%	2	3%	7	3%	possible
Crash course	9	11%	3	4%	22	9%	due to small
Total	81	100%	71	100%	227	100%	numbers

*For the calculation of the significance tests only the Improvement and Deterioration groups were compared. That is, the drivers 'without change' were not included in the analysis.
**These education levels roughly correspond to the following: VMBO/MAVO = Vocational Education / Lower Secondary Education; HAVO = Higher Secondary Ed.ucation; VWO = Pre-university Education.

7.3.5. Results of the driving diary

Differences between experienced and novice drivers

Experienced drivers reported driving on average 4.5 days of the week, novice drivers only drove about 3 days a week. Repeated Measures ANOVA on the corrected data (see Section 3.6.2) indicated that this difference between experienced and novice drivers was significant ($F_{1,415}$ = 64.87; p < .001) and large (η_P^2 = .14). Novice drivers report relatively more trips in the weekend than during the week ($F_{1,389}$ = 28.20; p < .001; η_P^2 = .07), drive more during leisure time, as compared to school/work related trips ($F_{1,390}$ = 23.93; p < .001;

η_P^2 = .06) and drive relatively more at night ($F_{1,396}$ = 26.60; p < .001; η_P^2 = .06) than experienced drivers.

The number of days driving, trips in the weekend or driving at night did not change over the two years of the study. The proportion of trips during leisure time decreased during the two years of the study ($F_{5,386}$ = 3.89; p < .01; η_P^2 = .05), however this decrease was the same for novice drivers and experienced drivers.

When novice drivers first started driving, they had passengers in their car more often than experienced drivers ($F_{1,390}$ = 24.74; p < .001; η_P^2 = .06). These passengers were more often peers while experienced drivers had family in the car more often ($F_{1,372}$ = 114.04; p < .001; η_P^2 = .24).

For experienced drivers, the proportion of trips with passengers did not change in the two years of the study; whereas, for the novice drivers, the proportion of trips with passengers decreased over time (see Figure 7.8). Repeated Measures ANOVA indicated a significant interaction effect ($F_{5,386}$ = 2.87; p < .05; η_P^2 = .04).

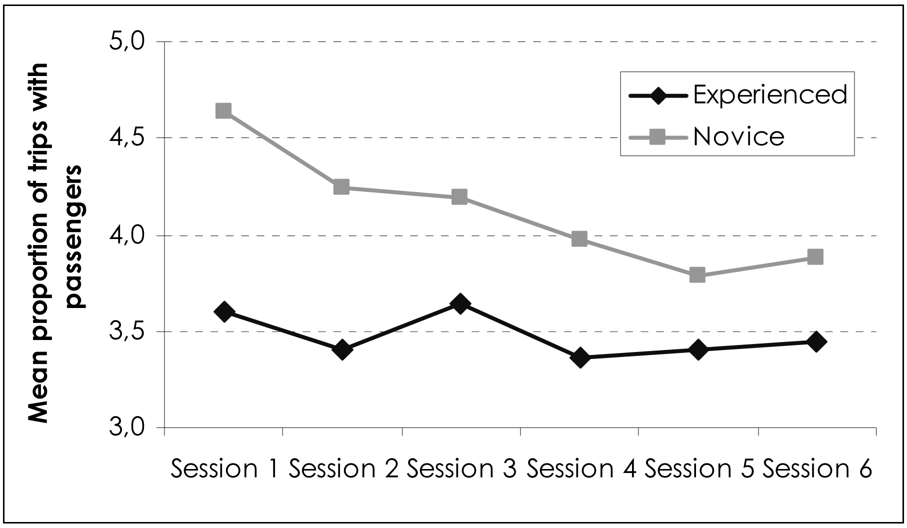

Figure 7.8. Mean proportion of trips with a passenger (ranging from 1= all trips alone to 7= all trips with passengers)

On average, 48% of the experienced drivers and 70% of the novice drivers reported speeding (driving 10 km/h faster than the speed limit) at least once in three weeks.

Figure 7.9 shows the aggregated 'interval' variable of the same diary question (driving 10 km/h faster than the speed limit), which was analysed with Repeated Measures ANOVA. This analysis showed that the difference between experienced and novice drivers was significant ($F_{1,397}$ = 27.12; p < .001; η_P^2 = .08) and revealed a small main effect of time ($F_{4,1734}$ = 4.80; p < .001; η_P^2 = .01). In addition, Repeated Measures ANOVA showed a small but significant interaction effect ($F_{4,1734}$ = 3.28; p < .01; η_P^2 = .01). Novice drivers showed a small increase of speeding over time, whereas the experienced drivers remained at the same level.

These effects were only found in the question about speeding within *urban* areas. There was no difference between novice and experienced drivers nor an effect over time, in driving 10 km/h faster than the speed limit in *rural* areas.

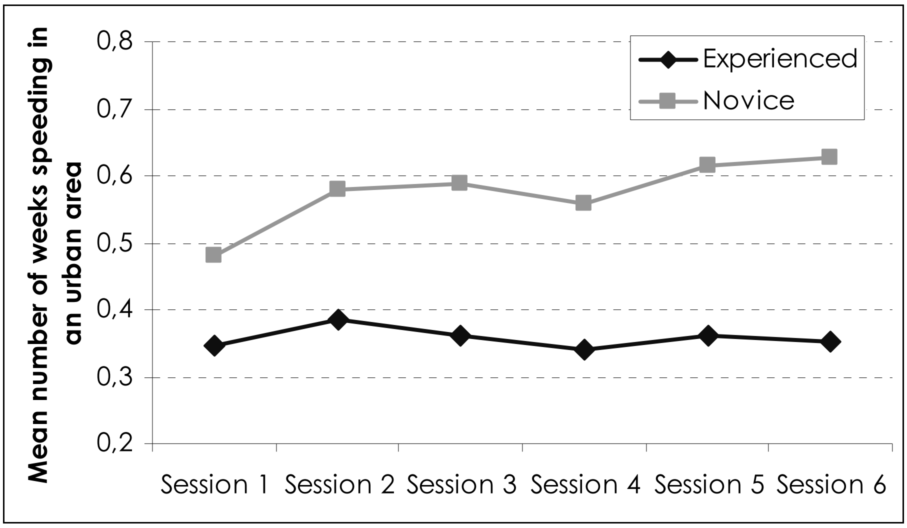

Figure 7.9. Mean number of weeks of a session in which participants drove 10 km/h faster than the speed limit in an urban area (ranging from 0 = no weeks to 3 = all three weeks of the session)

Differences between male and female novice drivers

To study the differences between male and female novice drivers only the data from the novice drivers were analysed. More specifically the data from the Novice I group was used, because these were complete series of data for all six sessions of the study.

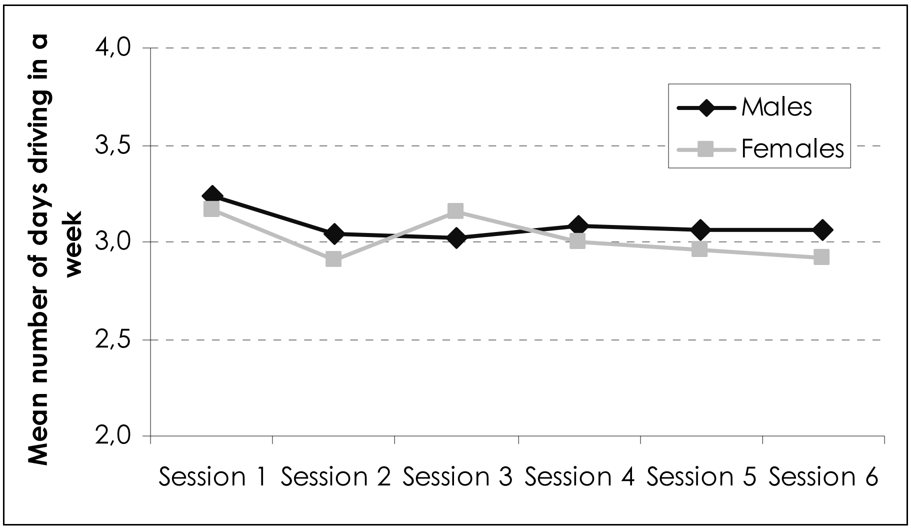

Figure 7.10. Mean number of days the novice male and female drivers drove in one week

As can be seen in Figure 7.10, there is little difference between novice male and novice female drivers in the number of days a week they drive, and this does not change over the course of the study. However, as Figure 7.11 shows, male drivers report driving significantly more hours in one week ($F_{1,256}$ = 12.16; p < .01; η_P^2 = .05). This suggests that male drivers make longer trips than female drivers.

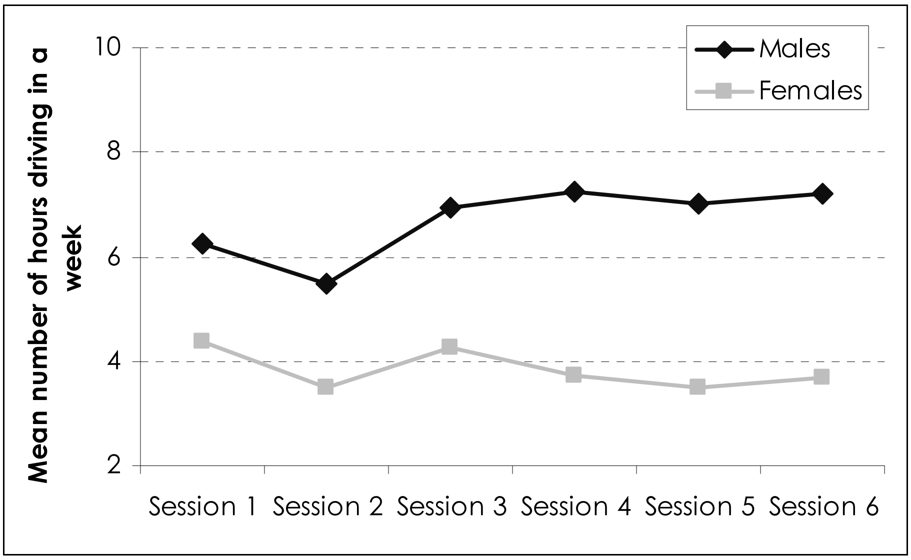

Figure 7.11. Mean number of hours the novice male and female drivers drove in one week

The variables on which novice drivers differed from experienced drivers (i.e. driving in the weekend, during leisure time, or with (peer) passengers) showed no differences between young males and females.

Young males and females did, however, differ in their speeding behaviour both in an urban ($F_{1,239}$ = 6.69; p < .05; η^2_P = .03) and rural area ($F_{1,239}$ = 5.29; p < .05; η^2_P = .02). On average, 74% of the novice males and 67% of the novice females reported speeding in an urban area at least once in three weeks. For rural areas the percentages were 84% and 69%, respectively. Both in urban and rural areas the percentage of male and female novice drivers reporting speeding, increased significantly over the two years of the study ($F_{4,996}$ = 8.34; p < .001; η^2_P = .03 and $F_{4,1065}$ = 2.51; p < .05; η^2_P = .01 respectively). There was no interaction effect.

Finally, young males and females differed in seatbelt use both in an urban ($F_{1,240}$ = 8.24; p < .01; η^2_P = .03) and rural area ($F_{1,239}$ = 5.44; p < .05; η^2_P = .02). On average, 18% of the novice males and 9% of the novice females reported driving without a seatbelt in an urban area at least once in three weeks. For rural areas the percentages were 10% and 2%, respectively. There were no changes over time in the percentages of seatbelt use.

7.4. Discussion

One of the main objectives of this research is to investigate how calibration develops over time. In this chapter the development of different components of the calibration model, *self-assessment of skills* and *adaptation to task demands* were described. Novice drivers were expected to start at a lower level compared to experienced drivers, and during the two years of the study improve their calibration skills.

Contrary to expectation, there seems to be no improvement after one year of driving concerning *self-assessment of skills*. Chapter 4 showed that novice drivers overestimate their driving skills more than experienced drivers during the driving assessment in 2006. This holds particularly for those drivers with low scores on the driving assessment. The comparison in the current chapter with the results of the driving assessment in 2007, showed that after one year, novice drivers show a similar overestimation of driving skills as the previous year. In particular, novice drivers with low scores (≤ 5.5) on the driving assessment in 2007 seem as confident as the drivers with low scores on the driving assessment in 2006.

Again, contrary to expectation, no relative improvement was found for the novice drivers in the first two years after licensing concerning the *adaptation to task demands*. The results for the Adaptation Test, which was designed to measure calibration, showed that both experienced and novice drivers improved somewhat over time. However, the amount of improvement for novice and experienced drivers did not differ.

Novice and experienced drivers did not differ in adaptation on the strategic level of the driving task (i.e. measured with the percentage of drivers cancelling an intended trip due to adverse conditions). Experienced drivers did report more cancelled trips in the first session of the study. However, this difference can be explained by the fact that the novice drivers only had a maximum of 2-4 weeks driving experience to report over, rather than the four months the experienced drivers reported over. During the remainder of the study the trend seems to be that both experienced and novice drivers report less and less cancelled trips. This seems to be a learning (or reporting) effect for which no clear explanation exists.

The results of the DBQ show a different pattern. Contrary to expectation, novice and experienced drivers reported the same amount of violations in the first questionnaire. In the following questionnaires the novice drivers reported more and more violations, while the experienced drivers remained at the same level.

This is an unexpected result as it does not correspond to the decreasing crash risk of young novice drivers, although similar patterns have been reported in previous research. For example Bjørnskau and Sagberg (2004) found that novice drivers report more violations with increased experience, as measured with the DBQ respectively one, five or nine months after licensing. Also a large cohort study in the UK found that reported violations, as measured with the DBQ, increased in the first three years after licensing (Wells, Tong, Sexton, Grayson & Jones, 2008).

These results, as well as the results from the current study, could be explained by using the concept of calibration. It can then be assumed that directly after licensing, novice drivers are not very confident about their driving skills and do not commit many traffic violations. After a certain amount of driving experience they probably perceive less risk in traffic and feel more confident about their driving skills, consequently they could also feel more comfortable to commit traffic violations. Nevertheless, this line of reasoning still provides no explanation for the decreasing crash risk of young novice drivers.

The results of the driving diary illustrate differences between novice and experienced drivers in the type of trips they make. Experienced drivers drive more, but novice drivers drive relatively more in the weekend, at night, during leisure time and with (peer) passengers. In addition, novice drivers report more incidences of speeding than experienced drivers.

However, similar to the results of the calibration variables discussed above, the reports of both experienced and novice drivers are relatively consistent over time. There is one exception: during the two years the novice drivers report declining rates of driving with passengers. In the first months after licensing novice drivers report much more trips with passengers than experienced drivers. At the end of the two year period of the study, this difference between novice and experienced drivers is much smaller.

Novice males and females differ on other variables than the total group of novice and experienced drivers. There were no differences between young males and females in how much they drove in the weekend, during leisure time, or how much they drove with (peer) passengers. Apparently novice males and novice females have similar 'novice driver' behaviour.

However, young males and females did differ in the length of their trips. Remarkably, novice males drive on just as many days of the week as novice females, but they report more *hours* driving in one week. So it appears that novice males drive longer trips. This is consistent with the results of Forsyth et al. (1995), who found that female drivers drove significantly less long distance trips (more than 100 miles in a single day) than male drivers in the first year of driving.

In addition, differences between novice males and novice females were found in how much they reported speeding (both in urban and in rural areas); and in how much they reported driving without a seatbelt (both in urban and in rural areas). For both speeding and seatbelt use, novice males reported that they had violated traffic rules more often than female drivers.

The fact that novice and experienced drivers differ on other variables than novice males and novice females, could suggest that the higher crash risk of novice males compared to novice females in the Netherlands is the result of a different mechanism than inexperience. If the reason why young males have a high crash risk is the same as why novice drivers in general have a high crash risk (but simply more severe), then young males and females would differ on the same variables as novice and experienced drivers. The current results could indicate that (Dutch) young male drivers have other unique characteristics; and that they especially have more problems with conforming to traffic rules, than novice female drivers and experienced drivers.

In Chapter 2, information on the crash risk of young, novice drivers was presented. Crash rates are the highest in the first months after licensing and drop substantially over the first two years of driving, with the most pronounced decline during the first six months or during the first 5000 kilometres of driving (OECD - ECMT, 2006). This means that, within this first period of independent driving, something relevant for crash risk changes within these drivers as a result of driving experience. The theory and results of previous chapters suggested that (development of) calibration could play a role in the crash risk reduction. However, the results of the current chapter appear to indicate that calibration did not develop (much) in the first two years of independent driving. The next chapter (Section 8.5) will extensively discuss possible explanations for the lack of development in our study.

8. Discussion

The objective of this thesis is to investigate the high crash risk of young novice drivers, and more specifically, to explore whether a development in calibration skills could explain the substantial decrease in crash risk in the first years after licensing.

There are many indications that with practice parts of the driving task develop (quickly) towards routine, that is driving (sub)tasks become automated. This automation of driving subtasks leads to a decrease in mental workload. However, lack of automation does not explain the high crash risk of young novice drivers completely. Since the driving task is ‘self-paced’, the driver can adjust the task demands (e.g. by reducing speed or increasing following distance), thus decreasing workload. In theory, a novice driver can use this strategy to overcome the limitations of his performance; a novice driver can decrease the task demands to fit his (deficient) level of automated driving.

This self-pacing aspect of the driving task was incorporated into the calibration model (see Figure 8.1), introduced in Chapter 2 and inspired by Brown (1989) and Fuller (2005), which describes that:

> *For safe driving, a driver needs to assess his own driving skills (A), weigh them against his perception of the complexity of the situation (B), and, as a final step, use the result of this balancing to adapt to task demands (C).*

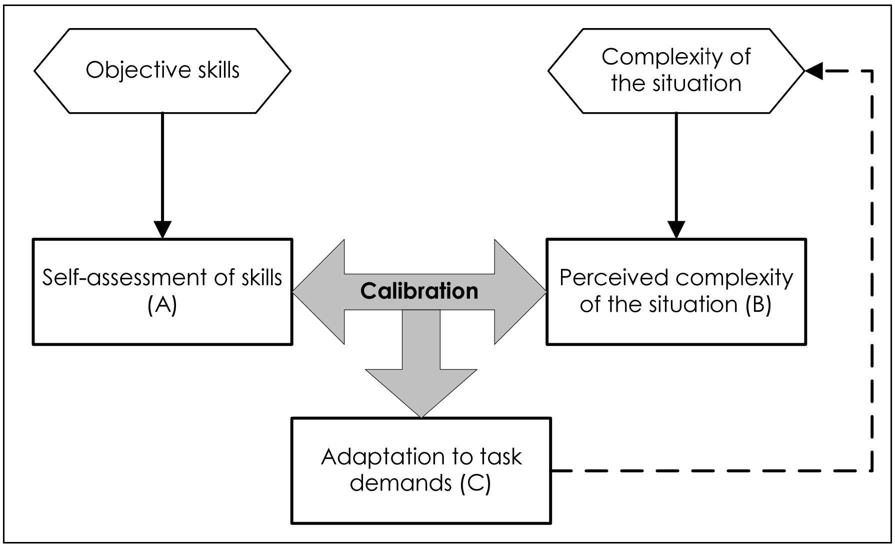

Figure 8.1. Model of the calibration process

This thesis aims to answer the following research questions:

1. *To what extent is poor calibration a contributing factor in the high crash risk of young novice drivers?*

- *a.* *Do young novice drivers overestimate their skills more than experienced drivers?*
- *b.* *Does an inadequate self-assessment of skills affect adaptation to task demands?*
- *c.* *Is there a relationship between the elements of the calibration model and self-reported crashes?*

2. *How can calibration be measured?*

3. *How does calibration develop over time?*

In order to answer these questions a group of young novice drivers was intensively followed during the first two years of their driving career. For comparison purposes, a group of experienced drivers followed the same procedure.

Every four months, the drivers completed one questionnaire and kept a driving diary for three weeks, this was called a 'session'. In total there were six sessions to complete. The questionnaire consisted of items on the self-assessed skills and confidence of the drivers, as well as how they perceive dangers in traffic. In addition, the questionnaire contained an abbreviated version of the DBQ and a newly developed Adaptation Test (see Chapter 5). In the driving diary participants described how much they had driven during the previous week, what the destinations of their trips had been and if they, for example, had had passengers in the car, violated traffic rules, etc.

In addition to this self-reported behaviour, the driving behaviour of a group of novice drivers was observed in an on-road driving assessment 0.5 and 1.5 years after they passed their driving test. The driving behaviour of a control group of experienced drivers was also assessed on these two occasions. In these driving assessments, professional examiners rated the drivers, amongst other things, on their 'ability to drive safely'.

8.1. Calibration

The results of this study provide support to the notion of calibration and the separate elements of the model (see Figure 8.1). The results support the expectation that young novice drivers overestimate their driving skills more than experienced drivers (see Section 8.1.1) and that inadequate self-assessment is connected to insufficient adaptation to task demands (see Section 8.1.2). Furthermore there are some indications that inadequate self-assessment resulted in more (self-reported) crashes (see Section 8.1.3).

Perceived complexity (element B from Figure 8.1) cannot be separated from the self-assessment of skills (see Section 2.4.3). Therefore, this element is discussed in relation to the other elements of the calibration model (e.g. incorporated in the Adaptation Test). Consequently, no specific relationship between perceived complexity and the other elements of the model was investigated.

8.1.1. Self-assessment of skills

Research question 1a. *The results indicate that novice drivers overestimate their driving skills more than experienced drivers (see Chapter 4).*

Two approaches were compared to investigate the (over)estimation of driving skills. In the first approach, drivers were asked to compare themselves to other drivers. The often reported finding, that the majority of drivers (about 80%) rate themselves to be better drivers than the average driver (Delhomme, 1996; McCormick et al., 1986; McKenna et al., 1991; Svenson, 1981), was not replicated in this study. Only about half of the experienced drivers viewed themselves as (much) better drivers than the average driver (51%) or their peers (41%).

More specifically, when novice drivers compare themselves to 'the average driver' on confidence and perceived danger in traffic, it seems that they are modest. That is, experienced drivers are more optimistic about their driving skills and risks in traffic than novice drivers. Compared to the average driver significantly more experienced drivers saw themselves as '(much) better drivers', having '(much) less risk in traffic', and being '(much) better in coping with hazards'.

