Optimisation of the practical driving test

Berichte der
Bundesanstalt für Straßenwesen

Mensch und Sicherheit Heft M 243 b

Optimisation of the practical driving test

by

Dietmar Sturzbecher
Susann Mörl
Jesko Kaltenbaek

University of Potsdam
Institute for Applied Research on Childhood,
Youth and the Family

Berichte der Bundesanstalt für Straßenwesen

Mensch und Sicherheit Heft M 243 b

Die Bundesanstalt für Straßenwesen veröffentlicht ihre Arbeits- und Forschungsergebnisse in der Schriftenreihe **Berichte der Bundesanstalt für Straßenwesen**. Die Reihe besteht aus folgenden Unterreihen:

A - Allgemeines
B - Brücken- und Ingenieurbau
F - Fahrzeugtechnik
M - Mensch und Sicherheit
S - Straßenbau
V - Verkehrstechnik

Es wird darauf hingewiesen, dass die unter dem Namen der Verfasser veröffentlichten Berichte nicht in jedem Fall die Ansicht des Herausgebers wiedergeben.

Nachdruck und photomechanische Wiedergabe, auch auszugsweise, nur mit Genehmigung der Bundesanstalt für Straßenwesen, Stabsstelle Presse und Öffentlichkeitsarbeit.

Die Hefte der Schriftenreihe **Berichte der Bundesanstalt für Straßenwesen** können direkt bei der Carl Schünemann Verlag GmbH, Zweite Schlachtpforte 7, D-28195 Bremen, Telefon: (04 21) 3 69 03 - 53, bezogen werden.

Über die Forschungsergebnisse und ihre Veröffentlichungen wird in der Regel in Kurzform im Informationsdienst **Forschung kompakt** berichtet. Dieser Dienst wird kostenlos angeboten; Interessenten wenden sich bitte an die Bundesanstalt für Straßenwesen, Stabsstelle Presse und Öffentlichkeitsarbeit.

Ab dem Jahrgang 2003 stehen **die Berichte der Bundesanstalt für Straßenwesen** (BASt) zum Teil als kostenfreier Download im elektronischen BASt-Archiv ELBA zur Verfügung.
http://bast.opus.hbz-nrw.de

Impressum

Bericht zum Forschungsprojekt FE 82.0345/2008:
Optimierung der Praktischen Fahrerlaubnisprüfung
Engl. Fassung (Optimisation of the practical driving test)

Fachbetreuung
Michael Bahr

Herausgeber
Bundesanstalt für Straßenwesen
Brüderstraße 53, D-51427 Bergisch Gladbach
Telefon: (0 22 04) 43 - 0
Telefax: (0 22 04) 43 - 674

Redaktion
Stabsstelle Presse und Öffentlichkeitsarbeit

Druck und Verlag
Fachverlag NW in der
Carl Schünemann Verlag GmbH
Zweite Schlachtpforte 7, D-28195 Bremen
Telefon: (04 21) 3 69 03 - 53
Telefax: (04 21) 3 69 03 - 48
www.schuenemann-verlag.de

ISSN 0943-9315
ISBN 978-3-95606-179-0

Bergisch Gladbach, Juni 2015

Abstract - Kurzfassung

Optimisation of the practical driving test

Within the overall system of novice driver preparation, the practical driving test plays an especially important role for the objective of improved driving safety: On the one hand, the test contents, assessment criteria and test results provide important orientation for the organisation of driving school training and the individual learning processes of the novice drivers (“control function”); on the other hand, the practical test serves to ensure that only novice drivers with adequate driving competence are entitled to participate in motorised road traffic (“selection function”).

The aim of the present project is to elaborate a scientifically founded model for a future, optimised practical driving test, together with a contextual and methodical (implementation) concept for its continuous maintenance, quality assurance and further development. In addition, the institutional structures of the test system, test methods and test procedures – including the necessary demand, assessment, documentation and evaluation standards – are to be described in a “System Manual on Driver Licensing (Practical Test)”.

As a first step, selected psychology-based driving competence models and the contents of training and test documents are to be analysed. The results of this analysis will then serve as the starting point for a discussion of possibilities to model and measure driving competence, and for the outlining of a driving competence model for the theoretical determination of appropriate test content. Subsequently, demand standards for an optimised practical driving test can be derived by applying action theory principles to the demands of motor vehicle handling, and thereby defined as minimum personal standards for driving test candidates. This elaboration is to take into account not only latest knowledge from the fields of traffic and test psychology, but also relevant stipulations in licensing regulations, international trends in the further development of test standards, and novice-specific accident causes and competence deficits.

A further outcome of the project – alongside theoretical-methodical foundations for optimisation of the practical driving test and for the draft of a system manual – is to be a “Catalogue of driving tasks (category B)”, in which the demand standards for the practical driving test are described in the form of situation-related driving tasks and situation-independent observation categories, as a means to specify the criteria for event-oriented performance assessment and overall competence evaluation. At the same time, criteria for the examiner's test decision are to be defined. This optimisation work will contribute, finally, to further development of the adaptive control strategy for the practical driving test.

To enable implementation of the further developed demand, assessment and documentation standards of an optimised practical driving test, a contextual and methodical concept for an electronic test report is to be presented, together with an ergonomically founded design proposal for both hardware and software. The computer-assisted documentation of test performance is intended to support the driving test examiner in planning of the course of a driving test and assessment of the candidate's driving behaviour. Furthermore, optimisation of the performance feedback to candidates and improved possibilities for scientific evaluation of the optimised practical driving test are expected. With regard to test evaluation, a fundamental model is to be described, which – alongside monitoring of the psychometric quality criteria within the framework of an instrumental evaluation – incorporates an evaluation of test results, product audits and the responses to candidate and driving instructor surveys. Finally, the possible influence of driver assistance and accident avoidance systems on the realisation of a driving test and on the assessment of test performance is to be discussed.

Optimierung der Praktischen Fahrerlaubnisprüfung

Die Praktische Fahrerlaubnisprüfung besitzt im Gesamtsystem der Fahranfängervorbereitung eine besondere Bedeutung für die Erhöhung der Verkehrssicherheit: Einerseits stellen die Prüfungsinhalte, Bewertungskriterien und Prüfungsergebnisse wichtige Orientierungspunkte für die Ausrichtung der Fahrschulausbildung und der individuellen Lernprozesse der Fahranfänger dar („Steuerungsfunktion“). Andererseits dient sie dazu, nur Fahranfänger mit ausreichender Fahrkompetenz zur motorisierten Teilnahme am Straßenverkehr zuzulassen („Selektionsfunktion“).

Das Ziel des vorliegenden Projekts besteht darin, ein wissenschaftlich begründetes Modell für eine künftige optimierte Praktische Fahrerlaubnisprüfung sowie ein inhaltliches und methodisches (Betriebs-) Konzept für ihre kontinuierliche Pflege, Qualitätssicherung und Weiterentwicklung zu erarbeiten. Weiterhin sollen die institutionellen Strukturen des Prüfungssystems sowie die Prüfungsverfahren und Prüfungsabläufe – einschließlich der

notwendigen Anforderungs-, Bewertungs-, Dokumentations- und Evaluationsstandards – in einem „Handbuch zum Fahrerlaubnisprüfungssystem (Praxis)“ beschrieben werden.

Zur Erreichung der Ziele werden zunächst ausgewählte verkehrspsychologische Fahrkompetenzmodelle sowie die Inhalte von Ausbildungs- und Prüfungsunterlagen analysiert. Darauf aufbauend werden Möglichkeiten zur Modellierung und Messung von Fahrkompetenz erörtert sowie ein Fahrkompetenzmodell zur theoretischen Bestimmung der Prüfungsinhalte skizziert. Auf dieser Grundlage werden dann die Anforderungsstandards der optimierten Praktischen Fahrerlaubnisprüfung aus handlungstheoretischen Anforderungsanalysen der Kraftfahrzeugführung hergeleitet und als personenbezogene Mindeststandards für Fahrerlaubnisbewerber definiert. Dabei werden – neben dem verkehrspädagogischen und testpsychologischen Erkenntnisstand – auch fahrerlaubnisrechtliche Vorgaben, internationale Trends bei der Weiterentwicklung der Prüfungsstandards sowie fahranfängerspezifische Unfallursachen und Kompetenzdefizite berücksichtigt.

Im Ergebnis des Projektes wird – zusätzlich zur theoretisch-methodischen Begründung der optimierten Praktischen Fahrerlaubnisprüfung und zu einem Entwurf für das Prüfungshandbuch – ein „Fahraufgabenkatalog (Fahrerlaubnisklasse B)“ vorgelegt, in dem die Anforderungsstandards der Prüfung im Sinne von situationsbezogenen Fahraufgaben und situationsübergreifenden Beobachtungskategorien beschrieben sowie darauf bezogene Kriterien für eine ereignisorientierte Leistungsbewertung und eine zusammenfassende Kompetenzbeurteilung festgelegt sind. Darüber hinaus werden Kriterien für das Treffen der Prüfungsentscheidung definiert. Diese Optimierungsarbeiten fließen schließlich in die Weiterentwicklung der adaptiven Steuerungskonzeption der Praktischen Fahrerlaubnisprüfung ein.

Zur Umsetzung der weiterentwickelten Anforderungs-, Bewertungs- und Dokumentationsstandards der optimierten Praktischen Fahrerlaubnisprüfung wird ein inhaltliches und methodisches Konzept für ein elektronisches Prüfprotokoll („e-Prüfprotokoll“) einschließlich eines hard- und softwareergonomisch begründeten Gestaltungsvorschlags vorgestellt. Durch die computergestützte Dokumentation der Prüfungsleistungen soll der Fahrerlaubnisprüfer künftig bei der Planung des Prüfungsablaufs und bei der Bewertung des Fahrverhaltens der Fahrerlaubnisbewerber unterstützt werden. Darüber hinaus werden eine Optimierung der Leistungsrückmeldung an die Bewerber und eine Verbesserung der Möglichkeiten für die wissenschaftliche Evaluation der optimierten Praktischen Fahrerlaubnisprüfung erwartet. Für die Prüfungsevaluation wird ein grundlegendes Modell beschrieben, das – neben der Kontrolle der psychometrischen Gütekriterien im Rahmen einer instrumentellen Evaluation – die Auswertung von Prüfungsergebnissen, von Produktaudits sowie von Bewerber- und Fahrlehrerbefragungen beinhaltet. Schließlich wird der mögliche Einfluss von Fahrerassistenz- und Unfallvermeidungssystemen auf die Prüfungsdurchführung und die Bewertung der Prüfungsleistungen diskutiert.

Contents

Acknowledgements

Annexes (in German language only) are accessible at the electronic archive ELBA under http://bast.opus.hbznrw.de

Acknowledgements

We would like to thank Ms. Bartelt-Lehrfeld from the Federal Ministry of Transport, Building and Urban Affairs, as well as Dr. von Braunschweig (Ministry of Economics, Transport, Agriculture and Viniculture of the Federal State of Rhineland-Palatinate), Mr. Buchardt (Ministry of Infrastructure and Agriculture of the Federal State of Brandenburg) and Mr. Kettenbach (Ministry of Building and Transport of the Federal State of North Rhine-Westphalia) as representatives of the federal states, for their valuable contributions to the project support group and their many constructive suggestions.

Further particular gratitude is due to Mr. Bahr and Mr. Willmes-Lenz, as technical advisors from the Federal Highway Research Institute, for their friendly and pertinent advice and support.

As representatives of the driving instructors, the Technical Examination Centres and their associated institutions, Mr. Biedinger (TÜV NORD Mobilität GmbH & Co. KG), Mr. Böhne (TÜV Rheinland Kraftfahrt GmbH), Mr. Friedel (TÜV | DEKRA arge tp 21), Mr. Glowalla (German Federation of Driving Instructor Associations, BVF), Mr. Isegrei (Central Military Vehicle Registration Office of the Bundeswehr), Mr. Kaup (TÜV SÜD Auto Service GmbH), Mr. Möhlke (TÜV Technische Überwachung Hessen GmbH), Dr. Petzholtz (TÜV | DEKRA arge tp 21), Mr. Radermacher (TÜV NORD Mobilität GmbH & Co. KG), Dr. Reiter (Association of Technical Inspection Agencies, VdTÜV), Dr. Schmidt (DEKRA Automobil GmbH) and Mr. Wagner (TÜV | DEKRA arge tp 21) also contributed to the successful work of the project support group, for which they, too, deserve our full thanks.

Contributors

The experts listed below contributed to the elaboration of the present report:

– Ms. Iris Menrath (Institute for Applied Research on Childhood, Youth and the Family, IFK) and
  Mr. Bernd Weiße (TÜV | DEKRA arge tp 21)
  supplied research results and manuscript drafts for parts of Chapters 2 and 3.

– Mr. Gerhard von Bressensdorf (German Federation of Driving Instructor Associations, BVF),
  Mr. Christoph Kleutges (TÜV Rheinland Kraftfahrt GmbH) and
  Mr. Mathias Rüdel (TÜV | DEKRA arge tp 21)
  commented the draft of the overall manuscript and accompanied the final review with their technical and professional advice.

Responsibility for the content of this publication lies with the authors.

1 Starting point and objective

1.1 Starting point

Road accident statistics at the end of the 1990s indicated a continuous decline in the numbers of persons injured or killed in road traffic; at the same time, however, it was shown that the risk of road traffic injury or death for novice drivers was still several times greater than for experienced drivers. Figure 1 below illustrates the situation for the years from 1998 to 2000 in an international context. The accident figures for 25- to 64-year-old drivers were relatively low in Germany over this period; novice drivers between 18 and 20 years of age, on the other hand, were unable to benefit to the same extent from the numerous measures which had contributed to a continuous enhancement of road safety in Germany since the 1970s.

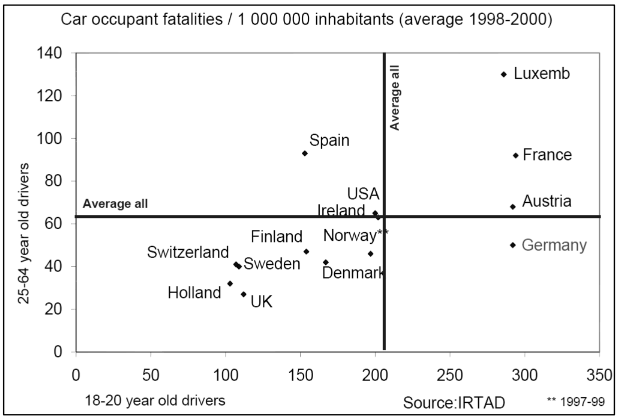

Fig. 1: Accident figures for car drivers of different age groups in international comparison (from: WILLMES-LENZ, 2008)

Under these circumstances, it was reasonable to ask whether the German system already exploits all possibilities to prepare novice drivers for independent participation in motorised road traffic. The Federal Highway Research Institute (Bundesanstalt für Straßenwesen, BASt) reacted to this question by establishing a series of projects to investigate individual aspects of novice driver preparation[1] with regard to their impact on road safety and their potential to reduce accident risk for novice drivers. The objective was to determine an optimum design for all relevant measures of novice driver preparation and to integrate such measures in the most appropriate manner in order to minimise novice driver accidents. These activities must also be viewed against the background of a road safety improvement programme initiated by the Federal Ministry of Transport, Building and Housing in 2001, in which improved road safety for novice drivers was declared one of seven key objectives (WILLMES-LENZ, 2002).

The driving licence test is an exceptionally important element within the overall system of novice driver preparation: On the one hand, in accordance with §17 of the Driving Licence Regulations (Fahrerlaubnisverordnung, FeV), it serves to ensure that only those licence applicants who are able to demonstrate a safe, environment-aware and energy-saving manner of driving are entitled to participate in motorised road traffic ("selection function"). On the other hand, the test contents, assessment criteria and test results provide important orientation for the organisation of driving school training and the individual learning processes of the novice drivers ("control function"), as the contents of training are defined and weighted according to the test demands, and the subsequent test provides feedback to the individual candidate on the level at which the requirements of motorised road traffic are already mastered and which possible deficits must still be tackled in the further course of novice driver preparation.

The project "Optimisation of driving licence testing", a component of the BASt safety research programme which was to be processed over the period from 2001 to 2004 by a consortium of the Technical Examination Centres for Motor Vehicle Traffic (Technische Prüfstellen für den Kraftfahrzeugverkehr), the bodies mandated to conduct and further develop driving licence tests in Germany, gave the start signal for a series of research and development activities aimed at strengthening the potential of the driving licence test as a road safety instrument[2]. It soon became clear, however, that adequate treatment of both the theoretical and practical aspects of the overall project topic was not feasible within the available timeframe. Consequently, the participants investigated firstly the potential for optimisation of the theoretical driving test, which at that time – in connection with growing trends towards computer-assisted test realisation – seemed most promising (BÖNNINGER & STURZBECHER, 2005). The practical driving test was addressed in a subsequent project, which was then financed by the Technical Examination Cen-

[1] "Novice driver preparation" is here understood to mean the entirety of all conditions and measures which are laid down in legislation or, beyond that, provided and used specifically in a particular cultural context to permit the learning of independent, safe and responsible driving of a motor vehicle in public road traffic and demonstration of the necessary knowledge and ability (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014).

[2] The historical development of the driving licence test and the interactions between the involved organisations are described in the report "The History of the Driving Test in Germany" (BÖNNINGER, KAMMLER & STURZBECHER, eds., 2009).

tres and processed between 2005 and 2008 by the TÜV | DEKRA Working Group "Technical Examination Centres in the 21st Century" (TÜV | DEKRA Arbeitsgemeinschaft der Technischen Prüfstellen im 21. Jahrhundert)[3]. The results of this project were published in a scientific report by the Federal Highway Research Institute (STURZBECHER, BÖNNINGER & RÜDEL, 2010) and comprise a description of the methodical foundations and possibilities for further development of the practical driving test. The project results at the same time represent the starting point for the present report on optimisation of the practical driving test; they are thus outlined briefly in the following.

According to STURZBECHER, BÖNNINGER and RÜDEL (2010), the practical driving test can be viewed from the methodical perspective as a work sample which enables the competence of the driving licence applicant to be examined and assessed by way of systematic behaviour observation within the framework of an adaptive test strategy. To permit optimum implementation of the accordingly derived test concept, the authors recommended the following steps for further development of the practical driving test:

– Formulation of situation-related, action-oriented demand standards in the form of driving tasks: In contrast to many other countries, a set of specific driving tasks already exists for the practical driving test in Germany and is presented in Annex 11 to the Examination Guidelines (Prüfungsrichtlinie, PrüfRiLi). The contents of this task catalogue, however, should for the future be modernised, streamlined, restructured and placed on a scientific foundation.
– Formulation of situation-independent, competence-oriented demand standards which, from the methodical point of view, can at the same time serve the driving test examiner as observation criteria (so-called "observation categories"): Observation categories are already described in Annex 3 ("Basic driving manoeuvres for Class B") and Annex 10 ("Demands on the test drive") to the Examination Guidelines. These observation categories should similarly be redefined on a systematic scientific basis, such that the full spectrum of the safety-relevant driving behaviour to be observed is covered as exhaustively and disjunctly as possible.
– Formulation of assessment standards: It is considered necessary to elaborate assessment and decision criteria which refer correspondingly to the optimised driving task catalogue and the newly formulated observation categories. To establish a reference between the driving tasks, observation categories and assessment criteria, the primarily expected behaviour and the applicable assessment criteria should then be described for each driving task; in this way, the demand standards of the practical driving test would be defined in essentially concrete form. The assessment criteria must be formulated as action- or event-oriented categories, and should moreover serve to record not only (driving) errors, but also positive aspects of the performance displayed by the candidate. At the same time, definitions are required for further competence-oriented assessment criteria which relate to the observation categories, but are nevertheless assignable to the event-oriented criteria. On the basis of the event- and competence-related assessment criteria, it is then possible to define certain minimum standards ("training standards", see below) which – in accordance with developmental and traffic psychology principles – would describe the (driving) behaviour or driving competence to be displayed by the novice driver with regard to public road safety; these minimum standards must subsequently be translated into decision criteria for determination of the test result.
– Formulation of documentation standards: As a final step, the scientifically revised methodology for the practical driving test, in other words the system of reformulated driving tasks, observation categories and assessment and decision criteria, must be transferred to an optimised electronic test report. As far as the structure and style of presentation are concerned, it seems expedient to retain the practice-proven multi-dimensional matrix which has already been used in the past by some Technical Examination Centres[4] and permits meaningful continuous documentation of the test performance in a user-friendly, computer-assisted form. Implementation of this recommendation requires research and development work to elaborate a practicable hardware and software solution for creation of the test report; this work must naturally be flanked by corresponding feasibility studies.

[3] For the sake of better legibility, the working group is hereafter referred to as "TÜV | DEKRA arge tp 21".

[4] The Technical Examination Centres have already been experimenting with matrix structures – comprising situation-related and situation-independent demand standards – for better documentation and evaluation of the test results since 1973 (see also Chapter 3).

An optimised test report can then serve as a basis for the concluding conversation between the candidate and the driving test examiner, for the further learning of the novice driver, and not least for evaluation of the test process.

- Elaboration and testing of an evaluation system serving quality assurance for the practical driving test: As the practical driving test belongs to a system of measures operating in the public interest to guarantee road safety, the legislator demands continuous quality assurance, for example formative and summative evaluation of the practical driving test. To this end, it is necessary to develop an evaluation system which – alongside the external and internal audits which have to date been performed predominantly by the Federal Highway Research Institute and the Technical Examination Centres – provides also for instrumental evaluation of the observation inventories used during driving tests, as well as continuous evaluation of the test results and the uniform realisation of multi-perspective customer surveys. Instrumental evaluation studies addressing the psychometric quality of the practical driving test have yet to be conducted, due to the lack of systematic and methodically founded demand, observation, assessment and decision standards, as well as expressive means to document test performance (test reports).

Over the course of the studies conducted between 2005 to 2008 to establish a pedagogical-psychological and methodical foundation for the practical driving test, it was furthermore determined that – beyond the above recommendations – the contribution of driving licence testing to road safety could be increased not only by raising its methodical quality, but also by making greater use of the test results as an instrument to control novice driver preparation. Consequently, the control function of testing was moved into the foreground alongside the selection function. In the context of the general school system, which, as an institution in the educational-sociological sense, displays certain similarities to the system of novice driver preparation (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014), the stronger focus placed on learning results as a basis for the steering of further development and quality assurance measures is referred to as "output control". Events at the end of the 1990s heralded a transition from input to output control[5], and a similar paradigm shift, with equivalent objectives, could also be observed in driver training at this time: Optimisation strategies were no longer concentrated solely on the framework conditions of educational processes (e.g. teaching plans, the further training and qualification of teaching staff, education administration), but instead above all on the outcomes of such processes, i.e. the learning results of the pupils. The theoretical concept which KLIEME and LEUTNER (2006) developed as a means to strengthen the procedures of output control in school education can also be applied to the benefit of novice driver preparation: On the one hand, it enables identification of the steps which are still outstanding in the purposeful further development of novice driver preparation; on the other hand, the project results presented in the current report can then be positioned within this further development.

How must we now proceed, in accordance with the school-related concept elaborated by KLIEME and LEUTNER (2006), to promote effective output control in novice driver preparation? The following points describe both the necessary steps and – in subsequent brackets – the contributions made to further progress by the present report:

1. Initial definition of a model of driving competence. (The discussion in the forthcoming Chapter 2 is intended to serve as a catalyst for further theory development in this direction, despite the fact that an actual driving competence model is still to be elaborated and validated empirically.)
2. Taking up the defined model of driving competence, training standards must be specified to describe the objectives of novice driver preparation in the form of desired learning results. These training standards could serve all organisations and individuals involved in novice driver preparation as a common target specification for the teaching and learning processes in their particular sphere of responsibility. (The demand and assessment standards presented in Chapter 3 bring forth so-called "driving task descriptions", which are attached to the present report as Annex 1. These descriptions indicate the learning achievements to be verified by way of the driving test in the form of driving tasks and the required behaviour to accomplish those tasks; they thus represent the aforementioned "desired learning results" and are consequently to be considered training standards. They are in future to provide a basis not only for driving li-

[5] The development was triggered by the poor learning achievements revealed for German school pupils in international comparisons ("PISA shock") and the resulting efforts to optimise the system of school education (BAUMERT et al., 1997; OECD, 2001; BAUMERT et al., 2001).

cence testing, but above all also for driver training, where they must be anchored in the corresponding curricula alongside complementary content-specific training standards.)

3. On the basis of the training standards, it is next necessary to develop methods to measure and verify the (partial) competences defined in the driving competence model. (In the present report, the practical driving test is detailed as such a method – in the sense of a systematic observation of driving behaviour by way of an adaptive test strategy; it aims specifically to assess practical components of driving competence and is to be viewed as an element of an overarching system of driving licence testing. To assess the verifiable partial aspects of driving competence in their entirety, it is necessary to apply also other test methods, such as the traditional theoretical driving test, as an example of a knowledge test; further methods, such as a traffic perception test, are still under development.)
4. The results obtained by way of the various methods to test driving competence must then be used to assess and improve the effectiveness of the educational processes and the individual forms of teaching and learning in novice driver preparation. The results of the practical driving test, for example, could contribute to an appraisal and optimisation of the effectiveness of driver training, albeit with certain provisos on account of the limited duration of the test and the exceptional conditions of the test situation. Within the framework of such optimisation processes, it is in consequence necessary to further develop also the driving competence model, the derived training standards and not least the test methods, which closes the loop of continuous, empirically based optimisation for an output-controlled system of novice driver preparation.

Further development of the practical driving test is not driven solely by developments in terms of pedagogical and psychological theories and methods or through evaluations of test results, however. Significant impetus is derived from innovations in the fields of computer and vehicle technology, which will in future impact the framework conditions for the practical driving test to a greater extent than ever before. This refers on the one hand to the availability of powerful tablet PCs, with which it has become possible to produce a meaningful electronic test report parallel to realisation of the actual test: Such electronic reports can be considered an essential technical prerequisite for effective and efficient evaluation of the test results; only in this way does it become feasible to implement output control relating to the practical aspects of novice driver preparation. Attention is drawn furthermore to the broad spectrum of driver assistance systems, the dynamic innovative development of which will presumably exert a strong influence on test conditions in the coming years. In the longer term, it is increasingly likely that such systems will even be able to relieve the driver of certain driving tasks; the prerequisites for safe participation in motorised traffic will change accordingly, and with them the competences required to master the demands. This circumstance must already be taken into account in deliberations on the future further development of driving licence testing. On the other hand, it is important that the practical driving test in its current form be adapted appropriately to the presence, use and possible benefits of driver assistance systems in the test vehicles.