This pattern changes when drivers compare themselves to their peers. In this case, novice drivers become more optimistic and are significantly more positive than experienced drivers about the risks in traffic. Experienced drivers, on the other hand, are more pessimistic when they compare themselves to their peers instead of 'the average driver'. The result that young drivers are more optimistic in comparison to their peers than when they compare themselves with 'the average driver' has been reported previously by Mathews and Moran (1986). However, the result that experienced drivers become more pessimistic when they compare themselves to their peers than to 'the average driver' is in contrast with the results of Horswill et al. (2004) who found no difference between the comparison with 'the average driver' or 'peer driver' for this group.

In the second approach, self-assessment of skills was compared with a more independent measure of these skills. That is, self-reported confidence about their driving skills was compared to the examiner's opinion of the participants' 'ability to drive safely' during the driving assessment. The results showed that the group of novice drivers who performed worst on the driving assessment (failed the test) reported the same confidence level as novice drivers who performed best on the driving assessment. In contrast, experienced drivers with the lowest scores on the driving assessment were also less confident about their driving skills.

The general finding of lack of correspondence between perceived driving skills and actual ability (see e.g. Mynttinen, Sundström, Koivukoski et al., 2009; Sundström, 2008a) was replicated by the current study. In addition, the current study showed that *novice* drivers, especially those with the lowest scores on the driving assessment, are even worse in assessing their driving skills than experienced drivers.

Comparing the results from these two approaches we conclude that novice drivers are not as optimistic about their driving skills as had been thought in the past. They seem to recognize that they are not as skilled (yet) as the average driver. However, when comparing their self-assessment with their actual behaviour there are indications that they overestimate their driving skills. For a valid indication of self-assessment of skills it is necessary to compare the drivers' assessment to a more independent measure of driving skills.

8.1.2. Effect of self-assessment of skills on adaptation to task demands

Research question 1b. *This research also showed that inadequate self-assessment, and specifically overestimation of skills, is connected to inadequate adaptation to task demands (see Chapter 6).*

To measure self-assessment of skills, drivers' self-reported confidence was compared with observed driving performance in an on-road driving assessment. The cross tabulations resulted in three groups: a) well-calibrated drivers (have the same opinion about their driving skills as the examiners); b) overconfident drivers (are confident about their driving skills, but had low scores on the driving assessment); and c) insecure drivers (are not confident about their driving skills, but had high scores on the driving assessment). Only half of the novice drivers were considered 'well calibrated' as opposed to 85% of the experienced drivers. As expected, a relatively large proportion of novices (24%) were considered 'overconfident'.

One of the characteristics of the calibration, described in Chapter 2, is that adaptation of task demands can take place on different levels of the driving task. This thesis distinguishes adaptation on the two highest levels of Michon's (1985) driving task hierarchy.

First, on the *strategic* level, the driver might, for example, avoid a difficult junction or choose (not) to drive in the dark. No evidence was found to suggest that inadequate self-assessment of skills affects adaptation on this level of the driving task. This could have been a result of the choice of questions. Only two questions were included, of which one seemed to have measured task adaptation due to congestion rather than perceived complexity, and was unusable as such and the other question was vulnerable to the uncontrolled element of opportunity.

Second, with respect to adaptation on the *manoeuvring* level, a driver might not sufficiently reduce speed or increase headway, when encountering more complex situations. Adaptation on this level was measured with the Adaptation Test (see Section 8.1.2). Results showed that overconfident drivers reported less instances of adaptation of driving speed to the complexity of traffic situations than well-calibrated drivers. Although insecure drivers seemed to perform somewhat worse on the Adaptation Test, the percentage of correct responses was not significantly different from the well-calibrated drivers.

Finally, self-reported violations were used as an indication of adaptation to task demands. This can be viewed as adaptation on both the strategic level, as it involves a strategic choice (not) to violate traffic rules, and on the manoeuvring level, as the specific situation has a large influence on driver behaviour (Reason, 1990). The expectation that inadequate self-assessment of skills is associated with (reporting) dangerous behaviour (irrespective of experience level), was supported by the finding that overconfident drivers generally reported more violating behaviour.

These results indicate that there is an effect of overconfidence on adaptation to task demands, at least on the manoeuvring level of the driving task.

Prior research has indicated calibration as a relevant factor for safe driving, and has suggested calibration (in theory) as an explanation for the high crash rate of young, novice drivers (e.g. Brown & Groeger, 1988; Gregersen, 1995; Mitsopoulos et al., 2006). The results of the current study provides evidence for this and an indication of *why* overestimation of skills affects crash risks; that is, at least on the manoeuvring level, overconfident drivers seem to adapt their behaviour less to the traffic situation.

8.1.3. Relationship with self-reported crashes

Research question 1c. *Indications were found that overestimation of skills resulted in more (self-reported) crashes (Chapter 6). However, no relationship between performance on the Adaptation Test and self-reported crashes was found (Chapter 5).*

As mentioned in the previous section, poor calibration has been theoretically linked to increased crash risk. Results of the current study support this theory and indicated a strong relationship between overestimation of skills and self reported crashes. Of the overconfident drivers, 50% reported a crash. This is significantly more than the well-calibrated and insecure drivers (respectively 18% and 10%).

In contrast, a recent study by Tronsmoen (2008) found that the better respondents perceived their driving ability to be, the lower their level of risk was (based on self-reported crashes). In this study by Tronsmoen, however, participants were asked to assess their driving skills ánd report their crash involvement in the same questionnaire. It is very likely that this diffused the causal relationship. If a driver had experienced a crash, it is likely that this negatively influenced his perception of his driving skills. In addition, people are inclined to give coherent answers. That is, it is probably difficult for a driver to report that he believes he is a very skilled driver, after reporting the occurrence of one (or more) crashes. In this chapter we reported on self-assessment scores obtained prior to the self reported crashes, which makes a causal relationship between self-assessment of skills and crash risk more plausible.

In the current study, a correlation between overestimation of skills and self-reported crash risk was found. However, an alternative explanation for this correlation cannot be ruled out. The construction of the calibration groups was based on a comparison of self-reported confidence ('driver confidence') and the examiners 'safe driving' scores. It could be argued that reported differences between calibration groups are solely caused by the differences in 'safe driving', and that 'driver confidence' did not contribute to the reported differences. For example, all overconfident drivers - by definition - scored low on the driving assessment. Moreover, the relationship between the driving assessment and self reported crashes also indicated that 50% of the drivers who scored low on the driving assessment (≤ 5.5) reported one or more crashes in the following year; and no relationship was found between self reported driver confidence and self reported crashes.

The merit of this alternative explanation is difficult to determine, because the number of participants is too small to have a reasonable power for tests of interaction between confidence and examiners' opinion. There were, for example, only six well-calibrated drivers who scored low on the driving assessment (but were correctly insecure about their driving skills). No interaction effects of reasonable size can be found with such small numbers.

Also, no correlation was found between performance on the Adaptation Test and self-reported crashes. Drivers with and without a self-reported crash in the two years of the study performed similarly on the Adaptation Test; with a mean Adaptation Test score of 40% for both groups (see also Section 8.4.2).

To sum up, it is difficult to conclusively determine the influence of calibration on (self-reported) crash liability. Although a strong correlation was found between overestimation of skills and self-reported crashes, this correlation is difficult to interpret due to the intervening correlation with performance on the driving assessment. Another confusing issue is that the expected correlation between self-reported crashes and performance on the Adaptation Test was not found.

8.2. How can calibration be measured?

Research question 2. *Calibration can be measured with the Adaptation Test. This test measures the effect of perceived complexity on adaptation to task demands and is able to differentiate between experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers (Chapter 5).*

The Adaptation Test consisted of 18 traffic scenes presented in two almost identical photographs (see Appendix B). The photographs differed in one single detail, which increased the complexity of the situation. As the photographs were presented randomly and participants could not return to previously viewed photographs, participants were kept unaware of the varying level of complexity.

The respondents were asked to assess at what speed they *would* drive in the depicted situations. A response was considered 'correct' if the reported speed was lower in the complex situation than in the corresponding simple situation. A higher speed in the complex situation was regarded to be as equally unwanted as no speed differences; in both cases the response was considered 'incorrect'. The analysis of the reported speed showed that, on average, drivers indeed reported a lower speed for the complex situation in sixteen out of eighteen situations.

The Adaptation Test was purposely not meant to measure a driver's hazard perception skill, but to take it one step further: What does the driver do when a hazard has been perceived. The Adaptation Test obtains an indication of adaptation behaviour (i.e. speed) as a result of a specific situation and a driver's beliefs about his own driving skills. For example, a driver may spot the danger (or hazard), but decide not to decrease speed because he believes to be capable of handling the situation. Because the focus was not on spotting the danger, there was no time constraint for participants to respond, as is often the case in (classic) hazard perception tests (see e.g. Groeger, 2000).

Analysis of the Adaptation Test scores showed that novice drivers performed worse on this Test than experienced drivers (i.e. reported a lower speed less often in the more complex situation). In addition, unsafe drivers and overconfident drivers, as identified by comparing self-reported confidence with the on-road driving assessment, performed worse on the Adaptation Test.

Although these results seem to support the validity of the Adaptation Test as an instrument to measure adaptation, other results seem to suggest otherwise. The expected correlation between performance on the Adaptation Test and self-reported crashes was not found. Another unexpected result involved males being significantly better at adapting speed to the situation than female drivers. This was an unexpected result, since (young) males have a higher crash risk than (young) female drivers (OECD - ECMT, 2006). However, a similar inconsistency has been reported previously with respect to the driving test, where young males have less trouble passing the test but have a higher crash risk than young females afterwards (e.g. Crinson & Grayson, 2005; Maycock, 2002).

Overall, it was concluded that the Adaptation Test can be used to measure adaptation of driving speed to the situation. The test is not very sensitive in differentiating between individual drivers (i.e. even experienced drivers did not report a decreased speed for the complex situation in more than half of the situations), and no correlation was found between the Adaptation Test scores and self-reported crashes. However, because of its consistent ability to differentiate between groups of drivers (experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers), the Adaptation Test was considered a useful method for assessing calibration.

8.3. Development of calibration over time

Research question 3. *There seems to be no improvement in calibration, neither in self-assessment of skills nor in adaptation to task demands, in the first two years of independent driving (Chapter 7).*

The different components of the calibration model, self-assessment of skills and adaptation to task demands were analysed over the two year period of the study. Novice drivers were expected to start at a lower level compared to experienced drivers, but also to improve during the two years of the study approaching the level of the experienced drivers.

Concerning *self-assessment of skills* there seems to be no improvement after one year of driving. The results of the first driving assessment, in 2006, showed that novice drivers tended to overestimate their driving skills more than experienced drivers. The results of the driving assessment in 2007, showed a similar pattern of confidence concerning their driving skills within the novice drivers. Particularly novice drivers with the lowest scores on the driving assessment in 2007 are similarly confident as the novice drivers with the lowest scores on the driving assessment in 2006.

Concerning the *adaptation to task demands*, no improvement relative to the experienced group was found for the novice drivers in the first two years after licensing. The results for the Adaptation Test, which was designed to measure calibration, showed that both experienced and novice drivers improved somewhat over time. However, the amount of improvement for novice and experienced drivers did not differ.

The results of the Driver Behaviour Questionnaire (DBQ) showed a different pattern. Contrary to what was expected, novice and experienced drivers reported the same amount of violations in the first questionnaire. In later sessions the experienced drivers reported fewer violations, but in contrast, the novice drivers reported more and more violations. This is an unexpected result as it does not correspond to the crash risk development of young novice drivers, but similar patterns have been reported in previous studies. For example, novice drivers have been found to report more violations (measured with the DBQ) with increased experience (Bjørnskau & Sagberg, 2004) and in the first three years after licensing (Wells et al., 2008).

8.4. Strengths and weaknesses of the research methods

8.4.1. Longitudinal study

Hawthorne effect

With this longitudinal design, it is possible that drivers who, on a regular basis complete questionnaires and diaries about traffic behaviour and traffic safety, will become more aware of the risks of driving than drivers who perhaps never even think about traffic safety.

In order to rule out a 'Hawthorne effect' (i.e. simply measuring certain behaviour may change that behaviour; Bouchet et al., 1996; Murray et al., 1988), the novice drivers were randomly divided into two groups. The Novice I group started with their first session in October 2005, while the remaining Novice II group was invited to participate in October 2005, but did not start completing questionnaires until May 2006.

The Novice I group and Novice II group were compared on every variable that was used in this study, but no significant differences were found. It is safe to say that there is no evidence suggesting a 'Hawthorne' effect in this study.

Drop-out

Despite the efforts to keep participants committed to this study, still some participants eventually dropped out. Out of 688 participants starting the study, 156 (23%) did not finish it (12% of the experienced drivers and 27% of the novice drivers).

This difference between novice and experienced drivers is presumably caused by the way participants were invited for the study. Because, experienced drivers had to actively sign up for the study, they were probably highly motivated participants. The novice drivers, on the other hand, were approached in person in the driving test centres right after passing their driving test. They did not have to do anything to sign up for the study other than leave their personal information. In addition, having just passed their driving tests, most of them were euphoric and probably could not forsee the implications of participation in a two-year study.

The experienced drivers who dropped out of the study did not significantly differ from the experienced drivers that finished the study. On the other hand, the novice drivers who dropped out of the study were significantly more males, with lower education, and who had needed less driving education to obtain their drivers' licence, than the novice drivers who completed the study.

It is difficult to assess the effect of this selective drop-out on the results of the study; especially in comparison with other studies. The response rate for the novice drivers was extremely high (92%); it can be argued that the drop-out in this group are people who usually do not even sign up for a study, in other words, the unknown non-response group. As with other questionnaire studies, it must be kept in mind that the sample in our study is not completely representative for the whole novice driver population.

8.4.2. The Adaptation Test

The Adaptation Test was considered to be a useful method to measure calibration because of its consistent ability to differentiate between groups of drivers (experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers). However, two points must be mentioned here. First of all, the test uses self-reported speed to measure adaptation to task demands which can differ from actual speed behaviour. However, using reported speed allows for complete control over the complexity level of the traffic situations. In the current study, we have accounted for the differences between reported speed and actual speed choice by basing the between group analyses on the difference between the reported speeds in the complex and simple traffic situation, rather than analysing absolute speeds.

Secondly, none of the groups performed very well on the test. Of all the subgroups, experienced drivers performed best on the test. Still, in more than half of the situations (55%) even experienced drivers did not report a lower speed for the more complex situation. A closer look at the situations revealed that often the same speed was reported for both the simple and complex situation. Despite explicit instructions that we were not asking for the official speed limit in the traffic situations, reported speed often fluctuated around the speed limit. It seems that for some traffic situations the road design determined speed choice more than the specific circumstances did (i.e. complexity of the situation). In addition, it is also possible to deal with the increased complexity without changing speed (i.e. increasing vigilance, preparedness to respond or with dropping subsidiary tasks). This response is not measured with the Adaptation Test.

It could be that not enough difference was created between the photographs of the simple and complex situation. Increasing this difference could indeed have led to more reported differences between the complex and simple version of the traffic situation. However, it is not necessarily the case that this would improve the test's ability to differentiate between novice and experienced drivers. When the difference in complexity would become more obvious, novice drivers could possibly also be better at recognizing this difference. Moreover, to make comparisons between reported speed possible, both the simple and complex situation needed to have the same priority situation and legal limit. With this restriction, the differences between the simple and complex photographs had to be subtle.

Despite these problems, when a different speed was reported, this was almost always in the correct direction, i.e. a decreased speed for the more complex situation.

Besides these theoretical issues, some inconsistencies in the results showed up. Although the difference was small (η_P^2 = .01), males were significantly better at reporting adapting speed to the situation than female drivers. This is unexpected, since (young) males have a higher crash risk than (young) female drivers (OECD - ECMT, 2006), and adapting speed to the traffic situation is assumed to be related to crash risk involvement. This result suggests that, apparently, male drivers 'know' when to adapt their speed to the situation better than female drivers, but do not always apply this knowledge in real traffic.

In addition, there was no difference between the number of males and females in the group of overconfident drivers. So, there appears to be a gender effect that cannot be explained by overestimation of skills. Unfortunately the sample was too small to find significant interaction effects between overestimation, gender and performance on the Adaptation Test.

Also, we did not find the expected correlation between lower scores on the Adaptation Test and increased self-reported crashes. A closer look at the reported crashes revealed that these mostly concern minor crashes (e.g. a fender bender in a parking spot). It is possible that these crashes were caused by poor vehicle handling skills, while the Adaptation Test measures a higher-order-skill: adaptation of speed to the traffic situation. Nevertheless, this is only speculating. It is impossible to assess in hindsight whether the reported crash was either caused by poor vehicle handling skills, or, for example, by losing control over the car because of an inappropriate speed choice or, indeed, by some other road user.

8.4.3. The on-road driving assessment

In the design of the study an on-road driving assessment was considered to be the most effective instrument to measure 'ability to drive safely'. However, there have been many reservations about the use of an on-road driving assessment to measure safe driving ability (see Senserrick & Haworth, 2005 for a review). Chapter 3 (Section 3.5.3) provides an extensive review of the concerns with respect to the validity and reliability of an on-road driving assessment or the official driving test. Take, for example, the discrepancy between young males having less trouble passing the driving test, but subsequently having a higher crash risk than their female counterparts afterwards (e.g. Crinson & Grayson, 2005; Maycock, 2002), indicating that the driving test is not very successful in distinguishing safe drivers from unsafe drivers (i.e. lack of validity). Another problem that has been suggested is the possibility of a driver behaving differently with or without an observer present (i.e. lack of reliability; e.g. Baughan et al., 2005).

Because of the existing doubts, some precautions were taken to control for validity and reliability related bias. For example, the examiners were explicitly informed that the group consisted of different types of drivers, some older who rarely drove, and some younger who drove on a regular basis (e.g. as a professional courier), so that they would not automatically rate young participants ('most likely novice drivers') worse than older participants. In addition, the participants were instructed not to mention anything about their prior driving experiences to the examiner.

Also, a small scale experiment was conducted (see Section 3.5.3) to assess the influence of a participant's appearance on the assessment of 'ability to drive safely'. This experiment revealed no clear bias due to the participants' appearances, however, it could not *exclude* the existence of one either.

The results do provide indications that the on-road driving assessment used in the current study is a valid one. First of all, there was a significant correlation between the driver's self-reported speeding behaviour from the questionnaire and the examiner's opinion about the speeding behaviour of that driver. Although, this does not say anything about their 'actual' behaviour in normal traffic (we do not have this information in the current study), it does give us the impression that the participants behaved regularly in the driving assessment. This is consistent with the results of a study by Quimby et al. (1999), in which a strong correlation was found between the examiners assessment of speeding behaviour and the behaviour that was monitored previously in a natural setting (i.e. actual speeding behaviour).

Another indication that the driving assessment used in this study was a valid method is the strong correlation between the examiner's assessment and the driver's self-reported crashes. Of the drivers who 'failed' the driving assessment in 2006 (scoring ≤ 5.5), 50% reported one or more crashes in the second year of the study. Of the drivers who 'passed' the driving assessment in 2006, only 15% reported one or more crashes in the second year of the study. This is also consistent with previous studies in which a correlation was found between crash liability and performance on a driving assessment (Grayson et al., 2003) or the driving test (Maycock & Forsyth, 1997).

8.4.4. Construction of calibration groups

The construction of the calibration groups, which were used to study the effect of self-assessment on adaptation to task demands, was based on a comparison of self-reported confidence in driving skills ('driver confidence' scores) and the examiner's assessment of 'ability to drive safely' ('safe driving' scores).

It may be argued that reported differences between calibration groups are caused solely by just one of these factors, for example 'safe driving' and that the other ('driver confidence') did not contribute to the reported differences. It is difficult to eliminate this alternative explanation, because the number of participants was too small to have a reasonable power for tests of interaction between confidence and examiners' opinion. As a result, no significant interaction effects were found.

However, we did not find any evidence that either 'driver confidence' or 'safe driving' could explain all the results. For example, the violation items of the DBQ indicated a significant main effect for 'driver confidence' but there was no significant main effect for 'safe driving'. In contrast, for the Adaptation Test, a significant effect for 'safe driving' was found but not for 'driver confidence'. These results suggest that the combination of 'safe driving' and 'driver confidence' (i.e. self-assessment of skills) explains more variation than both factors separately.

8.5. Explanations for lack of development in calibration

The objective of this thesis is to investigate the high crash risk of young novice drivers, and more specifically, to explore if a development in calibration skills could explain the substantial decrease in crash risk in the first years after licensing. Crash rates are the highest in the first months after licensing and drop substantially over the first two years of driving, with the most pronounced decline during the first six months or during the first 5000 kilometres of driving (OECD - ECMT, 2006). This means that, within this first period of driving, something relevant for crash risk changes within these drivers as a result of driving experience. The theory and results of previous chapters suggested that (development of) calibration could play a role in the crash reduction. So the question is now: Why has no improvement in calibration relative to the experienced group been found in the two years of our study?

Lack of relative development has also been reported in other studies on the attitude and behavioural development of novice drivers. For example Whelan et al. (2004) followed a group of experienced and novice drivers from before they received their learners' permit, then one year and two years later. The results show clear differences between novice and experienced drivers in their performance on a hazard perception test; experienced drivers generally performed better on these tests. However no difference was found in the development of performance; both novices and experienced drivers performed slightly better over time.

Also, a large cohort study in the UK with over 20,000 new drivers found that in the three years after passing the driving test self-reported attitude and behaviour measures were remarkably stable (Wells et al., 2008). The authors did find some changes in confidence ratings: drivers began their driving careers with relatively high levels of confidence in their driving ability, but after six months their confidence was much lower and did not change to any great extent in the subsequent time periods. Other than this, the authors reported that new drivers entered the driving population with fairly fixed ideas, and in addition, many of the changes that they did find were in an unexpected direction, that is, not in a direction that would benefit road safety.

Some alternative explanations for the lack of development in our study can be mentioned. First of all, it is possible that the *operationalisation* of calibration was not correct or phrased even more strictly, calibration does not exist. This explanation is not very likely since, in Part 2 of this thesis, almost all assumptions from the calibration model were confirmed. In addition, there is much literature suggesting that at least some of the theory on calibration is correct.

A second explanation for the lack of development can be the *longitudinal nature* of the study. A total of 23% of the participants that started the study dropped out at some point in time. Is it possible that the drop-out was selective, and not random? It is also possible that participants were led by the questionnaires, that they formed an attitude in the first questionnaire, and simply did not change their opinion.

Both of these explanations are also not very likely. As is described in Chapter 3 (Section 3.2.4) background characteristics do not differ much between participants who dropped out of the study and those who completed every questionnaire.

Also, the cohort study by Wells et al. (2008) mentioned earlier reported the same lack of development in attitudes. This study can be viewed as a semi-longitudinal study, as only 11% of the data was collected from participants who completed all four succeeding surveys. The rest of the information was from respondents completing either one or two or three questionnaires (not necessarily the first). The authors conclude that ‘attrition’ or drop-out was not a large factor in their study, but still little development in attitudes was discovered.

Furthermore, the results of the Novice II group, which started with the study a half year later, indicate that the longitudinal nature of the study did not bias the results. The Novice I and Novice II group differed on none of the variables, suggesting that the two questionnaires that the Novice I group completed previously had no effect on the responses on the third and following questionnaires.

A third explanation for the lack of development could be that, calibration resembles more a *trait* rather than a skill, which develops over a longer period of time. This would explain the difference between novice and experienced drivers; the experienced drivers have at least ten years more driving experience. Also the lack of development in the study would be explained; two years would simply be too short a period to measure the improvement of calibration skills.

A final explanation for the lack of development is that there is a *generation difference* between the novice and experienced drivers in our study. The experienced drivers had passed their driving test at least ten years before the start of our study, but most of them had held their license for about twenty years. This means that the novice and experienced group did not only differ in 10-20 years of driving experience, but also had a 10-20 year difference in society. The experienced drivers came from a different generation with a very different licensing system and road traffic environment.

Although the final two explanations seem to fit the results best; the consequence is that calibration may *not* be the most important explanation the drop in crash risk of young novice drivers, because crash statistics show the most improvement in the first months and calibration does not seem to improve in the first two years after licensing.

8.6. Gender differences

The results of the driving diary indicated that there were no differences between young males and females in how much they drove in the weekend, during leisure time, or how much they drove with (peer) passengers. This is inconsistent with previous research suggesting that the difference in crash risk of young males and females is partly caused by the type of trips males and females make (Forsyth et al., 1995). And it also does not support the suggestion that females use their drivers licence mostly to get from one place to the next, while males spend more times driving simply for the sake of driving (i.e. during leisure time; Forsyth et al., 1995).

However, Forsyth et al. also found that young females drove significantly less long distance trips in the first year of driving. This result was replicated in our study; young males reported driving on just as many days of the week as young females, but also reported driving more *hours* per week, suggesting that young males make longer trips.

No differences between novice males and novice females were found in how much they reported speeding (in both urban and rural areas); and how much they reported driving without wearing a seatbelt (in both urban and rural areas). Novice males reported that they had violated traffic rules more often than female drivers. These results are in line with previous research suggesting that young male drivers are more involved in risky driving (Begg & Langley, 2001; Ulleberg, 2002) and report more violations (Ozkan et al., 2006; Verschuur, 2003b) than their female counterparts.

Novice males and females differ on other variables than experienced drivers and the total group of novice drivers do. For example, where novice drivers reported more trips in the weekend, during leisure time, or with (peer) passengers than experienced drivers, no such differences were found between novice males and females. This may suggest that the higher crash risk of novice *males* compared to novice *females* in the Netherlands is the result of a different mechanism than inexperience. If the high crash risk of young males was caused by the same reason that novice drivers in general have a high crash risk (but that the 'novice driver' problem is simply much more severe with young males) the males and females would differ on the same variables than novice and experienced drivers.

The results of this study suggest that something else is going on with Dutch young male drivers; and that they especially have more problems with conforming to traffic rules than to novice female drivers and experienced drivers. This was already concluded by Begg and Langley (2001), who also found that by the age of 26 many male drivers had 'matured out' of this risky behaviour.

8.7. Implications for driver education and licensing

Calibration seems to be an important factor for safe driving; theory and results of this thesis suggests that it is. However, the results of this study also indicate that calibration skills are not substantially acquired in the first two years after licensing, and therefore, are probably not the best candidate as a skill that should be improved in driving education or licensing systems.

However, as this thesis has also demonstrated, experience is an important factor in the reduction of crash risk, perhaps even more important than age (see Section 2.1.2). There lies a paradox in this fact. In order to become a safe driver, a novice driver should gain experience, and should, hence, be exposed to this elevated risk. Until we have more understanding of how experience works, and which factors of experience are important for crash risk, the only thing we can do is let novice drivers gain as much driving experience as possible in a safe environment.

The current study indicates that novice drivers drive more at night than experienced drivers, and we know from previous research that novice drivers have an elevated risk when they drive at night (Gregersen & Nyberg, 2002, as cited in OECD - ECMT, 2006). The same holds for driving with passengers. Novice drivers drive relatively more trips with passengers in the car, and previous studies suggest that they are more at risk when they do (Brorsson et al., 1993; Preusser et al., 1998; Williams & Ferguson, 2002). From a safety point of view it seems reasonable to ban these privileges in the first period of driving, in order to try to protect young novice drivers and let them gain experience outside of these risk prone circumstances.

An even safer way to gain driving experience is to have a period of accompanied driving after the driving test. In Sweden the crash risk is 33 times lower when driving accompanied is compared to driving independently after passing the driving test (Gregersen & Nyberg, 2002). English crash studies indicated that accompanied driving is 20 times safer than driving alone after passing the driving test (Forsyth et al., 1995). Accompanied driving allows young novice drivers to safely acquire driving experience, while reducing their crash risk (even if we do not know exactly what the role of experience is).

8.8. Further research

Calibration has been identified as relevant and important for safe driving by many authors (e.g. Brown & Groeger, 1988; Deery, 1999; Fuller, 2008; Gregersen, 1995; Kuiken & Twisk, 2001; Mayhew & Simpson, 1995; Mitsopoulos et al., 2006). However, this is the first time that calibration has been studied systematically. The current study indicates that calibration is indeed relevant for safe driving and that novice drivers are worse at calibration than drivers with at least 10 years of driving experience. On the other hand, calibration does not appear to develop over the first two years of the driving career. Further research should indicate when and how calibration skills do develop.

More specifically, a number of interesting topics for further research can be mentioned here, that were beyond the scope of the current study. An example is the driving assessment as a method to measure safe driving. There has been much debate whether this method is valid and reliable (see Senserrick & Haworth, 2005 for a review). The current study indicated that the examiners' assessment, based on the drivers' performance in a half hour drive, was a fairly good predictor of crash liability. Because the driving test is such an important countermeasure in reducing traffic risk (e.g. Baughan et al., 2005) and has been under debate, it seems important to further investigate this predictive validity, for example with more participants.

To study the effects of the appearance of a participant on the assessment by the examiner, a small scale experiment was conducted (see Section 3.5.3). In this experiment, the driver's appearance was varied (younger or older) independent of actual driving performance (good or poor). Unfortunately, the CBR examiners were unable to participate in this experiment, and students of the University of Amsterdam had to serve as examiners. It would be very interesting to repeat this experiment with professional examiners or at least with professional driving instructors (as a surrogate for examiners). It is expected that these groups are even less biased by the drivers' appearance, than students without any assessment experience.

The results of the Adaptation Test show interesting results, but there is also room for improvement. The test could be extended with more situations, which would make the test more sensitive for differences between groups.

Another option would be to let participants take the test while wearing an eye-tracking device. This experiment would provide more information on where the participants focus their attention during the test, and if they indeed notice the small differences between the photographs.

Furthermore, the Adaptation Test could be upgraded to a driving simulator version, where the photographs are replaced with videos of simple and complex situations. In this case the participants would not have to report their driving speed in the situation, but actual driving speeds could simply be measured on the simulator.

Finally, this study revealed some interesting differences between young male and young female drivers that were beyond the scope of the current study. Most interesting is the fact that young males and females differ from each other on other variables than novice drivers differ from experienced drivers. This suggests that the problem with young *male* drivers is a different problem than the general young novice driver problem, and gives good reason for an additional study into the motives and behaviour of young male drivers.

8.9. Conclusions

The results of this study indicate that calibration is an important factor for safe driving, and that experienced drivers are better 'calibrated' than novice drivers. However, little evidence was found that calibration improves in the first two years after licensing. Therefore, calibration cannot explain the decrease in crash risk of young novice drivers during the first years of their driving career.

Novice drivers overestimate their driving skills more than experienced drivers. They are not as optimistic about their driving skills as had been thought in the past, and seem to recognize that they are not as skilled (yet) as the average driver. However, when comparing their self-assessment with observed behaviour there are indications that they overestimate their driving skills more than experienced drivers.

Inadequate self-assessment, and specifically overestimation of skills, is connected to inadequate adaptation to task demands. Overconfident drivers generally report more violating behaviours and less instances of adaptation of driving speed to the complexity of traffic situations. Prior research has indicated calibration as a relevant factor for safe driving, and has linked calibration to the high crash rate of young, novice drivers. The results of the current study give an indication of *why* overestimation of skills affects crash risks; that is because, at least on the manoeuvring level, overconfident drivers adapt their behaviour less to the traffic situation.

However, little evidence was found that calibration skills improve in the first two years after licensing. There are two plausible explanations. First, calibration is more like a *trait* rather than a skill, which develops over a longer period of time. This would explain the difference between novice and experienced drivers; the experienced drivers have at least 10 years more driving experience. It would also explain the lack of development in the study; two years would then be simply too short to measure the improvement of calibration skills.

It is also possible that the difference in calibration between novice and experienced drivers can be brought back to a *generation difference*. Since the experienced drivers were 10-20 years older than the novice drivers, they were from a different generation with a very different licensing system.

Although calibration has been suggested as a contributing factor in the high crash risk of young novice drivers, the results of this study indicate that the decrease in crash risk in the first two years after licensing *cannot simply* be explained with calibration. With respect to driving experience, we have to continue our search for “the X-factor”.

References

Åberg, L. & Rimmö, P.-A. (1998). Dimensions of aberrant driver behaviour. *Ergonomics, 41*, 39-56.

Anders, S., Huestegge, L., Skottke, E.-M., Müsseler, J. & Debus, G. (2006). Becoming an expert: Eye movements in static traffic scenes. In: *Proceedings of the 16th tri-annual World Congress on Ergonomics of the International Ergonomics Association (IEA)*. Aachen: RWTH Aachen University, Department of Psychology.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 349-406.

Arnett, J.J. (2002). Developmental sources of crash risk in young drivers. *Injury Prevention, 8 (Suppl II)*, p. ii17-ii23.

AVV - Adviesdienst Verkeer en Vervoer (2007). *Rijden onder invloed in Nederland in 1999-2006 -- ontwikkeling van het alcoholgebruik van automobilisten in weekendnachten.* In Dutch. Rotterdam: Ministerie van Verkeer en Waterstaat; Directoraat-Generaal Rijkswaterstaat.

Bailey, T.J. (2009). Self-awareness and self-monitoring -- important components of best educational practice for novice drivers. *Journal of the Australian College of Road Safety*, 44-51.

Baughan, C., Gregersen, N.P., Hendrix, M. & Keskinen, E. (2005). *Towards European standards for testing: Final report.* Brussels: Commission Internationale des Examens de Conduite Automobile CIECA.

Begg, D. & Langley, J. (2001). Changes in risky driving behavior from age 21 to 26 years. *Journal of Safety Research, 32*(4), 491-499.

Begg, D. & Langley, J. (2009). A critical examination of the arguments against raising the car driver licensing age in New Zealand. *Traffic Injury Prevention, 10*, 1-8.

Begg, D.J., Langley, J.D. & Williams, S.M. (1999). A longitudinal study of lifestyle factors as predictors of injuries and crashes among young adults. *Accident Analysis & Prevention, 31*, 1-11.

Bijleveld, C.C.J.H., Van der Kamp, L.J.T., Mooijaart, A., Van der Kloot, W.A., Van der Leeden, R. & Van der Burg, E. (1999). *Longitudinal data analysis: Designs, models and methods*. London: SAGE Publications Ltd.

Bjørnskau, T. & Sagberg, F. (2004). What do novice drivers learn during the first months of driving? Improved handling skills or improved road user interaction? In: Underwood, G. (Ed.), *Traffic and Transport Psychology: Theory and Applicaton. Proceedings of the 3rd International Conference of Traffic and Transport Psychology (ICTTP)* (pp. 129-140). Oxford: Elsevier Ltd.

Bolstad, C.A. & Hess, T.M. (2000). Situation Awareness and aging. In: Endsley, M.R. & Garland, D.J. (Eds.), *Situation Awareness analysis and measurement* (pp. 277-302). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bouchet, C., Guillemin, F. & Briançon, S. (1996). Nonspecific effects in longitudinal studies: Impact on quality of life measures. *Journal of Clinical Epidemiology, 49*(1), 15-20.

Brorsson, B., Rydgren, H. & Ifver, J. (1993). Single-vehicle accidents in Sweden: A comparative study of risk and risk factors by age. *Journal of Safety Research, 24*, 55-65.

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

Brown, I.D. (1997). How traffic and transport systems can benefit from psychology. In: Rothengatter, T. & Carbonell Vayá, E.J. (Eds.), *Traffic and Transport Psychology: Theory and Applicaton. Proceedings of the International Conference of Traffic and Transport Psychology (ICTTP)* (pp. 9-19). Oxford: Elsevier Science Ltd.

Brown, I.D. & Groeger, J.A. (1988). Risk perception and decision taking during the transition between novice and experienced driver status. *Ergonomics, 31*(4), 585-597.

Brown, I.D., Groeger, J.A. & Biehl, B. (1987). Is driver training contributing enough towards road safety? In: Rothengatter, J.A. & de Bruin, R.A. (Eds.), *Road users & traffic safety* (pp. 135-156). Assen, The Netherlands: Van Gorcum.

Brown, J.D. (1986). Evaluations of self and others: Self-enhancement biases in social judgements. *Social Cognition, 4*, 353-376.

Cantin, V., Lavallière, M., Simoneau, M. & Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis & Prevention, 41*(4), 763-771.

Carbonell, E.J. & Martín-del-Río, B. (2002). Some insights on how to work with human error in traffic behaviour. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 147-159). Oxford: Elsevier Science Ltd.

Carsten, O. (2002). Multiple perspectives. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 11-21). Oxford: Elsevier Science Ltd.

Chapman, P.R. & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27*, 951-964.

Chi, M.T.H. (1978). Knowledge structures and memory development. In: Siegler, R. (Ed.), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ: Erlbaum.

Chi, M.T.H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In: Sternberg, R.J. (Ed.), *Advances in the psychology of human intelligence* (pp. 7-76). Hillsdale, NJ: Erlbaum.

Clarke, D.D., Ward, P., Bartle, C. & Truman, W. (2006). Young driver accidents in the UK: The influence of age, experience, and time of day. *Accident Analysis & Prevention, 38*(5), 871-878.

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

Congdon, P. (1999). *Vicroads hazard perception test, can it predict accidents?* Report No. CR 99-1. Victoria, Australia: Australian Council for Educational Research.

Crinson, L.F. & Grayson, G.B. (2005). Profile of the British learner driver. In: Underwood, G. (Ed.), *Traffic and Transport Psychology: Theory and Application. Proceedings of the ICTTP 2004* (pp. 157-170). Oxford: Elsevier.

Crundall, D.E. & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41*(4), 448-458.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. & Brookhuis, K.A. (2007a). Do young novice drivers overestimate their driving skills? In: *Proceedings of the Young Researchers Seminar*. Brno, Czech Republic: CDV.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. & Brookhuis, K.A. (2007b). Overestimation of skills affects driver's adaptation to task demands. In: *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 39-45). Stevensen, Washington, USA.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. & Brookhuis, K.A. (2008). The development of a method to measure speed adaptation to traffic complexity: Identifying novice, unsafe, and overconfident drivers. *Accident Analysis & Prevention, 40*(4), 1524-1530.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. & Brookhuis, K.A. (2009). The Adaptation Test: The development of a method to measure speed adaptation to traffic complexity. In: *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Big Sky Montana, USA.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H. & Brookhuis, K.A. (to appear). Novice drivers' development 0.5 and 1.5 years after licensing; results of an on-road driving assessment. In: *Proceedings of the 4th International Conference on Traffic & Transport Psychology*.

De Waard, D. (2002). Mental workload. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 161-175). Oxford: Elsevier Science Ltd.

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

DeJoy, D.M. (1992). An examination of gender differences in traffic accident risk perception. *Accident Analysis & Prevention, 24*(3), 237-246.

Dekker, S.W.A. (2005). Ten questions about human error: A new view of human factors and system safety. Mahwah, NJ: Lawrence Erlbaum Associates.

Delhomme, P. (1996). The consequences of driver's evaluations of their driving abilities and their feeling of control over driving behaviour. In: *Conference report of the international conference Roadsafe '96 "influences affecting road user behaviour"* (pp. 107-118). London.

Delhomme, P. & Meyer, T. (2000). Risk taking and self-efficacy among young male drivers: Self-efficacy and changing task demands. In: *Proceedings of the International Conference on Traffic and Transport Psychology ICTTP*. Berne, Switzerland.

Detweiler, M. & Schneider, W. (1991). Modeling the acquisition of dual-task skill in a connectionist/control architecture. In: Damos, D.L. (Ed.), *Multiple-task performance* (pp. 69-99). London: Taylor & Francis.

Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons, Inc.

Drummond, A.E. (2000). Paradigm lost! Paradise gained? An Australian's perspective on the novice driver problem. In: *Proceedings of the Novice Driver Conference*. Bristol.

Endsley, M.R. (1995). Toward a theory of Situation Awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M.R. (2000). Theoretical underpinnings of Situation Awareness: A critical review. In: Endsley, M.R. & Garland, D.J. (Eds.), *Situation Awareness analysis and measurement* (pp. 3-32). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Engström, I., Gregersen, N.P., Granström, K. & Nyberg, A. (2008). Young drivers--reduced crash risk with passengers in the vehicle. *Accident Analysis & Prevention, 40*(1), 341-348.

Engström, I., Gregersen, N.P., Hernetkoski, K., Keskinen, E. & Nyberg, A. (2003). *Young novice drivers, driver education and training.* Report no. 491A. Linköping, Sweden: Swedish National Road and Transport Research Institute.

Falk, B. (2010). Do drivers become less risk-prone after answering a questionnaire on risky driving behaviour? *Accident Analysis & Prevention, 42*(1), 235-244.

Falkmer, T. & Gregersen, N.P. (2001). Fixation patterns of learner drivers with and without cerebral palsy (cp) when driving in real traffic environments. *Transportation Research Part F: Traffic Psychology and Behaviour, 4*(3), 171-185.

Farrand, P. & McKenna, F.P. (2001). Risk perception in novice drivers: The relationship between questionnaire measures and response latency. *Transportation Research Part F: Traffic Psychology and Behaviour, 4*, 201-212.

Forsyth, E., Maycock, G. & Sexton, B. (1995). Cohort study of learner and novice drivers: Part 3, accidents, offences and driving experience in the first three years of driving. TRL Report 111. Crowthorne: Transport Research Laboratory.

Fuller, R. (1984). A conceptualisation of driving behaviour as threat avoidance. *Ergonomics, 27*, 1139-1155.

Fuller, R. (1988). On learning to make risky decisions. *Ergonomics, 31*, 519-526.

Fuller, R. (2000). The task-capability interface model of the driving process. *Recherche Transports Securité, 66*, 47-59.

Fuller, R. (2002a). Human factors and driving. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 77-97). Oxford: Elsevier Science Ltd.

Fuller, R. (2002b). Psychology and the highway engineer. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 1-10). Oxford: Elsevier Science Ltd.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention, 27*(3), 461-472.

Fuller, R. (2007). Motivational determinants of control in the driving task. In: Cacciabue, P.C. (Ed.), Modelling driver behaviour in automotive environments. Critical issues in driver interactions with intelligent transport systems (pp. 165-188). London: Springer-Verlag.

Fuller, R. (2008). Driver training and assessment: Implications of the task-difficulty homeostasis model. In: Dorn, L. (Ed.), *Driver Behaviour and Training, Volume III. Proceedings of the 3rd Conference on Driver Behaviour and Training* (pp. 337-348). Aldershot: Ashgate Publishing Limited.

Fuller, R. & Santos, J.A. (2002). *Human factors for highway engineers*. Oxford: Elsevier Science Ltd.

Grayson, G.B. & Elliott, M.A. (2004). The attitudes and reported behaviours of novice drivers: Results from the Cohort II study. In: *Behavioural Research in Road Safety 2004, Fourteenth Seminar*: Department of Transport.

Grayson, G.B., Maycock, G., Groeger, J.A., Hammond, S.M. & Field, D.T. (2003). *Risk, hazard perception and perceived control.* TRL Report 560. Crowthorne, Berkshire: TRL Limited.

Gregersen, N.P. (1995). *Prevention of road accidents among young novice car drivers.* PhD Thesis, Linköping University, Linköping, Sweden.

Gregersen, N.P. (1996). Young drivers' overestimation of their own skill -- an experiment on the relation between training strategy and skill. *Accident Analysis & Prevention, 28*(2), 243-250.

Gregersen, N.P. & Bjurulf, P. (1996). Young novice drivers: Towards a model of their accident involvement. *Accident Analysis & Prevention, 28*(2), 229-241.

Gregersen, N.P. & Nyberg, A. (2002). *Lay instruction during driver training – a study on how it is carried out and its impact on road safety.* VTI Report 481 (in Swedish with English abstract). Linköping: Swedish National Road and Transport Research Institute (VTI).

Groeger, J.A. (2000). Understanding driving - applying cognitive psychology to a complex everyday task. Hove, East Sussex, UK: Psychology Press Ltd.

Groeger, J.A. (2006). Youthfulness, inexperience, and sleep loss: The problems young drivers face and those they pose for us. *Injury Prevention, 12*(suppl 1), i19-i24.

Groeger, J.A. & Brown, I.D. (1989). Assessing one's own and others' driving ability: Influences of sex, age, and experience. *Accident Analysis & Prevention, 21*(2), 155-168.

Gugerty, L.J. & Tirre, W., C. (2000). Individual differences in Situation Awareness. In: Endsley, M.R. & Garland, D.J. (Eds.), *Situation Awareness analysis and measurement* (pp. 249-276). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Harris, A. & Drummond, A. (1998). Research and development of a self-assessment tool for novice drivers. In: *Proceedings of the Road Safety Research, Policing and Education Conference*. Wellington, New Zealand.

Harrison, W.A. (2004). Investigation of the driving experience of a sample of Victorian learner drivers. *Accident Analysis & Prevention, 36*, 885-891.