The correlations between optimisation of the practical driving test and further development of the overall system of novice driver preparation have already been outlined. In conclusion, therefore, reference is here made to the BASt project "General concept for learner driver preparation". On the basis of the experience gained and new findings since the described paradigm shift around the turn of the century, this project seeks to promote further development of the individual elements of novice driver preparation, and to coordinate these elements in the most expedient manner to further reduce the risk of accident involvement for novice drivers (WILLMES-LENZ, GROßMANN & BAHR, 2011). The next steps for optimisation of the methodical quality of driving licence testing, as expounded in the present report, can be viewed in this context of a general concept for the further development of novice driver preparation in Germany. At the same time, however, the aforementioned general concept also addresses longer-term development objectives, for example the question as to optimum positioning of the practical driving test within the overall system of novice driver preparation in order to best satisfy its selection and control functions.

1.2 Objective

The focus of the present report is placed on a description of specific measures to optimise the methodical quality of the practical driving test. To achieve this objective, the initially presented recommendations of STURZBECHER, BÖNNINGER and RÜDEL (2010) are to be taken up and developed further. The intention is to elaborate a content- and method-related concept for continuous maintenance, quality assurance and further development of the practical driving test. The planned test concept should also enable better cross-references between contents and results of the practical driving test and the effectiveness of measures for novice driver preparation; in other words, the output control function is to be strengthened.[6] The procedure applied to elaborate this test concept during the course of the project, and subsequently the structure of the present report, can be summarised as follows:

– On the basis of an analysis of existing psychological models of driving competence, along with the contents of training and test documentation, possibilities for the modelling and measurement of driving competence are discussed, enabling ideas to be developed for a corresponding driving competence model (see Chapter 2).
– The demand situations which are addressed by competence theory considerations and must consequently be observed within the framework of the practical driving test are then operationalised in the form of driving tasks in accordance with previous studies conducted by McKNIGHT and ADAMS (1970a), McKNIGHT and HUNDT (1971a) and HAMPEL (1977). In addition, observation criteria related to partial competences (in the methodical sense of “observation categories” within the framework of systematic behaviour observation) are defined for these driving tasks. To enable an assessment of driving behaviour and of the (partial) competences demonstrated by test candidates through their mastering of the required tasks, furthermore, a set of event- and competence-oriented assessment criteria – and on this basis criteria for the decision as to whether or not the test is passed – is presented. Finally, implications of the optimised demand and assessment standards for the further development of an adaptive control concept for the practical driving test, and for the outstanding work to describe implementation standards in a methodical manual formulated according to pedagogical-psychological principles, are discussed (see Chapter 3).
– The methodical systematics of the practical driving test (driving tasks, observation categories, assessment criteria) and the overarching methodical documentation standards applicable to systematic behaviour observation then serve to define the special demands to be placed on an electronic test report, for which design recommendations are elaborated alongside corresponding proposals for a feasibility study (see Chapter 4).
– An electronic test report permits meaningful, transparent and objective documentation of the test performances of all driving test candidates, and thus also effective quality assurance by way of formative and summative evaluation of the practical driving test. To this end, an evaluation system is proposed (see Chapter 5) as a means to guarantee scientifically founded future further development of the practical driving test.
– In conclusion, recommendations are given with regard to the handling of driver assistance systems in the context of an optimised practical driving test (see Chapter 6).

[6] During the course of the project, the project objectives were expanded in the sense that the test concept to be elaborated was also to be presented in a draft for a “System Manual on Driver Licensing (Practical Test)”. This is seen to follow up on the corresponding “System Manual on Driver Licensing (Theory Test)”, which received the approval of the Federal/Regional Expert Committee “Driver Licensing and Driving Instructor Legislation” (BLFA-FE/FL) on 06.11.2008 and, at the behest of the federal transport ministry, has since served as a basis for realisation and further development of the theoretical driving test. These system manuals describe the test objectives, the institutions involved, the technical equipment to be used and procedures relating to data privacy, methodical implementation, evaluation and documentation in connection with driving licence testing; they thus represent an operational concept for the system of driving licence testing and contribute to quality-conformant test realisation in accordance with traffic-related policy and the relevant scientific and technical standards.

2 Theoretical foundations of an optimised practical driving test

2.1 Driving competence acquisition and driving competence models

A person who wishes to drive a motor vehicle independently in road traffic in Germany must furnish proof of the necessary driving and traffic competence. In accordance with the Driving Licence Regulations (FeV), the relevant practical driving skills are demonstrated within the framework of a practical driving test: In the course of a test drive of a certain limited duration in real traffic, the driving licence applicant must master a series of demands typically encountered in road traffic in the sense of driving tasks (including basic driving manoeuvres); this is intended to verify the candidate's ability to drive a motor vehicle safely. A detailed description of the current practical driving test in respect of the test model, test participants, test procedure, test contents, test methods, test documentation and quality assurance can be found in STURZBECHER, BIEDINGER et al. (2010).

For deliberations regarding the contents and methodical design of the practical driving test, it can be assumed that, in the "consecutive system" of novice driver preparation practised in Germany (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014), this test is conducted after basic formal driver training in a driving school and before substantial driving experience is gained by way of accompanied or solo driving in real traffic. It is against this background that the practical driving test must fulfil the control and selection functions which were already mentioned in the previous chapter: "This requirement can only be satisfied if the demand standards of the test are not formulated in the sense of demands to be met by an elaborated manner of driving, but instead derived from [answers to the questions of] (1) which components of driving competence are necessary for participation in motorised road traffic, (2) which of these components can be evaluated by a driving test, (3) which level of maturity of the verifiable components must be viewed as the minimum standard with regard to novice driver safety and can this level typically be attained during driver training, and finally (4) how can these minimum standards be operationalised in a methodically meaningful manner in the context of a driving test." (HAMPEL & STURZBECHER, 2010). The answering of these questions is an essential prerequisite for further development of the practical driving test. The first theoretical foundations in this direction were elaborated by STURZBECHER (2010) and by STURZBECHER and WEIßE (2011); these foundations represent the starting point for the following discussion and are to be expanded in the present chapter.

Concept of driving competence

The theoretical roots of this concept are to be found in action theory in the model of "vocational action competence", the function of which is to provide an object-referenced description of the demands of the vocational world. The underlying understanding of competence, which is still widespread even today, stems from the concept of personality described by ROTH (1971), who divides the action capabilities of the individual into the three competence components "domain competence", "personal competence" and "social competence"; this definition was later supplemented by "methodical competence" (KAUFFELD & GROTE, 2002; HEINRICH-BÖLL-STIFTUNG, 2004). Objections to this multi-dimensional approach to action competence lie in the apparently limited validity of a distinction between different subject areas and in the fact that such expressions of competence are based at least partially on the same personality traits (BREUER, 2003; HEINRICH-BÖLL-STIFTUNG, 2004). The model of "vocational action competence" nevertheless remains popular in competence research (ERPENBECK & HEYSE, 1999; FREY, 1999); for the purposes of driver licensing, however, is seems less promising on account of the aforementioned limitations.

According to STURZBECHER (2010), a better starting point for the theoretical description of driving and traffic competence is provided in the concept of competence presented by WEINERT (1999, 2001), which is in the meantime established in school education research and builds upon theoretical approaches from expertise research. Expertise research concerns itself primarily with studies of willing and capable experts and seeks to describe their action regulation by way of so-called "domains" (subject-related demand or action contexts); in doing so, it stresses the importance of subject-specific knowledge and practice-related experience for the acquisition of expertise. Correspondingly, the facet structure and context specificity of competence are for WEINERT (2001) essential aspects of the concept: With regard to the facet structure, he emphasises that competence should not be reduced to its cognitive components, as it also embraces far from negligible motivational components; "context specificity" is understood to mean that, functionally speaking, competences refer to certain classes of situation and can be seen to enable the individual to foresee and master situation-typical demands successfully. In total, the competence concept presented by WEINERT (ibid.) comprises seven facets; alongside the components "knowledge" "skill" and "ability", the aspects "understanding", "action", "experience" and "motivation" are also take into account.

If novice driver preparation is understood as a practice-oriented and experience-driven socialisation process and as an educational institution, then the sketched starting positions from competence theory can also be deemed valid here. Consequently, in conformity with WEINERT (2001), "driving competence" can be defined as those cognitive abilities and skills which are available to or can be learned by an individual as a means to solve certain problems in motorised road traffic, alongside the motivational, volitional and social readiness und ability to realise the problem solutions successfully and responsibly in variable traffic situations (STURZBECHER, 2010). The different facets of driving competence serve to aid specification of the training contents to be conveyed and the demands to be assessed in the driving licence test. As far as the processes to develop and test (driving) competence are concerned, KLIEME et al. (2007) derive two requirements from the context specificity and facet structure of competence, and these requirements must also be taken into account in the further development of driving licence testing: Each operationalisation of a competence must refer to specific classes of the demand situations, and the scope of demand situations must mirror a broad performance spectrum.

With regard to the demand or action context for which driving competence is conveyed and in which it must later be applied, it must furthermore be recognised that motorised road traffic constitutes a poorly defined or "lifeworld" domain with its changing conditions (e.g. weather conditions, traffic density) (STURZBECHER, 2010). Such lifeworld domains are characterised by a high level of complexity and dynamism, i.e. constantly varying demands subject to external influencing factors (GRUBER & MANDL, 1996). There are thus no rules or principles with equal validity for the response to all demand situations; it is rather that a specific problem solution strategy must be generated for each individual demand situation. Consequently, the possibilities for the testing of competences from lifeworld domains under standardised conditions are limited.

Driving competence acquisition

Transferable ("intelligent") knowledge suitable for flexible application, also in new situations, is to be viewed as the basis for all competence (BAUMERT, 1993). The acquisition of intelligent knowledge in a particular domain is best promoted by a mix of systematic and situated learning, i.e. learning in real-life situations (WEINERT, 1998). Basically speaking, two different forms of knowledge are distinguished: Declarative or factual knowledge, and procedural or action knowledge. These two forms of knowledge, however, are indivisible with regard to their acquisition and function, firstly because procedural knowledge builds upon a foundation of declarative knowledge, and secondly since the successful processing of complex tasks such as the driving of a motor vehicle demands an integrated utilisation of both declarative and procedural knowledge aspects in combination with the remaining elements of competence: "Qualification to perform a task means not only possessing the necessary declarative knowledge, but also prior acquisition of a cognitive system bringing together consciously accessible information, highly automated skills, intelligent strategies for knowledge application, a feel for the scope and quality of the available knowledge, positively realistic self-evaluation, and finally the action and learning motivation inherent to the individual's competence" (WEINERT, 1998, p. 111).

The aforementioned quote also points to the requirements to be met during the acquisition of driving competence. STURZBECHER and WEIßE (2011) follow the three-step model of competence or expertise acquisition (ANDERSON, 2001; GREENO, COLLINS & RESNICK, 1996; GRUBER & MANDL, 1996) and characterise the acquisition of driving competence as a process comprising three sequential stages: (1) A "cognitive stage", (2) an "associative stage" and (3) an "autonomous stage".

(1) At the "cognitive stage", declarative knowledge of the actions required for participation in motorised traffic is accumulated by way of instruction and independent study. This is a prerequisite for the reception of further relevant information and for the ability to classify and process this information in individual knowledge structures.

(2) At the subsequent "associative stage", the gathered stock of knowledge is then systematically improved and developed into action knowledge.

(3) The final "autonomous stage" serves perfection of the action knowledge; this enables ever

more rapid and precise application of this knowledge with ever fewer mistakes, and not least a reduction of the occupied attention and working resources.

The acquisition of driving competence thus begins with the systematic development of flexible, connectable and transferable knowledge of the circumstances of motorised road traffic. On this basis, it is then necessary to acquire the ability to apply the knowledge concerned effectively and in a manner appropriate to the situation, i.e. in the contexts of diverse traffic situations. The final outcome of this process is the experience-driven accumulation of a differentiated repertoire of problem- and situation-related action patterns, from which the appropriate driving behaviour can be retrieved either consciously or automatically. This third step thus serves above all to develop driving skills as the core of driving competence; corresponding starting points for a theoretical explanation are outlined by STURZBECHER (2010). These psychomotor driving skills are then combined with current information and factual knowledge to permit the successful mastering of traffic situations (see below).

It follows from the above sequential process of driving competence acquisition that the different forms of testing must be positioned according to the particular content to be assessed if they are to achieve their full control effect within the overall process of driving competence acquisition. The theoretical driving test, which addresses above all declarative knowledge, belongs at a relatively early stage of the acquisition process; the practical driving test, which is intended to verify (elaborated) action knowledge, on the other hand, should be placed after a substantial phase of driving experience and thus at the end of novice driver preparation.[7]

From the educational psychology perspective, more detailed definition of the aforementioned fundamental teaching/learning processes, as they are realised in the context of driving competence acquisition and driving school training, is guaranteed by teaching plans or training curricula.[8] A curriculum with binding stipulations to govern novice driver preparation, or even merely parts of the process such as driving instruction, does not yet exist in Germany, but is intended to be elaborated within the framework of the BASt project "Approaches for the optimisation of driving school training" (BREDOW, 2012). At present, the objectives and contents of driver training in Germany – and thus also of the components of driving competence to be developed – are defined solely with reference to licensing legislation in the Learner Driver Training Ordinance (Fahrschüler-Ausbildungsordnung, FahrschAusbO). To facilitate the implementation of these stipulations in practical driving instruction, a concept for a graduated training process was developed by the German Federation of Driving Instructor Associations (Bundesvereinigung der Fahrlehrerverbände, BVF) in the early 1980s. The corresponding didactic recommendations were subsequently a subject of constant further improvement and in 1993 formed the basis for "Curricular guidelines for practical training for car drivers" ("Curricularer Leitfaden – Praktische Ausbildung Pkw"; FISCHER et al, 2005; LAMSZUS, 2008), which the driving instructors are able to use on a voluntary basis. These guidelines also give descriptions of individual levels of driving competence acquisition and achievement; the following five stages are differentiated:

- "Basic stage" (e.g. elementary introduction to the motor vehicle, including the correct seat position and driving posture, the operation of individual vehicle controls, acquisition of basic psychomotor skills relating to gear-changing and driving off)
- "Supplementary stage" (e.g. degressive braking, estimating distances, basic driving manoeuvres, environment-aware and foresighted driving, traffic observation and hazard perception)
- "Advanced stage" (above all increasingly more complex driving manoeuvres in traffic, behaviour towards pedestrians and other road users)

[7] The practical driving test is currently taken at a relatively early point of the competence acquisition process, namely after approx. 1.5 to 3 months, i.e. at the beginning of the associative stage. Consequently, it permits assessment of only a minimum standard of driving competence, which is frequently still insufficient from the perspective of safe solo driving – as is indicated by the failure rate and the increased accident figures among novice drivers. The driving test thus controls only the initial phase of novice driver preparation. Within the framework of further development, therefore, the positioning of the practical driving test must be analysed and possibly reconsidered with a view to optimisation of its selection and control function.

[8] Curricula represent a much more comprehensive instrument by which to control training processes compared to teaching plans: Whereas teaching plans generally limit themselves to a catalogue of learning contents, curricula focus on training objectives, teaching/learning processes and their evaluation (TENORTH & TIPPELT, 2007). The learning contents are selected and structured on the basis of thematic and didactic considerations; at the same time, the relevant teaching/learning methods, teaching/learning media, learner assessment procedures and the general administrative and institutional conditions for curriculum development are described and founded scientifically (KELLY, 2009; MARSH, 2009; OLIVA, 1997).

- "Special training drives" (e.g. driving at higher speeds outside built-up areas and on motorways, driving in the dark)
- "Maturity and test stage" (e.g. preparation for the practical driving test).

To summarise, it can be said that the essential content components and levels of driving competence are already reflected in corresponding training specifications and design proposals, namely in the Learner Driver Training Ordinance and the BVF's curricular guidelines (GRATTENTHALER, KRÜGER & SCHOCH, 2009). Nevertheless, such legislative provisions and didactic recommendations are unable to offer more than mere inspiration for the elaboration of driver training curricula with appropriate foundations in educational psychology, including the associated contents and methods for driving licence testing, as they are not based on elaborated driving competence theory models. This should not be understood as criticism: Given the still unsatisfactory status of basic learning theory research relating to driving competence, few (empirically tested) models are available; the elaboration of such models is a task for cognitively oriented psychology, education science and specialised didactics.[9] In the same way as in the school education sector (see KLIEME et al., 2007), the development of control instruments for driving school training and licence testing has thus to date been based primarily on the experience of experts with particular qualifications in the field (i.e. here driving instructors and driving test examiners), and on their notions pertaining to practicable and systematic implementation.[10]

Driving competence models

Applying the initially described basic positions from competence theory to the concept of driving competence, it follows that driving competence models underlying an optimised practical driving test refer to typical demand situations of motorised road traffic and should reflect these situations as comprehensively as possible. Assuming that such driving competence models are to serve as a basis for the elaboration of training standards in the overall system of novice driver preparation, as explained in Chapter 1, they must furthermore fulfil the following functions identified by KLIEME et al. (2007):

- They must describe the content structure of those demands which novice drivers are expected to satisfy (components of driving competence)
- They must develop practical and scientifically founded notions of the possible graduations of driving competence, i.e. the competence levels which can be determined in the case of novice drivers (stages of driving competence).

In accordance with these two functions, KLIEME and LEUTNER (2006) distinguish competence structure and competence level models. Competence structure models are intended to answer the question as to "which and how many distinct dimensions of competence can be differentiated in a specific field". Competence level models, by contrast, refer to the "concrete situative demands which a person is able to master given a particular scope of competence" (p. 6). The two types of model thus relate to different aspects of the competence construct (content structure and attained levels), but they are not mutually exclusive and are ideally even complementary (KOEPPEN, HARTIG, KLIEME & LEUTNER, 2008). Consequently, it would be desirable for an ideal model of driving competence for the context of novice driver preparation to describe both the various content-based dimensions of driving competence (partial competences) and the possible levels of attainment of such partial competences among novice drivers.

Both the structure models and the level models of driving competence can be differentiated further. In the case of structure models, SCHECKER and PARCHMANN (2006) distinguish between (1) normative competence models, which indicate the (cognitive) prerequisites to be met by a learner to be able to solve tasks and problems in a particular subject or demand field, and (2) descriptive competence models, which describe "typical" patterns of (cognitive) prerequisites with which it is possible to map the behaviour of a learner when solving such subject- and demand-specific tasks and problems. In respect of level models, it must be taken into account whether the described levels of driving competence represent merely the possible manifestations of driving competence or also the steps leading to their acquisition; in the latter case, it is customary to speak of "competence acquisition models". It cannot be assumed, however, that the levels of a competence model will necessarily coincide with the described stages of competence acquisition; questions addressing competence acquisition models, in particular, are difficult to answer (KLIEME et al., 2007). Independently of the model type, theoretically developed competence models can initially only be considered hy-

[9] GRATTENTHALER, KRÜGER and SCHOCH (2009), for example, point out that neither the order of individual components nor the timeline of driving competence acquisition have yet been described with corresponding scientific foundations, and that this can be taken as explanation for differing arrangements of theory classes and driving instruction.

[10] In some cases, even scientific approaches elaborated by the Technical Examination Centres have not been applied (see Chapter 3).

potheses and must subsequently be verified empirically (KLIEME, 2004).

Already from this brief overview of the different types of competence model and their specific functions, it is clear that there can be no complete or general model of driving competence which fulfils all functions in ideal manner. This is true not least because models always serve a certain descriptive purpose and thus focus on very specific aspects of a phenomenon in reality. For the description of driving competence, it would naturally be desirable to develop model types with mutual references (i.e. to elaborate acquisition models for the various components of driving competence, for example); this, however, is a time-consuming and complex process, and is not to be placed in the foreground at this point. The objective of the present chapter is instead merely to systematise contents and methods and to provide theoretical foundations for the components of driving competence to be assessed by way of an optimised practical driving test in a normative driving competence model. Such a model of driving competence must initially be developed on the relatively general level of structure and acquisition aspects, and can only then be expanded into further detail. The first step is thus to determine the components of driving competence for which assessment is desirable and feasible within the framework of an optimised practical driving test. As a second step, it is then necessary to describe the demand situations to which test items are to refer, together with the levels of competence to be demonstrated in each case.

2.2 Classification of test demands in accordance with competence theory

Among the theoretical models which are used in the engineering and traffic sciences to describe content-based requirements relating to the control of a motor vehicle, the three-level model of driving demands developed by DONGES (1982) is especially conspicuous. This model corresponds with the three-level models of vehicle control of BERNOTAT (1970) and MICHON (1985); it is not a competence model, but it does incorporate action-oriented notions of the demand aspects of driving which can be taken as indication of certain driving competence prerequisites. Theoretical notions of the attainment levels and acquisition stages of driving competence, on the other hand, are to be found in the three-level model of goal-oriented performance by RASMUSSEN (1983). DONGES' (2009) combination of the two models (see Fig. 2), appears particularly interesting for the present purpose, because further development of the practical driving test requires that both the content structure dimensions and the levels and stages of driving competence acquisition be taken into account (see above).

DONGES (see right-hand side in Fig. 2) divides driving behaviour into different content-based actions relating to stabilisation, manoeuvring and navigation of the vehicle; each of these action categories requires corresponding components of driving competence. On the stabilisation level, the driver seeks to retain control over the motions of the vehicle under changing traffic conditions. To this end, he[11] makes whichever corrections are required to avoid losing control of the vehicle. This calls for timely recognition of any potential hazards (e.g. risk of skidding or oversteering when driving through bends), and correct selection and application of the appropriate reaction (e.g. acceleration, steering, braking). Actions on the manoeuvring level enable the driver to fulfil an overarching transport task. In doing so, he realises a planned journey route and matches his driving behaviour to the course of the road and any surrounding traffic. Central aspects of vehicle control at this level are driving manoeuvres such as cornering, overtaking, lane changing or the negotiation of junctions. All these manoeuvres must be performed safely and in compliance with the rules of the road, i.e. without endangering the vehicle or other road users, and without causing avoidable hindrance to others. The tasks for the driver on the manoeuvring level are thus to observe the traffic, to maintain an appropriate driving line, to adapt the vehicle speed and the distance to other vehicles to the situative circumstances, and not least to communicate with other road users. Before these manoeuvring demands can be tackled, the driver must usually determine a destination and driving route; it is nevertheless possible that the planning accomplished on this navigation level must be revised or adapted during the journey itself. The decision for a particular route, for example, must take into account factors such as the expected duration, which may vary at different times of the day, the purpose of the journey, possible intermediate destinations and the safety of the route (e.g. the probability that the road has been gritted in winter). During the journey, it may furthermore become necessary to seek an alternative (e.g. in case of hold-ups on the planned route), which then involves re-orientation.

[11] Wherever gender-specific nouns or pronouns are used in this report, this serves solely to maximise general legibility and is in all cases to be understood to refer to persons of both genders.

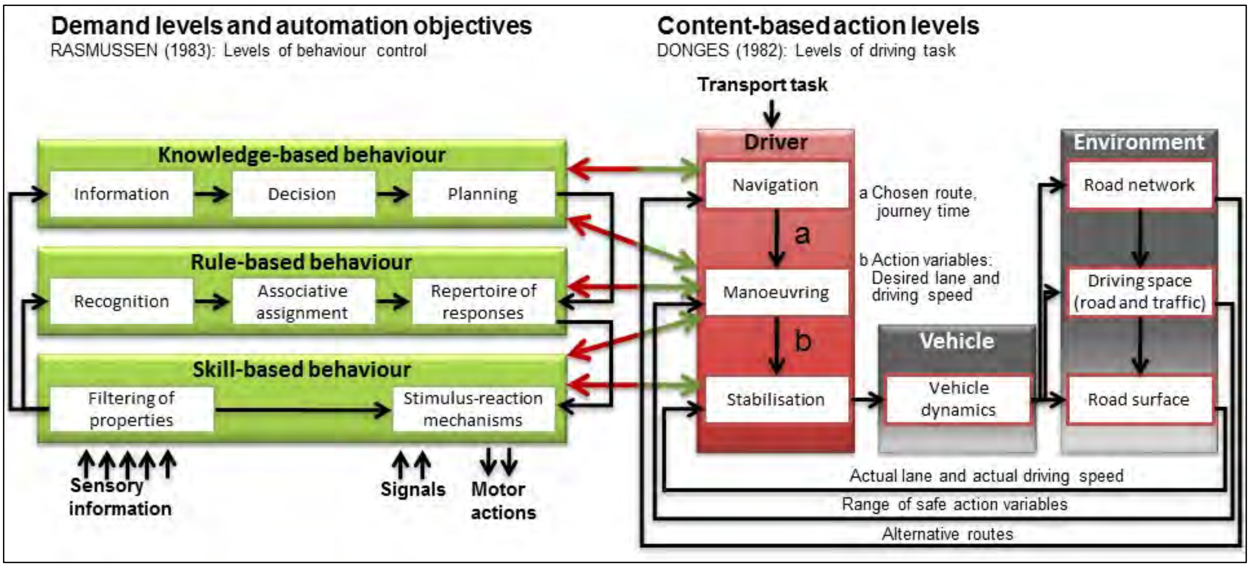

Fig. 2: Driving behaviour model (based on DONGES, 2009) combining content-based action demands (after DONGES, 1982) and automation levels (after RASMUSSEN, 1983)

Viewed overall, and as illustrated by the examples, it is possible to define action demands which permit the determination and assignment of corresponding driving competence prerequisites on all three levels of the model. These driving competence prerequisites must be developed in the course of driver training; a selection of the demands can then be assessed by way of the practical driving test in accordance with feasibility and road safety considerations.

RASMUSSEN (see left-hand side in Fig. 2) uses the underlying extent of automation to distinguish three different levels of behaviour control, namely “knowledge-based behaviour”, “rule-based behaviour” and “skill-based behaviour”. If the model is applied to driving behaviour, then it is characteristic for the – least automated – stage of knowledge-based behaviour that the driver must consciously recognise, evaluate and interpret the demands of a traffic situation. On this basis, the driver plans the appropriate next actions and the manner in which these actions are to be performed. In the case of rule-based behaviour, the driver applies one of a stored set of behaviour responses acquired through experience or practice, i.e. he already knows what is to be done as soon as a certain situation is recognised (in the sense of “if-then rules”); his driving behaviour is thus semi-automated. Skill-based behaviour, finally, is characterised by reflexive stimulus-reaction mechanisms (“routines”) which no longer require conscious control and are thus applied fully automatically. The automation of action sequences which are necessary in certain prototypical traffic situations (e.g. changing road lanes) facilitates mental information processing on the part of the driver and releases working capacities for the handling of demands at a superordinate level (e.g. navigation), for auxiliary actions (e.g. conversation with passengers) or for the processing of unexpected or new traffic situations (see also Chapter 6).