Harrison, W.A., Triggs, T.J. & Pronk, N.J. (1999). *Speed and young drivers: Developing countermeasures to target excessive speed behaviours amongst young drivers.* Report No. 159. Clayton, Australia: Monash University Accident Research Centre MUARC.

Hatakka, M., Keskinen, E., Gregersen, N.P., Glad, A. & Hernetkoski, K. (2002). From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(3), 201-215.

Herslund, M.-B. & Jørgensen, N.O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis & Prevention, 35*(6), 885-891.

Horrey, W.J., Lesch, M.F. & Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. *Accident Analysis & Prevention, 40*(2), 675-682.

Horrey, W.J., Lesch, M.F. & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of Safety Research, 40*(1), 7-12.

Horswill, M.S., Waylen, A.E. & Tofield, M.I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology, 34*(1), 177-195.

Houtenbos, M. (2008). *Expecting the unexpected. A study of interactive driving behaviour at intersections.* SWOV-Dissertationseries. PhD Thesis, Delft University, Leidschendam, The Netherlands.

Jaccard, J. & Becker, M.A. (2002). *Statistics for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth Publishing.

Jonah, B.A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis & Prevention, 29*(5), 651-665.

Kay, L., Bundy, A., Clemson, L. & Jolly, N. (2008). Validity and reliability of the on-road driving assessment with senior drivers. *Accident Analysis & Prevention, 40*(2), 751-759.

Keating, D.P. (2007). Understanding adolescent development: Implications for driving safety. *Journal of Safety Research, 38*(2), 147-157.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Kuiken, M.J. & Twisk, D.A.M. (2001). *Safe driving and the training of calibration.* Report no. R-2001-29. Leidschendam, the Netherlands: SWOV Institute for Road Safety Research.

Kweon, Y.-J. & Kockelman, K.M. (2003). Overall injury risk to different drivers: Combining exposure, frequency, and severity models. *Accident Analysis & Prevention, 35*(4), 441-450.

Mallon, K.L. (2006). *The influence of self-awareness of driving ability on on-road performance of persons with acquired brain injury.* PhD Thesis, Queensland University of Technology, School of Optometry, Brisbane, Queensland.

Martens, M.H. & Hoedemaeker, M. (2001). *Behavioural indicators for driver workload and comfort: A literature overview.* TNO-report TM-01-D019. Soesterberg, The Netherlands: TNO Human Factors.

Mathews, M.L. & Moran, A.R. (1986). Age differences in male drivers' perception of accident risk: The role of perceived driving ability. *Accident Analysis & Prevention, 18*, 299-313.

Matsuura, T. (2005). Are young male drivers more overconfident than older male drivers? *Japanese Journal of Applied Psychology, 30*(2), 79-86.

Maycock, G. (2002). *Novice driver accidents and the driving test.* TRL Report 527. Crowthorne, Berkshire: Transport Research Laboratory TRL.

Maycock, G. & Forsyth, E. (1997). Cohort study of learner and novice drivers. Part 4: Novice driver accidents in relation to methods of learning to drive, performance in the driving test and self assessed driving ability and behaviour. TRL report 275. Crowthorne: Transport Research Laboratory.

Maycock, G., Lockwood, C.R. & Lester, J.F. (1991). *The accident liability of car drivers.* TRL Research Report 315. Crowthorne, Berkshire: Transport Research Laboratory TRL.

Mayhew, D.R. & Simpson, H.M. (1995). The role of driving experience: Implications for the training and licensing of new drivers. Toronto, Ontario: Insurance Bureau of Canada.

Mayhew, D.R., Simpson, H.M. & Pak, A. (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis & Prevention, 35*(5), 683-691.

Mayhew, D.R., Simpson, H.M., Singhal, D. & Desmond, K. (2006). *Reducing the crash risk for young drivers.* Washington, D.C.: American Automobile Association AAA Foundation for Traffic Safety.

McCartt, A.T., Mayhew, D.R., Braitman, K.A., Ferguson, S.A. & Simpson, H.M. (2009). Effects of age and experience on young driver crashes: Review of recent literature. *Traffic Injury Prevention, 10*, 209-219.

McCormick, I.A., Walkey, F.H. & Green, D.E. (1986). Comparative perceptions of driver ability -- a confirmation and expansion. *Accident Analysis & Prevention, 18*, 205-208.

McKenna, F.P. (1993). It won't happen to me: Unrealistic optimism or illusion of control? *British Journal of Psychology, 84*, 51-65.

McKenna, F.P. & Crick, J.L. (1994). *Hazard perception in drivers; a methodology for testing and training.* Contractor Report; CR 313. Crowthorne, Berkshire: Transport and Road Research Laboratory TRRL.

McKenna, F.P. & Farrand, P. (1999). *The role of automaticity in driving.* Report No. PA3524/99. Crowthorne, UK: Transport Research Laboratory (TRL).

McKenna, F.P. & Horswill, M.S. (1999). Hazard perception and its relevance for driver licensing. *IATSS Research, 23*, 36-41.

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

McKenna, F.P., Waylen, A.E. & Burkes, M.E. (1998). *Male and female drivers: How different are they?* Berkshire, UK: THe University of Reading, Automobile Association AA Foundation for Road Safety Research.

McKnight, A.J. & McKnight, A.S. (2003). Young novice drivers: Careless or clueless? *Accident Analysis & Prevention, 35*, 921-925.

McPherson, S.L. & Thomas, J.R. (1989). Relation of knowledge and performance in boys' tennis: Age and expertise. *Journal of Experimental Child Psychology, 48*, 190-211.

Mesken, J., Lajunen, T. & Summala, H. (2002). Interpersonal violations, speeding violations and their relation to accident involvement in Finland. *Ergonomics, 45*(7), 469 - 483.

Michon, J.A. (1985). A critical view of driver behavior models: What do we know, what should we do? In: Evans, L. & Schwing, R.C. (Eds.), *Human behavior and traffic safety: Proceedings of a general motor symposium on human behavior and traffic safety* (pp. 516-520). Michigan: Plenum Press.

Mitsopoulos, E., Triggs, T. & Regan, M. (2006). Examining novice driver calibration through novel use of a driving simulator. In: *Proceedings of the SimTecT 2006 Simulation Conference: Challenges and Opportunities for a Complex and Networked World (SimTecT 2006)*. Melbourne, Australia.

Monárrez-Espino, J., Hasselberg, M. & Laflamme, L. (2006). First year as a licensed car driver: Gender differences in crash experience. *Safety Science, 44*, 75-85.

Mourant, R.R. & Rockwell, T.H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors, 14*(4), 325-335.

Murray, Å. (2003). Decreasing number of young licence holders and reduced number of accidents; a description of trends in Sweden. *Accident Analysis & Prevention, 35*, 841-850.

Murray, M., Swan, A.V., Kiryluk, S. & Clarke, G.C. (1988). The Hawthorne effect in the measurement of adolescent smoking. *Journal of Epidemiology and Community Health, 42*, 304-306.

Mynttinen, S., Sundström, A., Koivukoski, M., Hakuli, K., Keskinen, E. & Henriksson, W. (2009). Are novice drivers overconfident? A comparison of self-assessed and examiner-assessed driver competences in a Finnish and Swedish sample. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(2), 120-130.

Mynttinen, S., Sundström, A., Vissers, J., Koivukoski, M., Hakuli, K. & Keskinen, E. (2009). Self-assessed driver competence among novice drivers - a comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample. *Journal of Safety Research, 40*(4), 301-309.

Näätänen, R. & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention, 6*(3-4), 243-261.

Näätänen, R. & Summala, H. (1976). *Road user behaviour and traffic accidents*. Amsterdam: North-Holland Publishing Company.

NHTSA (2002). *Traffic safety facts 2002: Young drivers.* Report No. DOT HS 809 619. Washington, DC.: US Department of Transportation; National Highway Traffic Safety Administration (NHTSA).

Norman, D.A. (1981). Categorisation of action slips. *Psychological Review, 88*(1), 1-15.

Nyberg, A. & Gregersen, N.P. (2007). Practicing for and performance on drivers license tests in relation to gender differences in crash involvement among novice drivers. *Journal of Safety Research, 38*(1), 71-80.

Nyberg, A., Gregersen, N.P. & Wiklund, M. (2007). Practicing in relation to the outcome of the driving test. *Accident Analysis & Prevention, 39*(1), 159-168.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In: Boff, K.R., Kaufman, L. & Thomas, J.P. (Eds.), *Handbook of perception and human performance. Volume II, cognitive processes and performance*. New York: Wiley.

OECD - ECMT (2006). *Young drivers: The road to safety.* Paris: OECD - Organisation for Economic Co-operation and Development & ECMT - European Conference of Ministers of Transport; Joint OECD/ECMT Transport Research Centre.

Ozkan, T., Lajunen, T. & Summala, H. (2006). Driver Behaviour Questionnaire: A follow-up study. *Accident Analysis & Prevention, 38*, 386–395.

Parker, D., Reason, J.T., Manstead, A.S.R. & Stradling, S.G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38*(5), 1036-1048.

Parker, D., West, R., Stradling, S. & Manstead, A.S.R. (1995). Behavioural characteristics and involvement in different types of traffic accident. *Accident Analysis & Prevention, 27*(4), 571-581.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L. & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention, 38*(5), 887-894.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D.L., Blumenthal, J., Giedd, J.N., et al. (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science, 283*, 1908-1911.

Pelz, D.C. & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident Analysis & Prevention, 6*(1), 45-58.

Preusser, D.F. (2002). BAC and fatal crash risk. In: Mayhew, D.R. & Dussault, C. (Eds.), *Proceedings of the 16th International Conference on Alcohol, Drugs and Traffic Safety* (pp. 935-939). Québec: Société de l'assurance automobile du Québec.

Preusser, D.F., Ferguson, S.A. & Williams, A.F. (1998). The effect of teenage passengers on the fatal crash risk of teenage drivers. *Accident Analysis & Prevention, 30*, 217-222.

Quimby, A., Maycock, G., Palmer, C. & Grayson, G.B. (1999). *Drivers' speed choice: An in-depth study.* TRL Report 326. Crowthorne, Berkshire: Transport Research Laboratory.

Quimby, A.R. & Watts, G.R. (1981). *Human factors and driving performance.* TRRL Laboratory Report 1004. Crowthorne, Berkshire: Transport and Road Research Laboratory.

Ranney, T.A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis & Prevention, 26*(6), 733-750.

Rasmussen, J. (1986). Information processing and human-machine interaction; an approach to cognitive engineering. Amsterdam, the Netherlands: North-Holland.

Reason, J. (1990). *Human error.* Cambridge: Cambridge University Press.

Reason, J., Manstead, A., Stradling, S., Baxter, J. & Campbell, K. (1990). Errors and violations on the road: A real distinction? *Ergonomics, 33*(10/11), 1315-1332.

Renge, K. (1998). Drivers' hazard and risk perception, confidence in safe driving, and choice of speed. *IATSS Research, 22*(2), 103-110.

Risser, R. (1985). Behavior in traffic conflict situations. *Accident Analysis & Prevention, 17*(2), 179-197.

Ryan, G.A., Legge, M. & Rosman, D. (1998). Age related changes in drivers' crash risk and crash type. *Accident Analysis and Prevention, 30*, 379-387.

Sagberg, F. (1998). Month-by-month changes in accident risk among novice drivers. In: *Proceedings of the 24th International Congress of Applied Psychology*. San Francisco.

Sagberg, F. & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis & Prevention, 38*, 407-414.

Sayer, J.R., Devonshire, J.M. & Flannagan, C.A. (2005). *The effects of secondary tasks on naturalistic driving performance.* Report no. UMTRI-2005-29. Ann Arbor, Michigan, USA: The University of Michigan, Transportation Research Institute.

Searleman, A. & Herrmann, D. (1994). *Memory from a broader perspective*. Singapore: McGraw-Hill, Inc.

Senserrick, T. & Haworth, N. (2005). *Review of literature regarding national and international young driver training, licensing and regulatory systems.* Report no. 239. Clayton, Victoria: Monash University Accident Research Centre, MUARC.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127-190.

Shinar, D., Meir, M. & Ben-Shoham, I. (1998). How automatic is manual gear shifting? *Human Factors, 40*(4), 647-654.

Siegrist, S. (ed.) (1999). Driver training, testing and licensing - towards theory-based management of young drivers' injury risk in road traffic. Results of eu-project GADGET, guarding automobile drivers through guidance, education and technology, Work Package 3. Berne: Schweizerische Beratungsstelle für Unfallverhütung, BfU.

Skaar, N.R. & Williams, J.E. (2005). Gender differences in predicting unsafe driving behaviors in young adults. In: *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. pp. 363 - 369). Rockport, Maine, USA: University of Iowa.

Sowell, E.R., Thompson, P.M., Holmes, C.J., Jernigan, T.L. & Toga, A.W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience, 2*(10), 859-861.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Sullman, M.J.M., Meadows, M.L. & Pajo, K.B. (2002). Aberrant driving behaviours amongst New Zealand truck drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(3), 217-232.

Summala, H. (2002). Behavioural adaptation and drivers' task control. In: Fuller, R. & Santos, J.A. (Eds.), *Human factors for highway engineers* (pp. 189-200). Oxford: Elsevier Science Ltd.

Summala, H. & Näätänen, R. (1988). The zero-risk theory and overtaking decisions. In: Rothengatter, T. & de Bruin, R. (Eds.), *Road user behaviour: theory and research* (pp. 82-92). Assen, the Netherlands: Van Gorcum.

Sundström, A. (2008a). Construct validation and psychometric evaluation of the self-efficacy scale for driver competence. *European Journal of Psychological Assessment, 24*(3), 198-206.

Sundström, A. (2008b). Self-assessment of driving skill - a review from a measurement perspective. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(1), 1-9.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143-148.

SWOV - Fact sheet (2008). *Young novice drivers.* Leidschendam, the Netherlands: Institute for Road Safety Research SWOV.

Taylor, D.H. (1964). Driver's galvanic skin response and the risk of accident. *Ergonomics, 7*, 439-451.

Triggs, T.J. & Regan, M.A. (1998). Development of a cognitive skills training product for novice drivers. In: *Proceedings of the Road Safety, Policing and Education Conference*. Wellington, New Zealand.

Tronsmoen, T. (2008). Associations between self-assessment of driving ability, driver training and crash involvement among young drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(5), 334-346.

Turetschek, C. & Risser, R. (2008). Comparison of novice drivers in Austria and the Czech Republic with and without the use of intelligent speed adaptation. In: Dorn, L. (Ed.), *Driver Behaviour and Training, Volume III. Proceedings of the 3rd Conference on Driver Behaviour and Training* (pp. 311-321). Aldershot: Ashgate Publishing Limited.

Twisk, D.A.M. (1995). *Factors contributing to the high accident liability of novice drivers and the role of driver training.* Report no. D-95-9. Leidschendam, the Netherlands: SWOV Institute for Road Safety Research.

Twisk, D.A.M. & Stacey, C. (2007). Trends in young driver risk and countermeasures in European countries. *Journal of Safety Research, 38*(2), 245-257.

Ulleberg, P. (2002). Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. *Transportation Research Part F: Traffic Psychology and Behaviour, 4*(4), 279-297.

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

Underwood, G., Chapman, P., Wright, S. & Crundall, D. (1997). Estimating accident liability. In: Rothengatter, T. & Carbonell Vayá, E.J. (Eds.), *Traffic and transport psychology: Theory and application* (pp. 247-258). Amsterdam: Pergamon.

Van der Hulst, M. (1999). *Adaptive control of safety margins in driving.* PhD Thesis, Rijksuniversiteit Groningen, Groningen.

Van der Zwan, S. (2009). *De invloed van leeftijd op de beoordeling van rijvaardigheid.* Afstudeerscriptie 0248347; in Dutch. Amsterdam: Universiteit van Amsterdam.

Van Elslande, P. & Faucher-Alberton, L. (1997). When expectancies become certainties: A potential adverse effect of experience. In: Rothengatter, T. & Carbonell Vayá, E.J. (Eds.), *Traffic and transport psychology: Theory and application* (pp. 147-159). Amsterdam: Pergamon.

Van Leyden Sr., J. (1993). *Psychologische functieleer* (3rd ed.). Houten/Zaventem, the Netherlands: Bohn Stafleu Van Loghum.

Verschuur, W.L.G. (2003a). The Driver Behaviour Questionnaire in the Netherlands. In: *Proceedings of the 16th ICTCT Workshop*. Soesterberg, The Netherlands.

Verschuur, W.L.G. (2003b). *Onderzoek uitgevoerd met de 'Driver Behaviour Questionnaire'; eindrapportage.* Werkgroep Veiligheid/R-03/68; in Dutch. Leiden, The Netherlands: Rijksuniversiteit Leiden.

Verschuur, W.L.G. & Hurts, K. (2008). Modeling safe and unsafe driving behaviour. *Accident Analysis & Prevention, 40*(2), 644-656.

Vlakveld, W.P. (2004). New policy proposals for novice drivers in the Netherlands. In: *Behavioural Research in Road Safety, Fourteenth Seminar*: Department of Transport.

Vlakveld, W.P. (2005). Jonge beginnende automobilisten, hun hoge ongevals risico en maatregelen om het ongevalsrisico terug te dringen. Een literatuurstudie. Report no. R-2005-3; in Dutch. Leidschendam, the Netherlands: SWOV Institute for Road Safety Research.

Waller, P.F., Elliott, M.R., Shope, J.T., Raghunathan, T.E. & Little, R.J.A. (2001). Changes in young adult offence and crash patterns over time. *Accident Analysis & Prevention, 33*(1), 117-128.

Walton, D. (1999). Examining the self-enhancement bias: Professional truck drivers' perceptions of speed, safety, skill and consideration. *Transportation Research Part F: Traffic Psychology and Behaviour, 2*, 91-113.

Waylen, A.E., Horswill, M.S., Alexander, J.L. & McKenna, F.P. (2004). Do expert drivers have a reduced illusion of superiority? *Transportation Research Part F: Traffic Psychology and Behaviour, 7*, 323-331.

Wegman, F.C.M. & Aarts, L.T. (eds.) (2006). *Advancing sustainable safety: National road safety outlook for 2005-2020.* Leidschendam, the Netherlands: SWOV Institute for Road Safety Research.

Weinstein, N.D. & Lyon, J.E. (1999). Mindset, optimistic bias about personal risk and health-protective behaviour. *British Journal of Health Psychology, 4*, 289-300.

Wells, P., Tong, S., Sexton, B., Grayson, G. & Jones, E. (2008). *Cohort II: A study of learner and new drivers; volume 1 – main report.* London: Department for Transport.

Whelan, M., Senserrick, T., Groeger, J., Triggs, T. & Hosking, S. (2004). *Learner driver experience project.* Report No. 221 Victoria, Australia: Monash University Accident Research Centre MUARC.

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

Wilde, G.J.S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis, 2*, 209-258.

Wilde, G.J.S. (1988). Risk homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissension in recent reactions. *Ergonomics, 31*, 441-468.

Williams, A.F. (2003). Teenage drivers: Patterns of risk. *Journal of Safety Research, 34*, 5-15.

Williams, A.F. & Ferguson, S.A. (2002). Rationale for graduated licensing and the risks it should address. *Injury Prevention, 8 (Suppl II)*, ii9-ii16.

Summary

Young, novice drivers have a higher crash rate than drivers from all other age categories. Crash rates are highest in the first months after licensing and drop substantially over the first two years of unsupervised driving, with the most pronounced decline during the first six months or during the first 5000 kilometres of driving. There are basically two factors associated with the high crash risk: young age and lack of experience. Although young age is an important factor, crash studies suggest that the decrease in risk is more strongly related to gaining *experience* than to biological maturation. All novice drivers, irrespective of age, show an exponentially decreasing crash risk in the first years of their driving career. Therefore, this thesis focuses on how experience reduces crash risk over time, and which relevant processes are involved.

Chapter 2 provides a theoretical background on the subject of young novice drivers and more specific on the development of (driving) experience. There are many indications that with practice parts of the driving task develop into a routine; that is, driving (sub)tasks become automated. This automation of driving subtasks leads to a decrease in mental workload. However, lack of automation does not explain the high crash risk of young novice drivers completely. Since the driving task is 'self-paced', the driver can adjust the task demands (e.g. by reducing speed or increasing following distance), thus decreasing workload. In theory, a novice driver can use this strategy to overcome the limitations of his performance; a novice driver can decrease the task demands to fit his (deficient) level of automated driving.

This self-pacing aspect of the driving task is incorporated into the calibration model, inspired by Brown (1989) and Fuller (2005), which states that:

> *For safe driving, a driver needs to assess his own driving skills, weigh them against his perception of the complexity of the situation, and, as a final step, use the result of this balancing to adapt to task demands.*

To study if calibration could explain the drop in crash risk of young novice drivers in the first years of their driving careers, three research questions are formulated in Chapter 1:

1. To what extent is poor calibration a contributing factor in the high crash risk of young novice drivers?
   a. Do young novice drivers overestimate their skills more than experienced drivers?
   b. Does an inadequate self-assessment of skills affect adaptation to task demands?
   c. Is there a relationship between the elements of the calibration model and self-reported crashes?

2. How can calibration be measured?

3. How does calibration develop over time?

Chapter 3 describes the methods that were used to monitor the development of calibration. A group of young novice drivers was intensively followed from the moment of licensing over a period of two years. During this period, participants completed questionnaires, kept a driving diary and participated in an on-road driving assessment.

The questionnaire contained items on self-assessment of skill and perceived risks in traffic. To monitor the internal processes of calibration, the Adaptation Test was developed and administered in the questionnaire. In the driving diary, drivers reported on the trips that they made and the situations encountered in traffic. Drivers reported, for example, how much they had driven, if they had driven at night, with or without passengers, and if they had consumed any alcohol before driving.

To compare the reported experiences with actual driving performance, a subgroup participated in an on-road driving assessment. Driving skills were assessed on two occasions (in 2006 and 2007) in order to detect any changes over time as a result of driving experience.

Chapter 4 describes how *self-assessment of skills* can be measured. More specifically, it studies whether novice drivers overestimated their driving skills more than experienced drivers. Two approaches for measuring self-assessment of driving skills were used. In the first (more traditional) approach the drivers' assessments of their own skills were compared to the group average. Results of this approach suggest that novice drivers are modest about their skills. That is, experienced drivers are more optimistic about their driving skills and risks in traffic than novice drivers.

In the second approach, the driver's assessment was compared to a more independent measure of driving skills (the assessment of an examiner). The results of this approach indicate that novice drivers overestimate their driving skills more than experienced drivers (Research question 1a). Especially novice drivers with a low score for 'safe driving' on the driving assessment are more confident than their performance would suggest. The experienced drivers with the lowest scores on the driving assessment are less confident about their driving skills. It was concluded that a comparison with an independent measure of driving skills is necessary to measure over-estimation (or under-estimation) of skills.

Chapter 5 describes the development of the Adaptation Test, which measures *adaptation to task demands* as a function of *perceived complexity of the situation.* The Adaptation Test consists of 18 traffic situations presented in two almost identical photographs, that differ in one single detail which increases the situation's complexity. The direction of the difference in reported speed between the two pictures was used as an indication of drivers' inclination to adapt their speed to the complexity of the traffic situation.

The results show that novice drivers performed worse on the Adaptation Test (i.e. less often reported a lower speed in the more complex situation) than experienced drivers. In addition, unsafe drivers and overconfident drivers, as identified in the on-road driving assessment, performed worse on the Adaptation Test. However, the expected correlation between performance on the Adaptation Test and self-reported crashes was not found.

Because of its consistent ability to differentiate between groups of drivers (experienced versus novice drivers, safe versus unsafe drivers, and well-calibrated versus overconfident drivers), it was concluded that the Adaptation Test is effective at measuring adaptation of driving speed to the complexity of the situation (Research question 2).

Chapter 6 describes if and to what extent inadequate *self-assessment of skills* affects *adaptation to task demands*. The results showed that overconfident drivers reported significantly more violating behaviour than well-calibrated and insecure drivers. Overconfident drivers also showed significantly less instances of adaptation to traffic complexity, as measured with the Adaptation Test. Finally, 50% of the overconfident drivers reported one or more crashes in the second year of the study which was significantly more than the well-calibrated and insecure drivers (respectively 18% and 10%).

The results presented in Chapter 6 indicate that overconfidence is related to inadequate adaptation to task demands (Research question 1b). The relationship between overconfidence and the number of self-reported crashes (Research question 1c) was less conclusive. Although a strong correlation was found between overestimation of skills and self-reported crashes, this correlation is difficult to interpret due to the intervening correlation with performance on the driving assessment. Also the expected correlation between self-reported crashes and performance on the Adaptation Test was not found.

Chapter 7 studies the development of (the separate elements of) calibration over time. *Self-assessment of skills* (Chapter 4) was compared from one year to the next and the development in *adaptation to task demands* (Chapter 6) was described on six occasions during two years. The scores on the Adaptation Test (Chapter 5) during the two years of the study were compared for the novice and experienced drivers. In addition, this chapter reports the results of the driving diary that the participants kept during six periods over two years of the study.

The results showed that, although a difference was found between novice and experienced drivers in their calibration skills (as was already found in previous chapters), the expected relative *improvement* in calibration skills of the novice drivers was not found in the two years after receiving their driver's licence (Research question 3).

The results of the driving diary suggest that novice drivers differ from experienced drivers with respect to the type of trips they make. Novice drivers drive more at night, during weekends, during leisure time and with passengers. In addition, novice *male* drivers drive faster and without seatbelts more often than novice *female* drivers. There are no indications that these trip characteristics change much in the first two years after licensing.

Chapter 8 sums up all the findings from the previous chapters to answer the three research questions formulated in the introduction and discusses the strength and weaknesses of this study. For example, with respect to the longitudinal design of the study, no indications were found for a 'Hawthorne effect' (i.e. simply measuring certain behaviour may change that behaviour) or for the unavoidable drop-out of participants having influenced the results of this study considerably.

Regarding the Adaptation Test, there were some irresolvable weaknesses. For example, the fact that the Adaptation Test relies on self-reported speed behaviour which can differ from actual speed behaviour. And secondly, the finding that none of the groups performed very well on the test. Of all the subgroups, experienced drivers performed best on the test. Still, in more than half of the situations (55%) even experienced drivers did not report a lower speed for the more complex situation. Therefore, the Adaptation Test is not very sensitive in discriminating between individual drivers.

With respect to the on-road driving assessment, results suggest that this is a valid method to measure ‘ability to drive safely’. Firstly, because a significant correlation was found between the driver’s self-reported speeding behaviour from the questionnaire and the examiner’s opinion about the speeding behaviour of that driver. And secondly, a strong correlation was found between the examiner’s assessment and the driver’s self-reported crashes. Of the drivers who scored lowest on the driving assessment in 2006 (scoring ≤ 5.5), 50% reported one or more crashes in the second year of the study. Of the drivers with higher scores (scoring > 5.5) in 2006, only 15% reported one or more crashes in the second year of the study.

Finally, the construction of the calibration groups was based on a comparison of self-reported confidence in driving skills (‘driver confidence’ scores) and the examiner’s assessment of ‘ability to drive safely’ (‘safe driving’ scores). It may be argued that reported differences between calibration groups are caused solely by just one of these factors, for example ‘safe driving’ and that the other (‘driver confidence’) did not contribute to the reported differences. Although no evidence for this alternative explanation was found, it could also not be eliminated because the number of participants was too small to have a reasonable power for tests of interaction between confidence and examiners’ opinion.

Chapter 8 discusses the implications for driver education and licensing and suggests some topics for further research. Because the results of this study indicate that calibration skills are not substantially acquired in the first two years after licensing, these are probably not skills that could be improved in driving education or within licensing systems. Further research should indicate when and how calibration skills do develop.

Nevertheless, as this thesis has also demonstrated, the development of experience is an important factor in the reduction of crash risk. Until we have more understanding of how experience works, and which factors of experience are important for crash risk, the only thing we can do is let novice drivers gain as much driving experience as possible in a safe environment. This can be achieved by limiting the privileges of young novice drivers (e.g. limit on passengers, driving in the dark) or by introducing a period of accompanied driving after licensing.

This thesis indicates that the examiner's assessment, based on the driver's performance in a half hour drive, is a fairly good predictor of crash liability. Because the driving test is such an important measure in reducing traffic risk and has been under debate, it seems important to further investigate this predictive validity of the driving test. A small scale experiment with college students (see Section 3.5.3), studied the effects of the appearance of a driver on the assessment of his driving skills by an observer. It would be very interesting to repeat this experiment with professional examiners or at least with professional driving instructors (as a surrogate for examiners).

Finally, Chapter 8 discusses several explanations for the lack of development of calibration in this study, of which two seem most plausible. First, calibration could be more like a *trait* rather than a skill, which develops over a longer period of time. This would explain the difference between novice and experienced drivers; the experienced drivers have at least ten years more driving experience. It would also explain the lack of development in the study; two years would then be simply too short to measure the improvement of calibration skills.

It is also possible that the difference in calibration between novice and experienced drivers can be brought back to a *generation difference*. Since the experienced drivers were 10-20 years older than the novice drivers, they were from a different generation with a very different licensing system.

Although calibration has been suggested as a contributing factor in the high crash risk of young novice drivers, the results of this study indicate that the decrease in crash risk in the first two years after licensing *cannot simply* be explained with calibration. With respect to driving experience, we have to continue our search for "the X-factor".

Samenvatting

Jonge, onervaren automobilisten hebben een grotere kans om bij een ongeval betrokken te raken dan automobilisten uit alle andere leeftijdscategorieën. Het ongevalsrisico is het hoogst in de eerste maanden na het behalen van het rijbewijs, en daalt vervolgens aanzienlijk tijdens de eerste twee jaar van zelfstandig rijden. De grootste afname in het risico vindt daarbij plaats gedurende de eerste zes maanden of tijdens de eerste 5000 kilometer. Er zijn grofweg twee factoren die te maken hebben met dit hoge ongevalsrisico: de jonge leeftijd en het gebrek aan rijervaring. Hoewel de jonge leeftijd een belangrijke factor is, laten ongevallenstudies zien dat de afname in ongevalsrisico meer te maken heeft met de ontwikkeling van rijervaring dan met biologische ontwikkeling. Alle onervaren automobilisten laten, onafhankelijk van hun leeftijd, een exponentiële afname zien van het ongevalsrisico in de eerste jaren van hun rijcarrière. Dit proefschrift richt zich dan ook op de vraag hoe de ontwikkeling van rijervaring bijdraagt aan de afname van het ongevalsrisico en welke relevante processen hierbij een rol spelen.

In Hoofdstuk 2 worden de theoretische achtergronden van de jonge onervaren automobilisten uiteengezet, en wordt dieper ingegaan op de ontwikkeling van (rij)ervaring. Er zijn aanwijzingen dat, met oefening, verschillende onderdelen van de rijtaak routine worden. Dat wil zeggen: deze (sub)taken worden automatismen, wat weer zorgt voor een afname in mentale belasting tijdens het autorijden. Echter, gebrek aan automatismen bij jonge, onervaren automobilisten kan het hoge ongevalsrisico niet volledig

verklaren. Omdat autorijden 'self-paced' is kan een automobilist de taakzwaarte zelf aanpassen (bijvoorbeeld door snelheid te minderen of volgafstand te vergroten) en daarmee de mentale belasting verminderen. In theorie kan een onervaren automobilist deze strategie gebruiken om zijn (gebrekkige) rijvaardigheid te compenseren.

Dit 'self-pacing' aspect van de rijtaak is onderdeel van het kalibratiemodel, geïnspireerd door Brown (1989) en Fuller (2005), dat stelt dat:

> *Om veilig te kunnen autorijden, moet een automobilist zijn eigen rijvaardigheid inschatten, dit afwegen tegen de waargenomen complexiteit van de verkeerssituatie, en tot slot, het resultaat van deze afweging gebruiken om de taakzwaarte aan te passen.*

Om te onderzoeken of kalibratie de afname in ongevalsrisico bij jonge onervaren automobilisten in de eerste jaren van hun rijcarrière kan verklaren worden in Hoofdstuk 1 drie onderzoeksvragen geformuleerd:

1. In hoeverre is gebrekkige kalibratie een factor in het hoge ongevalsrisico van jonge onervaren automobilisten?
   a. Overschatten jonge onervaren automobilisten hun rijvaardigheid meer dan ervaren automobilisten?
   b. Beïnvloedt een onjuiste inschatting van rijvaardigheid de mate waarin taakzwaarte wordt aangepast?
   c. Is er een relatie tussen de elementen van het kalibratiemodel en het aantal zelfgerapporteerde ongevallen?

2. Hoe kan kalibratie worden gemeten?

3. Hoe ontwikkelt kalibratie zich over de tijd?

Hoofdstuk 3 beschrijft de methoden die gebruikt zijn om de ontwikkeling in kalibratie te onderzoeken. Een groep jonge onervaren automobilisten is twee jaar intensief gevolgd vanaf het moment dat zij hun rijbewijs behaalden. Tijdens deze periode vulden de deelnemers vragenlijsten in, hielden ze een rittendagboek bij en namen ze deel aan een rijvaardigheidsrit.

De vragenlijst bestond uit vragen over de inschatting van rijvaardigheid, en de waargenomen risico's in het verkeer. Om de interne processen van kalibratie direct te meten is de Adaptatietest ontwikkeld en opgenomen in de vragenlijst. In het rittendagboek beschreven automobilisten wat ze meemaakten in het verkeer, bijvoorbeeld hoeveel ze reden, of ze

’s nachts hadden gereden, met of zonder passagiers en of ze alcohol hadden genuttigd voor de rit.

Om het zelfgerapporteerde gedrag te kunnen vergelijken met de werkelijke rijvaardigheid heeft een deel van de automobilisten in de studie een rijvaardigheidsrit gereden. Hierin werd op twee momenten in de tijd (in 2006 en 2007) de rijvaardigheid beoordeeld om zo de ontwikkeling als gevolg van toegenomen rijervaring te kunnen meten.

Hoofdstuk 4 beschrijft hoe de *inschatting van rijvaardigheid* kan worden gemeten. Hierbij wordt met name onderzocht of jonge onervaren automobilisten hun rijvaardigheid meer overschatten dan ervaren automobilisten, waarbij twee manieren om inschatting van rijvaardigheid te meten worden vergeleken. In de eerste (meer traditionele) benadering wordt de inschatting van rijvaardigheid vergeleken met het groepsgemiddelde. Resultaten van deze benadering laten zien dat onervaren automobilisten bescheiden zijn over hun rijvaardigheid. Dat wil zeggen, ervaren automobilisten zijn optimistischer over hun rijvaardigheid en de risico’s in het verkeer dan onervaren automobilisten.

In de tweede benadering worden de inschattingen van de bestuurder vergeleken met een meer onafhankelijke maat van die rijvaardigheid (het oordeel van een examinator). De resultaten van deze benadering laten zien dat jonge onervaren automobilisten hun rijvaardigheid meer overschatten dan ervaren automobilisten (Onderzoeksvraag 1a). Vooral onervaren automobilisten met een lage score voor ‘veilig rijden’ op de rijvaardigheidsrit hebben meer zelfvertrouwen dan de prestatie op de rit zou rechtvaardigen. De ervaren automobilisten met de laagste scores op de rijvaardigheidsrit hebben minder vertrouwen in hun rijvaardigheid. Er wordt geconcludeerd dat een vergelijking met een onafhankelijke maat van rijvaardigheid noodzakelijk is om overschatting (of onderschatting) van de rijvaardigheid te meten.

Hoofdstuk 5 beschrijft de ontwikkeling van de Adaptatietest, die de *aanpassing van taakzwaarte* meet als een functie van de *waargenomen complexiteit van de verkeerssituatie*. De Adaptatietest bestaat uit 18 verkeerssituaties, elk weergegeven op twee vrijwel identieke foto’s. Elk paar foto’s verschilt op één klein detail dat de complexiteit van de verkeerssituatie verandert. Het verschil in zelfgerapporteerde snelheid tussen de twee foto’s wordt gebruikt als een indicatie voor de mate waarin de automobilist zijn snelheid aanpast aan de complexiteit van de verkeerssituatie.

De resultaten tonen aan dat onervaren automobilisten slechter presteren op de Adaptatietest (ze rapporteren minder vaak een lagere snelheid in de meest complexe verkeerssituatie) dan ervaren automobilisten. Hiernaast blijken onveilige en overmoedige automobilisten, die als zodanig zijn geïdentificeerd in de rijvaardigheidsrit, slechter te presteren op de Adaptatietest. Echter, de verwachte correlatie tussen prestatie op de Adaptatietest en aantal zelfgerapporteerde ongevallen werd niet gevonden.

Omdat de Adaptatietest consequent in staat blijkt te zijn groepen automobilisten te onderscheiden (ervaren versus onervaren automobilisten, veilige versus onveilige automobilisten en goed gekalibreerde versus overmoedige automobilisten) wordt geconcludeerd dat de Adaptatietest een effectieve methode is om de aanpassing van snelheid op de complexiteit van de verkeerssituatie te meten (Onderzoeksvraag 2).

Hoofdstuk 6 beschrijft hoe de *inschatting van rijvaardigheid* de *aanpassing van taakzwaarte* beïnvloedt. De resultaten tonen aan dat overmoedige automobilisten significant meer overtredingen rapporteren dan goed gekalibreerde en onzekere automobilisten. Overmoedige automobilisten vertonen ook significant minder aanpassing van taakzwaarte aan de complexiteit van de verkeerssituatie zoals gemeten door de Adaptatietest. Tot slot rapporteert 50% van de overmoedige automobilisten één of meer ongevallen in het tweede jaar van de studie. Dit is significant meer dan de goed gekalibreerde en onzekere automobilisten (resp. 18% en 10%).

De resultaten van Hoofdstuk 6 tonen aan dat overschatting van de rijvaardigheid samenhangt met onvoldoende aanpassing van taakzwaarte (Onderzoeksvraag 1b). De correlatie tussen overschatting van rijvaardigheid en aantal zelfgerapporteerde ongevallen (Onderzoeksvraag 1c) is minder eenduidig. Hoewel een sterke relatie is gevonden tussen overschatting van de rijvaardigheid en zelfgerapporteerde ongevallen, is het lastig deze resultaten te interpreteren vanwege de overlappende correlatie met prestatie op de rijvaardigheidsrit. Bovendien is ook de verwachte correlatie tussen zelfgerapporteerde ongevallen en prestatie op de Adaptatietest niet gevonden.

In Hoofdstuk 7 wordt de ontwikkeling van (de verschillende elementen van) kalibratie over de tijd gevolgd. *Inschatting van rijvaardigheid* (Hoofdstuk 4) in het ene jaar wordt met die in het volgende jaar vergeleken en *aanpassing van taakzwaarte* (Hoofdstuk 6) wordt op zes momenten gedurende de twee jaar beschreven. De scores op de Adaptatietest (Hoofdstuk 5) van ervaren en onervaren automobilisten worden voor de twee jaar van de studie vergeleken. Hiernaast rapporteert Hoofdstuk 7 ook de resultaten van het rittendagboek dat de deelnemers op zes momenten gedurende de twee jaar van de studie hebben bijgehouden.

De resultaten tonen aan dat, hoewel een verschil in kalibratie wordt gevonden tussen onervaren en ervaren automobilisten, de verwachte relatieve *ontwikkeling* in kalibratie van de onervaren automobilisten niet is gevonden in de twee jaar na het behalen van het rijbewijs (Onderzoeksvraag 3).

Volgens het rittendagboek maken onervaren automobilisten andere ritten dan ervaren automobilisten. Onervaren automobilisten rijden relatief meer 's nachts, in het weekend, in hun vrije tijd en met passagiers. Hiernaast rijden *mannelijke* onervaren automobilisten vaker te snel en zonder autogordels dan *vrouwelijke* onervaren automobilisten. Er zijn aanwijzingen dat deze ritkenmerken niet veel veranderen in de twee jaar na het behalen van het rijbewijs.

Hoofdstuk 8 vat alle resultaten van de eerdere hoofdstukken samen en beantwoordt de drie onderzoeksvragen. Hoofdstuk 8 bespreekt ook de sterke en zwakke punten van deze studie. Zo wordt bijvoorbeeld over de longitudinale opzet van de studie geconcludeerd dat er geen aanwijzingen zijn gevonden voor een 'Hawthorne effect' (door het meten van bepaald gedrag wordt het gedrag zelf beïnvloed). Ook zijn er geen aanwijzingen gevonden dat de onvermijdelijke uitval van deelnemers in de studie de resultaten aanzienlijk heeft beïnvloed.

Wat de Adaptatietest betreft is er wel een aantal onopgeloste zwakke punten. Ten eerste gaat de Adaptatietest uit van zelfgerapporteerd snelheidsgedrag, dat zou kunnen afwijken van werkelijk snelheidsgedrag. Ten tweede heeft geen van de groepen heel goed gepresteerd op de test. Van alle subgroepen scoren de ervaren automobilisten het hoogst. Toch rapporteren ook zij in meer dan de helft van de gevallen (55%) geen lagere snelheid in de meer complexe verkeerssituatie. Dit betekent dat de Adaptatietest niet erg gevoelig is in het onderscheiden van individuele automobilisten.

Met betrekking tot de rijvaardigheidsrit tonen de resultaten dat dit een valide methode is om 'veilig rijden' te meten. Ten eerste werd een significante correlatie gevonden tussen het gerapporteerde snelheidsgedrag in de vragenlijst en het oordeel van de examinator over het snelheidsgedrag. Ten tweede werd een sterke relatie gevonden tussen het oordeel van de examinator en het aantal gerapporteerde ongevallen van de deelnemers. Van de automobilisten die het laagst scoorden op de rijvaardigheidsrit in het eerste jaar (score ≤ 5,5) rapporteerde 50% een of meer ongevallen in het tweede jaar van de studie. Van de automobilisten met hogere scores in het eerste jaar (score > 5,5) rapporteerde slechts 15% een of meer ongevallen in het tweede jaar van de studie.

Het samenstellen van de kalibratiegroepen, ten slotte, was gebaseerd op een vergelijking van 'gerapporteerd vertrouwen' in de eigen rijvaardigheid en het oordeel van de examinator op 'veilig rijden'. Het zou kunnen zijn dat de verschillen tussen de kalibratiegroepen volledig veroorzaakt worden door slechts een van deze twee factoren, bijvoorbeeld 'veilig rijden' en dat de andere factor ('gerapporteerd vertrouwen') niet bijdraagt aan de gevonden verschillen. Hoewel geen bewijs voor deze alternatieve verklaring is gevonden, kan zij ook niet volledig worden uitgesloten omdat het aantal deelnemers te klein was voor voldoende statistische 'power' om de interactie tussen gerapporteerd vertrouwen en het oordeel van de examinator te onderzoeken.

In Hoofdstuk 8 worden ook de gevolgen van deze studie voor de rijopleiding en examinering besproken, en worden suggesties gedaan voor vervolgonderzoek. Omdat de resultaten van dit onderzoek aantonen dat kalibratievaardigheden niet hoofdzakelijk worden ontwikkeld in de eerste twee jaar na het behalen van het rijbewijs, zijn dit waarschijnlijk geen vaardigheden die kunnen worden aangeleerd in de rijopleiding of bevorderd door examinering. Vervolgonderzoek zou kunnen uitwijzen hoe en wanneer kalibratievaardigheden zich dan wel ontwikkelen.

Desalniettemin, zoals deze studie ook aantoont, is de ontwikkeling van rijervaring een belangrijke factor in de afname van ongevalsrisico. Totdat we beter begrijpen hoe rijervaring werkt, en welke aspecten van rijervaring belangrijk zijn voor het ongevalsrisico, is het enige wat we kunnen doen ervoor zorgen dat onervaren automobilisten zo veel mogelijk rijervaring opdoen in een zo veilig mogelijke omgeving. Dit kan worden bereikt door het aantal privileges voor jonge onervaren automobilisten te beperken (bijvoorbeeld door een limiet op het aantal passagiers of een nachtelijk rijverbod) of door het invoeren van een periode van begeleid rijden na het behalen van het rijbewijs.

Deze studie laat zien dat het oordeel van de examinator, op basis van een rijvaardigheidsrit van een half uur, een behoorlijk goede voorspeller is voor zelfgerapporteerd ongevalsrisico. Omdat het rijexamen zo'n belangrijke maatregel is voor de verkeersveiligheid, maar ook een waar wel eens kritiek op is, lijkt het een logische stap om deze voorspellende validiteit verder te onderzoeken. In een kleinschalig experiment met studenten (zie Paragraaf

3.5.3), werden de effecten onderzocht die het uiterlijk van een bestuurder had op de inschatting van zijn rijvaardigheid door een beoordelaar. Het zou zeer interessant zijn dit experiment te herhalen met professionele examinatoren of op z'n minst met professionele rijinstructeurs (als vervangers van de examinatoren).