In conclusion, it remains to be said that the driving behaviour model presented by DONGES (2009) comprises elements of both driving competence structure and driving competence level models, and that these aspects are furthermore integrated with each other: The cross-references between the three content-based demand levels (DONGES, 1982) and the three behaviour control and automation levels (RASMUSSEN, 1983) give an indication of the automation (or competence) levels on which a driver should normally handle the various contextual demands of driving. It can be seen, for example, that it is hardly possible to automate navigation, and that this occurs mainly on the level of knowledge-based behaviour. This results not least from the fact that journey destinations, routes and road conditions will only seldom reoccur in identical combination; navigation is thus generally subject to conscious control and ties up mental capacities. By contrast, the handling (or manoeuvring) of a vehicle, which includes above all vehicle control, traffic observation, communication with other traffic participants, speed regulation and positioning of the vehicle in a given traffic situation, is accomplished more or less automatically by an experienced driver in the form of skill-based behaviour; it nevertheless requires also knowledge- and rule-based behaviour at times – above all in unusual or unexpected traffic situations. This means that vehicle handling can be automated to a certain degree, but must also be controlled consciously in part, and that a novice driver must first acquire the possible automation level through driving practice. Behaviour on the stabilisation level, finally, is based on skills; it becomes automated over time and then occupies only a small proportion of working memory capacity.

Alongside the numerous starting points to be found in the model presented by DONGES (2009) for the elaboration of a driving competence model serving to systematise driving test contents, STURZBECHER and WEIßE (2011) already identified one important limitation, namely the lack of a content-based demand or competence component which takes into account the social context of driving (e.g. relevant social values and norms) and its interactions with individual attitudes.[12] The authors (ibid.) thus proposed the incorporation of such components and – after critical analysis of the potential of the GADGET model for a description of driving test requirements – concluded that, "for driving licence testing, the relatively clearly discriminated demand levels 'Stabilisation level', 'Manoeuvring level', 'Navigation level' and 'Value level' provide a useful starting point for the determination of content-based driving demands, and thus also of content-based components of driving competence. A corresponding structure model thus starts with a basic operational demand level (the stabilisation level); this base supports a tactical level (the manoeuvring level, where the operational elements are arranged meaningfully into driving manoeuvres according to situative demands), followed by a strategic level (the navigation level where driving is planned) and an overarching value level. Consequently, these four demand levels should be used to systematise content for the demand standards of the practical driving test" (ibid., p. 23-24).

GRATTENTHALER, KRÜGER and SCHOCH (2009) establish a bridge between content-related structural components of driving competence on the one hand, and the learning psychology mechanisms and stages of driving competence acquisition on the other, by describing driving competence as action knowledge which can be subdivided into three knowledge forms or stages of acquisition, namely "explicit knowledge", "implicit knowledge" and "process knowledge":

1. In the context of long-term knowledge, "explicit knowledge" is understood to mean factual or declarative knowledge. This comprises semantic or abstract knowledge of concepts, objects, facts, conditions or rules, as well as episodic or situated knowledge, in the form of situation prototypes and action scripts serving as central elements of top-down action planning. The descriptor "explicit" indicates that this knowledge can generally be reported and thus also conveyed by way of verbal instruction.
2. "Implicit knowledge" refers to procedural components of long-term knowledge, which are acquired in the form of motor schemata and further differentiated by the feedback loops of action effect, environment perception (above all visually) and proprioception. The action result is compared with the action planning, and modification of the action process is initiated where the result deviates from the objective. The descriptor "implicit" indicates that this knowledge is normally unsuitable for reporting: Even if a particular action sequence – whether driving a motor vehicle or simply tying a shoe lace – is mastered perfectly, it is normally not possible to explain exactly how psychomotor procedures are accomplished and which specific points must be taken into account. Consequently, implicit knowledge cannot be acquired by way of instruction alone, and is instead dependent on more or less intensive practice and the gathering of experience under changing action conditions. The outcome is a set of (automated) psychomotor skills serving to realise the given action.
3. "Process knowledge", finally, integrates explicit and implicit knowledge: To permit successful mastering of the demands of constantly changing and more or less familiar traffic situations, factual knowledge relating to these situations is activated[13] and combined with psychomotor skills; this also presumes pertinent resource control and self-evaluation.

[12] With his "hierarchical model of driving competence acquisition", KESKINEN already integrated a fourth value-referenced component into the accepted three-value model of vehicle control (see above) in the mid-1990s. This thinking was later incorporated into the so-called "GADGET matrix" ("Guarding Automobile Drivers through Guidance, Education and Technology"; HATAKKA, KESKINEN, GREGERSEN & GLAD, 1999), in which the content-based demand levels "Vehicle manoeuvring", "Mastering of traffic situations", "Goals and context of driving" and "Goals for life and skills for living" are distinguished. The four described levels are combined with the three dimensions "Knowledge and skills", "Risk-increasing factors" and "Self-evaluation". This matrix provides a structural definition framework for driving competence and is used in countries such as Norway, Sweden and the Netherlands to determine training contents and teaching/learning forms (and thus implicitly also test contents and forms of testing) for novice driver preparation.

[13] STURZBECHER and WEIßE (2011) describe the relevant action processes through recourse to the information processing model presented by DODGE (1982) and on this basis derive possibilities for the operationalisation of test contents relating to hazard perception and avoidance.

The content-based demands or action levels of driving behaviour developed above in accordance with DONGES (1982, 2009) can now be combined with the psychological prerequisites described by GRATTENTHALER, KRÜGER and SCHOCH (2009) in the sense of general competence components or types of knowledge, alongside a supplementary motivation component, to form a competence structure model (see Fig. 3). This competence structure model permits specification and classification of the content-related components of driving competence to be assessed by different forms of testing; at the same time, it offers a reference for determination of the appropriate form of acquisition or knowledge. The model serves furthermore as a basis for the definition of test contents and corresponding test tasks, which must subsequently be described in the form of training or test standards (see below). Particularly in the case of the practical driving test, the test tasks refer in turn to those traffic situations which can be mastered by way of appropriate (driving) behaviour. It can also be derived from the model that explicit knowledge and motivation are more relevant for the mastering of tasks on the higher demand level; implicit knowledge, on the other hand, is of greater significance for the mastering of demands on the lower levels.

On which areas of driving competence should the (optimised) practical driving test focus in respect of content? Both the legislative framework and the basic ecological conditions of the practical driving test – which are equally relevant in the context of psychological testing – suggest that content for this form of testing should be based primarily and predominantly on the manoeuvring level. In accordance with § 2 (5) of the Road Traffic Act (Straßenverkehrsgesetz, StVG), and likewise § 17 (1) of the Driving Licence Regulations (FeV), a safe and environment-aware manner of driving a motor vehicle, i.e. appropriate performance of the driving manoeuvres necessary to realise a selected route, are to be viewed as the core objective of the test to be conducted by the driving test examiner.[14] Driving licence legislation also contains concrete provisions relating to the test contents in the form of a catalogue of driving tasks for the practical driving test; in terms of content, the situation-related driving tasks anchored therein (e.g. changing lanes, overtaking and passing, negotiating crossroads and junctions) can be assigned to the manoeuvring level. In addition, situation-independent, behaviour-referenced demand standards are specified in the sense of components of driving competence to be assessed, and again relate essentially to the manoeuvring level (e.g. “traffic observation”, “speed regulation”); this point is to be taken up again later (see Chapter 3).

Compared to the manoeuvring level, the value level plays an insignificant role with regard to the demands of the practical driving test: The values and attitudes of a driving licence applicant, and similarly his motivation to comply with the requirements of driving in an appropriate manner, can hardly be assessed in the test situation, because all candidates aim to pass the test and will thus behave in conformity with the demands – possibly in contrast to their behaviour in later solo driving. Consequently, verification of the motivational prerequisites for driving competence is above all a task for medical-psychological assessments of the fitness to drive and not for the driving test.

On the navigation level, too, significant proof of driving competence on the part of the candidate cannot be expected from the practical driving test, not least because the driving test examiner – as in a number of further European countries (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014) – specifies the test route and test demands step by step in accordance with an adaptive test strategy, in order to be able to optimise his basis for assessment (STURZBECHER, BIEDINGER et al., 2010). Even though navigation is not explicitly operationalised as a test requirement, navigation tasks may be set on occasions – insofar as the candidate declares that he possesses corresponding local knowledge. The stabilisation level also offers only limited possibilities for the assessment of driving competence: Elementary skills relating to operational vehicle control are naturally considered prerequisites for successful completion of the practical driving test; given the careful and foresighted manner of driving which is expedient in the test situation, however, circumstances which could potentially lead to a loss of control (referring, for example, to the longitudinal and transverse stability of the test vehicle) will only seldom arise (see Chapter 6).

[14] More precise requirements are specified in §§ 15 to 18 and especially in Annex 7 FeV, in which the essential demands placed on driving tests are detailed. Part 2 of Annex 7 FeV, in particular, contains stipulations of the test subject matter, the test duration, the test vehicle and the manner of realisation and evaluation of the practical driving test. The Examination Guidelines and their Annexes 2 to 12 contain complementary stipulations (e.g. a catalogue of driving errors and descriptions of the basic driving manoeuvres and test drive).

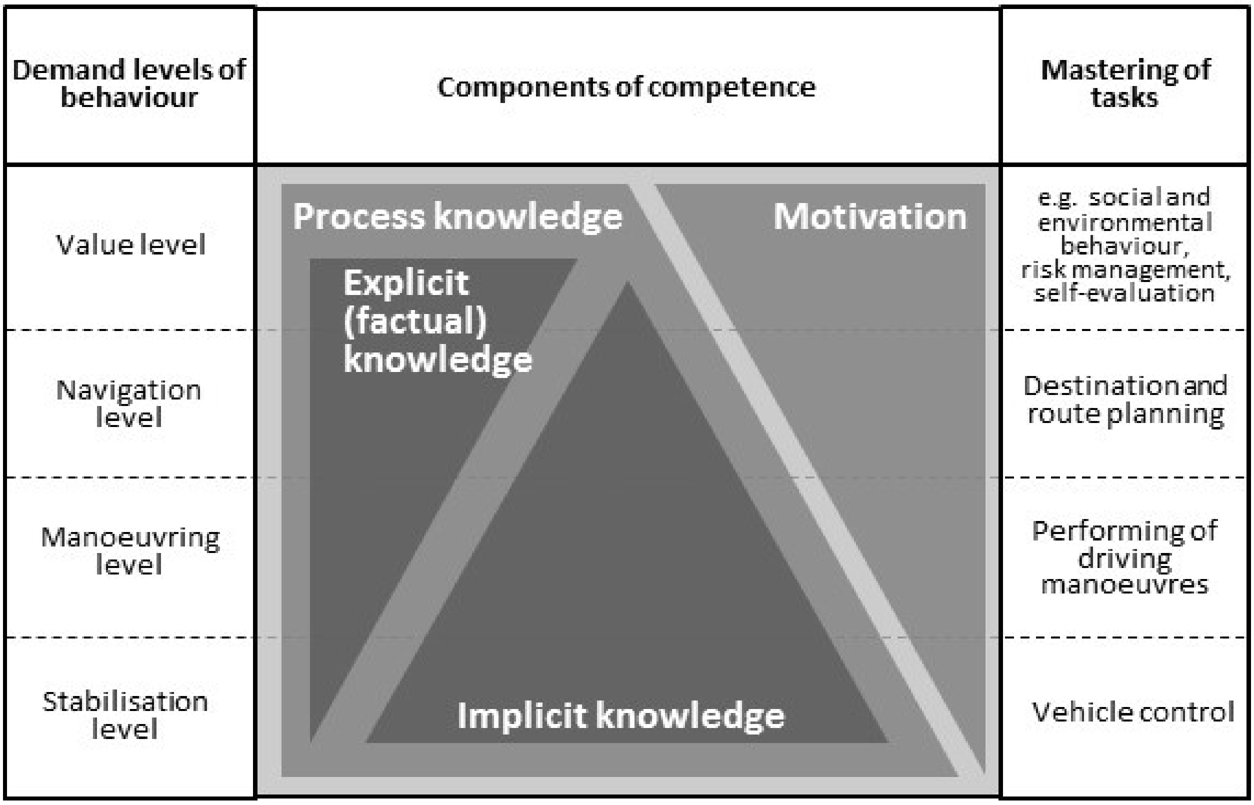

Fig. 3: Structure model of the content-based demand levels and psychological components of driving competence (from STURZBECHER & WEIßE, 2011)

It must here be pointed out that, although the different forms of knowledge specified in the aforementioned model of driving competence suggest a certain order of competence acquisition, they do not represent stages of acquisition or competence levels in the narrower sense. Stages of acquisition comprise a hierarchical system of competence levels within a domain; the systematics for description of the acquisition stage, however, may vary depending on the particular domain concerned. The descriptions build upon the assumption that a person who has reached higher stages of acquisition will also reliably master the demands of lower stages (KLIEME et al., 2007). Each stage of acquisition is defined by way of cognitive processes and actions at a level which a person has acquired or possesses at the stage concerned, but not at lower stages. Each stage of acquisition can thus be assigned characteristic tasks with different degrees of difficulty; a learner with the corresponding level of competence should then be able to solve these tasks reliably. Such possibilities for assignment would be extremely desirable for the elaboration of tasks for the driving test, but seem hardly attainable in the context of driver licensing, because driving instruction in real traffic – in contrast to school education with its predominantly graduated process of competence acquisition – demands the parallel acquisition of competence components such as traffic-specific knowledge, psychomotor skills and cognitive abilities relating to hazard perception. Furthermore, it can be assumed that, even with regard to generally familiar traffic situations (e.g. overtaking), special situation-related circumstances could give rise to new forms of action demand, such that partial competences (e.g. maintaining a certain driving line) which are already fully mastered in other situations may need to be improved or acquired anew. To describe such learning processes and progress during driver training, GRATTENTHALER, KRÜGER and SCHOCH (2009) propose the following spiral model of (driving) competence acquisition (see Fig. 4).

The spiral model presented below combines notions relating to content structure components and the acquisition mechanisms and levels of driving competence. At the same time, it illustrates why there can be no static model for the levels of driving competence, and that it initially only seems possible to outline a minimum competence level for the solo driving of a motor vehicle in real traffic (see Chapter 3). The model shows that the authors apparently assume the same content-based demand and action levels which are suggested in the present report (see above); however, they use the broader term "attitude level" in place of the narrower concept of a “value level” preferred by STURZBECHER and WEIßE (2011).

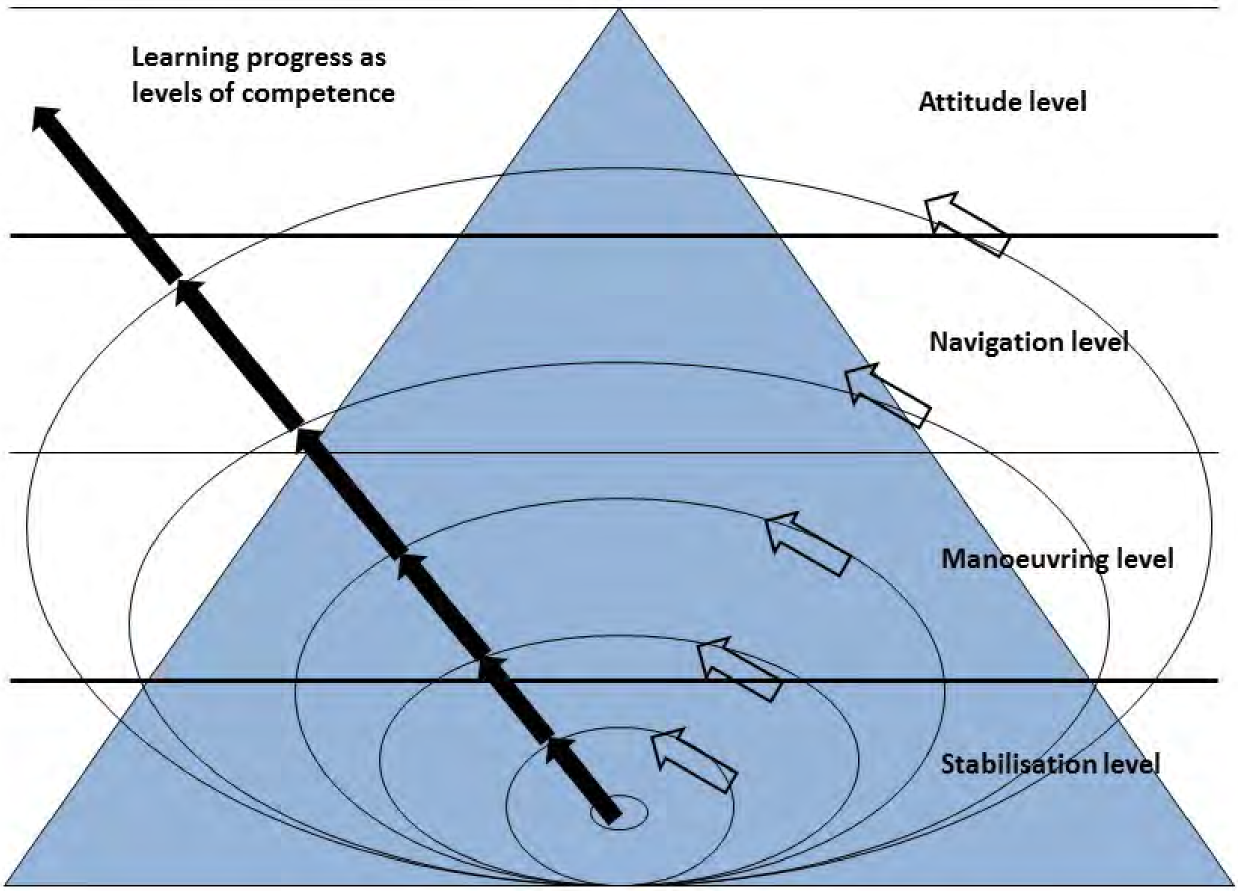

Fig. 4: Competence acquisition model depicting learning progress (from GRATTENTHALER, KRÜGER & SCHOCH, 2009, p. 89)

2.3 Training and test standards

Chapter 1 sketched the necessary steps for transition from the current situation of input control to a future system of output control in novice driver preparation. This was followed – as presented in the previous sections – by the elaboration of competence models, which, on a general level, describe the training and test contents (structure model) and define the levels of partial driving competence to be achieved (level or process model), as a basis for the elaboration of training standards and associated test standards.

What is to be understood by "training standards"? Taking the perspective of novice driver preparation as a whole, training standards must meet the following requirements:

– They must specify binding training objectives for the overall system of novice driver preparation and thereby establish references between driver training and licence testing.

– On the basis of the training objectives, they must determine the essential driving and traffic competences which driving licence applicants must have acquired within the framework of novice driver preparation. In this way, they define the social training mandate which is to be fulfilled by the institution of "novice driver preparation".

– They must describe the minimum levels of driving competence which novice drivers should have acquired at the transitions between individual phases of novice driver preparation. These levels must be described in adequately concrete terms to permit translation into driving tasks and assessment within the framework of driving tests.

To aid the development of training standards, KLIEME et al. (2007) formulated seven quality traits which characterise good training standards – also for the field of novice driver preparation:

(1) Subject specificity (i.e. they refer closely to particular learning contents)

(2) Focus (i.e. they concentrate on core areas of learning, rather than attempting to cover the full spectrum of the field with all its side branches)

(3) Cumulativity (i.e. they refer to the competences which must have been acquired up to a certain point of time within the learning process; they thus target cumulative, systematically integrated learning)
(4) Binding applicability (i.e. they represent minimum competence demands which are expected of all learners)
(5) Differentiation (i.e. they describe not only the minimum competence level, but also further higher levels, and thus illustrate possible paths of competence development)
(6) Comprehensibility (i.e. they are formulated clearly, concisely and in an understandable manner)
(7) Feasibility (i.e. they represent a challenge for the learner, but can nevertheless be attained with realistic effort).

The demand that training standards should specify different levels of competence already indicates that it is possible to formulate different forms of standards: Following on from the classification of school education standards in KLIEME et al. (2007), "minimum standards" in the context of novice driver preparation refers to the basic competences which all driving licence applicants should possess at the time of the practical driving test in order to be permitted to drive a motor vehicle solo in public road traffic. "Regular standards", on the other hand, describe a typical level of competence which is achieved by an average learner (but consequently not by all), while "maximum standards" represent the highest level which is generally beyond the potential achievement of all but a few learners. To safeguard road safety in public traffic, it is imperative to specify minimum standards, as they focus the attention of driving instructors and driving test examiners on any safety-relevant performance weaknesses displayed by novice drivers. This does not exclude the possibility of defining higher demands in the sense of regular or maximum standards as the objectives for more advanced learning processes (e.g. for accompanied driving); in fact, this is even indispensable within the framework of a driver training curriculum. Compared to other standards, however, the description of minimum standards is of decisive importance not simply for road safety, but also for the structuring of driver training, quality assurance instruments and the demands to be met in a driving test, and is thus urgently necessary.[15]

Viewed overall, training standards for novice driver preparation hold a key position within the entirety of all efforts to improve road safety: They define the minimum scope of competences which novice drivers must have acquired up to a certain point of time in their driving career, and can thus serve all persons and organisations involved in novice driver preparation as a common goal description and quality specification for the teaching/learning processes for which they are each individually responsible. The standards are at the same time an important tool for quality assurance and quality development. By focussing the training curricula and test specifications on essential aspects, finally, they provide guidance for both learners and instructors, and are able to contribute significantly to training equality and the achievement orientation of the training system.

On the basis of the above descriptions of a general system framework from the perspectives of competence theory and traffic-related pedagogical didactics, it should now be possible to derive specific, competence-oriented demand standards and assessment criteria in the form of test standards for an optimised practical driving test in the following Chapters 3 to 5. The determination and scientific founding of these specific test standards – and in the narrower sense of the test tasks for driving licence testing – is naturally only one very limited aspect of the process to elaborate training standards for novice driver preparation; in a subsequent step, these standards must be cross-referenced to driver training and anchored – alongside the supplementary content provided by training standards – in the corresponding curricula.

2.4 Summary

The theoretical and (optimised) practical driving tests are two different forms of testing within a more comprehensive methodical concept for the verification of driving and traffic competence. Within the framework of this concept, and possibly in combination with a future, yet to be developed test of competences relating to traffic perception and hazard avoidance (see STURZBECHER and WEIßE, 2011), they should ideally complement each other with regard to test contents and the assessed components of competence, and should in doing so each compensate the methodical deficits and limitations of the other form of testing. This can only be achieved by way of corresponding

[15] In the school education sector, too, only "regular standards" have been formulated to date (MUSZINSKI et al., 2009). One reason is seen in the danger that the formulation of minimum standards without robust, specific level models and (time-consuming) task-related empirical testing or validation could lead to the over- or underchallenging of pupils (KMK, 2005a).

competence theory models, with which it is possible to specify and describe the content-related facets of driving competence which are to be assessed by each form of testing, and the acquisition level which is required in each case. Such models were introduced in the present chapter.

The (optimised) practical driving test offers unique and irreplaceable opportunities for the assessment of driving and traffic competence in a practice-relevant environment: During a driving test in real traffic, the candidate must continuously observe and assess changing traffic situations with regard to their further development and hazard potential; irrespective of a certain action pressure, he must then react appropriately and with the due foresight. The appropriateness of the candidate's reactive and anticipatory behaviour can thus be judged under realistic conditions by the driving test examiner. Last but not least, the awareness of actual hazards and possible loss of control in real traffic exposes the candidate to specific test conditions which – compared to testing in a driving simulator – seem to render the driving test indispensable as a source of ecologically valid proof for the level of driving competence attained within the framework of novice driver preparation. To be able to exploit this assessment potential in a methodically sound manner, it is necessary to elaborate training and test standards. The overarching demands placed on such standards have also been examined in the present chapter.

The next step is now to establish and expand a methodical and content-referenced foundation for the aforementioned test standards, in other words to develop a psychometric model, measuring concept and measuring method from the driving competence model. In line with the previously described theoretical considerations, according to which competences must always be operationalised with reference to (demand) situations, it is first important to determine the variables which are to be assessed in the practical driving test as indicators of driving competence, and the demand situations in road traffic in which they can be observed reliably and validly. On this basis, driving tasks must be described and combined with observation requirements and assessment criteria for the driving test examiner. For these challenges to be met successfully in the following Chapter 3, there are two aspects which must both be taken into account: The instrumental-methodical aspect refers to the character of the method used for measurement, systematic driving behaviour observation and the ensuing demands in respect of process design. The technical, content-related aspect results from the action or demand domain "road traffic" and means that the test standards must be determined through analysis of the demands of real traffic as they relate to action theory. In doing so, due consideration must be given to the existing scientific traditions and research results, as well as to progressive developments in national and international test practice.

3 Contextual and methodical design of an optimised practical driving test

3.1 Demands placed on work samples and systematic behaviour observations

From the methodical perspective, the (optimised) practical driving test can be viewed as a process-oriented, competence-referenced diagnostic work sample, through which the relevant components of practical driving competence are determined and assessed by way of "systematic behaviour observation" (STURZBECHER, BÖNNINGER & RÜDEL, 2010): During a test drive, the test candidate is faced with driving tasks representing the (minimum) requirements for participation in motorised road traffic, while the driving test examiner observes the handling of these demands systematically on the basis of defined observation categories and assesses the attained level of driving competence by way of an adaptive test strategy. The test procedure must satisfy the usual quality criteria defined in test psychology, namely objectivity, reliability and validity (see Chapter 5); for methodical improvement of this procedure, it is necessary to consider the quality demands which are placed on work samples and systematic behaviour observations in general, and on their use within the framework of adaptive competence testing in particular. The general test psychology principles associated with these questions are thus to be outlined first in the following, before determination and founding of the special methodical architecture of the practical driving test in subsequent chapter sections.

Work samples serving competence diagnosis and principles of their design

Work samples are a common means to judge performance in the field of personnel assessment. The abilities of the candidate are assessed by way of the behaviour displayed in the processing of standardised tasks deemed to be representative for the subject area ("domain") concerned (SCHULER & FUNKE, 1995). According to KANNING, a work sample serves to "simulate important elements of occupational activity and subjects the candidate's behaviour and work results in these situations to systematic observation. A work sample possesses the greatest individual validity of all personnel assessment methods" (2004, p. 425). Successful completion of a work sample requires a certain degree of domain-specific competence; consequently, work samples should only be used if the persons involved already possess basic (professional) skills (ibid.).