Tot slot geeft Hoofdstuk 8 nog een aantal verklaringen voor het feit dat er geen ontwikkeling in kalibratie is gevonden in deze studie, waarvan er twee het meest waarschijnlijk lijken. Ten eerste zou kalibratie meer een soort *karaktertrek* dan een *vaardigheid* kunnen zijn, die zich over een langere periode ontwikkelt. Dit zou verklaren waarom er wel een verschil wordt gevonden tussen ervaren en onervaren automobilisten; de ervaren automobilisten hebben minstens tien jaar meer rijervaring. Het zou ook het gebrek aan ontwikkeling in de studie verklaren; twee jaar is gewoonweg te kort om verbetering in kalibratie te kunnen meten.

Het is ook mogelijk dat het verschil in kalibratie tussen onervaren en ervaren automobilisten niets anders is dan een *generatieverschil*. De ervaren automobilisten zijn ongeveer tien tot twintig jaar ouder dan de onervaren automobilisten in deze studie; ze komen uit een andere generatie met een compleet andere rijopleiding en rijbewijssysteem.

Hoewel kalibratie wordt genoemd als een verklarende factor voor het hoge ongevalsrisico van jonge onervaren automobilisten, laten de resultaten van deze studie zien dat de afname in het ongevalsrisico in de eerste twee jaar na het behalen van het rijbewijs *niet eenduidig* verklaard kan worden door kalibratie. Met betrekking tot rijervaring zullen we voorlopig nog op zoek blijven naar 'de X-factor'.

About the author

Saskia de Craen was born in Voorburg on May 16th in 1977. After obtaining her Atheneum diploma from the Alfrink College in Zoetermeer she began her study of Psychology at Leiden University. In 2002 she received her master's degree within the Division of Methodology and Psychometrics after a study of the aspects of K-means cluster analysis.

Since 2000, Saskia has been working with SWOV Institute for Road Safety Research, first as an intern, and as a full-time researcher since 2002. She has worked on several traffic safety subjects, such as speed enforcement and practical driver training for moped drivers. In 2004 she worked on the NovEv project; a European project in which the effects of a Second Phase driver training were evaluated. This project introduced her to the subject of young novice drivers and was the start of this PhD study.

Appendix A - Driver Confidence Questionnaire[20]

What is your opinion about traffic risks?

- Traffic is very dangerous
- Traffic is quite dangerous
- Neutral
- Traffic is quite safe
- Traffic is very safe

How confident are you concerning your own driving skills?

- Very confident
- Confident
- Neutral
- Somewhat insecure
- Very insecure

When you compare yourself with the average driver…

… do you think you are

- A much better driver
- A better driver
- Not a better nor a worse driver
- A worse driver
- A much worse driver

… what is your risk of being involved in a crash

- Much smaller
- Smaller
- The same
- Higher
- Much higher

[20] This is a translation of the original Dutch questionnaire which can be found at the end of Appendix A

… How skilled are you in coping with the dangers of traffic?

- Much more skilled
- More skilled
- The same
- Less skilled
- Much less skilled

When you compare yourself with peer drivers…

… do you think you are

- A much better driver
- A better driver
- Not a better nor a worse driver
- A worse driver
- A much worse driver

… what is your risk of being involved in a crash

- Much smaller
- Smaller
- The same
- Higher
- Much higher

… How skilled are you in coping with the dangers of traffic?

- Much more skilled
- More skilled
- The same
- Less skilled
- Much less skilled

Wat is jouw mening over het gevaar in het verkeer?
- Het verkeer is héél gevaarlijk
- Het verkeer is best gevaarlijk
- Neutraal
- Het verkeer is best veilig
- Het verkeer is héél veilig

Hoeveel zelfvertrouwen heb je als het gaat om je rijgedrag?
- Heel veel zelfvertrouwen
- Veel zelfvertrouwen
- Neutraal
- Een beetje onzeker
- Heel erg onzeker

Als je jezelf vergelijkt met de gemiddelde automobilist…
…vind je dat je
- Veel beter rijdt
- Beter rijdt
- Niet beter / niet slechter
- Slechter rijdt
- Veel slechter rijdt

…wat is het risico dat je betrokken raakt bij een auto-ongeluk?
- Veel minder groot
- Minder groot
- Even groot
- Groter
- Veel groter

…hoe vaardig ben jij in het omgaan met de gevaren in het verkeer?
- Veel beter
- Beter
- Hetzelfde
- Slechter
- Veel slechter

Als je jezelf vergelijkt met je leeftijdsgenoten …

…vind je dat je

- Veel beter rijdt
- Beter rijdt
- Niet beter / niet slechter
- Slechter rijdt
- Veel slechter rijdt

…wat is het risico dat je betrokken raakt bij een auto-ongeluk?

- Veel minder groot
- Minder groot
- Even groot
- Groter
- Veel groter

…hoe vaardig ben jij in het omgaan met de gevaren in het verkeer?

- Veel beter
- Beter
- Hetzelfde
- Slechter
- Veel slechter

Appendix B - The eighteen situations of the Adaptation Test


Situation B1. Bus stop


Situation B2. Bicycle crossing street


Situation B3. Signalling light

Situation B4. Crossing pedestrian

Situation B5. Bicycle riders

Situation B6. Braking on highway

Situation B7. School children

Situation B8. Weather

Situation B9. Round-a-bout

Situation B10. Pedestrian crossing

Situation B11. Football on road

Situation B12. Bended road (not sign.)

Situation B13. Road worker

Situation B14. Bicyclists standing or riding (sign. In wrong direction)

Situation B15. Rain coat

Situation B16. Horse

Situation B17. Clear or less clear right-of-way rule

Situation B18. Unclear view

Appendix C - Driver Behaviour Questionnaire[21]

Below you will find a list of mistakes, errors, and offences that are committed in traffic. It contains **twenty-four situations** of this type of behaviour. **For each situation we would like to know how frequently you experience it.**

When filling in the survey you have to give your personal opinion. As it is impossible to give precise answers, we are only interested in your personal judgement. This means that you shouldn't think about an answer too long. Base your judgements on your recollections **over the past four months**. Ór, if you have recently passed the driving exam, on what you can remember about your driving lessons.

Filling in the survey takes approximately 15 minutes.

The range is from 0 to 5:
0 = Never 1 = Occasionally 2 = Regularly 3 = Often 4 = Very often 5 = Nearly all the time

[21] These are the items of the original abbreviated DBQ (Parker, Reason et al., 1995). The Dutch version of the DBQ that was administered in this study (Verschuur, 2003a; Verschuur & Hurts, 2008) can be found at the end of Appendix C.

		NEVER					NEARLY ALL THE TIME
1.	Attempt to drive away from traffic lights in third gear	0	1	2	3	4	5
2.	Become impatient with a slow driver in the outer lane and overtake on the inside	0	1	2	3	4	5
3.	Drive especially close to the car in front as a signal to its driver to go faster or get out of the way	0	1	2	3	4	5
4.	Attempt to overtake someone that you hadn't noticed to be signalling a right turn	0	1	2	3	4	5
5.	Forget where you left your car in car park	0	1	2	3	4	5
6.	Switch on one thing, such as the headlights, when you meant to switch on something else, such as the wipers	0	1	2	3	4	5
7.	Realise that you have no clear recollection of the road along which you have just been travelling	0	1	2	3	4	5
8.	Cross a junction knowing that the traffic lights have already turned against you	0	1	2	3	4	5
9.	Fail to notice that pedestrians are crossing when turning into a side street from a main road	0	1	2	3	4	5
10.	Angered by another driver's behaviour, you give chase with the intention of giving him/her a piece of your mind	0	1	2	3	4	5
11.	Misread the signs and exit from a roundabout on the wrong road	0	1	2	3	4	5
12.	Disregard the speed limits late at night or early in the morning	0	1	2	3	4	5

		NEVER					NEARLY ALL THE TIME
13.	On turning left, nearly hit a cyclist who has come up on your inside	0	1	2	3	4	5
14.	Queueing to turn left onto main road, you pay such close attention to the main stream of traffic that you nearly hit the car in front	0	1	2	3	4	5
15.	Drive even though you realise that you may be over the legal blood-alcohol limit	0	1	2	3	4	5
16.	Have an aversion to a particular class of road user, and indicate your hostility by whatever means you can	0	1	2	3	4	5
17.	Underestimate the speed of an oncoming vehicle when overtaking	0	1	2	3	4	5
18.	Hit something reversing that you had not previously seen	0	1	2	3	4	5
19.	Intending to drive to destination A, you 'wake up' to find yourself on the road to destination B, perhaps because the latter is your more usual destination	0	1	2	3	4	5
20.	Get into the wrong lane approaching a roundabout or a junction	0	1	2	3	4	5
21.	Miss 'Give Way' signs, and narrowly avoid colliding with traffic having the right of way	0	1	2	3	4	5
22.	Fail to check your rear-view mirror before pulling out, changing lanes, etc.	0	1	2	3	4	5
23.	Get involved in unofficial 'races' with other drivers	0	1	2	3	4	5
24.	Brake too quickly on a slippery road, or steer the wrong way into a skid	0	1	2	3	4	5

Hieronder vind je een lijst van fouten, vergissingen en overtredingen in het verkeer. **Vierentwintig situaties** van dit soort gedrag staan vermeld**. Bij elke situatie wordt gevraagd, hoe vaak je dit overkomt.**

Bij het invullen van de vragenlijst gaat het om je persoonlijke oordeel. Het is niet mogelijk om een precies antwoord te geven, wij zijn alleen geïnteresseerd in je algemene indruk. Denk dus niet te lang na over de vraag. Baseer je oordelen op wat je kunt herinneren van **de afgelopen vier maanden**. Óf, als je net je rijbewijs hebt gehaald, wat je kunt herinneren van je rijlessen.

Het invullen van de vragenlijst duurt ongeveer 15 minuten.

De hokjes lopen van 0 naar 5:
0 = Nooit 1 = Een enkele keer 2 = Regelmatig 3 = Vaak 4 = Heel vaak 5 = Bijna altijd

		NOOIT					BIJNA ALTIJD
1.	Bij een verkeerslicht proberen om in de verkeerde versnelling weg te rijden, omdat je vergeten bent om naar de eerste versnelling te schakelen	0	1	2	3	4	5
2.	Er genoeg van krijgen om achter een langzaam rijdende automobilist op de linker baan te zitten en deze rechts inhalen	0	1	2	3	4	5
3.	Met opzet een voorligger op korte afstand volgen om aan de bestuurder duidelijk te maken dat deze óf harder moet rijden, óf naar rechts moet gaan	0	1	2	3	4	5
4.	Iemand proberen in te halen waarvan je niet in de gaten had, dat hij aangaf links af te slaan	0	1	2	3	4	5
5.	Vergeten waar je de auto hebt geparkeerd in een parkeergarage	0	1	2	3	4	5
6.	Iets aanzetten (bijv. de richtingaanwijzer) terwijl je iets anders aan wilde zetten (bijv. de ruitenwissers)	0	1	2	3	4	5
7.	Tot de ontdekking komen, dat je geen duidelijke herinnering hebt van het stuk weg dat je net gereden hebt	0	1	2	3	4	5
8.	Een kruispunt oversteken, terwijl je weet dat het verkeerslicht net op rood is gesprongen	0	1	2	3	4	5
9.	Niet in de gaten hebben dat er voetgangers aan het oversteken zijn, wanneer je rechts af wilt slaan van een hoofdweg naar een zijweg	0	1	2	3	4	5
10.	Zo kwaad worden over iemands rijgedrag, dat je deze persoon achterna gaat om hem / haar eens goed te laten merken wat je ervan vindt	0	1	2	3	4	5
11.	Een verkeersbord verkeerd lezen en daardoor op een rotonde een verkeerde afslag nemen	0	1	2	3	4	5
12.	De snelheidslimiet negeren als het 's ochtends vroeg of 's avonds laat is	0	1	2	3	4	5

		NOOIT					BIJNA ALTIJD
13.	Terwijl je rechts afslaat, bijna een fietser aanrijden die rechts van je reed	0	1	2	3	4	5
14.	Terwijl je in de rij staat om rechts af te slaan naar de hoofdweg, zo sterk op het aankomende verkeer letten, dat je bijna op de voorligger botst	0	1	2	3	4	5
15.	Naar huis rijden, terwijl je weet dat je meer hebt gedronken dan wettelijk is toegestaan	0	1	2	3	4	5
16.	Een uitgesproken hekel hebben aan een bepaald type weggebruiker en dit op elke mogelijke manier laten blijken	0	1	2	3	4	5
17.	De snelheid van een tegemoetkomende auto onderschatten, terwijl je aan het inhalen bent	0	1	2	3	4	5
18.	Bij het achteruit rijden tegen iets aan rijden dat je niet had gezien, omdat je vergat je spiegels te gebruiken	0	1	2	3	4	5
19.	Je bent op weg en betrapt jezelf erop, dat je de verkeerde route volgt, waarschijnlijk omdat je gewend bent deze route te rijden	0	1	2	3	4	5
20.	Verkeerd voorsorteren als je een kruispunt of een rotonde nadert	0	1	2	3	4	5
21.	Een verkeersbord, dat aangeeft dat je voorrang moet verlenen, over het hoofd zien en daardoor bijna in botsing komen met verkeer dat voorrang heeft	0	1	2	3	4	5
22.	Vergeten om in de spiegels te kijken voordat je wegrijdt, inhaalt en dergelijke	0	1	2	3	4	5
23.	Betrokken raken bij een wedstrijdje om wie het hardst kan rijden met andere automobilisten	0	1	2	3	4	5
24.	Te hard remmen op een gladde weg of verkeerd sturen, zodat je slipt	0	1	2	3	4	5

Appendix D - The Driving Diary[22]

Welcome to your weekly diary. Here we ask you to answer some brief questions about your driving behaviour in the past week, and we ask you to describe the most notable event of the past week.

We ask you to describe the period of precisely one week. For example, if it is now Tuesday afternoon, we would like you to describe the events from Tuesday afternoon of last week up to the present moment.

All questions are about the journeys in which you yourself were driving the car. This means that, for instance, examples of what happened when you were a passenger, or something that happened to you while you were cycling are not to be included.

Success!

On how many days did you drive in the past week? … days
A total of how many hours did you approximately drive in the past week? … hours

In the past week, did you ever consider driving, but decided not to because it did not seem sensible (for example because it was already dark, or because it was raining)?

- Yes
- No

The respondent, who says not to have driven, has completed filling in the weekly diary after these questions.

[22] This is a translation of the original Dutch questionnaire which can be found at the end of Appendix D

How many hours did you drive in the weekend and how many on weekdays?
(Tick a box. If you tick the box on the far left, you have only driven in the weekend. If you tick the middle box, you say you have driven approximately the same number of hours in the weekend as on weekdays.

All hours on the weekend	1	2	3	4	5	6	7	All hours on weekdays

What was the purpose of the journeys?

All from home to school/ work (or during work)	1	2	3	4	5	6	7	Leisure only

During how many journeys did you carry passengers?

I made all journeys alone	1	2	3	4	5	6	7	I had passengers in every journey

If you made journeys with passengers, were they friends/peers or others?

Only friends / peers	1	2	3	4	5	6	7	Others

In the past week, have you ever …

	YES	NO
Driven on a weekend night?	0	0
Driven after having drunk one or more glasses of alcohol?	0	0
Driven at least 10 km/h faster than the max. speed:		
On urban roads?	0	0
On rural roads?	0	0
Driven without wearing the safety belt:		
On urban roads?	0	0
On rural roads?	0	0
Made a short detour to avoid a busy or complex traffic situation?	0	0

So far the brief questions about the previous week. We would now like to ask you to describe the most notable situation of the past week.

> *This means a traffic situation that gave you a (brief) fright or that you didn't expect. This can only be a situation in which you yourself drove the car. A situation can be a crash or a near-crash, but that needn't be the case. It can also be something in traffic that you didn't expect or a situation you think you handled well.*

Describe the situation
What happened? How did it happen?

Why did you choose this situation?
- It taught me something
- It gave me a fright
- I found it very notable
- It came unexpected
- Other, namely: ____________________

BLOCK A: DESCRIPTION OF THE SITUATION (MORE DETAILED)

What other road users were involved?
- Nobody (just me)
- Slow traffic (cyclists, pedestrians)
- Motorized traffic
- Both slow and motorized traffic

Where did it take place?

- In an urban area
- In a rural area, motorway
- In a rural area, other
- Unknown

More detailed description of the location:

- Intersection without traffic lights
- Intersection wíth traffic lights
- Intersection (not certain whether there were traffic lights)
- Roundabout
- Exit (a parking, for example)
- Straight stretch of road
- Bend
- Parking spot
- Approach/exit motorway/secondary road
- Unknown
- Other, namely: ___________________________________

At what speed were you driving? … km/h

Where there circumstances that made the situation more complicated? (you can tick more than one answer)

- Weather conditions (rain, fog, etc.)
- Cars parked at the side of the road
- Dense vegetation along the side of the road
- Much traffic on the road
- Condition of the road surface (for example slippery due to rain or sleet, or pot-holes due to neglected maintenance)
- Too many lanes
- You had to reverse
- Dark
- Other, namely: __________________________
- No, it was a simple situation

Did you carry passengers?

- No, I was in the car alone
- Yes, friends or peers
- Yes, family
- Yes, someone else

BLOCK B: RESPONSIBILITY FOR THE SITUATION

What caused the situation?

- You did
- Another road user
- External factors (Factors you cannot influence, for example slipperiness, a ball or pet suddenly shooting onto the road)
- You and another road user
- You and external factors
- External factors and another road user
- Other, namely: ____________________

Do you think you acted well before the situation occurred?

- Yes
- No
- I don't know

Is there anything you could have done better?

BLOCK C: HOW DID YOU HANDLE THE SITUATION?

Did the situation lead to a crash?

- Yes
- No
- Unknown

What did you do?
(you can tick more than one answer)

- Nothing
- Adapt speed
- Swerve or adjust steering
- Gesture, signal
- Something different, namely: ________________________

Did it work?

- Yes
- No
- I don't know

What did the other person do?

- Nothing
- Adapt speed
- Swerve or adjust steering
- Gesture, signal
- Something different, namely: ______________________________

Did it work?

- Yes
- No
- I don't know

Do you think you handled the situation well?

- Yes
- No
- I don't know

Is there anything you could have done better?

Welkom bij het weekboek. In dit gedeelte stellen we je een paar korte vragen over je rijgedrag van de afgelopen week en vragen we je de meest opvallende gebeurtenis van de afgelopen week te beschrijven.

We vragen je de periode van precies één week te beschrijven. Bijvoorbeeld als het nu dinsdagmiddag is, willen we graag dat je de gebeurtenissen vanaf vorige week dinsdagmiddag tot en met nu beschrijft.

Alle vragen hebben betrekking op ritten waar jijzelf de auto bestuurde. Dus bijvoorbeeld ervaringen die je als passagier beleefde, of iets dat je op de fiets is overkomen tellen niet mee.

Succes!

Op hoeveel dagen van de afgelopen week heb je gereden? … Dagen

Hoeveel uren heb je de afgelopen week, ongeveer, bij elkaar gereden? … Uur

Heb je in de afgelopen week wel eens overwogen om te gaan rijden, maar dit toch niet gedaan omdat dit je niet verstandig leek (bijvoorbeeld omdat het al donker was, of regende)?

- Ja
- Nee

De deelnemer die zegt niet te hebben gereden, is na deze vragen klaar met het weekboek.

Hoeveel uren heb je in het weekend of doordeweeks gereden?
(Kruis een hokje aan. Als je het meest linkse hokje aankruist heb je alleen in het weekend gereden. Als je het middelste hokje kiest, geef je daarmee aan dat je ongeveer evenveel uren in het weekend als doordeweeks hebt gereden.

Alle uren in het weekend	1	2	3	4	5	6	7	Alle uren doordeweeks

Wat was het doel van de ritten?

Allemaal van huis naar school/ werk (of tijdens het werk)	1	2	3	4	5	6	7	Alleen vrije tijd

Tijdens hoeveel ritten zaten er passagiers in je auto?

Ik reed alle ritten alleen	1	2	3	4	5	6	7	Tijdens alle ritten had ik passagiers

De ritten dat er passagiers in de auto zaten, waren dit dan vrienden/leeftijdsgenoten of anderen?

Alleen vrienden / leeftijdsgenoten	1	2	3	4	5	6	7	Anderen

Heb je in de afgelopen week wel eens…

	JA	NEE
In een weekend-nacht gereden?	0	0
Gereden na het drinken van één of meer glazen alcohol?	0	0
Minstens 10 km. harder gereden dan de max. snelheid:		
Binnen de bebouwde kom?	0	0
Buiten de bebouwde kom?	0	0
Zonder gordel gereden:		
Binnen de bebouwde kom?	0	0
Buiten de bebouwde kom?	0	0
Een stukje omgereden om een drukke of complexe verkeerssituatie te vermijden?	0	0

Dit waren de korte vragen over de afgelopen week. We willen je nu vragen de meest opvallende situatie van de afgelopen week te beschrijven.

> *Hiermee bedoelen we een situatie in het verkeer waar je even van schrok of wat je niet verwachtte. Het gaat alleen over situaties waarin jijzelf de auto bestuurde. Een situatie kan bestaan uit een ongeval of een bijna-ongeval, maar het hoeft niet. Het kan ook iets in het verkeer zijn dat je niet verwachtte of een situatie die je volgens jou goed hebt opgelost.*

Beschrijf de situatie
Wat gebeurde er? Waardoor kwam het?

Waarom heb je deze situatie gekozen?

- Ik heb hier wat van geleerd
- Ik ben geschrokken
- Ik vond het erg opvallend
- Ik vond het onverwacht
- Anders, namelijk: ____________________

BLOK A: BESCHRIJVING VAN DE SITUATIE (MEER SPECIFIEK)

Welke andere verkeersdeelnemers waren erbij betrokken?