Where a work sample is to be designed as a method of testing, there are various preparations to be made. It is first necessary, for example, to perform a demand analysis for the domain concerned, in order to identify its central work tasks and possible solution strategies. To this end, the overall process of task handling is divided into its constituent action steps. There are various ways to achieve this. One particularly suitable possibility, according to KANNING (2004), is the "Critical Incident Technique" developed by FLANAGAN (1954). "Critical incidents" are here understood as situative tasks (in the sense of problems), the mastering of which is decisive for the success (aptitude) or lack of success (lacking aptitude) of the person performing the task. To be able to use the method, it is necessary to identify (key) situations in which the behaviour or personal traits which represent the aptitude or lack of aptitude to complete tasks successfully can be distinguished with maximum discrimination; this is achieved through observations, expert interviews and analyses of objective data. The identification of appropriate key situations is followed by determination of the forms of behaviour or personal traits which lead to successful or unsuccessful handling of the task. In addition, scales must be constructed with meaningful graduations to describe the observed behaviour. Documentation of the test procedure and of proper observance of the specified methodical standards, finally, requires the development of observation and evaluation sheets with which it is possible to assess the extent to which the demands are satisfied.

Once all these preparations have been completed, the second step is to design the work sample in the narrower sense. It is here important to ensure that the selected tasks mirror the reality of the (work) situations as closely as possible. Whichever criteria are applied when selecting and designing the test situations, however, it will never be possible to achieve full correspondence between the demands of the work sample and those of working reality. On the other hand, such correspondence is not actually necessary: It is more important that the test procedure should cover all central and performance-relevant aspects of typical (occupational) demand situations (or the results of the demand analysis), rather than attempting to reflect every detail of daily (work) activity. Especially in the case of complex work samples, as required for the testing of multi-faceted competences in the field of vocational training, for example, it is necessary to elaborate guidelines in which the phases of testing, the time specifications for individual phases, and the criteria and implementation procedures for the assessment of performance within the individual phases are defined (BÄHR & WEIBERT, 2010).

The aforementioned general demands relating to methodically appropriate work sample design are naturally also applicable in the case of the practical driving test, as a special and relatively complex form of work sample. The elaboration of a test method must thus proceed in the manner described.

Systematic behaviour observation as a methodical instrument for competence assessment by way of work samples

In the scientific context, observations are understood as examination methods serving the purpose-oriented and methodically controlled perception of an entity under analysis, for example an object, event or process (HÄCKER & STAPF, 2004). In accordance with a defined observation procedure, selected aspects of an observation subject can be placed at the focus of attention, observed, recorded and evaluated in a variety of ways. Where the subject of the observation is (human) behaviour, it is customary to speak of "behaviour observation"; this is the fundamental method used in education psychology diagnosis (INGENKAMP & LISSMANN, 2008). Behaviour observation permits the assessment and evaluation of directly visible social and performance behaviour; according to BORTZ and DÖRING (2006), it is recommended, in particular, where verbal self-portrayal on the part of the assessment candidate can be expected to result in conscious or unconscious falsification of the relevant behaviour.[16] Against this background, the method of behaviour observation appears suitable also for competence assessment, as a person's purpose-oriented behaviour and actions can be assumed to indicate corresponding action competence (KAUFHOLD, 2006).

"Systematic behaviour observation"[17] is a special form of behaviour observation characterised by binding specifications regarding the subject of the observation, the observation situation, the observation environment and observation categories, as well as common definitions in respect of realisation of the observation and assessment of the observed behaviour. These specifications and definitions are described in an observation system, together with the "language" to be used for documentation (see Chapter 4). They provide for control of the observation situation, which is a necessary prerequisite for the scientifically based objective verification of examination questions on the basis of observation data (FIEGUTH, 1977). Such observation systems "for the coordination and systematisation of individual acts of observation and for simultaneous recording of the resultant information" (GRÜMER, 1974, p. 40) delimit the behaviour to be observed in the interest of methodical quality. Their uniform and correct application must nevertheless be learned and practised by the user within the framework of observer training. Observers must possess fundamental domain knowledge in the field under observation, to ensure that they can correctly interpret their observations and reach a proper assessment. Finally, systematic behaviour observation should control the manner in which observers actually work with the observation system, and the classic quality criteria defined in test psychology (objectivity, reliability and validity) must also be evaluated (FIEGUTH, 1977; see Chapter 5).

To safeguard the psychometric quality of the method, systematic behaviour observation must be structured on the basis of adequately standardised demands. It is furthermore necessary to define observation categories, which KANNING (2004) describes as situation-independent classes of observation subject. The observation categories should be limited to a meaningful number, and they should cover the whole spectrum of the behaviour to be observed as exhaustively and disjunctly as possible (FISSENI, 2004). The function of observation categories is to guide human perception in the purpose-oriented search for information to support a subsequent assessment and decision (orientation and structuring function); this also serves to relieve the observer. Depending on the question to be answered, the observation categories may be more or less abstract in nature (KANNING, 2004). Generally speaking, the instrumental quality of an observation method is proportional to the precision with which the behaviour to be observed and its observation categories are defined (FISSENI, 2004). Last but not least, the proper realisation of systematic behaviour observation requires specific assessment criteria as a basis for interpretation of the observations.

On the basis of a comparative examination of published research on observation methodology, KÖTTER and NORDMANN (1987) elaborated a planning and control process which is intended to

[16] Such falsification could be assumed if a driving licence were to be granted on the basis of applicants' self-assessment of their driving competence.

[17] The terms "standardised observation" and "structured observation" are also used with the same meaning in the literature of the field.

secure the methodical quality of observation results where – as in the case of the practical driving test, which is conducted many thousands of times each year – comparable results are the objective for extensive observation series. This planning and control process comprises three steps, which together describe the proper designing and optimisation of an observation method. If these three steps are applied to the (optimised) practical driving test, as suggested by STURZBECHER (2010), this identifies the following methodical and content-related challenges, which are to be overcome in the course of further development of the test:

(1) Firstly, adequate development of an optimised practical driving test requires the elaboration of a concept to show how the specification and arrangement of a selection of different observation or traffic situations with sufficiently standardised demands and particular relevance for road safety would serve to structure and control the test procedure in such a manner, that it yields meaningful (i.e. reliable and valid) information on the driving competence of the test candidate. The structuring and control concept for the optimised practical driving test can be divided into two parts, which are described and explained more closely in the following: The "driving task concept" attributable to McKNIGHT and ADAMS (1970a) facilitates the content-based definition and "portioning" of test demands by way of situation-related "driving tasks" and the situation-independent components of driving competence which can be observed during the performance of those tasks, while the "circular model of an adaptive test strategy" presented by STURZBECHER, BIEDINGER et al. (2010) governs arrangement of the driving tasks and the actions of the driving test examiner when conducting the test.

(2) Secondly, optimisation of the practical driving test requires elaboration of a concept for adequate assessment of the test performance and driving competence demonstrated in the observation situations, and for documentation of the data acquired. In this context, it seems expedient to turn to the aforementioned "work sample" (EBBINGHAUS & SCHMIDT, 1999) and "systematic behaviour observation" (INGENKAMP & LISSMANN, 2008; FISSENI, 2004; KANNING, 2004) concepts, which are already well developed and in widespread use in educational and psychological diagnosis. These concepts support explanation of how observation methods are to focus on the relevant subject aspects and permit professionally adequate assessment of the test performance. At the same time, the concept of systematic behaviour observation represents the methodical starting point for determination of the demands relating to documentation of the observation and assessment data. In connection with the optimised practical driving test, these demands are to be implemented in an electronic test report (see Chapter 4).

(3) Thirdly, a discriminating evaluation methodology is required to permit high-quality psychometric observation and test results to be gained from the optimised practical driving test. This refers to the elaboration and uniform application of appropriately content-referenced and methodically sound assessment and decision criteria. Here, too, methodical concepts relating to work samples and systematic behaviour observation are valuable contributions to optimisation of the practical driving test: They serve description of the necessary transformation of observed behaviour into event- and competence-oriented assessments, as well as the interpretation and concentration of these assessments leading to an unambiguous test decision.

The planning and control process which was elaborated by KÖTTER and NORDMANN (1987) as a basis for the methodically reflected development and improvement of observation methods – as applied to the case of the practical driving test in the above explanations – thus represents the fundamental action brief for the pending optimisation of the test. Within the framework of this optimisation, the demand and observation standards, and likewise the assessment and decision criteria of the practical driving test, are to be provided with a scientific foundation and further developed in appropriate manner.

Assessment and decision criteria as means for objective test realisation

Assessment criteria are specifications of how displayed performance is to be assessed. In the case of tests addressing the performance of individual candidates, they serve to limit the scope of judgement granted to the examiner and raise the test objectivity. To this end, different and maximally disjunct levels of test performance are defined, each of which must – in the context of systematic behaviour observation – relate to abstracted or exemplary observable aspects of behaviour. It must be possible for each element of recorded performance to be assigned unambiguously to the various assessment or performance levels, so as to safeguard a high degree of evaluation objectivity.

On the basis of the defined assessment levels, decision criteria can be formulated to specify more or less precisely how certain assessments are to translate into particular test decisions, depending on the contexts and frequency of those assessments. The prescribed decision alternatives, as well as the specified decision relevance of individual assessments, are guided by overarching objectives and value standards, as well as the anticipated, desired and undesirable consequences of the test decision. The implications of the test decision determine which forms of misjudgement can be tolerated to which extent (BORTZ & DÖRING, 2006): Where the false classification of unsuitable candidates as "suitable" (so-called "alpha risk") is deemed to be particularly dangerous on account of the possible consequences, the decision criteria must be stricter; this, however, will almost always entail a larger proportion of suitable candidates being wrongly assessed as "unsuitable" (so-called "beta risk").

With regard to the test decisions, a distinction is made between selection decisions (in the sense of assignment to one of the groups "suitable/passed" or "unsuitable/failed") and ranking or classification decisions, where candidates are assigned judgement alternatives associated with particular achievements or limitations. The difference between ranking and classification decisions lies in the fact that a ranking decision – as in the case of selection – is based on a single aggregated value, whereas a classification decision is derived from multivariate assessment constellations ("profiles") (WIECZERKOWSKI & ZUR OEVESTE, 1978).

Assessment and decision criteria must guarantee content validity and sound foundations in order to permit methodically robust decisions. The necessary foundation can be derived from subject-specific content analysis in respect of the actions to be assessed, from detailed observations by suitable persons or from expert opinions.[18] Furthermore, assessment and decision criteria must be formulated in accordance with a reference standard or benchmark. In this context, it is possible to distinguish individual (ipsative), social (collective) and practical (criterion-oriented) reference standards (PARADIES, WESTER & GREVING, 2005; RHEINBERG, 2008): While individual reference norms relate to the development progress of the person under assessment, social reference standards establish a correlation to the distribution of the expected performance in a representative social sample (so-called "reference population"). In the case of criterion-oriented reference standards, the observed performance is compared with previously defined (minimum) performance standards. Assessment on the basis of individual or social reference standards is always to be rejected in connection with concluding tests where the test decision concerns the granting of entitlements (which also applies to the case of the practical driving test).[19]

When elaborating assessment criteria relating to a practical reference standard, it is necessary to define the aspects of behaviour which are relevant for a particular performance assessment criterion. Both displayed behaviour and the failure to display a given aspect of behaviour may be taken into account. Besides a dichotomous assessment ("behaviour displayed"/"behaviour not displayed"), it is also possible to differentiate the quality of the observed performance – as described above – by specifying an intermediate level within the scope of the assessment criteria. Such levels can be represented by way of a numerical (e.g. assessment of ability from "0" or "none" to "10" "fully developed")[20], verbal (e.g. school-style grades from "very good" to "insufficient"), graphic (e.g. in the form of a thermometer) or symbolic scale (e.g. the symbols "+ plus" and "– minus") (ROTH & HOLLING, 1999). If some or all of the scale levels are assigned behaviour descriptions (possibly enriched with additional behaviour examples – so-called "anchor examples"), then we can speak of a behaviourally anchored rating scale (so-called "Thurstone scale"). In many cases, it is considered

[18] In the assessment of vocational aptitude, demand profiles are created for certain work tasks on the basis of activity or job analyses; a demand profile here represents a summary of those personal traits which are critical for success in the given work (NERDINGER, BLICKLE & SCHAPER, 2011). To this end, persons who successfully practise the activities associated with a particular job (or are assumed to hold the required qualification) are surveyed to identify their personal traits and observed systematically during the performance of certain activities. Occasionally, the actions of persons who do not possess the required qualifications are also analysed (i.e. those who do not hold the job in question or else seem unsuitable for this job).

[19] BÄHR and WEIBERT (2010) illustrate this with a particularly vivid example: "Who would gladly have their car brakes repaired by a mechanic who has perhaps made the best of his limited talents during his training, but objectively has never mastered the repairing of brakes? Society must be able to rely on an objective assessment aligned to the demands of the later field of work, i.e. a criterion." (p. 154).

[20] All tests organised by the Chambers of Commerce and Industry (IHK) and Chambers of Craft Trades (HWK) are assessed on the basis of a criterion-oriented 100-point assessment scale in accordance with the model examination regulations elaborated by the Federal Institute for Vocational Education and Training (BIBB, 2007). To better illustrate the result to the candidate, six score ranges are defined to correspond with the six-level school grading system. The assessment of test performance is thus differentiated far beyond a mere "passed" or "failed" decision; the candidate's performance resources are classified precisely and in readily comprehensible fashion within the given spectrum.

sufficient to provide a precise linguistic definition of the two poles of a scale, and on this basis to apply the assumption of scale linearity. This then constitutes a so-called “number scale”.

At the end of a single-candidate test, the examiner must finally derive a test decision from a series of individual assessments on the basis of specified decision criteria. There are different ways to reach such a decision. One possibility is simple totalling of all individual assessments; alternatively, certain assessments could be weighted to take into account the particular relevance of specific assessment contents. For the determination of the test result, two complementary methods can be distinguished, although both must effectively lead to the same result: With the “deduction method”, the candidate begins the test with a certain (“full”) number of points; errors lead to points being deducted in accordance with the assessment criteria and the specified assessment key. With the “cumulative method”, by contrast, the candidate starts out with a score of zero and receives points (or error points) for each correct (or incorrect) aspect of observed behaviour. Further deviating or additional procedures are possible alongside these two basic methods. It could be specified, for example, that certain minimum criteria must be satisfied to pass the overall test, or else that certain exclusion criteria must not apply. As a further alternative, it could also be possible to compensate poor performance in one area with good or very good performance in another.

Generally speaking, it is necessary to provide a professionally founded specification of the minimum overall performance which is necessary to pass the test. This often takes the form of a so-called “cut-off value”, which indicates either the minimum number of points which must be achieved or the maximum number of errors which must not be exceeded. The distinction between a “still adequate level of competence” and a “no longer adequate level of competence” is nevertheless one of the most difficult test decisions. The degree of interpretation objectivity which can be attained is dependent on the extent to which exact and binding decision rules are (or can be) defined: Clear rules promote interpretation objectivity, whereas scope of judgement will generally impair the objectivity.

Adaptive control concepts as means to secure the validity of competence tests

Single-candidate tests such as driving licence tests or the final tests at the end of vocational training are to be viewed as procedures by which the required competences are to be demonstrated, measured and assessed (BEINER, 1982). Corresponding test certificates – e.g. a driving licence or vocational qualification – are then issued on the basis of the demonstrated competence; they lend utility value to competences and represent the “hard currency” of educational institutions (SEVERING, 2011). Generally, the recorded level of performance in educational institutions is determined to a large extent by the corresponding test demands (BÄHR & WEIBERT, 2010), wherein the demands are defined by way of the set tasks, and the quality of their fulfilment is judged in accordance with assessment and decision criteria (see above).

In educational institutions such as “novice driver preparation”, which integrate various forms of teaching/learning and testing (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014), tests possess several functions (ROST, 2010): They serve on the one hand to determine whether test candidates have acquired certain competences, and whether they are in a position to apply those competences (“learning assessment function”). Within this process, the test contents and methods of performance assessment influence the contents, didactic methods and structures of the training (“control function” or “backwash effect”) and exert pressure on the learner to realise learning activities (“disciplinary function”). At the same time, the passing of a test is associated with the granting of certain privileges (“entitlement function”) which are denied to unsuccessful candidates (“selection function”). Tests are furthermore able to provide information to candidates on the level of performance achieved (“feedback function”).

Competence tests usually base their overall assessment of competence on several individual measurements or test tasks, which can each be viewed as valid indicators for the components of competence concerned (BÜHNER, 2011), i.e. the competence components are “operationalised” by way of the test tasks. During the period of the test, the examiner sets these tasks for the test candidate in a certain order, in accordance with his chosen test strategy; in this way, the examiner controls the course of the test. With regard to the control concept, the examiner may apply either a fully standardised linear test strategy, or else an adaptive test strategy: In the case of a linear test strategy, the tasks are set in a strictly defined order. Where, by contrast, the course of the test is adapted continuously to the performance displayed by the candidate or to changing situative conditions, we can speak of an adaptive test strategy. The objective of an adaptive test method lies in validation and/or refinement of the assessment of a test candidate's performance level. Even with an adaptive test strategy, the fundamental methodical procedures are already predefined; their actual implementation, however, can be varied flexibly in accordance with both the data already gathered and data still to be acquired (RETTIG & HORNKE, 2000).[21]

Oral tests are a good example for adaptive testing (FREY, 2008); they can be seen as the oldest (EBBINGHAUS & SCHMIDT, 1999) and in the (school) education system as the most widespread method of performance monitoring (INGENKAMP & LISSMANN, 2008). Such tests should begin with simple questions (ROLOFF, 2002), as an initial positive experience promotes motivation and increased self-assurance on the part of the test candidate. JÜRGENS and SACHER (2008) identify five process steps, in the sense of a general framework for test realisation, which the examiner must follow – and to a certain extent implement simultaneously – during the course of an oral test:

1. Listen and determine the objective correctness of the answers given by the candidate
2. Judge whether the level of performance displayed by the candidate corresponds to the requirements, or whether the candidate is over- or underchallenged
3. Plan new questions – already during the candidate's answer to the present question – which take into account the level of performance displayed so far
4. Receive and interpret relational messages from the candidate, so as to be able to design the test accordingly
5. Communicate own relational messages and – in order to avoid misunderstandings or undesired reactions on the part of the candidate – check that they are received and interpreted correctly.

One meaningful methodical compromise between

- a linear control concept which is fully standardised with regard to the constituent tasks and their order within the test, as would be desirable for reasons of objectivity, on the one hand, and
- an intuitive, adaptive control concept which is characterised by the determination of new tasks during the actual course of the test, and thus facilitates situation-oriented test control, on the other hand,

would be a partially standardised (or criterion-driven) adaptive control concept. Under such a concept, the test items are selected from a stipulated catalogue of tasks which have previously been validated in respect of the competences to be tested and can be adapted appropriately to the current situation (JÜRGENS & SACHER, 2008). A specimen solution should exist for each such task (ROLOFF, 2002). To aid proper interpretation of the task solutions during the test, it is necessary to provide implicitly or – with a view to test quality – preferably explicitly formulated assessment criteria as a basis for planning of the further course of the test and for the final test decision. The availability of empirically founded task catalogues and assessment criteria raises the objectivity, reliability and validity of tests of competence.

It must be mentioned at this point that the model of an optimised practical driving test outlined by STURZBECHER, BÖNNINGER and RÜDEL (2010) already represents such a methodical compromise in the sense of a partially standardised, criterion-driven test control process, and that the concept of an "adaptive test strategy for the practical driving test" (see below) – in the same way as the sequence model described by SACHER (2008) for adaptive oral tests (see above) – offers a framework standard for test realisation.

The following sections now sketch the essential methodical foundations for an optimised practical driving test. These foundations are to be evaluated with regard to the extent to which they satisfy the general demands placed on professional work samples, systematic behaviour observation, adequate assessment and decision criteria, and not least adaptive test control concepts. It is furthermore to be determined whether the various steps of the aforementioned planning and control process (KÖTTER & NORDMANN, 1987) have been taken into due account in the optimisation measures implemented to date.

3.2 Driving tasks as situation-related demand standards

As already described in detail above, methodically demanding work samples require that the work process under assessment be broken down into its essential constituent steps, which can then be set as tasks for the test candidate. Consequently, appropriate segmenting of the actions to be performed by a driver (i.e. technical preparation and

[21] Adaptive competence assessment methods enjoy a long tradition in the field of performance diagnosis under the designation "answer-dependent tests": Already at the beginning of the 20th century, BINET and SIMON used a test procedure "tailored" to the individual candidate to measure the intelligence of children (WEISS, 1985). Since the mid-1990s, computer-based adaptive test methods have gained widespread popularity (above all in the USA) (FREY, 2008; STEINER, 2009).

completion of a drive, alongside the driving process itself) into individual tasks (including driving tasks and basic driving manoeuvres) was a focal topic for methodical optimisation from the very beginning of work to develop scientific foundations for the practical driving test[22]. Respect is due to HAMPEL (1977) for having recognised the potential value of the driving task concept[23] developed by McKNIGHT and ADAMS (1970a, 1970b) for the elaboration of structuring demand standards for the practical driving test in Germany; he also embraced proposals for a situative understanding of driving tasks and for a driving task catalogue, which had been presented, for example, by JENSCH, SPOERER and UTZELMANN (1978) within the framework of their "traffic behaviour theory" as a means to better integrate hazard aspects into novice driver preparation. As the outcome of further research and development tasks (HAMPEL & KÜPPERS, 1982), a catalogue of driving tasks gradually took shape over the period up to 1987 and essentially remains applicable today as a foundation for test demands (HAMPEL & STURZBECHER, 2010). Both the basic theoretical positions of this driving task concept and the methodical task implementation were taken up once more by STURZBECHER, BÖNNINGER and RÜDEL (2010) in connection with the pending fundamental optimisation of the current practical driving test. The authors understand "driving tasks" to mean prototypical ("exemplary") classes of similar traffic situations to be mastered[24] and present suggestions for modernisation and restructuring of the driving tasks set in the practical driving test.

Before we can apply the theoretical and methodical thoughts of McKNIGHT and ADAMS (1970a; 1970b), McKNIGHT and HUNDT (1971a; 1971b), HAMPEL (1977) and STURZBECHER, BÖNNINGER and RÜDEL (2010) as a foundation for further optimisation of the practical driving test, however, it is necessary to determine whether and to what extent the methods used by McKNIGHT and ADAMS (1970b) in the elaboration of driving tasks and the correspondingly founded specification of driving tasks for the German practical driving test by HAMPEL and KÜPPERS (1982) meet today's scientific standards for the construction of work samples and systematic behaviour observation. In the following, it is explained why the former studies, in particular, provide robust theoretical and methodical starting points for the imminent reform of the practical driving test in Germany; the elaboration and contents of the present reform proposal for an optimised catalogue of driving tasks can then be presented on this basis.

Elaboration of the driving (driver) task concept at the beginning of the 1970s

The first pedagogically oriented studies[25] relating to the tasks (including driving tasks)[26] to be mastered by a driver were conducted by McKNIGHT and ADAMS (1970a; 1970b). Their overarching objective was to define the necessary goals and contents of driving school instruction by way of a comprehensive and detailed empirical analysis of the demands of motorised road traffic – in the sense of a description of "good driving behaviour" – and on this basis to elaborate a training curriculum (including instruments to measure levels of achievement) for the acquisition and demonstration of driving competence.[27] Their basic assumption was that – as similarly in other areas of education – it was first necessary to specify concrete, hierarchically structured objectives for driver training and to monitor the attainment of these objec-

[22] In Germany, intensive studies of the processes of driving licence testing from the perspective of test psychology began essentially in the mid-1970s; the background, course and results of the developments of that time are described in detail by HAMPEL and STURZBECHER (2010).

[23] It is to be pointed out that system analyses pertaining to road traffic, the construct of traffic situations and the concept of driving tasks as a method for the segmenting of driving processes are not used exclusively in the context of driver licensing (v. BENDA, 1985). The description and classification of situative traffic demands in the form of driving tasks also plays an important role in the field of accident research. FASTENMEIER and GSTALTER (2003), for example, developed their "Situational Analysis of the Behavioural Requirements of Driving Tasks" (SAFE) within the project "Driver Behaviour and Human-Machine Interaction (FVM)", which was in turn a component of the research initiative "Intelligent Traffic and User-Oriented Technology" (INVENT).

[24] The similarity of these driving situations refers to their outward structures (traffic conditions, persons involved, actions), to the situation-related (test) demands which must be met by the driving licence applicant, and to the action sequences necessary to master the situation.

[25] Previous driving task analyses had generally limited themselves to investigation of the psychological correlations between sensory stimuli (e.g. seeing, hearing) and the driver's behaviour in respect of vehicle control.

[26] McKNIGHT and ADAMS (1970b) define "tasks" as "units of work to be performed", i.e. sequences of actions geared to the attainment of a particular goal. Their use of the term "driver's tasks" already indicates that this is understood to include tasks which cannot be deemed driving actions or "driving tasks" in the narrower sense, for example tasks associated with the preparations for actual driving (assembling of vehicle documents, securing of loads).

[27] The reports presented by McKNIGHT and ADAMS (1970a, 1970b) and McKNIGHT and HUNDT (1971a, 1971b) were results of research studies conducted by the Human Resources Research Organization on behalf of the National Highway Traffic Safety Administration (http://www.humrro.org/corpsite/). The safety relevance of professional "driver education" in high schools was at that time a subject of controversial debate in the USA; there was substantial public interest in questions relating to the effectiveness of such offers compared to the less cost-intensive alternative of lay training.

tives, as a prerequisite for subsequent consideration of whether attainment actually contributes to road safety. In this connection, the authors also point out that training objectives cannot be derived from accident records alone, in the same way that goal attainment cannot be read from accident statistics: Accidents are rare occurrences which are in most cases triggered by a combination of different causes (e.g. personality deficits, lack of driving experience, adverse weather or road conditions, incorrect behaviour which could not be compensated by other road users); even if a correlation can be established between driving instruction and accidents, it remains unclear which aspects of driver training are safety-relevant.