- Niemand (alleen ikzelf)
- Langzame verkeersdeelnemers (fietsers, voetgangers)
- Gemotoriseerd verkeer
- Langzaam en gemotoriseerd verkeer

Waar gebeurde het?
- Binnen de bebouwde kom
- Buiten de bebouwde kom, snelweg
- Buiten de bebouwde kom, overig
- Onbekend

Meer specifieke beschrijving van de locatie:
- Kruispunt zonder stoplichten
- Kruispunt mét stoplichten
- Kruispunt (onbekend of er stoplichten waren)
- Rotonde
- Uitrit (bijvoorbeeld van parkeerplaats)
- Recht stuk weg
- Bocht
- Parkeerplaats
- Op/Afrit snelweg/provinciale weg
- Onbekend
- Anders, namelijk: ________________________________

Hoe hard reed je ongeveer? … km/uur

Waren er omstandigheden die de situatie moeilijker maakten?
(je mag meerdere antwoorden aankruisen)
- Weersomstandigheden (regen, mist, enz.)
- Geparkeerde auto's langs de weg
- Dichte begroeiing langs de weg
- Drukte op de weg
- Staat van het wegdek (bijvoorbeeld glad door regen of ijzel, of gaten door slecht onderhoud)
- Veel voorsorteer stroken
- Je moest achteruit rijden
- Donker
- Anders, namelijk: ______________________
- Nee, het was een eenvoudige situatie

Had je passagiers in de auto?
- Nee, ik was alleen in de auto
- Ja, vrienden of leeftijdsgenoten
- Ja, familie
- Ja, iemand anders

BLOK B: VERANTWOORDELIJKHEID VOOR DE SITUATIE

Waardoor kwam het?

- Door jou
- Door een andere verkeersdeelnemer
- Door externe factoren (Factoren waar je geen invloed op hebt, bijvoorbeeld gladheid, een bal of huisdier dat plotseling de weg op schiet)
- Door jou en een andere verkeersdeelnemer
- Door jou en externe factoren
- Door externe factoren en een andere verkeersdeelnemer
- Anders, namelijk: ____________________

Vind je dat je goed hebt gehandeld voordat de situatie zich voordeed?

- Ja
- Nee
- Weet niet

Is er iets dat je beter had kunnen doen?

BLOK C: HOE BEN JE MET DE SITUATIE OMGEGAAN?

Leidde de situatie tot een ongeval?

- Ja
- Nee
- Onbekend

Wat deed jij?
(je mag meerdere antwoorden geven)

- Niets
- Snelheid aanpassen
- Uitwijken of bijsturen
- Gebaren maken, signalen afgeven
- Iets anders, namelijk: ________________________

Hielp het?

- Ja
- Nee
- Weet niet

Wat deed de ander?

- Niets
- Snelheid aanpassen
- Uitwijken of bijsturen
- Gebaren maken, signalen afgeven
- Iets anders, namelijk: ______________________________

Hielp het?

- Ja
- Nee
- Weet niet

Vind je dat je goed hebt gereageerd op de situatie?

- Ja
- Nee
- Weet niet

Is er iets dat je beter had kunnen doen?

Appendix E - Estimated data in the driving diary

	Number of days		Number of hours		Trips during weekend		Trips during leisure time		Trips with passengers		Trips with peers		Did not drive, adverse conditions	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Complete series	514	95%	514	95%	424	78%	423	78%	422	78%	346	64%	514	95%
Corrected														
One value estimated	16	3%	16	3%	59	11%	59	11%	59	11%	97	18%	15	3%
Two values estimated	8	2%	7	1%	21	4%	22	4%	22	4%	39	7%	8	1%
Complete series after correction	538	99%	537	99%	504	93%	504	93%	503	93%	482	88%	538	99%
To many missing	3	1%	4	1%	37	7%	37	7%	38	7%	59	11%	4	1%
Total	541	100%	541	100%	541	100%	541	100%	541	100%	541	100%	541	100%

	Avoided difficult intersection		Driving at night		Driving after consumption alcohol		Speeding in urban area		Speeding in rural area		Wearing a seatbelt in urban area		Wearing a seatbelt in rural area	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Complete series	422	78%	422	78%	423	78%	423	78%	422	78%	422	78%	421	78%
Corrected														
One value estimated	56	10%	59	11%	59	11%	61	11%	58	11%	59	11%	59	11%
Two values estimated	36	7%	28	5%	32	6%	30	6%	34	6%	33	6%	35	7%
Complete series after correction	514	95%	509	94%	514	95%	514	95%	514	95%	515	95%	515	95%
To many missing	27	5%	32	6%	27	5%	27	5%	27	5%	26	5%	26	5%
Total	541	100%	541	100%	541	100%	541	100%	541	100%	541	100%	541	100%

Appendix F - Reported driving behaviour from the driving diary

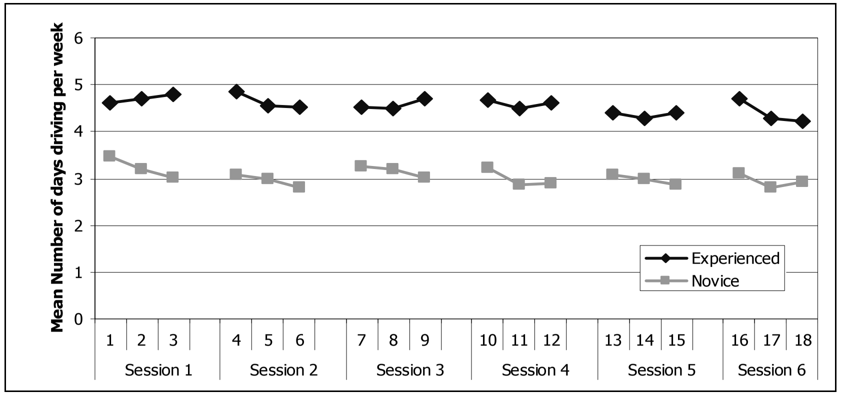

Figure F.1. Mean number of days the participants reported driving in one week

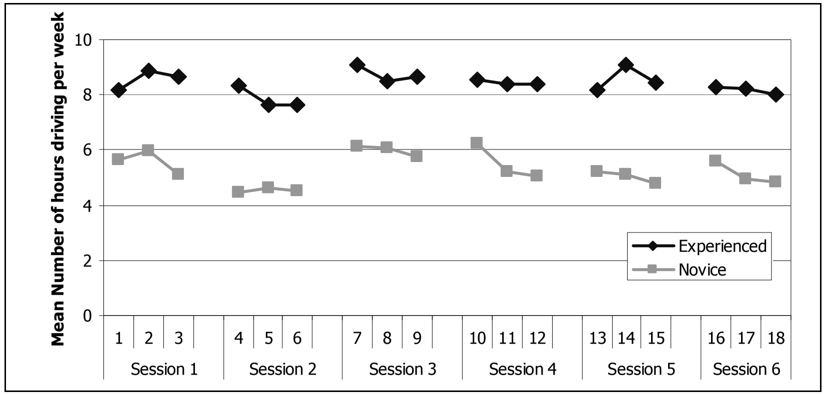

Figure F.2. Mean number of hours the participants reported driving in one week

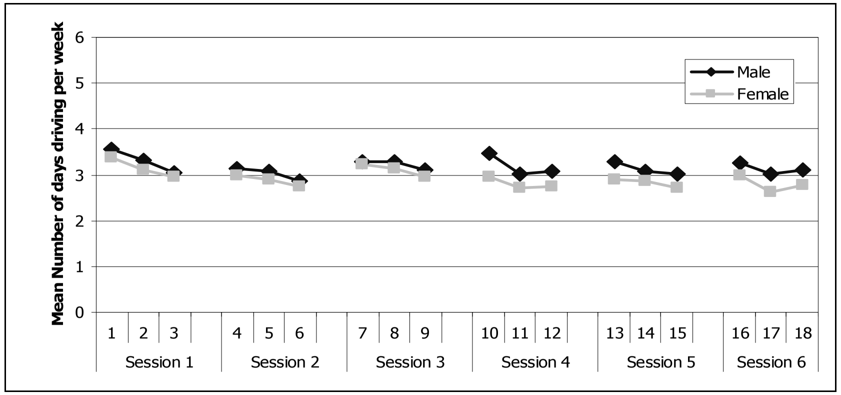

Figure F.3. Mean number of days the *novice* drivers reported driving in one week

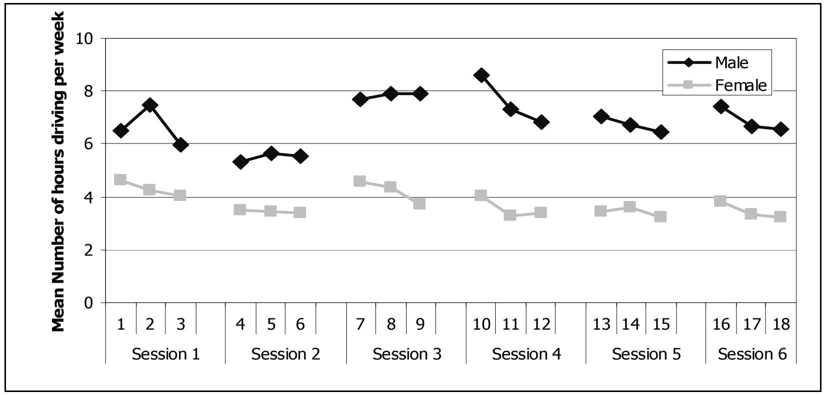

Figure F.4. Mean number of hours the *novice* drivers reported driving in one week

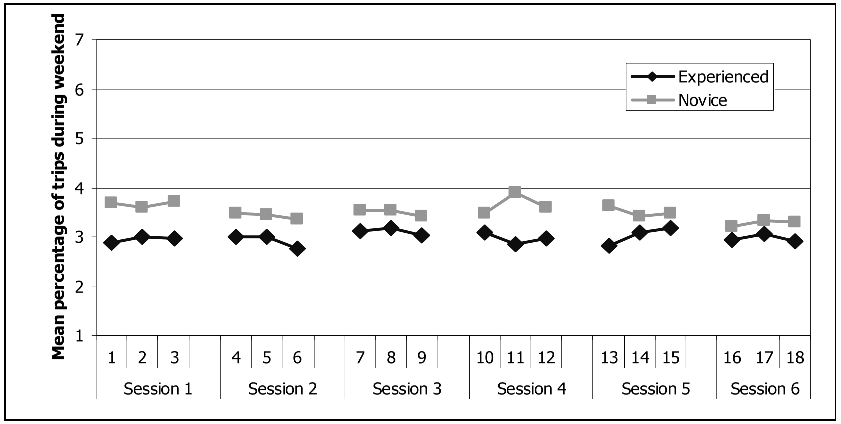

Figure F.5. Mean proportion of trips in the weekend (ranging from 1 = no trips in weekend to 7 = all trips during the weekend)

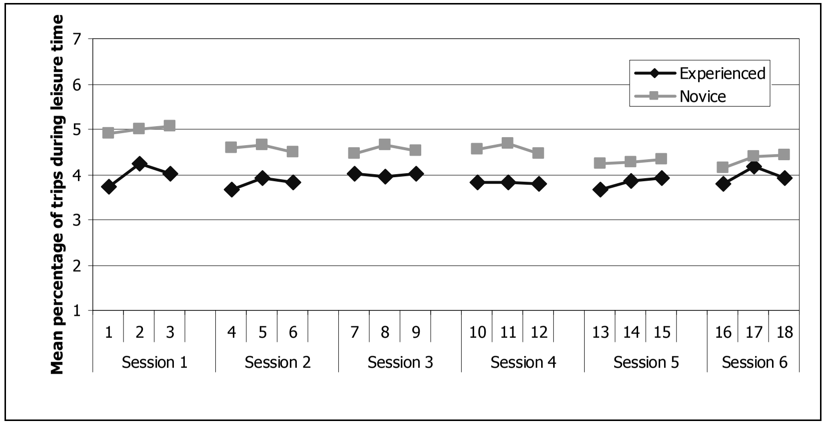

Figure F.6. Mean proportion of trips during leisure time (ranging from 1 = all trips are school/work related, to 7 = all trips are during leisure time)

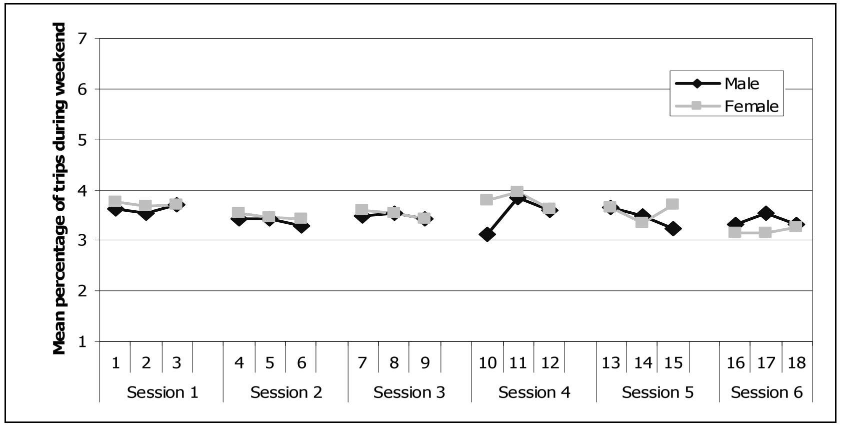

Figure F.7. Mean proportion of trips in the weekend (ranging from 1 = no trips in weekend to 7 = all trips during the weekend) for *novice* drivers

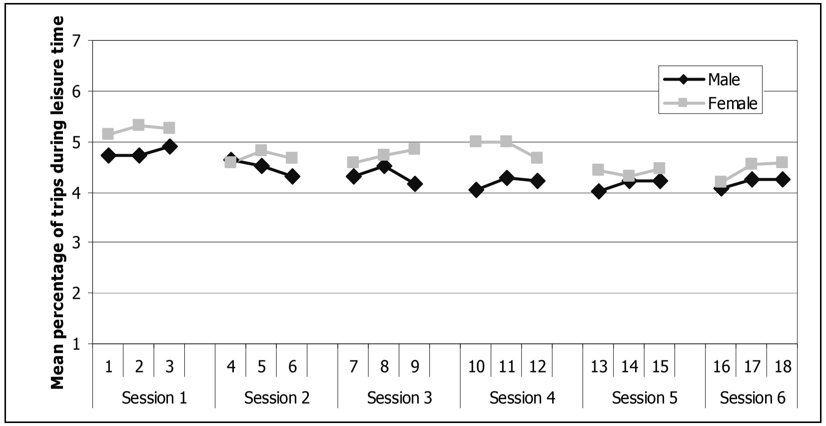

Figure F.8. Mean proportion of trips during leisure time (ranging from 1 = all trips are school/work related, to 7 = all trips are during leisure time) for *novice* drivers

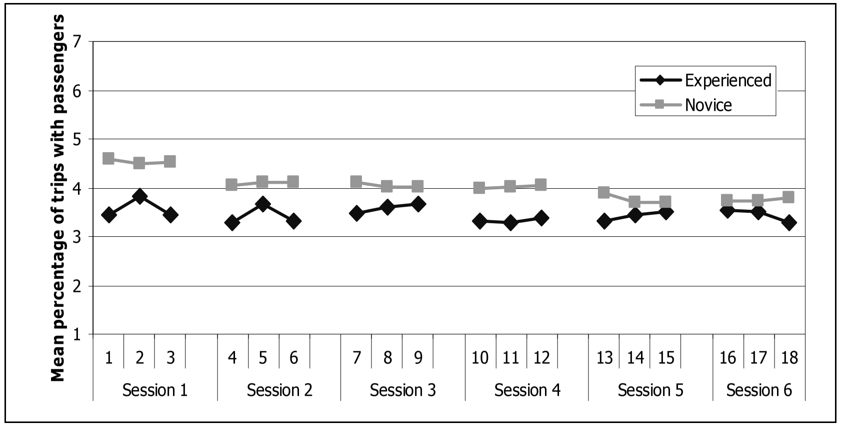

Figure F.9. Mean proportion of trips with a passenger (ranging from 1= all trips alone to 7= all trips with passengers)

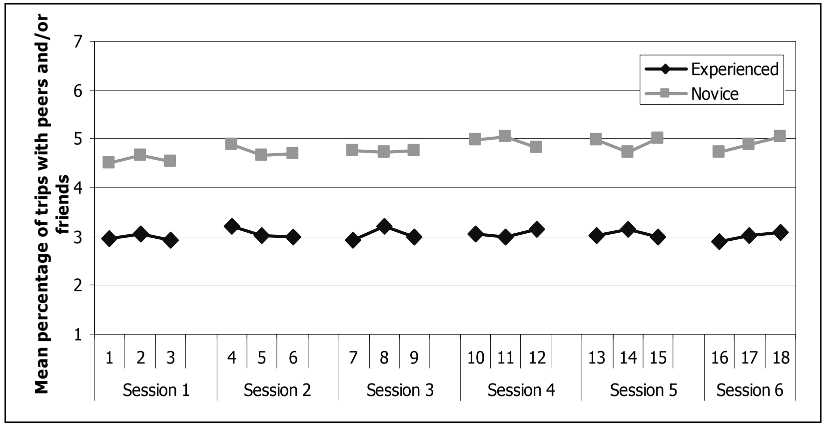

Figure F.10. Mean proportion of trips with peers (ranging from 1 = trips with passengers that are not peers to 7 = trips with peers)

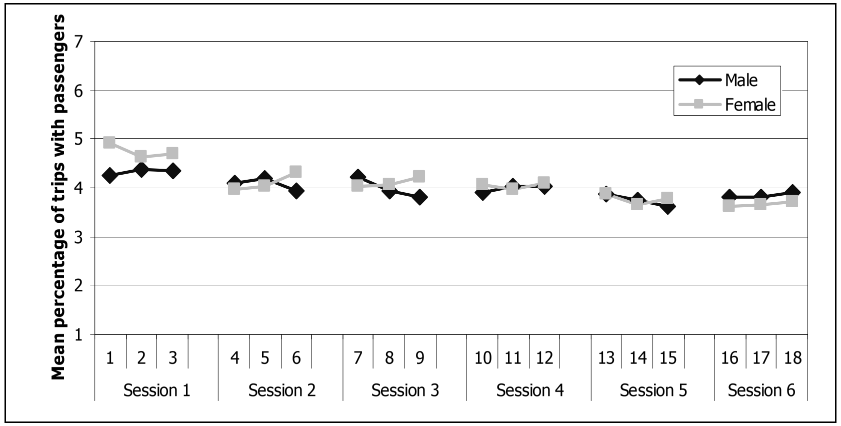

Figure F.11. Mean proportion of trips with a passenger (ranging from 1= all trips alone to 7= all trips with passengers) for *novice* drivers

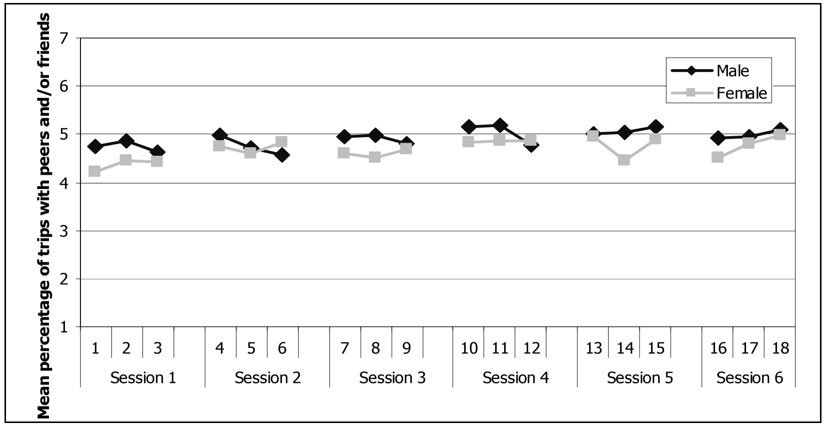

Figure F.12. Mean proportion of trips with peers (ranging from 1 = trips with passengers that are not peers to 7 = trips with peers) for *novice* drivers

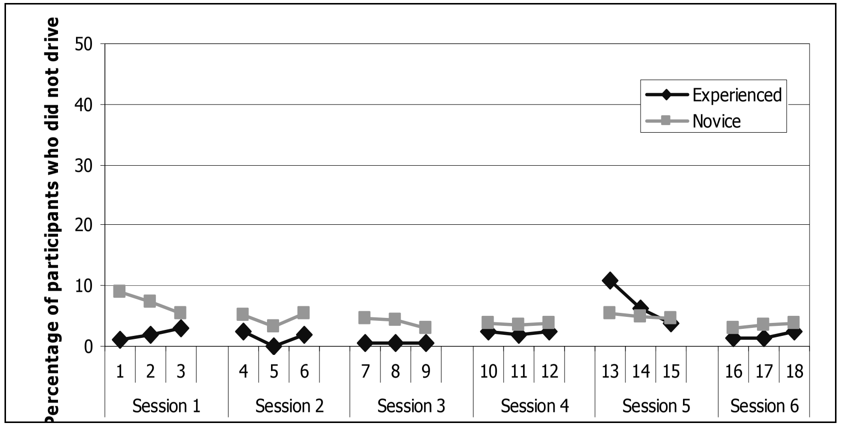

Figure F.13. Percentage of participants who reported not driving due to adverse circumstances at least once in the preceding week

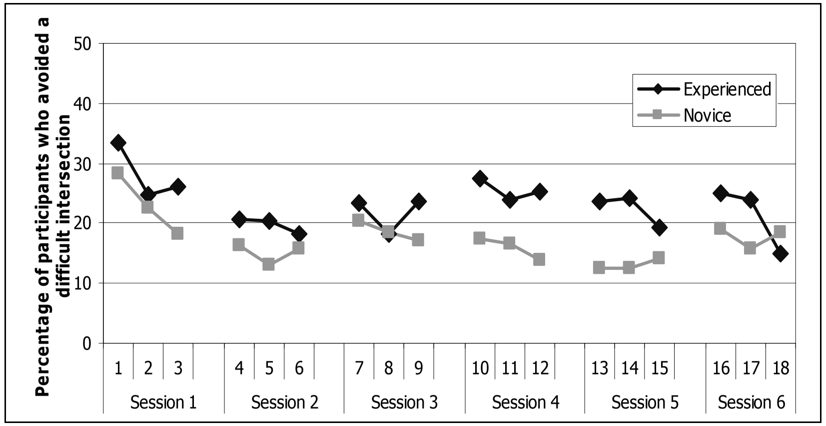

Figure F.14. Percentage of participants who reported avoiding a difficult intersection at least once in the preceding week