The first stage of the process defined by McKNIGHT and ADAMS (1970b) as a means to determine relevant training objectives was to break down the various actions to be performed by the driver and thereby to identify prototypical tasks; this stage comprised three successive steps:

1. "Task analysis": This theoretical-analytical step[28] served scientifically founded dissection of the overall activity "driving" into separate, more or less complex action sequences ("tasks"). This involved (a) a comprehensive search to identify those traffic situations which demand a behavioural reaction on the part of the driver, (b) determination of the correspondingly correct or appropriate behaviour, (c) subsequent structuring and concentration of the diverse aspects of behaviour to produce a smaller number of more manageable, complex action patterns ("driver's tasks"), and (d) systematic, theory-based determination and description of the elementary actions and action sequences ("subtasks") which make up such an action pattern.
2. "Criticality evaluation": This second, empirical step provided an assessment of the significance of the necessary vehicle control actions and action sequences for safe and efficient[29] driving.
3. "Development of task descriptions": On the basis of the described actions and criticality evaluation, meaningful (driving) task descriptions were elaborated and compiled into a task catalogue.

(re 1) The starting point for determination of (driving) tasks which can be taken as demands to be met by a novice driver was a detailed theoretical analysis of all behaviour-induced characteristics of the system (or domain) "road traffic", which comprises the components "driver", "own vehicle", "road used", "surrounding traffic" and "natural environment". By way of a systematic literature review covering over 600 traffic-related publications[30], and after eliminating redundancies in the data, a final list of approx. 1,000 such behaviour-relevant system characteristics was obtained.

A systematic analysis of all these characteristics, together with a number of combinations ("interacting characteristics")[31] selected on the basis of logical considerations, then yielded approx. 1,500 examples of essential or frequently arising behaviour with which drivers can react appropriately to the demands of typical traffic situations. To simplify handling of the revealed diversity of behaviour, the individual aspects were subsequently structured and organised into 45 more complex action patterns ("driver's tasks").[32] To this end, aspects of behaviour which referred either to the same objective or to the mastering of the same category of situation were grouped together under a single heading.

The next step was to divide the identified action sequences into so-called "Off-road behaviour" and "On-road behaviour". The elements in each of these groups were then further categorised on the basis of their overall action objective (e.g. overtaking) or else in accordance with temporal or spatial situation characteristics (e.g. driving at night). The group of "Off-road behaviour" was subdivided into the three categories "Pre-driving behaviour" (e.g.

[28] Empirical approaches to the determination of safety-relevant driving behaviour sequences, such as observations of driving behaviour or driver surveys, were deemed inefficient or useless research strategies by the authors, because observations, in their opinion, would always only reflect a small extract from the diversity of driving behaviour, and the quality of surveys would be impaired by memory deficits; furthermore, it would in both cases remain unclear whether recorded aspects of behaviour were performed adequately from the point of view of safety, as would be significant for the determination of learning objectives.

[29] McKNIGHT and ADAMS (1970b) took into account not only road safety objectives, but also, for example, socially relevant ecological objectives and selected individual aspects (e.g. technical condition of the vehicle, costs of driving), which were expected to be topics of professional driving instruction; peripheral activities (e.g. improving the vehicle's appearance, financing), on the other hand, were excluded.

[30] The literature survey included instructional texts used in driver training, accident statistics, "critical incident reports", engineering studies, studies from the field of behavioural research, work analyses and teaching films.

[31] The authors understand "interacting characteristics" as multiplicative combinations, where the interactions between individual situation characteristics call for behaviour which goes beyond that occasioned by a mere additive combination of the behaviour associated with each individual characteristic.

[32] The authors point out that clear-cut assignments were not always possible and that the purpose of the grouping was merely to enable pragmatic structuring and the simplified identification of required information; the determined task structure, for example, was not (yet) intended to reflect any inherent structure in driving behaviour.

planning, loading), "Maintenance and servicing" (e.g. routine maintenance, periodical technical inspections) and "Legal responsibilities" (e.g. behaviour following an accident, obligation to carry a driving licence and vehicle papers). The group of "On-road behaviour" was similarly split into three categories (the third of which was further divided into four subgroups):

- The category "Basic control" comprised tasks with no reference to a specific traffic situation, namely tasks serving vehicle operation and control of movement of the vehicle (e.g. starting the engine, pulling away, accelerating, stopping).
- The tasks assigned to the category "General driving" were similarly situation-independent, but in contrast to those in the first category need to be performed continuously and parallel to vehicle operation during driving (e.g. observation, navigation).
- The category "Situational behaviour", finally, referred to vehicle handling and manoeuvring in specific typical traffic situations, and was thus further divided into four subgroups in accordance with different characteristics of the road traffic system (see above): (1) "Traffic-induced behaviour" (e.g. parking, overtaking), (2) "Road-induced behaviour" (e.g. choice of driving lane, negotiation of bends), (3) "Environmentally induced behaviour" (driving in certain weather conditions, night driving) and (4) "Vehicle-induced behaviour" (e.g. towing, behaviour in case of breakdowns).

The fourth step of "task analysis", finally, was of special importance for the quality of the analysis results: While the first two steps had contributed to a – methodically original – heuristic strategy to identify traffic-relevant behaviour, and the third step could be viewed as a plausible structuring strategy, the concluding step constituted a systematic analysis of the identified action patterns or "tasks" in accordance with domain-specific theoretical considerations, with the aim of depicting all sub-tasks and individual actions necessary for proper and correct performance of a task not only in their entirety, but where possible also with quantitative standards (e.g. specification of a safe distance to be observed also as a definitive number of metres).[33] In this way, the individual elements of behaviour identified by the heuristic process could be verified and supplemented with reference to the pursued objective and validated overall with regard to their function as components of the "driver's tasks". This analysis step was designed with a very broad scope, so as to ensure that no safety-relevant actions could be overlooked. Consequently, almost 1,700 specific actions were described as being necessary to drive a motor vehicle. These actions were arranged in a hierarchy comprising tasks, subtasks and individual actions.

(re 2) Within the framework of the criticality evaluation, the described tasks, subtasks and individual actions were assigned a criticality index. This was seen as a means to support driving instructors in their structuring of the training and the prioritisation of learning objectives. The criticality evaluation was performed in the form of expert rankings[34], for which a total of 100 experts were recruited from the fields of driver education, licensing, traffic safety promotion and traffic law enforcement. The approx. 1.500[35] actions and action sequences to be evaluated were divided into 300 randomly drawn groups of 25 elements each, meaning that each element was included in five differently composed groups and was thus evaluated five times. Each expert subsequently received three separate envelopes by post, each of which contained (1) a group of 25 actions or action sequences to be evaluated, (2) precise descriptions of each element of behaviour (i.e. the result of the demand analysis), (3) instructions on the intended procedure of criticality evaluation, and (4) additional information to assist evaluation from literature reviews and over 1,000 accident analyses. The task for the experts was to judge the elements of a given group with regard to their criticality and to arrange them in corresponding order (from "1" = "most critical" to "25" = "least critical")[36]. As a result, a numerical

[33] To break the action sequence (i.e. driving task) "Overtaking" down into its constituent subtasks, the first step was to view the basic process elements (e.g. decision to overtake, preparation for overtaking, change of lane, passing the other vehicle, return to the original lane). Additionally, in a further step, variants of overtaking dependent on a particular traffic situation were analysed (e.g. situations subject to different traffic regulations). In this way, it was possible to build up an objective description of the whole driving task "Overtaking" and to define performance standards for its corresponding subtasks. The more or less complex action components contributing to performance of the overall action sequence (e.g. setting of indicators, use of mirrors) at the same time represent the behaviour to be observed and – above all in the context of a driving test – assessed. McKNIGHT and ADAMS (1970a, 1970b) identified 214 such actions merely for the process of overtaking.

[34] The relative merits of rating and ranking methods were investigated in advance by way of a pilot study; the results favoured the use of a ranking process. It was determined furthermore that each evaluator could only properly rank a maximum of 25 actions or elements of behaviour.

[35] A number of the total of 1,700 actions and action sequences were grouped together, as they were considered parts of an integral process and thus a similar level of criticality could be expected.

[36] The five action-specific ranking values were transformed into normalised scores (mean = 0; standard deviation = 10; -20 = "least critical"; +20 = "most critical"). Subsequently, these normalised scores were averaged to obtain a "criticality index"

criticality index could be assigned to each task, subtask or individual action[37] required to drive a motor vehicle (Fig. 5).

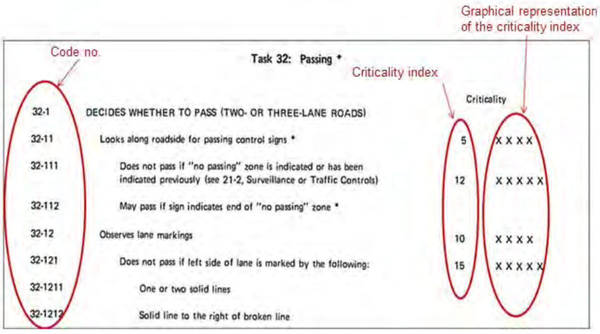

Fig. 5: Example of the evaluation results showing the criticality index for individual actions contributing to the driving task "overtaking" (McKNIGHT & ADAMS, 1970a)

It is to be noted that the judgements of criticality returned independently by the contributing experts displayed astounding similarity. Nevertheless, and despite the attempt to establish an objective context for the experts' evaluation by providing scientific (accident) statistics, the results of the evaluation are subject to a certain degree of subjectivity; even so, they are deemed sufficiently robust, given the intended purpose and the methodical care with which they were obtained.

(re 3) On the basis of the demand analysis, finally, comprehensive descriptions were elaborated for all 45 (driving) tasks determined (McKNIGHT & ADAMS, 1970a). These descriptions incorporated not only the results of the criticality evaluations for the identified subtasks and individual actions, but also a diversity of further research results. In most cases, the task descriptions could be supplemented with additional scientific information on

– typical driver performance ("performance information"),
– the limits of driver capabilities ("performance limits"),
– the criticality and significance of certain elements of behaviour and traffic situations, e.g. frequency of accidents at junctions ("criticality information"),
– perceptual, motor or cognitive processes during driving ("skills"), and
– the individual's action motivation or qualification, e.g. understanding of why certain actions must be performed ("knowledge").

It is to be pointed out that these (driving) task descriptions by McKNIGHT and ADAMS (1970a) did not yet specify any performance standards to be attained during training (or assessment criteria to be applied in the context of learner assessment or testing).

McKNIGHT and HUNDT (1971a) elaborated performance and assessment standards within the framework of a second phase of the further development of driver education in the USA, specifically in conjunction with the elaboration of performance-oriented learning objectives and instruments for the evaluation of learning achievement in the context of a training curriculum (RILEY & McBRIDE, 1974). The evaluation instruments relating to learner assessment[38] later offered starting points for scientifically founded optimisation of the practical driving test – also in Germany; for this reason, they are to be described in further detail at this point, together with the corresponding elaboration process.

For the elaboration of learning objectives for their training curriculum, and likewise of corresponding test contents for the evaluation instruments, McKNIGHT and HUNDT (1971b) referred directly to the demands and criticality evaluations obtained within the framework of the task analysis (see above): In connection with the criticality evaluations, the authors had asked the experts to assess additionally whether an element of behaviour should represent a learning objective in driver education. Behaviour which the experts described commonly as both relevant for training purposes and of high criticality was automatically defined as

for each of the 1,500 elements. Alongside the numerical index, the level of criticality was indicated graphically as a certain number of "x" symbols: "(x)" represents "-20 to -12", "(xx)" represents "-11 to -4", "(xxx)" represents "-3 to +3", "(xxxx)" represents "+4 to +11", "(xxxxx)" represents "+12 to +20" (see Fig. 5).

[37] The authors believed it to be especially important that the criticality evaluation should refer not only to tasks and subtasks, but also to the – more or less critical – individual actions: The criticality of an overall task is in the end dependent on the number and criticality of its constituent actions.

[38] It should be noted that these instruments for learner assessment cannot be equated to a driving test (McKNIGHT never concerned himself explicitly with the development of a driving test). Learner assessment within the framework of a curriculum in this case served primarily to steer the learning processes and to verify attainment of all relevant learning objectives; to this end, different methods can be used at different stages of the learning process. A state-run (driving licence) test, by contrast, stands at the end of a training process, or at least at the end of a significant stage of training, and there realises above all a selection function relating to the granting of extended (mobility) entitlements. This places increased psychometric quality demands on the instrument, and furthermore entails additional political expectations (e.g. "test equality"); at the same time, time limitations and the associated cost constraints must also be taken into account. Consequently, a driving licence test can only address a small selection of learning objectives which are deemed particularly relevant with regard to certain specified criteria (e.g. road safety). The test contents and assessment specifications for a driving test thus represent an essentially standardised subset of the demand and assessment standards for learner assessment and are derived on the basis of similar criteria and procedures.

a "performance objective"; behaviour which was considered relevant but less critical, or even uncritical, on the other hand, was classified as a knowledge prerequisite for the attainment of learning objectives. All other relevant behaviour whose criticality had been assessed differently by individual experts was discussed by a committee of driving instructors with regard to possible inclusion as a learning objective. The outcome was a catalogue of performance-oriented learning objectives which could be considered relevant for driver training on the basis of systematic expert assessments, wherein each objective was assigned to one of five groups in accordance with the criticality determined at the stage of task analysis. Ten examples were selected at random from each of these groups, and the five groups of 10 elements were again presented to 48 driving instructors to assess their significance for safe and efficient driving and for successful completion of driver education. This enabled not only validation of the "performance objectives", but at the same time also the specification of test contents and criticality-oriented minimum standards to be achieved by learner drivers.

Subsequently, evaluation instruments were derived from the elaborated learning objectives and assessment standards (McKNIGHT & HUNDT, 1971a, 1971b). These evaluation instruments included a written "Knowledge Test" with 105 test items in multiple-choice format, including also questions on legal regulations, vehicle maintenance and journey planning, and a two-part practical "Performance Test" comprising a "Driving Fundamentals Test" referring to basic vehicle control and a so-called "Driving Situations Test".[39] The "Driving Fundamentals Test" was to be realised in a traffic-free or at least low-traffic environment and comprised nine tasks: "Pre-driving vehicle inspection", "Starting" and "Starting on an incline", "Accelerating", "Gear shifting" and "Use of gears", "Parking", "Turning" and "Stopping". It can be noted at this point that the tasks of the "Driving Fundamentals Test" possess similarity to the basic driving manoeuvres required within the framework of today's practical driving test. The tasks of the "Driving Situations Test", on the other hand, exemplified the more complex demands of typical situations encountered in real day-to-day traffic. In this context, McKNIGHT and HUNDT (1971a, 1971b) distinguished between categories of situations which could be planned by either the test administrator or candidate (e.g. merging into traffic, overtaking, the negotiation of bends and junctions, motorway driving, bridges and tunnels) and those categories of situations which could not be planned as they are dependent on the behaviour of other road users (e.g. pedestrians, cyclists, or oncoming, preceding, overtaking and parked vehicles) or traffic conditions (e.g. road surface, weather conditions). As a basis for the "Driving Situations Test", a catalogue of situations was established – similarly to the "driving tasks" defined by HAMPEL and KÜPPERS (1982) or STURZBECHER, BIEDINGER et al. (2010) – together with descriptions of the behaviour necessary to master the corresponding task (performance standards or assessment criteria).[40] One important difference between the two tests was that the number of tasks to be solved in the "Driving Fundamentals Test" was always the same, whereas the scope of the "Driving Situations Test" varied due to the unplanned situations. No studies were conducted to evaluate the reliability and validity of the tests; they were nevertheless considered valid on account of the expert contributions to content elaboration (see above). An overview of the driving tasks of the "Driving Situations Test" described by McKNIGHT and HUNDT (1971a) can be found in Table 1.

Returning to the original question: To what extent do the studies conducted by McKNIGHT and others in the 1970s, i.e. analysis of the demands placed on a driver by participation on motorised road traffic and the corresponding specification of learning objectives as the basis for a driving test, represent a solid foundation for (further) optimisation of the practical driving test as a work sample and systematic behaviour observation from today's perspective? When seeking to answer this question, due consideration must be given to the initially outlined methodical demands relating to work or driving sample design.

It is beyond doubt that McKNIGHT and ADAMS (1970a, 1970b) were successful in reducing the overall process of driving to its individual action components and tasks on the basis of domain-specific criteria and acceptable methodical re-

[39] To assist realisation of the two practical tests, procedural instructions and a catalogue of tasks with corresponding assessment possibilities were compiled into a so-called test booklet. The test administrator could then use this booklet to record observations and to assess the specified aspects of behaviour in each relevant situation. The mastering of individual situations was to be assessed with "pass" or "fail"; subsequently, the individual observations were to be compacted to obtain an overall assessment, taking into account the different criticality indices of the situations concerned, and thus a decision on passing or failing of the test as a whole.

[40] This being a learning-objective-referenced test which served primarily to determine the novice driver's learning deficits and learning progress, the candidate was also expected to answer questions during the drive (e.g. whether the distance to an oncoming vehicle was still sufficient to permit overtaking, or whether a parking space was long enough to park the vehicle).

search strategies, subsequently in structuring the identified components according to different levels of complexity, and finally – by way of their criticality criterion – in identifying those key demands which are particular determinants for successful mastering of a situation and provide an indication of competence on the part of the driver. With their likewise methodically and professionally demanding determination of learning objectives, McKNIGHT and HUNDT (1971a, 1971b) then involved domain experts to uncover the aspects of driving behaviour which lead to successful or unsuccessful mastering of the set tasks. The same authors also developed methodical standards and observation/assessment sheets as evaluation instruments permitting assessment of the novice driver's level of goal attainment within the framework of driver education. Thus, all the work described Chapter 3.1 above as the first step in proper work sample design had been accomplished successfully. The second step, namely construction of the work sample in the narrower sense, was not taken however, since the elaboration of a driving licence test was not a project goal. There was consequently no attempt to determine those typical demand situations which are most performance-relevant and at the same time suitable for testing under circumstances limited by cost and capacity constraints on the one hand and the feasibility of planning on the other. The studies of the aforementioned authors nevertheless play a valuable vanguard role for the second design step; we will return to this point when explaining the present proposals for optimisation of the practical driving test.

Use of the driving task concept in the German system of driver licensing

Following the enactment of corresponding codification principles in Prussia, it became necessary to hold an official permit to drive a motor vehicle from 15th December 1900.[41] On this basis, "the Berlin police ordinance on the operation of motor vehicles, which had been conceived as a model for broader regulations, came into force [on 15th April 1901]. … According to this ordinance, permission to operate a motor-driven vehicle was granted only to persons who had obtained confirmation from an authority, a driving school under the auspices of an authority or an officially recognised expert to verify that they were fully acquainted with the handling of the vehicle, possessed knowledge of the traffic regulations and displayed the character traits deemed to be prerequisites. … At this time, practical driving skills were to be demonstrated, for example, by way of simple driving exercises in the courtyard of the police headquarters" (STURZBECHER et al., 2009, p. 40ff.; FACK, 2000).

The most important step towards nationwide standardisation of the practical driving test was taken with the passing of a "Motor Vehicle Traffic Act" on 3rd May 1909 and the associated "Ordinance on Motor Vehicle Traffic" of 3rd February 1910: From then on, prospective drivers were required to complete simple driving exercises, such as passing obstacles, braking, reversing or turning, in the manner of what would today be termed "basic driving manoeuvres". Furthermore, candidates were to demonstrate their fitness to drive, the necessary calmness to operate a motor vehicle and a minimum presence of mind during a test drive in real traffic of moderate density (FACK, 2000). These test demands were refined and developed further above all with the "Motor Vehicle Traffic Ordinance" of 1923, and again by way of the Examination Guidelines of 20th January 1934: Both situation-related driving tasks (e.g. "Encountering and overtaking horse-drawn vehicles", "Turning into other roads") and situation-independent tasks (e.g. "Changing speed", "Safe traffic observation", "Estimating distances") were now stipulated for the test drive in real traffic (STURZBECHER et al., 2009).

The first scientific efforts to elaborate methodical foundations and possibilities for optimisation of the practical driving test began in Germany in the mid-1970s (HAMPEL et al., 2009).[42] The fundamental significance of the work done by McKNIGHT and others was also acknowledged in certain aspects at this time. The starting point for the conceptional treatment was recognition of the fact that the test demands to be fulfilled by the driving licence applicant were described in different forms and with partially divergent content in different statutory regulations: "This situation makes it somewhat difficult to define the test subject unambiguously in all details. This requires considerable interpretation on the part of the examiner. Standardisation and, in particular, more detailed specification of the test subject matter thus appears urgently necessary" (HAMPEL, 1977, p. 45). At the same point, it is explained why examination guidelines cannot be deemed an adequate basis for demand standards in the sense of psychological testing: "Actions which are spread relatively unsystematically across all situations … are named as test de-

[41] Previously, drivers had merely received a manual issued by the manufacturer to describe vehicle operation and did not need to furnish any proof of driving competence (FACK, 2000).

[42] Experts in the field demanded reformation of the practical driving test more emphatically than modification of the theoretical test in the 1970s as a means to reduce novice driver accident involvement (HAMPEL et al., 2009).

mands. Alongside, there are situation-related demands …. Given such undifferentiated use of categories with different dimensions, the repeated overlapping of characteristics is inevitable. … In our opinion, the specifications in the examination guidelines are insufficient to provide an exact definition of the required behaviour. There is a lack of clear and unambiguous task descriptions for the novice driver" (ibid., p. 46). It seems that the problems addressed are still essentially unsolved: "From today's perspective, both the elaboration of a closed and robust methodical foundation for the practical driving test and the mutual adaptation of the correspondingly developed methodical test standards, on the one hand, and the already existing legal test standards, on the other hand, would appear to be still outstanding" (STURZBECHER, BIEDINGER et al., 2010, p. 71).

To further attainment of the set goal and to derive demand standards for the practical driving test, the BASt commissioned HAMPEL (1977) to conduct broad research aimed at documentation of those methods of driving behaviour observation which had been developed both at home and abroad primarily for the assessment of fitness to drive and "could claim to cover the whole scope of driving behaviour" (ibid., p. 157), alongside investigation of the different "scientific approaches, with the objective of determining the extent to which they could be relevant for routine testing" (ibid., p. 5). HAMPEL found the results to be rather sobering, however: "From an overview of the different approaches, it can be seen that they are geared predominantly to the forecasting of fitness to drive and the aim of identifying problematic drivers, or else that they build upon questions which permit only limited conclusions to be drawn with regard to the proving value of driving tests" (ibid., p. 118). The efforts to gain new knowledge of the demand and performance structures of driving behaviour by subjecting the findings of different driving behaviour observations to explorative factor analysis, and on this basis to derive demand standards for the practical driving test, also failed to yield satisfactory results: "Cautious judgement indicates that the existing factor structures should rather be taken to reflect the opinions of competent observers on the complex of driver behaviour. It is our belief, that such condensed information … must not be confused with a direct representation of the actual behaviour of vehicle drivers" (ibid., p. 125).

The sought mirror of ideal driver behaviour actually already existed at this time in the results of the task analysis conducted by McKNIGHT and ADAMS (1970a). It is true that HAMPEL (1977) included the work of these authors in his research – alongside the "Road Test" described by McGLADE (1960, 1963), which was based exclusively on situation-related driving tasks – and also recognised their value for driving school instruction[43]; the significance of this method for the description of demand criteria for the German practical driving test, however, was underestimated: "One limitation lies in the only conditionally comparable American education system. This basically applies also to the test framework proposed by McKNIGHT (1974)" (HAMPEL, 1977, p. 79).

As objection to the assessment made at that time, it must be said that the task analysis was geared to the elaboration of learning objectives and the implementation of corresponding curricula; despite possibly divergent curricular expectations for the German system of driving school training, it would thus also have been possible to build upon the demand analysis performed by McKNIGHT and ADAMS (1970a) in Germany. This can be considered all the more true against the background of the essentially identical fundamental demands which are placed on drivers in all technically advanced Western industrial countries. Consequently, and again contrary to the opinion of HAMPEL[44], the learning objectives and evaluation instruments elaborated by McKNIGHT and HUNDT (1971a) also provide an acceptable starting point for the further development of driving school training and learner assessment in Germany. With regard to the optimisation of testing, the value of such proposals is actually recognised by HAMPEL (1977) elsewhere:

1. HAMPEL (1977, p. 91) acknowledges that, with the evaluation instruments proposed by McKNIGHT and HUNDT (1971a) for the assessment of practical driving skills, "it is guaranteed that the whole scope of required skills is covered".
2. The vehicle control tasks (or basic driving manoeuvres) described by McKNIGHT and HUNDT (1971a) as elements of their "Driving Fundamentals Test" are seen as "a noteworthy suggestion" by HAMPEL (ibid., p. 104).
3. HAMPEL summarised in the conclusion of his research report that, "for objectivisation of the driving test … a content-referenced and task-

[43] HAMPEL (1977, p. 77) refers, for example, to the assessment of JENSCH, SPOERER and UTZELMANN (1977), who describe McKNIGHT's work as "the most broadly expanded approaches to driver education".

[44] HAMPEL (1977, p. 119) writes that test tasks could be derived as a "representative sample from the 'universe' of all learning tasks where the learning objectives are defined so specifically and in such detail that they can be operationalised directly as test tasks. McKnight's system is one example in this respect. A corresponding solution is still to be found for the German context."

based analysis of driver behaviour from which test demands could be derived" was indispensable, and that such an analysis should be "initiated with priority" in Germany (ibid., p. 160); At the same time, however, he conceded that "corresponding analyses also exist in the international field. They should be evaluated, supplemented where necessary, and transferred to the context of the Federal Republic" (ibid., p. 144).

If we consider the study results and test methodology proposals published by HAMPEL (1977) from today's perspective, they appear to represent the most important programmatic contribution to further development of the German practical driving test in the 20th century – despite the aforementioned inconsistencies and the limitations described in the following. First of all, let us consider the limitations and merits of HAMPEL's work with reference to the definition of appropriate demand and observation standards (his statements on the elaboration of detailed assessment and decision criteria are to be discussed elsewhere):

1. HAMPEL (1977) fails to make an adequate distinction between instruction-oriented methods of learner driver assessment, as elaborated in exemplary manner by McKNIGHT and HUNDT (1971a), and a driving licence test, where different methodical demands apply (see above; footnote 38). A complete description of the learning objectives is not sufficient as "unambiguous orientation for the examiner" (HAMPEL, 1977, p. 144). Safe driving under night-time conditions, for example, represents a highly safety-relevant and thus important learning objective; the driving instructor must promote and assess attainment of the correspondingly necessary competences by the novice driver, and must take his findings from such learner assessment into account in his planning of the further course of training. Even so, this learning objective is not reflected accordingly in the content of the practical driving test, because the legislator – for cost and capacity reasons – is averse to demanding that the driving licence applicant take an (additional) night-time driving test. In the course of driver training, therefore, it is possible to assess the mastering of specified learning objectives under very different driving conditions (e.g. different lighting and visibility conditions, traffic density, weather conditions) on different occasions; this is not generally possible during the driving test, however. Consequently, the learning objectives for driver training cannot be treated automatically as test contents; it is rather the case that narrowed test contents must be derived from the learning objectives (or the corresponding instruments of learner assessment) by way of appropriately founded criteria.
2. HAMPEL (1977) offers no solution for the question as to how the mastering of unforeseeable or unplannable demands and traffic situations "which arise essentially from the actions of other road users, weather conditions and the changing of traffic signals" (p. 102) can be taken into account in the practical driving test. McKNIGHT and HUNDT (1971a) also offer elaborated observation schemata for learner assessment in such situations. For the practical driving test, however, this is not a workable solution: As long as test paradigms (i.e. specified standardised demands) and the political requirement of test equality[45] (both of which necessitate equal test demands for all candidates) remain paramount, unplanned demands cannot be taken into account in the assessment of driving licence testing. From the perspective of road safety, on the other hand, this seems unacceptable: Do we really wish to permit a driving licence applicant who displays serious driving errors in poor visibility or when the road is wet – i.e. unplannable test conditions – to drive solo? HAMPEL recommends that, in case of "distinctive weather situations", the weather conditions should be recorded "to enable appropriate assessment" (ibid., p. 67); how this is to be reconciled conceptually with a test paradigm, however, remains unclear.
3. The demand standards of the practical driving test should not – as HAMPEL (1977) believes – "be further developed in the form of internal professional guidelines", which are "already available" (ibid., p. 144). It instead seems desirable to elaborate and publish a transparent manual of psychological test methods, which, alongside demand standards, could also contain implementation regulations, as well as

[45] "Test equality" is not a defined category in psychological testing. It thus appears more expedient to work with the corresponding concept of "population-specific equivalence": This means that no target group for a test must be disadvantaged by virtue of special characteristics which are independent of the subject of the test. The results of a knowledge test, for example, must not be dependent on the gender of the test candidates, but instead solely on their intelligence. In the case of a learning-objective-referenced test, the difficulty of a task plays no role with regard to its reasonableness, provided the task is valid and refers to actually significant learning. In the context of psychological testing, therefore, no candidate in a learning-objective-referenced test is entitled to expect particularly simple or – compared to other candidates – equally difficult tasks.

assessment and decision criteria for the practical driving test, for example.

4. From the overarching perspective, it can be viewed as an important merit for HAMPEL (1977) that he was the first researcher to conclude that driving behaviour observations, according to the outcome of intensive empirical examination of the contemporary state of research, could not be considered objective (test) methods in the strict sense and were furthermore unsuitable to fulfil this role: “As long as driving tests are conducted in real traffic, it seems that full standardisation is impossible” (ibid., p. 5). Even so, HAMPEL still remained faithful to the test paradigm in 1977: The practical driving test should also satisfy claims of “standardisation in the sense of normalisation according to the rules of classic test theory” (ibid., p. 143). His recommendations thus target an approximation of objective test conditions through the locality-specific elaboration of standardised test route sections; this path, however, has still not been followed to date – probably for reasons of practicability – and is in our opinion insufficient in itself as a test strategy for proper driving competence assessment.[46]

5. With regard to the establishing of candidate-referenced demand standards, HAMPEL (1977) set a new pattern for the field with his demand for specification of “a certain catalogue of driving tasks which are to be performed during the course of the drive”; such a rule would also be in line with the test stipulations (as they were applicable at that time). These concrete, typical driving tasks, which, according to HAMPEL, every candidate should be required to master in several instances and under changing framework conditions, were derived from an analysis of learning objectives (ibid., p. 150). HAMPEL reached this conclusion via recognition of the fact that – despite its impressive objectivity – detailed, event-oriented determination of the proper fulfilment of individual, elementary behaviour demands according to dichotomous assessment criteria (“correct” versus “false”), as demanded by McKNIGHT and HUNDT (1971a), would place excessive demands on the examiner in terms of observation, assessment and documentation of the test: “A more modest, but perhaps more realistic concept, in our opinion, is that of BARTHELMESS, which is limited to the description of six selected situations in which driving skills should be proven” (HAMPEL, 1977, p. 99); such a situation-oriented approach would also be most conducive to test objectivity, as shown by the empirical findings of SCHUBERT and EDLER (1965). HAMPEL (1977) thus focussed the demand standards for the practical driving test – albeit without explicit reference – on the categories of “Situational behaviour” and “Basic control” described by McKNIGHT and HUNDT (1971a): While situation-related driving tasks should be placed in the foreground of the test drive in real traffic, fundament control tasks (or basic driving manoeuvres) were to be demonstrated at the beginning of the test, preferably on a separate test ground. These thoughts led to formulation of a proposal for a driving task catalogue by TÜV Rheinland in 1977; at the same time, similar task catalogue proposals by TÜV Bayern and by JENSCH, SPOERER and UTZELMANN (1978) were taken up (HAMPEL & STURZBECHER, 2010). An overview of the driving tasks defined for this catalogue can be found in Table 1.

6. Neither the examination guidelines nor the concepts of McKNIGHT and ADAMS (1970a) or HAMPEL (1977) contain an explicit, theoretically founded and methodically practicable proposal for a structural description of the relationship between situation-related and situation-independent demands, or for possible implementation of these distinct demands in the practical driving test. In the examination guidelines – as already criticised by HAMPEL (1977) – both forms of demand were still found side by side. In McKNIGHT and ADAMS (1970a), the two types of demand stand unstructured and unconnected in the three categories “Basic control” (situation-independent demands relating essentially to vehicle operation), “General driving” (similarly situation-independent demands, e.g. observation) and “Situational behaviour” (situation-related demands). This results in a certain indistinctness (JENSCH, SPOERER & UTZELMANN, 1977) and also fails to take into account the regular recurrence of the situation-independent demands when handling the situation-related demands. This circumstance must be described in a structural concept and instrumentalised to reduce the complexity of observations and judgements during the practical driving test. This challenge is in part still unsolved, although HAMPEL (1977) recog-

[46] Such standardised routes are used in a number of concepts for driving behaviour observation serving assessment of the fitness to drive, whose approaches HAMPEL (1977, p. 157) rightly felt to be inapplicable “under the conditions of a routine test”.

nised the problem: He points out that the determined methods of driving behaviour analysis use very different observation and judgement categories[47] and rely on different methods of assessment (rating scales or alternative assessments). As regards the observation categories, HAMPEL follows v. KLEBELSBERG (1970) and distinguishes “primary characteristics which are accessible to direct observation (e.g. adaptation of the engine revs)” and “secondary characteristics which require conclusions to be drawn from other observations (e.g. concentrated driving)”. Within the framework of a comparison, he notes: “In the judgement systems used for driving tests, on the other hand, concrete stipulations relating to directly observed behaviour, i.e. primary characteristics, are clearly dominant” (HAMPEL, 1977, p. 93 ff.). At the end of his comparative discussion, HAMPEL reaches the following conclusion: “It is generally difficult to imagine how, in driving tests where the result has serious consequences for the candidate, the judgements could be based on mere description of the impressions gained by the examiner. The candidate will hardly be satisfied with the opinion that his driving was ‘careless’, and will instead want to know how exactly this carelessness was manifested. … The consequence is that only primary characteristics are suitable for use in driving tests. Secondary characteristics can only serve to round off the picture” (1977, p. 94). Elsewhere, however, we find indication of how the problem could be solved: In the aforementioned driving task proposal of TÜV Rheinland (see above), the situation-independent demands are described – somewhat inappropriately – by HAMPEL (1977, p. 140ff.) as “behaviour” or “actions” in the sense of observation categories (even though this term is not actually used) and arranged in a matrix of “driving situations and behaviour” for the documentation of test performance. HAMPEL and STURZBECHER (2010, p. 57) later referred to this development as the “origin” of the concept of driving tasks and observation categories in driver training and testing in Germany; the observation categories are to be discussed further in the next chapter.

7. Finally, it remains to be ascertained that the reform proposals put forward by HAMPEL (1977) contained few theoretically and methodically founded recommendations with regard to an appropriate test strategy or implementation rules for the practical driving test in the sense of instructions for the examiner; we will return to this topic in Chapter 3.5. This is not intended as criticism: Given the fact that the academic community had only just become aware of the theoretical and methodical gaps in the scientific foundations of the driving test at that time, and since the prerequisites for processing of these deficits had only just been established (HAMPEL et al., 2009), the main purpose of the study by HAMPEL (1977) – in line with the intentions of the BASt – was to identify contents for a necessary research and development programme to optimise the practical driving test, rather than to provide answers to all open questions, some of which are still unanswered today.

From today's perspective, it can be noted that the derivation of driving tasks from the learning objectives of driving training, which was demanded by HAMPEL (1977) as a precondition for further efforts to objectivise the practical driving test, failed to materialise and has still not been realised successfully to date. His central goal, namely to establish an candidate-oriented demand standard, in other words a catalogue of driving tasks to be performed by all candidates, was also abandoned. Instead, the traffic policy decision makers in German driver licensing at the end of the 1970s resolved to seek “as far as possible merely solutions within the framework of the existing provisions” (HAMPEL & KÜPPERS, 1982, p. 14). Rather than a candidate-referenced demand standard with driving tasks, they chose to define merely a task-based demand profile for relevant traffic environments (“test locations”[48]) within the framework of test location guidelines: “The immediate aim was thus a reorganisation of the regulations and guidelines already applicable to the driving test, and not the development of a whole new methodology. … Where new provisions were necessary, the study group gathered the opinions of traffic experts from

[47] HAMPEL (1977) uses the term “judgement categories” to describe the aspects of the candidate's test performance which are to be assessed by the examiner. Before assessment, however, these aspects must first be observed; accordingly, the present report uses instead the term “observation category”, as is customary in conjunction with descriptions of observation processes in today's methodology literature.

[48] Test locations – according to the legal definition – are built-up areas which, by way of their road network, the existing traffic signs and installations and their traffic density and structure, permit the testing of essential driving procedures. Test locations are designated as such by the responsible supreme state authority, an office stipulated by that authority or the office responsible under federal state legislation; the practical driving test can also be conducted in the surroundings of test locations. The driver licensing authority specifies the location at which a candidate must take the test (§ 17 (3) FeV).

the Technical Examination Centres, the driving instructors and representatives of scientific research nominated by the BASt. Compared to a systematic analysis of learning objectives and driving tasks, this was only the second-best solution; the involvement of all the affected institutions, however, did bring the advantage of greater acceptance for the ensuing guidelines on test locations[49]" (HAMPEL & STURZBECHER (2010, p. 59).

Despite the fact that, according to the study conducted by HAMPEL and KÜPPERS (1982), driving tasks were now no longer referred to the individual candidate or to each individual test, and were instead to be specified with regard to their suitability for testing at different test locations, it remained to be asked, which driving tasks should actually be tested. The original objective, namely to elaborate training- and safety-relevant driving tasks for the practical driving test, was as topical as it had ever been, as was HAMPEL's (1977, p. 90ff.) appropriate recommendation that a solution should not be based on "existing non-systematic collections of characteristics", but rather on systematically elaborated task analyses supported by both expert judgements and empirical validity checks. HAMPEL and KÜPPERS (1982) nevertheless chose a different, three-stage approach: The first step was a document analysis covering all the fundamental road traffic legislation which contained stipulations relating to the test drive, e.g. Road Traffic Regulations (StVO), Road Traffic Licensing Regulations (StVZO), Examination Guidelines. This was then supplemented, in a second and third step, with locality-referenced demands taken from national and international publications in the fields of training and testing. The outcome was a list of 53 location-specific demand criteria.[50] It remains unclear, however, why this eclecticist approach was preferred over a systematic scientific task analysis: There were possibly doubts as to whether the task analysis by McKNIGHT and ADAMS (1970a) and the driving task catalogue of the "Driving Situations Test" by McKNIGHT and HUNDT (1971a) could be transposed to German road traffic conditions, or perhaps the focus on the demand lists of German institutions was expected to raise acceptance among the national academic community.

The ensuing list of demands was distributed to 234 experts involved in driver licensing (examiners, driving instructors, traffic engineers and traffic psychologists) with the request to assess whether the essential learning demands of driving competence acquisition were covered; amendments and alternative formulations were expressly welcomed. In addition, the experts were asked to gauge the significance of these demands for the practical driving test. The outcome was a collection of 18 driving tasks for which corresponding local conditions should be encountered with a specified minimum frequency[51] to constitute a satisfactory test location. As a final step, validation of this "standard demand profile for test locations" (HAMPEL & KÜPPERS, 1982, p. 90) was sought within the framework of field testing at 35 random locations; its essentially unchanged contents are still today the basis for the situation-related demand standards of the practical driving test. A detailed description of the – minor – changes to this demand catalogue over the period from 1987 to the present day can be found in STURZBECHER, BIEDINGER et al. (2010).

If we compare the list of driving tasks which was proposed by TÜV Rheinland as a basis for further development of the systematics of demand standards in 1977 (HAMPEL & STURZBECHER, 2010) with the driving task catalogue for the Driving Situations Test of McKNIGHT and HUNDT (1971a), it can be noted that all the driving tasks of the TÜV Rheinland proposal – with minor formal deviations – are also to be found in the driving test described by McKNIGHT and HUNDT (1971a) (see Table 1). Among the planned situations described in the US demand catalogue, the categories "Off-road driving", "Bridges or tunnels", "Hills" and "Emergency planning" are missing from the German proposal. This seems plausible and is not problematic: The aforementioned driving tasks are relatively uncommon challenges in daily road traffic or else only typical for certain regions of Germany; consequently, they may be suitable as optional components for incorporation into driving school training (and learner assessments) on a regional basis, in accordance with local traffic risks, but should not be designated elements of a uniform nationwide practical driving test. At this point, the

[49] These test location guidelines had become necessary as the increasing levels of motorisation in Germany in the 1960s and 1970s had led to ever wider deviations in traffic density – and thus also in the test demands – between individual test locations (MÖRL, KLEUTGES & ROMPE, 2008).

[50] If the driving task list elaborated by HAMPEL and KÜPPERS (1982) is compared with those of McKNIGHT and ADAMS (1970a) or McKNIGHT and HUNDT (1971a), it can be seen that many of the demands coincide. Nevertheless, HAMPEL and KÜPPERS (1982) only specified the source "McKnight" for three demands. This was probably a consequence of the order of their source reviews, where international sources were considered last.

[51] The corresponding frequency specifications for the "Driving Situations Test" of McKNIGHT and HUNDT (1971a) were calculated by way of systematically determined criticality indices and expert assessments, whereas those of the German counterpart were based exclusively on expert recommendations.

aforementioned distinction between evaluation instruments for driver education, as elaborated by McKNIGHT and HUNDT (1971a), and a practical driving test, as HAMPEL (1977) had in mind, comes to bear.

In connection with the unplanned situations, the HAMPEL proposal also differs from the US demand catalogue in that it dispenses with the categories "Parked vehicles", "Preceding vehicles", "Oncoming vehicles" and "Overtaking vehicles". These driving tasks are hardly avoidable in the urban road traffic environment which can reasonably be expected at all German test locations; they are assessed either in connection with other driving tasks (e.g. the driving task "Passing" includes driving past parked vehicles) or as situation-independent demands. There is thus no need for separate stipulation of these demands for the practical driving test. Furthermore, McKNIGHT and HUNDT (1971a) specify the situations "Traffic signals", "Road surface conditions" and "Weather conditions", which are likewise not to be found in the HAMPEL proposal. This, too, seems plausible, or is at least not to be considered a deficit: For the German demand catalogue, the passing of signal-controlled crossroads and junctions represents a (likewise scarcely avoidable and thus not explicitly listed) special instance of the driving task "Observance of the rules of right-of-way"; "Road surface conditions", and even more so "Weather conditions", can hardly be varied within the framework of a driving test, and thus cannot be taken into account systematically in the test demands – in contrast to the situation of learner assessment during driver training, which may take place at different times and in different traffic environments.

Driving Situations Test (McKNIGHT & HUNDT, 1971a)	**Driving task proposal by TÜV Rheinland** (1977)
Planned situations	
Entering and leaving traffic	Driving off, stopping; merging into traffic
Curves	
Simulation of evasive action	Passing and overtaking other road users
Overtaking	
Intersections (crossing, left turns, right turns)	Turning across oncoming traffic Observance of the rules of right-of-way
Freeways	Motorways and high-speed roads
Off-road driving	
Bridges or tunnels	
Hills	
Emergency planning	
Unplanned situations	
Changing lanes	Use of road lanes
Pedestrians and cyclists	Pedestrians and cyclists
Special vehicles	Buses and rail-borne vehicles
Parked vehicles	
Preceding vehicles	
Oncoming vehicles	
Overtaking vehicles	
Traffic signals	
Road surface conditions	
Weather conditions	

Tab. 1: Comparative overview of the driving task catalogue of the Driving Situations Test (McKNIGHT & HUNDT, 1971a) and the task proposal elaborated by TÜV Rheinland in 1977 (HAMPEL & STURZBECHER, 2010)

All in all, the TÜV Rheinland proposal (HAMPEL & STURZBECHER, 2010) of 1977 – as a subset of the driving task catalogue of McKNIGHT and HUNDT (1971a) – thus stands wholly in the tradition of the criterion-referenced, systematic and scientifically robust demand analysis, learning objective elaboration and test content determination (in the sense of learning assessment) conducted by McKNIGHT and his colleagues. TÜV Rheinland had in 1977 effectively reduced the test contents defined for the "Driving Situations Test" (in its function as an evaluation instrument for learner assessment in driving schools) to those aspects of content which are methodically meaningful in the context of a practical driving test. It can no longer be reconstructed, how and according to which criteria he reached his driving task proposal; the result, however, appears to be professionally plausible and scientifically sound. A systematic empirical validation based on appraisals by (German) experts and the results of trial implementations, which HAMPEL (1977) demanded for driving task catalogues, was apparently not (or no longer) performed, as those responsible had later – as HAMPEL writes (see above) – chosen the "second-best" variant for the determination of demand standards: Collection of an overall set of driving demands with reference to particular locations (primarily those already existing in Germany) and majority-based selection of the relevant driving tasks by branch experts.

HAMPEL and KÜPPERS (1982) provide a more detailed description of the procedures used to determine driving tasks for the "standard demand profile for test locations" and of the empirical studies conducted to evaluate the suitability of these demands for testing at different test locations. Over the period up to finalisation in the Examination Guidelines of 1987, the demands to be satisfied by test locations were modified in certain details in order to "find acceptable solutions to the inevitable conflicts between the aim of establishing the desirable test conditions and the actual circumstances of local traffic conditions". The objectives and approaches were thus "clearly determined by the political framework … The project results, subject to certain amendments, were later incorporated into the corresponding guidelines by the responsible committees" (HAMPEL & STURZBECHER, 2010, p. 59). If we compare the 1977 driving task proposal by TÜV Rheinland (HAMPEL & STURZBECHER, 2010) with the driving task catalogue of the "Standard demand profile for test locations" by HAMPEL and KÜPPERS (1982), then the latter appears suboptimal and seems to represent a backward step (see Table 2). For example, the categories "Passing", "Overtaking" and "Rail-borne vehicles", which were well founded in McKNIGHT and HUNDT (1971a) and similarly demanded by HAMPEL (1977), were now missing after expert appraisal of the 53 driving tasks contained in the original catalogue of HAMPEL and KÜPPERS (1982); roundabouts were no longer mentioned explicitly. The driving task catalogue incorporated into the Examination Guidelines in 1987 was later modified slightly on several occasions. One important change was the addition of "Driving outside built-up areas (with possibilities to overtake)", which once more made reference to "Overtaking" as a possible driving task. The currently applicable catalogue of driving tasks can be found in Annex 11 ("Demands on the test location and its surroundings") to the Examination Guidelines.

Driving task proposal by TÜV Rheinland (1977)	**"Standard demand profile for test locations"** (HAMPEL & KÜPPERS, 1982)
Driving off, stopping; merging into traffic	Driving off and merging into moving traffic from the kerbside Entering (merging into) in priority roads
	Driving outside built-up areas (bends and blind spots)
Passing and overtaking other road users	
Turning across oncoming traffic Observance of the rules of right-of-way	Turning left on roads with oncoming traffic Passing crossroads … - with the priority rule "give way to the right" - with a stop sign - controlled with light signals - and junctions where the priority road turns away to the right or left
Motorways and high-speed roads	Motorways, high-speed roads
	Driving on … - roads with road markings - roads with a traffic density of at least 100 vehicles per hour - one-way streets - roads with two or several marked lanes for one direction
Use of road lanes	Changing between road lanes
Pedestrians and cyclists	Approaching and passing pedestrian crossings Turning right/left with special consideration for cyclists (e.g. parallel cycle lane)
Buses and rail-borne vehicles	Passing public transport stopping points

Tab. 2: Comparative overview of the driving task proposal elaborated by TÜV Rheinland in 1977 (HAMPEL & STURZBECHER, 2010) and the "Standard demand profile for test locations" (HAMPEL & KÜPPERS, 1982)

The still applicable and since 1987 practically unchanged catalogue of driving tasks was a subject of content analysis and methodical evaluation by STURZBECHER, BIEDINGER et al. (2010) as part of the project "Practical Driving Test – Foundations and Possibilities for Optimisation" conducted by the working group "TÜV ÐEKRA arge tp 21". This analysis of the driving task list illustrated "the necessity of its restructuring and further development, as it reveals both content redundancy and methodical inconsistencies, the elimination of which would facilitate test organisation, test observation, test assessment and test decisions" (STURZBECHER, BIEDINGER et al., 2010, p. 96). It was also noted that the driving tasks varied considerably with regard to their complexity and level of abstraction, and that no distinction was made between situative behaviour demands (e.g. "Changing between road lanes") and general conditions (e.g. "Driving outside built-up areas"). Further criticism referred to the fact that the driving tasks were not yet defined in the sense of demand standards to be met by the driving licence applicant and were moreover inadequately described. Extending the evaluation to include Annex 10 to the Examination Guidelines, finally, it can be faulted that the demand catalogue fails to distinguish between situation-related demands in the sense of driving tasks and situation-independent demands (in the sense of observation categories, see the following Chapter 3.3). The need for optimisation of the methodical architecture of testing (see above) – as identified by HAMPEL (1977) – is thus still outstanding.

On the basis of the aforementioned critical findings, BÖNNINGER et al. (2010, p. 173) demand "streamlining, restructuring and modernisation of the table of driving tasks", which could then be handled as a candidate-oriented minimum demand standard.[52] A relatively small number of safety-relevant driving tasks needed to be "formulated with a similar degree of complexity and sufficient generalisation", so as to permit implementation at every test location. The draft for a new scientifically founded catalogue of driving tasks should be based on "both expert ratings and a traffic-psychology-oriented demand analysis, leading in turn to an improved driving task list which can then be verified empirically in respect of its practicability at a representative selection of test locations." Last but not least, in view of the increasing volume of cross-border traffic in Europe, a driving task catalogue optimised in this way should be aligned with EU stipulations, and could then contribute to harmonisation of the European test systems (ibid.).

Driving task catalogue for an optimised practical driving test

The above considerations and knowledge served the present project as a starting point for optimisation of the driving task catalogue, as suggested by BÖNNINGER et al. (2010). At the same time, following a demand expressed by HAMPEL (1977), the project tackled a detailed description of the driving tasks, along with appropriately task-referenced observation categories and assessment/decision criteria, as a means to raise the psychometric quality of the practical driving test. Elaboration of the optimised catalogue content was based on the corresponding driving task catalogues of the Driving Situations Test (McKNIGHT & HUNDT, 1971a) and TÜV Rheinland (HAMPEL & STURZBECHER, 2010), as well as the reform proposals of STURZBECHER, BIEDINGER et al. (2010); in addition, it was considered necessary to take into account EU stipulations and international standards (see below), as well as the current research addressing novice-typical driving competence deficits and the principal causes of accidents involving novice drivers; we will return to this point in the driving task descriptions.

To facilitate further development of the driving task catalogue, the scientific procedures and (safety-relevant) criteria outlined in previous report sections in conjunction with the demand analysis by McKNIGHT and ADAMS (1970a, 1970b) and the elaboration of a Driving Situations Test by McKNIGHT and HUNDT (1971a, 1971b) were reconstructed and identified as a sound working basis, alongside the TÜV Rheinland driving task proposal of 1977 (HAMPEL & STURZBECHER, 2010). A comparative content analysis of these initial publications yielded a draft for an optimised driving task catalogue founded on the aforementioned empirical research and development studies from the 1970s and 1980s. This draft was then discussed in the so-called "project support group"[53] and developed into a reform proposal for a future catalogue of driving tasks (see Table 3).

[52] The recommendations relating to the design and assessment of basic driving manoeuvres are not taken up in the present report. It is nevertheless recommended, within the framework of the pending reforms, to further develop the assessment standards on the basis of the proposals by BÖNNINGER et al. (2010, p. 173): "It is necessary to streamline the assessment criteria applicable to the basic driving manoeuvres and to reduce their significance for the test decision to the level and handling defined for simple errors."

[53] Fundamental questions relating to the content and structures of the future driving tasks, observation categories and assessment/decision criteria were discussed in a so-called "project support group" comprising representatives of the federal ministry responsible for traffic, the federal states, the Federal Highway Research Institute (BASt), the German Federation of Driving Instructor Associations (BVF), the Technical Examination Centres, the Bundeswehr, the working group TÜV | DEKRA arge tp 21, the Association of Technical Inspection Agencies (VdTÜV), the University of Potsdam, the Institute for Applied Research on Childhood, Youth and the Family (IFK) and the Institute for Prevention and Road Safety (IPV).