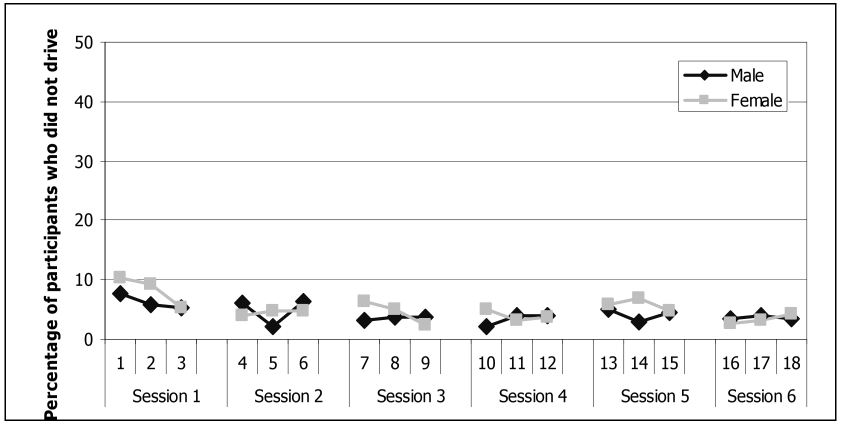

Figure F.15. Percentage of *novice* drivers who reported not driving due to adverse circumstances at least once in the preceding week

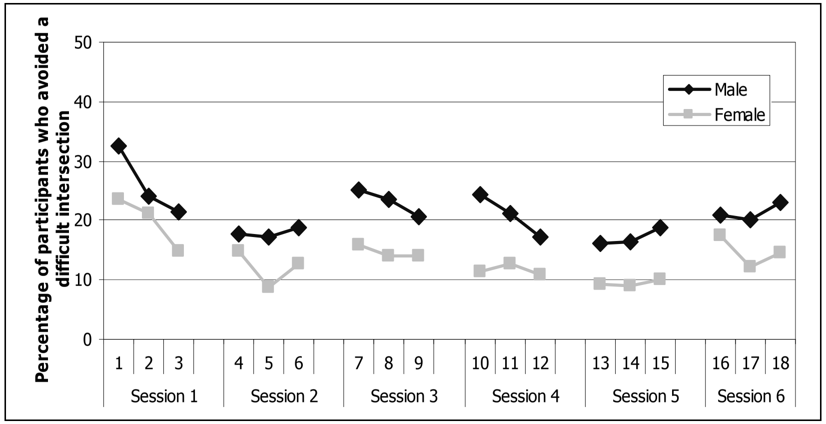

Figure F.16. Percentage of *novice* drivers who reported avoiding a difficult intersection at least once in the preceding week

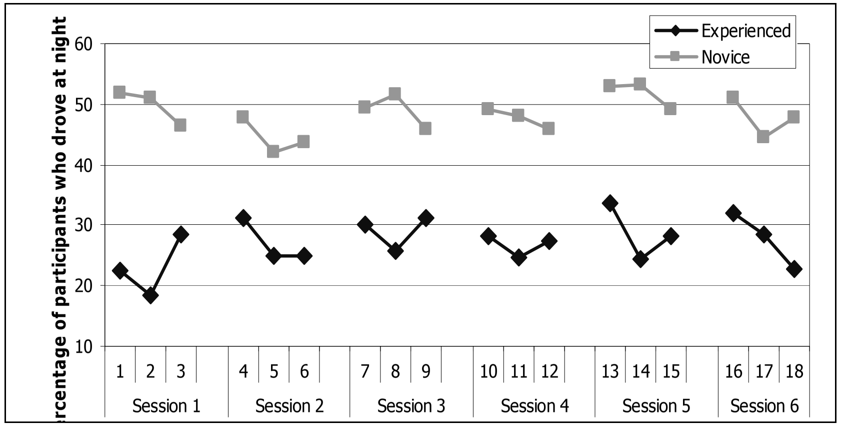

Figure F.17. Percentage of participants who reported driving at night at least once in the preceding week

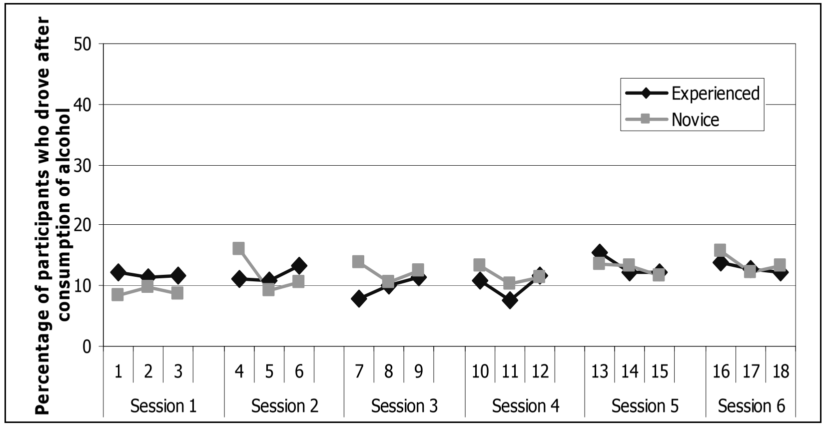

Figure F.18. Percentage of participants who reported driving after the consumption of alcohol at least once in the preceding week

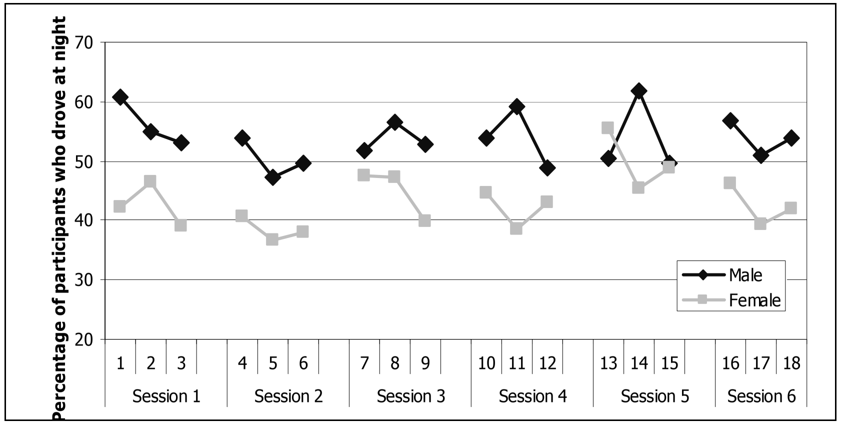

Figure F.19. Percentage of *novice* drivers who reported driving at night at least once in the preceding week

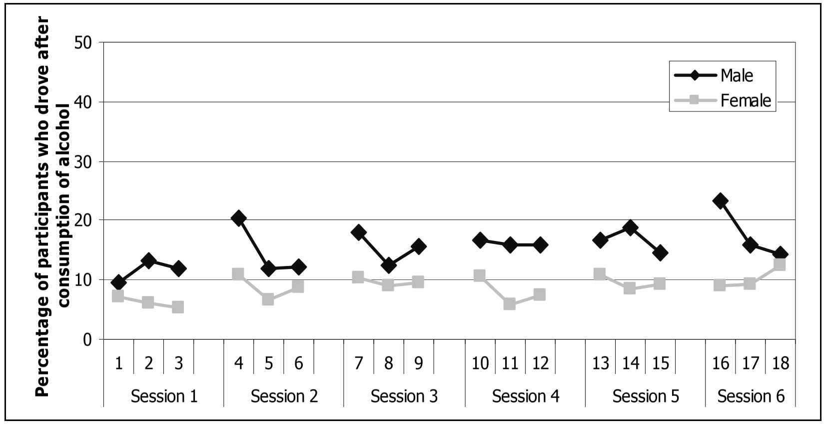

Figure F.20. Percentage of *novice* drivers who reported driving after the consumption of alcohol at least once in the preceding week

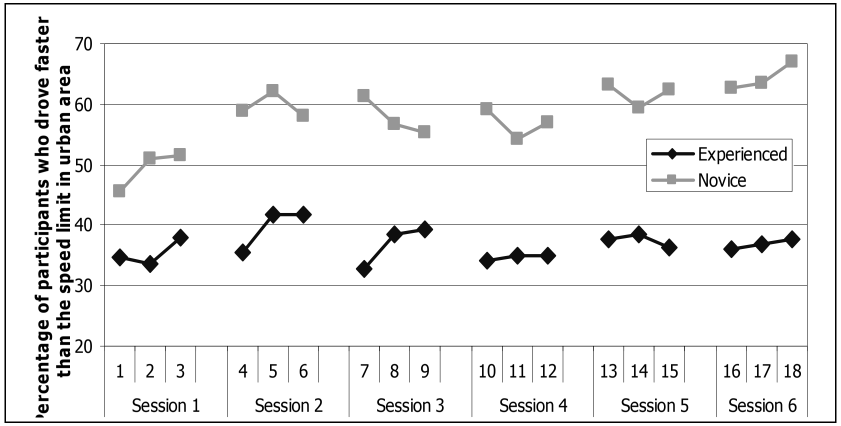

Figure F.21. Percentage of participants who reported driving 10 km/h faster than the speed limit in an urban area at least once in the preceding week

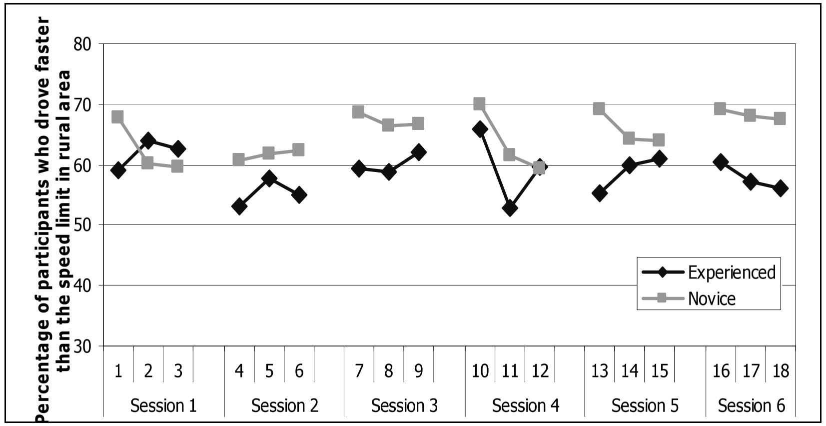

Figure F.22. Percentage of participants who reported driving 10 km/h faster than the speed limit in a rural area at least once in the preceding week

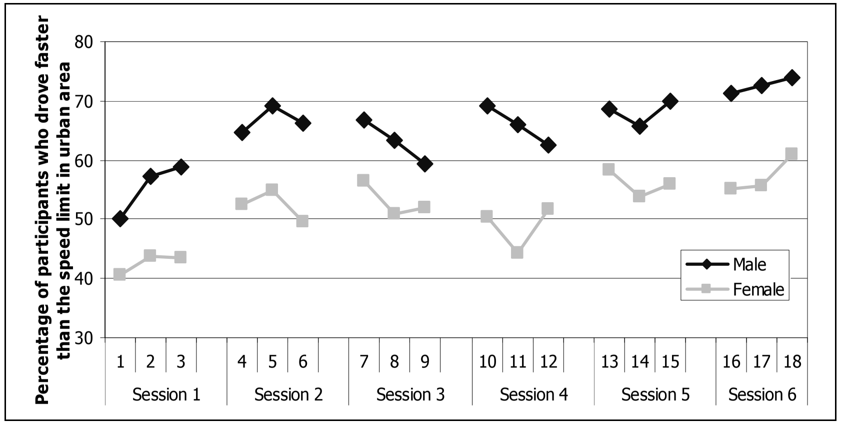

Figure F.23. Percentage of *novice* drivers who reported driving 10 km/h faster than the speed limit in an urban area at least once in the preceding week

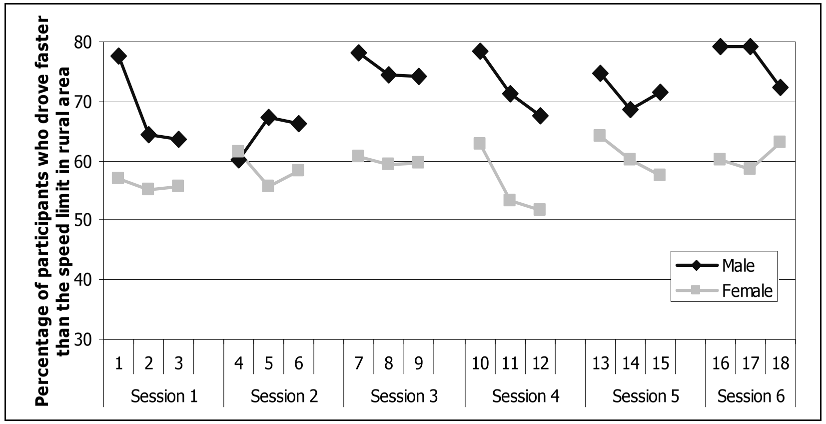

Figure F.24. Percentage of *novice* drivers who reported driving 10 km/h faster than the speed limit in a rural area at least once in the preceding week

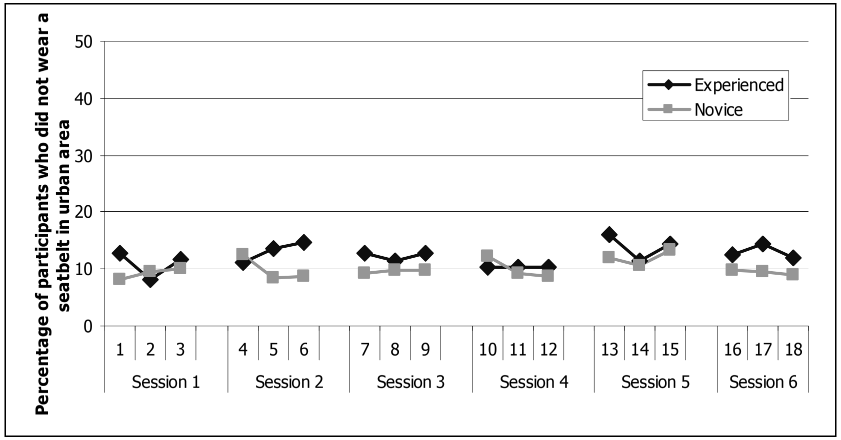

Figure F.25. Percentage of participants who reported driving without wearing a seatbelt in an urban area at least once in the preceding week

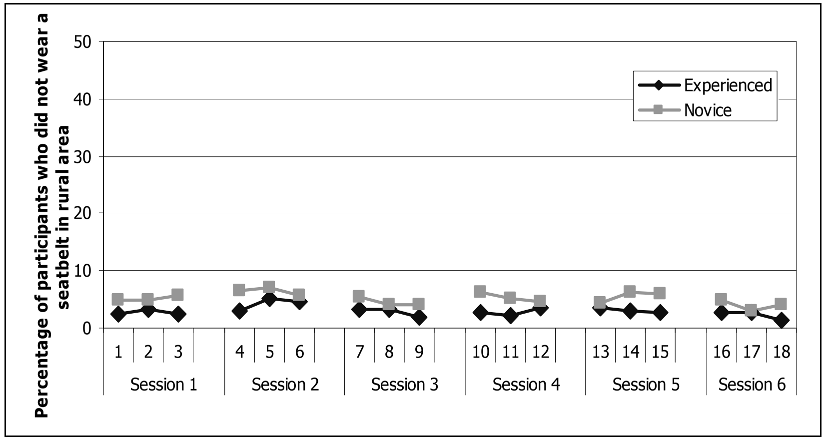

Figure F.26. Percentage of participants who reported driving without wearing a seatbelt in a rural area at least once in the preceding week

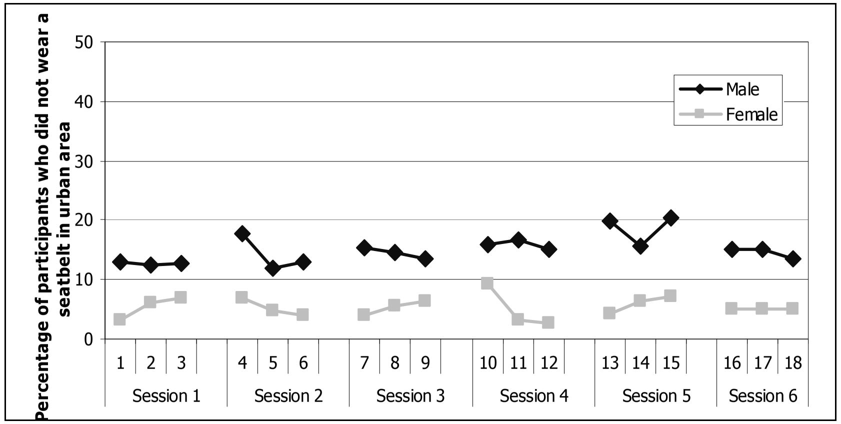

Figure F.27. Percentage of *novice* drivers who reported driving without wearing a seatbelt in an urban area at least once in the preceding week

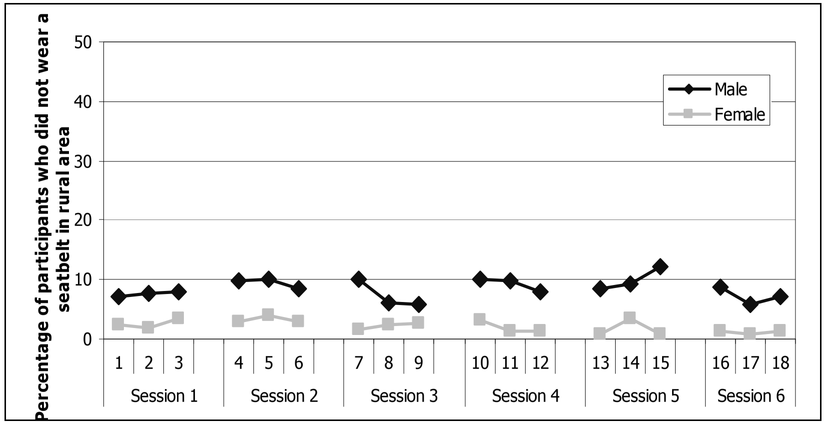

Figure F.28. Percentage of *novice* drivers who reported driving without wearing a seatbelt in a rural area at least once in the preceding week

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL series. See for a complete overview the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Craen, S. de, *The X-factor: A longitudinal study of calibration in young novice drivers*, T2010/2, March 2010, TRAIL Thesis Series, the Netherlands

Tarau, A.N., *Model-based Control for Postal Automation and Baggage Handling*, T2010/1, January 2010, TRAIL Thesis Series, the Netherlands

Knoop, V.L., *Road Incidents and Network Dynamics: Effects on driving behaviour and traffic congestion*, T2009/13, December 2009, TRAIL Thesis Series, the Netherlands

Baskar, L.D., *Traffic Control and Management with Intelligent Vehicle Highway Systems,* T2009/12, November 2009, TRAIL Thesis Series, the Netherlands

Konings, J.W., *Intermodal Barge Transport: Network Design, Nodes and Competitiveness,* T2009/11, November 2009, TRAIL Thesis Series, the Netherlands

Kusumaningtyas, I., *Mind Your Step: Exploring aspects in the application of long accelerating moving walkways,* T2009/10, October 2009, TRAIL Thesis Series, the Netherlands

Gong, Y., *Stochastic Modelling and Analysis of Warehouse Operations,* T2009/9, September 2009, TRAIL Thesis Series, the Netherlands

Eddia, S., *Transport Policy Implementation and Outcomes: the Case of Yaounde in the 1990s*, T2009/8, September 2009, TRAIL Thesis Series, the Netherlands

Platz, T.E., *The Efficient Integration of Inland Shipping into Continental Intermodal Transport Chains. Measures and decisive factors,* T2009/7, August 2009, TRAIL Thesis Series, the Netherlands

Tahmasseby, S., *Reliability in Urban Public Transport Network Assessment and Design*, T2009/6, June 2009, TRAIL Thesis Series, the Netherlands

Bogers, E.A.I., *Traffic Information and Learning in Day-to-day Route Choice,* T2009/5, June 2009, TRAIL Thesis Series, the Netherlands

Amelsfort, D.H. van, *Behavioural Responses and Network Effects of Time-varying Road Pricing,* T2009/4, May 2009, TRAIL Thesis Series, the Netherlands

Li, H., *Reliability-based Dynamic Network Design with Stochastic Networks,* T2009/3, May 2009, TRAIL Thesis Series, the Netherlands

Stankova, K., *On Stackelberg and Inverse Stackelberg Games & their Applications in the Optimal Toll Design Problem, the Energy Markets Liberalization Problem, and in the Theory of Incentives*, T2009/2, February 2009, TRAIL Thesis Series, the Netherlands

Li, T., *Informedness and Customer-Centric Revenue*, T2009/1, January 2009, TRAIL Thesis Series, the Netherlands

Agusdinata, D.B., *Exploratory Modeling and Analysis. A promising method to deal with deep uncertainty,* T2008/17, December 2008, TRAIL Thesis Series, the Netherlands

Kreutzberger, E., *The Innovation of Intermodal Rail Freight Bundling Networks in Europe. Concepts, Developments, Performances,* T2008/16, December 2008, TRAIL Thesis Series, the Netherlands

Taale, H., *Integrated Anticipatory Control of Road Networks. A game theoretical approach*, T2008/15, December 2008, TRAIL Thesis Series, the Netherlands

Li, M., *Robustness Analysis for Road Networks. A framework with combined DTA models,* T2008/14, December 2008, TRAIL Thesis Series, the Netherlands

Yu, M., *Enhancing Warehouse Performance by Efficient Order Picking*, T2008/13, October 2008, TRAIL Thesis Series, the Netherlands

Liu, H., *Travel Time Prediction for Urban Networks*, T2008/12, October 2008, TRAIL Thesis Series, the Netherlands

Kaa, E.J. van de, Extended Prospect Theory. *Findings on Choice Behaviour from Economics and the Behavioural Sciences and their Relevance for Travel Behaviour*, T2008/11, October 2008, TRAIL Thesis Series, the Netherlands

The Netherlands TRAIL Research School
Delft University of Technology
Eindhoven University of Technology
Erasmus University Rotterdam
Radboud University Nijmegen
University of Twente

P.O. Box 5017
2600 GA Delft
The Netherlands
Telephone: +31 (0)15 27 860 46
Telefax : +31 (0)15 27843 33

www.rsTRAIL.nl

TUDelft

Technische Universiteit
Eindhoven
University of Technology

University of Twente
The Netherlands

Radboud University Nijmegen

ISBN: 978-90-5584-127-1