Driving Situations Test (McKNIGHT & HUNDT, 1971a) (excerpt, see above)	**Driving task proposal by TÜV Rheinland** (1977)	**Proposal for optimisation of the driving task catalogue** (2012)
Entering and leaving traffic	Driving off, stopping; merging into traffic	Joining/leaving traffic and changing lanes
Changing lanes	Use of road lanes	
Curves		Curves and connecting roads
Simulation of evasive action	Passing and overtaking other road users	Passing and overtaking
Overtaking		
Intersections (crossing, left turns, right turns)	Turning across oncoming traffic Observance of the rules of right-of-way	Crossroads and junctions
		Roundabouts
Special vehicles	Buses and rail-borne vehicles	Rail-borne vehicles
Pedestrians and cyclists	Pedestrians and cyclists	Pedestrians
		Cyclists
Freeways	Motorways and high-speed roads	

Tab. 3: Comparative overview of the driving task catalogue of the Driving Situations Test (McKNIGHT & HUNDT, 1971a), the task proposal elaborated by TÜV Rheinland in 1977 (HAMPEL & STURZBECHER, 2010) and the present reform proposal for optimisation of the practical driving test

In the member states of the EU – and thus also in Germany – the minimum demand and implementation standards for the practical driving test are dictated by the EU Directive on Driving Licences 2006/126/EC of 20th December 2006, Annex II, paragraph 7.4 "Behaviour in traffic", according to which applicants must perform all the following actions in normal traffic situations, in complete safety and taking all necessary precautions:

"7.4.1. Driving away: after parking, after a stop in traffic, exiting a driveway;

7.4.2. Driving on straight roads; passing oncoming vehicles, including in confined spaces;

7.4.3. Driving round bends;

7.4.4. Crossroads: approaching and crossing of intersections and junctions;

7.4.5. Changing direction: left and right turns; changing lanes;

7.4.6. Approach/exit of motorways or similar (if available): joining from the acceleration lane; leaving on the deceleration lane;

7.4.7. Overtaking/passing: overtaking other traffic (if possible); driving alongside obstacles, e.g. parked cars; being overtaken by other traffic (if appropriate);

7.4.8. Special road features (if available): roundabouts; railway level crossings; tram/bus stops; pedestrian crossings; driving up-/downhill on long slopes;

7.4.9. Taking the necessary precautions when alighting from the vehicle" (EUROPEAN PARLIAMENT & EUROPEAN COUNCIL, 2006, L 403/43).

If we compare the stipulations of the EU Directive on Driving Licences with the driving task proposal elaborated for a future practical driving test in Germany within the framework of the present project (see Table 4), then it can be noted that the reform proposal – subject to correspondingly detailed description of the driving tasks (see below) – satisfies all the essential demands formulated by the EU. The only requirements missing from the reform proposal are "Driving up-/downhill on long slopes" (7.4.8) and "Taking the necessary precautions when alighting from the vehicle" (7.4.9): Driving on inclines can only be tested in certain regions, and thus – in the same way as the situation category "Hills" in McKNIGHT and HUNDT (1971a) – cannot be part of a task catalogue which is binding for all driving test candidates; the taking of necessary precautions when alighting from the vehicle, on the other hand, is not to be deemed a driving task from the assumed standpoint in the context of test psychology, because it does not involve manoeuvring of the vehicle. Nevertheless, this demand is naturally a meaningful test requirement and is also stipulated in Germany within the framework of the test element "Technical completion of the drive".[54] With the special mention given to cyclists in the driving task catalogue, the German reform proposal goes beyond the stipulations of the EU Directive on Driving Licences.

[54] Requirements relating to "Technical completion of the drive" are specified in paragraph 16 of Annex 10 to the Examination Guidelines: "At the end of the test drive, the vehicle/vehicle combination is to be parked in compliance with the traffic regulations, such that it can be loaded or unloaded safely and such that persons are able to get into or out of the vehicle safely." The same paragraph also indicates the assessment criteria to be applied. Correct technical completion of the drive thus involves:

- Securing of the vehicle against rolling away by engaging a gear and/or applying the parking brake (use of both methods when parking on a gradient)
- In the case of vehicles with automatic transmission, securing against rolling away in accordance with the manufacturer's recommendations (owner manual)
- Securing against unauthorised use
- Observation of the traffic before and when opening the vehicle door.

Driving task stipulations of the EU Directive on Driving Licences (2006)	Proposal for optimisation of the driving task catalogue (2012)
7.4.1 Driving away: after parking, after a stop in traffic, exiting a driveway **7.4.6** Approach/exit of motorways or similar (if available) **7.4.6** Joining from the acceleration lane; leaving on the deceleration lane **7.4.5** Changing lanes	Joining/leaving traffic and changing lanes
7.4.3 Driving round bends **7.4.2** Driving on straight roads	Curves and connecting roads
7.4.2 Passing oncoming vehicles, including in confined spaces **7.4.7** Overtaking/passing: overtaking other traffic (if possible); driving alongside obstacles, e.g. parked cars; being overtaken by other traffic (if appropriate)	Passing and overtaking
7.4.4 Approaching and crossing of intersections and junctions **7.4.5** Changing direction: left and right turns	Crossroads and junctions
7.4.8 Roundabouts	Roundabouts
7.4.8 Railway level crossings; tram stops	Rail-borne vehicles
7.4.8 Pedestrian crossings; bus stops	Pedestrians
	Cyclists

Tab. 4: Comparative overview of the driving task catalogues of the EU Directive on Driving Licences and the present reform proposal for optimisation of the practical driving test in Germany

It is to be noted that the EU Directive on Driving Licences describes merely minimum requirements relating to the tasks for driving test examiners and the methodical quality of driving tests. It thus cannot be excluded that, given the traditionally significant differences in the training and test structures in individual EU member states, valuable input could be found for the elaboration of test standards – including driving tasks – beyond the task specifications of the EU directive. For this reason, too, the plan to search for safety-relevant driving tasks in the systems of training and testing in use in other countries can be deemed a promising approach.

After comparison of the driving test systems of the aforementioned 36 countries, it can in general be assumed that stipulations relating to the driving tasks to be set during a test drive exist in almost all countries. There are nevertheless significant differences with regard to the formulation of demands and the degree of discrimination: While some countries specify merely the driving task “Crossroads” (e.g. Great Britain, Ireland), for example, others (e.g. Finland, Victoria) indicate different types of crossroads (priority to traffic from the right, controlled by light signals, controlled by traffic signs) which must be incorporated into a driving test; in Victoria, the types of crossroads are even further subdivided according to the number of road lanes. In respect of the level of detail in driving task specifications, the countries can thus be classified into two groups:

(1) The first group comprises those countries which specify only general demands to be met by the test drive or test route; driving tasks are only outlined very roughly or else not defined explicitly at all: In countries such as Belgium, Estonia and Luxembourg, for example, it is merely stipulated that the test route must offer an adequate diversity of driving situations, in order to be able to test the most important aspects of driving behaviour under different conditions. In some cases, it is also required that the test route includes roads both within and outside built-up areas (e.g. in France and Luxembourg). Further stipulations refer to the length of the test route (e.g. in Portugal and Ireland) or to the traffic density (e.g. in Poland).

(2) The second group of countries (e.g. Finland, Austria, Switzerland, Sweden) can be characterised in that relatively detailed demand standards are described, either in the sense of candidate-oriented driving tasks or (as in Germany) as demands to be met by test locations. Depending on the road safety relevance of the individual driving tasks, specifications may also be made as to the desired or required frequency with which driving tasks are encountered during a single driving test or over a certain number of tests (e.g. Sweden).

From the starting point of these two groups of countries, it seems expedient to take a closer look at those countries in the second group which have in the past implemented essential further developments in their systems of novice driver preparation as a means to enhance their effectiveness with regard to improved novice driver safety. These reform-oriented countries include above all Finland, the Netherlands, Norway and Sweden.[55]

[55] GENSCHOW, STURZBECHER and WILLMES-LENZ (2014) write in this context: “These reform-oriented countries … play an important role for the further development of safety-enhancing measures and for the testing and introduction of innovative approaches. Following their elaboration and testing of a series of reform projects over the past two decades, and in view of the topicality for the European discussion of novice driver safety, attention is here drawn especially to the countries Finland, Norway, Sweden, the Netherlands and Austria.” The

To this end, Table 5 below identifies – by way of comparison with the driving task catalogue for a future practical driving test for Germany – the driving tasks which are to be found in the test stipulations or training curricula of these North and West European countries, where driver training is geared very strongly to the so-called “GDE matrix”.

In Table 5, the driving tasks which are only found in the training curriculum but not in the test report in the countries Finland, the Netherlands, Norway and Sweden are highlighted in bold type: It is presumably the case that these driving tasks are not binding contents of the driving task catalogue for the practical driving test. In the comparison with the reform proposal for a future optimised practical driving test in Germany, it is shown that all driving tasks from the training curricula of the reform-oriented European countries are also to be found – without exception – in the driving task catalogue of the German reform proposal, leaving aside the requirement of driving on motorways and in different traffic environments (Norway), which, from our standpoint based on test psychology principles, does not constitute a driving task in the narrower sense (even though it naturally represents a meaningful test demand; we will return to this point later). It can thus be assumed that the proposed future German driving task catalogue not only complies with the stipulations of the EU Directive on Driving Licences, but also corresponds to training and testing practice in the reform-oriented European countries.

Finland and Norway, on the other hand, forego the explicit formulation of a driving task “Overtaking”, as well as driving tasks relating to behaviour towards pedestrians and cyclists; the same basically applies in the Netherlands, except that mention is there made of pedestrian crossings. The absence of driving tasks relating to traffic situations involving pedestrians and cyclists in the practical driving test in Norway and Finland could be due to the fact that such tasks cannot be tested for all candidates in the less densely populated regions of those countries. With regard to the driving task “Overtaking”, it could furthermore be presumed that, firstly, driving licence applicants are expected to display particular caution during the test, and secondly, that overtaking in dense traffic is seen to demand a special level of driving competence which the candidate has usually not yet attained on account of his limited driving experience. Consequently, many driving licence applicants will rightly forego overtaking manoeuvres during the driving test, which may then be interpreted as desirable realistic awareness of his still limited driving competence on the part of the test candidate. For traffic safety reasons, it is thus perhaps not always meaningful to demand the driving task “Overtaking”, especially in the case of adverse weather conditions; we will also return to this point later.

driving tasks in use in Austria, likewise a reform-oriented country (see above), were not analysed further, because situation-related and situation-independent demands are there strongly intermixed in the test report and the specifications relating to performance assessment thus cannot be reconstructed from the test report alone.

Germany	Finland	Netherlands	Norway	Sweden
Joining/leaving traffic and changing lanes	Changing lanes	Joining and leaving motorways **Changing lanes**	Changing lanes and vehicle positioning	Joining the traffic on main roads Leaving the traffic on main roads Staying in lane/ changing lanes
Curves and connecting roads		Driving on straight and winding roads		Driving on narrow and winding roads
Passing and overtaking	**Overtaking**	**Overtaking other road users or passing obstacles**	**Overtaking and being overtaken**	Passing stationary vehicles Overtaking
Crossroads and junctions	Junctions and crossroads Crossroads controlled by light signals Crossroads with a priority direction Crossroads without priority direction	Approaching and passing crossroads	**Passing crossroads** Turning right and left at crossroads	Crossroads Crossroads with light signals
Roundabouts	Roundabouts		Roundabouts	Roundabouts
Rail-borne vehicles		Behaviour at special road features such as railway level crossings, pedestrian crossings, bus stops	**Railway level crossings**	Tram and railway level crossings
Pedestrians	**Pedestrians**		**Pedestrians**	Vulnerable road users
Cyclists	**Cyclists**		**Cyclists**	
	Trunk roads Motorways		Driving in different traffic environments (… motorways)	Motorways/trunk roads

Tab. 5: Situation-related driving tasks in the reform-oriented European GDE countries in comparison to the German reform proposal

It was already mentioned above (see footnote 12) that the systems of novice driver preparation in North and West European countries are often based on the so-called “GDE matrix” (Goals of Driver Education), a collection of hierarchical learning objectives for driver training which has also gained in significance at European level as a result of a broadly founded EU project to improve novice driver safety (HATAKKA, KESKINEN, GREGERSEN & GLAD, 1999). This development has been accompanied by an expansion of the objectives and content of driver training to include the promotion of safety-oriented attitudes. In this way, driver training acquires an educational, value-building purpose, the fulfilment of which, however, can hardly be verified in a methodically satisfactory manner within the framework of a driving test. In overseas countries, by contrast, driver training has traditionally followed a different approach which “describes an extended preparatory period of practical driving experience under low-risk conditions as a decisive form of qualification leading to the acquisition of safe driving and traffic competence. This perspective has been implemented comprehensively in the GDL (Graduated Driver Licensing) systems on the North American continent and in Australia/Oceania” (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014). The reform-oriented GDL countries in which the systems of measures leading to the granting of a driving licence have been optimised by elaborating innovative training curricula include New Zealand, Canada (Quebec), Australia (Victoria) and the USA. It thus also seems expedient to seek inspiration for further development of the German catalogue of driving tasks in the task specifications of these curricula. The driving tasks revealed are presented in Table 6 – again in comparison to the reform proposal for a future practical driving test for Germany.

Germany	New Zealand	Canada (Quebec)	Australia (Victoria)	USA (ADTSEA)
Joining/leaving traffic and changing lanes	Joining traffic from the kerbside Joining high-speed roads	Changing lanes	Joining traffic from the kerbside Joining dense traffic Changing lanes	Joining traffic from the roadside Leaving traffic by mounting the kerb Joining and leaving motorways Changing lanes
Curves and connecting roads	Curves Driving straight on single- and multiple-lane roads	Curves Driving straight	Roads with curves and straight sections	Curves Driving straight
Passing and overtaking	Overtaking		Passing Overtaking	Overtaking and being overtaken Overtaking and being overtaken on two-lane roads
Crossroads and junctions	Turning on roads with priority signs Turning at light signals	Crossroads	Passing crossroads Turning at crossroads Observing rules of priority, priority signs and light signals	Approaching crossroads Approaching multiple-lane crossroads Turning at crossroads Multiple turning lanes Separate lanes for left turns
Roundabouts	Roundabouts		Roundabouts	
Rail-borne vehicles				Sharing the road with trains and public transport
Pedestrians		Vulnerable road users, e.g. pedestrians, cyclists, motorcyclists	Recognising and reacting appropriately to potential hazards such as pedestrians and cyclists	Recognising and reacting appropriately to pedestrians
Cyclists				Recognising and reacting appropriately to cyclists
Motorways	Driving outside built-up areas	Driving in different environments (rural, urban, residential areas, and motorways)	Motorways	Driving on motorways Driving in rural environments Driving in urban environments

Tab. 6: Situation-related driving tasks in the framework curricula of reform-oriented GDL countries in comparison to the German reform proposal

It is to be noted that the driving task catalogue of the (non-binding) US American ADTSEA curriculum[56] also includes the demands "Handling advanced technologies (ESP)" and "Meeting, following and being followed by other vehicles on single- and two-lane roads (space management)". The handling of technologies, however, is according to our definitions not a driving task. With regard to "space management", this appears to represent an elementary, situation-independent demand. As such, it can be deemed important in the context of driver training, but in our opinion should no longer possess independent significance as test content by the time of the driving test.

[56] Stipulations relating to the necessity of formal driving school training vary between the individual US states, as do the curricular used, where appropriate. In most states, driver training is prescribed at least for certain age groups, or else associated with certain incentives (earlier granting of a learner driving licence possible, or fewer required hours of accompanied driving). The states which demand the completion of formal driver training nevertheless follow different approaches: Some elaborate curricula for the whole state, whereas others delegate this responsibility to local institutions (e.g. school districts) and instead define a legal framework; others again provide only this legal framework (CHAUDHARY, BAYER, LEDINGHAM & CASANOVA, 2011). Curricula are currently in use in 33 states (HIGHWAY SAFETY CENTER, Indiana University of Pennsylvania, 2009). In 2006, the American Driver and Traffic Safety Education Association (ADTSEA) elaborated a national model curriculum which each state can use and amend as it sees fit. The third revised version of the curriculum was published in July 2012; at that time, 15 states were already using the curriculum either as a whole or in parts.

The evaluation of Table 6 with the driving task catalogues of important GDL countries confirms the impression which was already gained from the analysis of training and test practice in the reform-oriented GDE countries: The German reform proposal for a future driving task catalogue is not only founded on replicable theoretical and methodical principles, but only corresponds to the – similarly scientifically founded – learning objectives and test contents for driver training and the practical driving test in countries with progressive systems of novice driver preparation. Through the analysis of all 36 countries considered (see above), it can be ascertained that driving tasks relating to the content categories "Changing lanes", "Driving through curves", "Passing and Overtaking" and "Negotiating crossroads and junctions" (including or supplemented by "Turning right and left") are to be found not only in all reform-oriented countries, but also in the majority of all other countries; these are furthermore traffic situations which frequently lead to accidents involving novice drivers (see below).

In some countries, test requirements are specified in the form of driving tasks which cannot actually be planned within the framework of the practical driving test, or else cannot be realised in all cases because they are dependent on regional circumstances (e.g. driving uphill/downhill), the time of day (e.g. driving in the dark) or weather conditions (e.g. driving on slippery roads or in the rain). In view of this limited practicability, the mastering of such situative test conditions should not be prescribed in regular driving tasks[57] for driving licence tests, irrespectively of whether their assessment appears desirable from the professional perspective. It is nevertheless expedient to follow the recommendation given by HAMPEL (1977, p. 67), namely "reasonable assessment … to record this type of additional demand"; we will return to this point in Chapter 3.4 ("Assessment and decision criteria").

Another aspect of variable test conditions is the requirement to drive in different traffic environments. Different rules and demands apply when driving through densely populated residential areas or driving on a motorway, for example. Consequently, task categories such as "Motorways" (or "Motorways and high-speed roads"), "Driving in rural and urban environments", "Driving within and outside built-up areas" or also more generally "Driving in different traffic environments" are specified as test requirements in a number of countries (see Table 5 and Table 6). In our opinion, the operationalisation of such demands as independent driving tasks seems suboptimal and inadequately distinct from the methodical and professional perspectives: Generally speaking, the essence of such test demands is that certain driving manoeuvres are to be performed at different speeds (e.g. changing lanes) or with special consideration given to particular groups of road users (e.g. playing children). It can thus be recommended – especially where, as in the present reform proposal, due consideration for vulnerable road users is already defined explicitly as a driving task – to specify different environmental conditions for the testing of selected driving tasks. In this connection, the aforementioned project support group reached the conclusion that, in future, certain driving tasks of the practical driving test should be performed under different framework conditions, where possible. These framework conditions should be described by way of a typology of traffic environments, road design features and speed recommendations based on the existing stipulations[58]. Agreement was reached on the following three categories:

1. Roads which can be used up to a maximum speed of 50 km/h (typically roads within built-up areas)
2. Roads which can be used up to a maximum speed between 50 and 100 km/h (typically roads outside built-up areas)
3. Roads which can be used at maximum speeds in excess of 100 km/h (typically motorways and similarly constructed roads).

Finally, it is conspicuous from the comparative analysis of national demand catalogues for the practical driving test, that many countries fail to distinguish clearly between situation-related demand standards (in the sense of driving tasks) and fundamental, situation-independent demands (e.g. vehicle operation, traffic observation or speed adaptation) which must be satisfied – albeit in varying manners in individual cases – to satisfy every driving task (see Chapter 3.3 "Observation categories

[57] STURZBECHER, BIEDINGER et al. (2010) discuss a possible need to test regionally specific demands against the background of test equity, and summarise that higher safety objectives "could possibly override the requirement of uniformity". They see the solution to this contradiction not in the elaboration of regionally varied candidate-oriented task catalogues, but rather in regional specifications relating to test locations and traffic environments: "The fundamental, elementary driving tasks would remain identical for all test candidates, and routes featuring special regional hazards would be used for the driving test in the whole country. Where is the disadvantage compared to the present solution, whereby it is a matter of random chance whether such route sections are included in the chosen test route?" (p. 98).

[58] The Examination Guidelines stipulate that the practical driving test is always to be conducted "within built-up areas" and "outside built-up areas (paragraph 5.8). It is furthermore intended that around half of the actual driving time is to be devoted to test route sections outside built-up areas, including motorways or similarly constructed roads, where possible (paragraph 5.9).

as situation-independent demand standards"). This methodical problem, which was already recognised by HAMPEL (1977), has thus remained essentially unsolved to this day.

Description of the driving tasks

From the analyses thus far, it can be recognised that a complete, professionally adequate and systematic description of demand standards for the practical driving test (including differentiation of candidate-oriented, situation-related driving tasks, situation-independent observation categories and local test prerequisites) has not yet been accomplished in Germany (as also in other driver licensing systems). To tackle these challenges within the framework of the present BASt project, a special working group "Driving tasks" ("AG Fahraufgaben") was founded. The experts and scientists appointed to this working group were representatives of the Federal Highway Research Institute (BASt), the Federation of Driving Instructor Associations (BVF), the Technical Examination Centres and the Bundeswehr, alongside the working group TÜV | DEKRA arge tp 21 and various other scientific institutions.[59]

To facilitate action-related description of the eight driving tasks defined by the project support group, together with their associated assessment criteria, the first step was to establish a scientific information base[60]. This collection contained all information found on the form and sequence of the ideally displayed behaviour actions for each driving task. Statements on the necessity of training, the forms of realisation within the framework of driving school training and the necessity of assessment during the practical driving test were also recorded. Finally, driving-task-related research was conducted into novice-specific accident causes and competence deficits: HAMPEL (1977) already suggested that, when elaborating demand standards, the chosen driving tasks should take into account the most frequent accident situations, and in particular those involving novice drivers. The historical sources yielded numerous substantiated findings on levels of difficulty and the hazard criticality of traffic situations and driving tasks (see above). As the safety relevance of individual driving tasks is subject to change over time (e.g. due to technical advances or changed road designs), however, particular value was attached to the evaluation of newer studies[61] on novice-specific competence deficits and accident causes.

Overall, the analysis has revealed that driving at inappropriate speed and problems with correct vehicle positioning, especially in curves and when overtaking, are the most common novice-specific competence deficits and accident causes (CAVALLO, BRUIN-DEI, LAYA & NEBOIT, 1989; JAMSON, 1999). Fatal accidents are particularly frequent when novice drivers are faced with these driving tasks on roads outside built-up areas – in other words at relatively high speeds; given the high accident potential, such driving manoeuvres should also be tested under the above traffic conditions. A further novice-specific source of accidents was identified in lost control over the vehicle, especially when turning at crossroads or changing lanes (DUNCAN, WILLIAMS & BROWN, 1991; ELLINGHAUS & STEINBRECHER, 1990). This all points to a necessity to test driving tasks which involve merging into and leaving traffic flows, changing lanes, negotiating crossroads and turning into side roads. Precisely these driving tasks are to be found in the described reform proposal for the German driving task catalogue.

As the final step, all content-related knowledge drawn from the aforementioned information base (ideal sequences of actions when mastering a driving task, safety relevance, training necessity, assessment necessity) was condensed into corresponding driving task descriptions. To this end, a scientifically founded draft for each driving task description was elaborated from the collected material, and subsequently discussed as a basis for further development by the branch experts in the working group "Driving tasks". The objective was to reflect the most important, safety-relevant demands of driving in real traffic as exhaustively and

[59] Members of the working group were: Michael Bahr (BASt), Arne Böhne (TÜV Rheinland), Gerhard von Bressensdorf (BVF), Peggy Frommann (Institute for Applied Research on Childhood, Youth and the Family - IFK), Peter Glowalla (BVF), Marcellus Kaup (TÜV SÜD), Christoph Kleutges (TÜV Rheinland), Susann Mörl (Institute for Prevention and Road Safety - IPV), Michael Palloks (IFK), Dr. Wilhelm Petzholtz (TÜV | DEKRA arge tp 21), Rolf Radermacher (TÜV NORD), Mathias Rüdel (TÜV | DEKRA arge tp 21), Dr. Andreas Schmidt (DEKRA), Stefan Sick (Bundeswehr – Central Military Vehicle Registration Office), Prof. Dr. Dietmar Sturzbecher (University of Potsdam), André Wagner (IFK). The working group met on 20 occasions over the period from September 2010 to January 2012.

[60] The information base evolved from an analysis of historical sources (McKNIGHT & ADAMS, 1970a; McKNIGHT & HUNDT, 1971a; HAMPEL, 1977; JENSCH, SPOERER & UTZELMANN, 1978; HAMPEL & KÜPPERS, 1982), international research results (see above), the stipulations of German driving licence legislation (Driving Licence Regulations, Examination Guidelines), the curricular guidelines of the German Federation of Driving Instructor Associations, and training materials from driving school publishers (Degener-Verlag, Verlag Heinrich Vogel).

[61] These studies included, above all: BARTL and HAGER (2006); BRAITMAN et al., (2008); GRATTENTHALER, KRÜGER and SCHOCH (2009); McCARTT et al. (2009); STATISTISCHES BUNDESAMT (2010).

disjunctly as possible in a modern catalogue of driving tasks, to describe these tasks together with event-oriented assessment criteria, i.e. with reference to driving errors or above-average driving performance, and to specify essentially error-free mastering of these driving tasks (see Chapter 3.4 "Assessment and decision criteria") as the minimum standard for the practical driving test. The outcome was a candidate-oriented driving task catalogue comprising the eight previously outlined driving tasks, each formulated as an action-related demand standard (see below).

How did the experts proceed when describing the driving tasks? The first section of each description consisted of a definition of the driving task concerned and a "fundamental action algorithm"[62] outlining the steps required to solve the task. To this end, the basic actions to be taken by the candidate to complete the particular driving task were depicted as a schematic diagram; the driving task was thus characterised by way of its typical action objectives and situational attributes. The component actions were arranged in the order in which they usually occur in practice, based on a prototypical "standard situation"; there are nevertheless certain actions which are performed simultaneously or follow on from each other without a distinct transition. In a few cases, the overall driving task was divided into separate subtasks within the framework of driving task definition. Given the similarities in the action sequences, and thus also the similar action and test demands placed on the candidate, these subtasks are recombined for the concluding competence-referenced assessment of the practical driving test as a whole (see below), but they must nevertheless be described separately and assessed with regard to particular events to facilitate test control and documentation of the candidate's performance. Event-related performance documentation serves to objectivise and found the summary competence assessment. Furthermore, individual subclasses of situation were described for each driving task or subtask, insofar as this was necessary to designate frequently occurring, still relatively complex subcategories of traffic situation where the candidate is required to demonstrate modified behaviour compared to the standard situation.

In the second section of the driving task descriptions, it was specified in each case, what is to be expected of the test candidate when performing the driving task concerned, specifically with regard to the five observation categories "Traffic observation", "Vehicle positioning", "Speed adaptation", "Communication" and "Vehicle control/Environment-aware driving" (see Chapter 3.3). These action-related minimum demands must always be satisfied, i.e. independently of any subclasses of situation. Where such subclasses are associated with modified or additional action demands, the description is supplemented to include these demands. Finally, the assessment criteria are listed for each individual observation category. These criteria are event-oriented assessment criteria which apply independently of situation subclasses. The spectrum of event-oriented assessment criteria covers "Normal performance"[63], "Examples for above-average performance", "Examples of simple errors" and "Serious errors". The recording and consideration of positive aspects of performance was to date merely recommended in the Examination Guidelines, but not stipulated as a binding requirement (see Chapter 3.4).

In the following, the driving tasks are to be described in brief; detailed driving task descriptions and the related assessment criteria can be found in Annex 1 to the present report.[64] The brief descriptions focus on the defined subtasks, where appropriate, and the corresponding subclasses of the relevant driving situations:

(1) "Joining/leaving traffic and changing lanes": This driving task refers in the broadest sense to lane-changing manoeuvres, and is divided into the subtasks "Joining traffic", "Leaving traffic" and "Changing lanes". It is generally necessary to change lanes when driving on motorways or similarly constructed roads and on other multiple-lane roads. For the subtask of "Joining traffic", the situation subclasses "Joining traffic in situations without special requirements", "Joining traffic without a merging lane or from a shortened merging lane (e.g. in connection with road works)" and "Joining traffic from a merging lane shared with exiting

[62] Algorithms are schematic action sequences leading to the solution of a problem. In this sense, driving competence is understood as the potential for problem-solving in motorised traffic, in the context of which the driving situations to be mastered are to viewed – in educational psychology terms – as the "problems".

[63] No explicit examples of "normal performance" are given, as this is understood to correspond to the action-related minimum demands described for each driving task.

[64] This driving task catalogue is the draft proposal elaborated by the project working group "Driving tasks" as per 28.02.2012. The working group met over the period from September 2010 to January 2012. Any differences between the driving tasks as described in the annex and in the main body of the present report result from the fact that the report authors introduced a small number of amendments (e.g. the name of driving task 1) after the final meeting of the working group. Such amendments must be taken into account when continuing development of the driving task catalogue within the framework of a revision project.

traffic" were specified. For the subtask of "Leaving traffic", the situation subclasses "Leaving traffic in situations without special requirements", "Leaving traffic without a merging lane or via a shortened exit lane (e.g. in connection with road works)", "Exit lanes with traffic signs (hazard signs, maximum permissible speed, yellow road markings)", "Leaving traffic via an exit lane shared with merging traffic" and "Leaving traffic in special traffic situations (e.g. tailbacks) were defined. With regard to changing lanes, the examiner is to distinguish between the subclasses "Changing lanes in low-density traffic", "Changing lanes in high-density traffic", "Moving into a lane into which it is possible to change from both sides" and "Zip-merging".

(2) "Approaching and negotiating curves and driving on connecting road sections": This driving task comprises the two subtasks "Curves" and "Connecting roads", wherein "Driving on connecting road sections" refers to the driving between concrete driving tasks (including the connecting road sections between two curves). Alongside normal curves, curves on mountain roads are defined as a separate situation subclass. For the driving on connecting road sections, a distinction is to be made between "Outside built-up areas" and "Within built-up areas".

(3) "Passing obstacles and overtaking driving or waiting vehicles": This driving task comprises the two subtasks "Passing obstacles" and "Overtaking", the fundamental action demands for which are in many respects similar. With regard to overtaking, it was already pointed out that this driving manoeuvre can be counted one of the most dangerous traffic situations and should for this reason also be tested, where possible. As likewise noted above, however, the driving test examiner should only require demonstration of this driving task where appropriately low-risk opportunities – measured against the usual training level of novice drivers – arise during the test drive. Two situation subclasses exist for the subtask "Passing", namely "Without priority rules" and "With priority rules"; for "Overtaking", the subclasses "Roads where the lanes for oncoming traffic must be used to overtake", "Overtaking single-track vehicles" and "Overtaking multiple-track (slow-moving) vehicles with high substructure and/or wide load" were defined.

(4) "Passing crossroads and junctions and turning right or left at crossroads and junctions": This driving task divides into the subtasks "Passing crossroads and junctions", "Turning right at crossroads and junctions" and "Turning left at crossroads and junctions". In these contexts, "Priority for traffic from the right", "With signs indicating priority", "With light signals" and "Controlled by a police officer" serve as situation subclasses.

(5) "Negotiating roundabouts": For this driving task, it is to be documented during the test drive whether the situation encountered belongs to the subclass "Roundabout comprising a single lane" or the subclass "Roundabout comprising several lanes".

(6) "Approaching and passing railway level crossings, approaching trams, and overtaking and being overtaken by trams": The two subtasks can be summarised into a single driving task "Rail-borne vehicles". Railway level crossings are subclassified according to the situations "Controlled crossing", "Uncontrolled crossing" and "Level crossing with special circumstances (port areas, presence of railway staff, stop-and-go traffic)". For the subtask "Trams", the situation subclasses "Tram travelling in the same direction on one lane of the road", "Tram travelling in the opposite direction on one lane of the road", "Tram travelling in the middle of the road" and "Tram turning off into another road".

(7) "Approaching and passing bus and/or tram stops, approaching and passing pedestrian crossings, approaching and passing pedestrians": This driving task comprises three subtasks which refer to pedestrians; the traffic environments "Crossroads and junctions" and "Roundabouts" are left aside, however, as it is more practicable to address consideration for pedestrians directly in connection with those driving tasks. For the subtask "Approaching and passing bus and/or tram stops", the situation subclasses "Buses and trams stopping at the right kerbside", "Bus and tram stops in the centre of the road", "Approaching school/public transport buses with warning indicators flashing" and "Stopping school/public transport buses with warning indicators flashing" were defined. Driving behaviour with regard to the subtask "Approaching and passing pedestrian crossings" can be documented by way of the two situation subclasses "Pedestrian crossing with additional signs" and "Pedestrian crossing without additional signs". The subtask "Approaching and passing pedestrians", finally, comprises the situation sub-

classes “Pedestrians crossing the road” and “Children”.

(8) “Approaching and passing cyclists”: This driving task takes into account all cyclists encountered aside from crossroads and junctions and other than in connection with passing and overtaking manoeuvres. The situation subclasses here include “Cyclists on the same road” and “Cyclists crossing the direction of traffic”.

As already mentioned at the beginning of the present chapter, the described driving task catalogue represents a selection from the entirety of all prototypical demand situations arising in road traffic, and comprises those demands which are not only relevant for road safety, but also generally suitable for testing within the time frame and under the regional circumstances of the driving test. There are many further driving tasks which must be taught additionally in the course of driver training and there gauged by way of learner assessment; the corresponding demand standards cannot be stipulated by the test system, however, and must instead be specified in the form of a framework curriculum. It is finally to be noted that the possibilities of the practical driving test remain limited: As the test drive is conducted in real traffic, it is not possible – e.g. for cost or planning reasons – to actually test all demands identified as relevant from the purely professional perspective. It appears hardly feasible, for example, to assess the candidate's driving competence in hazardous situations in the context of the practical driving test. In future, however, such components of competence could be addressed within the framework of supplementary simulative forms of testing, such as the “hazard perception tests” which are already in use in some countries (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014).

3.3 Observation categories as situation-independent demand standards

The practical driving test represents a complex performance assessment process. The focus of this process is placed on observation and assessment of the driving behaviour displayed by a driving licence applicant during a drive in real traffic. To measure driving performance, the driving test examiner applies the method of “systematic behaviour observation” (see above) and plays a decisive role in this process: He himself serves as part of the measuring instrument (FIEGUTH, 1977) and must follow a maximally controlled and uniform procedure in order to gather meaningful observations as a basis for systematic assessment of the test candidate's driving performance. This observation task is accomplished simultaneously with the processing of his corresponding planning and documentation tasks (see Chapter 3.5 “Control concept”); this circumstance limits the observation capacities of the driving test examiner.

In the context of systematic behaviour observation, the observer possesses two important “adjusting screws” with which to raise the uniformity and professional significance of his observations: Firstly, he can structure and plan the observation situation in accordance with appropriate demand standards; in this way, he maximises the probability of being able to observe behaviour useful for assessment. The driving test examiner is here supported by the availability of a driving task catalogue (see Chapter 3.2). Secondly, the observer can concentrate his attention on a precise and targeted search for the particular information required in subsequent assessment and decision processes, and thus objectivise his information processing. To this end, he uses observation categories to focus his perception on essential elements of behaviour and thereby to reduce the complexity of the subject under observation. Such observation categories relieve the observer provided they are limited to a meaningful number and cover the whole spectrum of the behaviour to be observed as exhaustively and disjunctly as possible (KANNING, 2004).

In the case of the practical driving test, the driving test examiner uses observation categories – e.g. “traffic observation” and “vehicle positioning” – to narrow the scope of his observation activities to those aspects of candidate behaviour which are important for an objective assessment of driving performance and a corresponding test decision (STURZBECHER, BIEDINGER et al., 2010). The observation categories thus provide an exact specification of what is to be taken into account by the driving test examiner with regard to the individual driving tasks and traffic situations. This serves on the one hand to define the actions expected of the driving test examiner, but at the same time describes demands to be met by the test candidate; observation categories thus possess a double function. As observation categories are selected such that they can be observed in connection with every driving task and in every corresponding traffic situation (e.g. proper vehicle operation is required in every traffic situation), they represent situation-independent, behaviour-related demand standards which – from the candidate's perspective – supplement the catalogue of driving tasks. Observation categories furthermore facilitate the specification of assessment criteria: If the test-relevant aspects of candidate behaviour are already identified in the form of observation categories, it is easier to determine the conditions under which this behaviour is to be considered inadequate. Observation categories can thus serve to structure assessment criteria relating to test performance (e.g. driving errors, particularly good performance) and establish references to the driving tasks. In this way, they enable an efficient documentation of test performance in a test report.

Given the double function of observation categories as demand standards for both the driving test examiner and the test candidate, they can be operationalised in different manners. On the one hand, they can be formulated as task instructions for the examiner, with specification of those aspects of candidate behaviour to which particular attention should be paid. On the other hand, alongside this instructional, behaviour-oriented description, it is also possible – if the driving licence applicant is seen as a protagonist in the practical driving test – to elaborate competence-referenced definitions, because the behaviour demands and associated performance expectations (or assessment criteria) contained in observation categories at the same time specify the elements and areas of competence to be demonstrated by the candidate.

The use of situation-independent demand standards and observation categories in the German system of driver licensing

Since when has the test psychology concept of "observation categories" been in use in the (German) system of driver licensing, and how has this concept developed in the meantime? The concept of "observation categories" still played no role in the demand analysis for the activity of "driving" by McKNIGHT and ADAMS (1970a): The authors sought to provide the fullest possible description of the actions associated with driving a motor vehicle in the broadest sense, thereafter to dissect the process of driving in the narrower sense, and finally to weight the identified tasks and subtasks in accordance with their significance for road safety (see above). They were not (yet) concerned with questions of how to properly record – i.e. observe – and assess the action sequences which were deemed especially significant for smooth driving and road safety. It was only in connection with the subsequent elaboration of evaluation instruments by McKNIGHT and HUNDT (1971a) that – albeit minimal – attention was given to this topic.

The inductive approach taken by McKNIGHT and ADAMS (1970a) produced a broad spectrum of traffic-related action patterns; little was done to analyse the content correlations between such actions, however, and despite the subsequent general categorisation into groups, this yielded no real hierarchical system. In the category "Basic control" under "On-road behaviour", for example, the task "Accelerating" stands on equal level alongside "Speed control", presumably because the authors based their grouping on driving speed as an aspect of vehicle operation rather than a decision in the sense of speed adaptation, which would have belonged more appropriately to the category of situation-independent demands under the heading of "General driving". At the same time, the categorisation ignored the overlapping of situation-independent and situation-related driving tasks, and likewise their different forms and levels of complexity: In practice, for example, the general, situation-independent driving task "Observation" is always a component of the situation-related driving task "Changing lanes"; driving tasks related to traffic conditions, such as "Urban driving" or "Highway driving", are much more complex in nature than a driving task "Observation" and also constitute a much less distinct demand standard on account of their variability. Such correlations and differences were barely touched upon in either the task analysis by McKNIGHT and ADAMS (1970a) or the later elaboration of evaluation instruments by McKNIGHT and HUNDT (1971a), and they were certainly not used in any way to aid construction of an efficient observation method: McKNIGHT and ADAMS themselves emphasised their heuristic, pragmatic approach to categorisation, and McKNIGHT and HUNDT (1971a) were also apparently more concerned with matters of test organisation than theoretical criteria for the structuring of driving tasks, because

- the tasks of the category "Basic control" are examined on a test ground in the Driving Fundamentals Test,
- the tasks of the category "Situational behaviour" are reflected in the Driving Situations Test and performed during the drive in real traffic, and
- the tasks of the category "General driving" are operationalised indirectly as variable framework conditions (e.g. urban traffic, rural traffic) or as assessment criteria in connection with the situation-related driving tasks (e.g. observation errors).

The described peculiarities of the demand systems proposed by McKNIGHT and ADAMS (1970a) or McKNIGHT and HUNDT (1971a) must not be considered methodically problematic where the intention is solely to develop training curricula and evaluation instruments for learner assessment. As soon as they are to serve as the starting point for the process of a driving licence test, however, the quality standards applicable to psychological testing acquire much greater importance: Given its significance for both public safety and the individual citizen (e.g. restriction of access to mobility, time and financial burdens), it is expected that the contents of a driving licence test have been defined sharply and with systematic structure as a basis for the desired methodical reliability and content validity, and that they can be examined and assessed in an economically efficient manner. It thus seems only logical that HAMPEL (1977) should open scientific study addressing the practical driving test with an analysis of existing test psychology approaches, with the objective of sharpening definitions of the subjects of observation: He compared the methodical systems of seven common forms of traffic psychology observation relating to driving behaviour – including the Road Test by McGLADE (1965), which was in use for testing in the USA, and the Driving Situations Test proposed by McKNIGHT and HUNDT (1971a) – and analysed the corresponding subjects of observation by way of tabular overviews and factor analyses, though without identifying a practicable solution. The term "observation category" was not defined explicitly, however; in fact, varying terminology was used[65] and referred above all to both situation-independent demands and situation-related driving tasks.

HAMPEL (1977) took a big step towards the formulation of appropriate observation categories when he analysed the recording of test performances and the corresponding modes of assessment from the perspective of practicability. With regard to the desirable complexity of observation units, for example, he found that the recording and assessment of elementary driving behaviour in the manner proposed for the Driving Situations Test by McKNIGHT and HUNDT (1971a) would be beneficial for the psychometric quality of the test, but would at the same time place excessive burdens on the examiner (see above). On the other hand, he warned against overly complex observation units, especially where they already entail psychological interpretation or generalisation: "It is generally difficult to imagine how … in driving tests where the result has serious consequences for the candidate, the judgements could be based on mere description of the impressions gained by the examiner. The candidate will hardly be satisfied with the opinion that his driving was 'careless', and will instead want to know how exactly this carelessness was manifested. He will presumably also be entitled to assert this claim by way of legal action, where necessary. The consequence is that only primary characteristics are suitable for use in driving tests. Secondary characteristics can only serve to round off the picture" (HAMPEL, 1977, p. 94). In accordance with this finding, HAMPEL (ibid.) noted that concrete references to directly observed behaviour – i.e. primary characteristics (see above) – were also clearly dominant in the judgement systems used for driving tests.

From the aforementioned analyses, HAMPEL rightly concluded that adequately complex situation-related driving tasks which can be assessed – without psychological interpretation – on the basis of determined errors represent suitable subjects for observation. Accordingly, he demanded the operationalisation of test tasks "as typical driving situations which are to be handled by the test candidate" (1977, p. 158); finally, as already described in Chapter 3.2, he elaborated a catalogue of driving tasks (HAMPEL & KÜPPERS, 1982). With regard to the parallel documentation of task-specific driving performance by way of an "EDP-ready form for the recording of test results in accordance with VdTÜV Notice 731", however, one significant problem was revealed: Given the diversity of possible (incorrect) behaviour to be observed and assessed in conjunction with performance of the driving tasks, it is not practicable for the driving test examiner to retain the entirety of his observations in his working memory, let alone to document these observations in a list-style report form. What is needed here is pre-structuring of the possible test performance observations (both driving errors and characteristics of positive performance) into situation-independent action contexts in which such performance could be displayed (e.g. vehicle operation, traffic observation). These overarching action contexts represent observation categories: They permit efficient classification, referencing and recording of the driving performance in accordance with the set driving tasks; in this way, they relieve and provide orientation for the driving test examiner. For this function to be realised, the number of categories must be strictly limited to the most important general driving demands: "The greater the number of elements of behaviour which must be observed concurrently or in close succession, the more difficult it becomes to record observations adequately and without omissions. According to the experience gained from trials, it seems hardly feasible for more than

[65] HAMPEL (1977) speaks of "observation categories" (p. 78) when referring to "driving behaviour analysis" after v. KLEBELSBERG (1970), but of "assessment criteria" (p. 92) in the comparative tabular overview, and later of "behaviour categories" (p. 178), "tasks" (p. 180) or "behaviour attributes" (p. 182) in conjunction with other methods.

four characteristics of driving behaviour to be observed and subsequently recorded with adequate precision in a typical traffic situation" (KROJ & PFEIFFER, 1973, p. 21ff.).

Given the presented correlations, it seems plausible that situation-independent demands standards with the function of observation categories were first to be found – albeit without being designated as such – in the matrix-style test report forms[66] used by individual Technical Examination Centres between the mid-1970s and 1996: It was only through the cross-referenced, multi-dimensional arrangement of

– situation-related demand standards grouped according to typical traffic situations ("driving tasks") and
– assessment standards grouped according to typical, situation-independent action contexts ("observation categories")

that it was possible to document driving performance in a more efficient manner, namely in the cells of the ensuing matrix. Such arrangements were used for the first time in the previously mentioned "TÜV Rheinland draft for a matrix with which to record driving errors", which was presented in 1977; from the methodical point of view, therefore, this matrix "is to be considered the origin of driving tasks and observation categories in the current-day meaning in driver training and testing in Germany" (HAMPEL & STURZBECHER (2010, p. 57).[67] Unfortunately, HAMPEL and KÜPPERS (1982) later neglected to follow up this point in their work to further develop the practical driving test, for example when they conducted empirical studies (albeit without the founding of a demand analysis) to determine the demand situations or driving tasks which should be mastered by candidates in the practical driving test in Germany. On the other hand, their commission was merely "to compile practicable demand criteria for the locations at which driving tests are conducted" (ibid. p. 13), in other words to describe the local prerequisites for the realisation of driving tasks. Consequently, neither observation categories nor assessment criteria were placed at the focus of discussion within the course of their investigations.

It was not until 2008, that STURZBECHER, BIEDINGER et al. (2010) tackled the task of deriving scientifically founded observation categories. To this end, within the framework of their project "Practical driving test – Foundations and possibilities for optimisation", they subjected the (still applicable) stipulations relating to situation-independent demands in driver licensing legislation – specifically Annex 7 FeV, 2.1.5 and the Annexes 3 (Basic driving manoeuvres) and 10 (Test drive) to the Examination Guidelines – to thorough content analysis, in order to be able to describe the desired "driving behaviour of the test candidate in appropriate safety-oriented content and, from the methodical point of view, by way of a correspondingly limited number of clearly discrete categories, without departing from the observation standards prescribed by the German legislation on driver licensing" (ibid., p. 109). The starting point for their critical appraisal was the wish, jointly with experts from the Technical Examination Centres, to identify methodical streamlining and restructuring potential, and thereby to condense the specified observation contents into representative and maximally discrete categories of driving behaviour.

The content analysis revealed that the 15 observation categories anchored in Annex 7 to the Driving Licence Regulations and in Annexes 3 and 10 to the Examination Guidelines contained references to all essential, safety-relevant behaviour which was to be demonstrated by the candidate during a driving test and observed accordingly by the driving test examiner; the categories could nevertheless be structured more systematically and more efficiently. On this basis, STURZBECHER, BIEDINGER et al. (2010) proposed five observation categories which were to take the place of the current specifications, namely "Traffic observation", "Vehicle positioning" (possibly distinguishing "Use of the road" and "Safety margins"), "Speed adaptation", "Communication and adaptation to traffic" and "Vehicle control" (possibly distinguishing "Action sequences" and "Environment-aware driving").

From the legal perspective, the proposal was welcomed almost immediately. JAGOW (2010, p. 147), for example, commented: "This new concept is to be welcomed, because the observation categories would then in future be differentiated clearly from the test tasks, and furthermore appear both less complex and easier to handle." It could furthermore be shown that the recommended observation categories were very similar in terms of scope and contents to those implemented in methodically progressive driver licensing systems in a number of other European countries (STURZBECHER, MÖRL & GENSCHOW, 2010). At the same

[66] A specimen of the "EDP-ready form for the recording of test results in accordance with VdTÜV Notice 731" can be found in HAMPEL (1977, p. 47); a specimen of the DEKRA report used up to 1996 is presented in HAMPEL and STURZBECHER (2010, p. 69).

[67] A detailed account of the historical development of observation categories in connection with the elaboration of "driving error catalogues" in the 1970s can be found in HAMPEL and STURZBECHER (2010).

time, however, the limitations of the content analysis became apparent, and thus, in turn, the necessity of further scientific treatment before the concept could be anchored in driver licensing legislation: "Modernisation of the observation contents also appears to be quite evidently necessary, but requires not only the reviewing and – insofar as necessary and meaningful – reorganisation of the current definitions, but also a scientific analysis of present-day traffic demands; this is neither feasible nor intended in the present context, but remains outstanding" (STURZBECHER, BIEDINGER et al., 2010, p. 109). To facilitate implementation of these proposals, BÖNNINGER et al. (2010, p. 173) recommend that the findings be taken "as a basis for expert ratings and a traffic-psychology-oriented demand analysis, in order to further perfect this list of observation categories."

Observation categories for an optimised practical driving test

The above thoughts and findings, along with the derived recommendations for further development, were taken up in the present project. As regards content, a starting point for the elaboration of future observation categories was provided by the corresponding lists of continuously applicable, situation-independent demands presented by McKNIGHT and ADAMS (1970a; "On-road behaviour", categories "Basic control" and "General driving"), by TÜV Rheinland (HAMPEL & STURZBECHER, 2010) and in the proposal developed by STURZBECHER, BIEDINGER et al. (2010). In addition, as for the elaboration of the driving task catalogue, due consideration was given to the stipulations of EU regulations and international standards, as well as to the latest state of research into novice-typical driving competence deficits and the principal causes of accidents involving novice drivers. The draft for an optimised category list elaborated on this basis was subsequently discussed in the project support group (see above) and developed into a reform proposal for future observation categories (see Table 7).

A comparison between the reform proposal presented here and the aforementioned category list of STURZBECHER, BIEDINGER et al. (2010) shows that they differ only slightly in the category designations used, but not in terms of content. The observation categories "Traffic observation", "Vehicle positioning" and "Vehicle control" are essentially identical; in the latter case, the proposal that demands relating to environment-aware driving be operationalised as aspects of vehicle control (STURZBECHER, BIEDINGER et al., 2010) was followed, as a corresponding manner of driving is always inseparable from vehicle control actions such as gear changing, accelerating and braking. For this reason, it appeared expedient to combine both aspects in a single observation category "Vehicle control/Environment-aware driving". The category designation "Speed adaptation" was preferred over the previously used designation "Speed regulation" in order to emphasise that speed control in this sense serves not least to integrate the vehicle into an overall traffic flow, in other words to adapt driving speed – within the framework of the applicable speed limits – to that of other motorised road users. To minimise content overlaps between the observation categories, the aspects of adaptation to traffic was subsequently deleted from the observation category "Communication". Overall, it can be said that the observation categories recommended in the draft by STURZBECHER, BIEDINGER et al. (2010) were subjected to further sharpening in respect of the addressed content.

It should be mentioned at this point that the possible specification and description of a further observation category "Observing right-of-way" was a subject of long, heated discussion among the experts in the project support group, not least because the EU Directive on Driving Licences stipulates a similar category – 9.3.4 "Priority/giving way" – as an observation standard (see below). The conclusion reached through this discussion, however, was that such an observation category would not be equally applicable to all driving tasks (e.g. "Approaching and negotiating curves and driving on connecting road sections"); it would thus not necessarily represent a situation-independent demand standard. Consequently, it would be better to reflect the associated elements of behaviour – in the sense of driving errors or failure to observe rules – in assessment criteria for the relevant driving tasks, (see Chapter 3.4).

If we compare the observation categories of the present reform proposal with the situation-independent action demands derived from their scientific driving task analysis by McKNIGHT and ADAMS (1970a), then many points of coincidence can be found (see Table 7). All observation categories of the reform proposal possess content equivalents under the "On-road behaviour" categories "Basic control" and "General driving" of the demand catalogue elaborated by McKNIGHT and ADAMS (1970a). It is true that the latter authors describe the tasks "Navigation" and "Compensating for physical limitations" as further elements of these categories, but these demands can be deemed to be properly relevant only for driver training, rather than for the practical driving test: During the test, the candidate usually receives more or less concrete driving instructions from the examiner, and independent navigation is thus essentially unnecessary; furthermore, it is generally reasonable to expect the candidate to display a rather cautious, risk-avoiding and safety-oriented manner of driving as compensation for the driving risks associated with his current level of training.

McKNIGHT and ADAMS (1970a) also name the situation-independent action demands "Urban driving", "Freeway driving" and "Highway driving" as "General driving" tasks; the classifications of our proposal, however, treat these demands more appropriately as framework conditions under which the realisation of (situation-related) driving tasks should be varied in the course of testing (see above). Vehicle positioning is not mentioned explicitly as an independent demand by McKNIGHT and ADAMS (1970a) and is instead operationalised, without exception, in all tasks which refer to driving manoeuvres. It is thus de facto nevertheless present as an essential situation-independent driving demands and – in accordance with our demand classifications and terminology – as an observation category. Overall, therefore, the present reform proposal is supported by the methodically careful and empirically oriented study approach of McKNIGHT and ADAMS (1970a).

Situation-independent action demands – Task analysis (McKNIGHT & ADAMS 1970a)	**Situation-independent action demands – TÜV Rheinland proposal** (1977)	**Reform proposal relating to observation categories** (2012)
Surveillance	Traffic observation	Traffic observation
Operationalisation in relevant driving tasks (e.g. changing lanes, negotiating curves, negotiating intersections, passing)	Road area use Safe distance to other road users	Vehicle positioning
Speed control *(additional operationalisation in relevant driving tasks)*	Driving speed too slow or too fast	Speed adaptation
Reacting to traffic	Use of indicators	Communication
Pre-operative procedures, starting, accelerating, steering, stopping, backing up, skid control	Vehicle handling	Vehicle control / Environment-aware driving
Navigation		
	Obstructing or endangering of other road users	
	Intervention by the driving instructor	
	Observance of traffic signs or traffic rules	

Tab. 7: Comparative overview of the situation-independent action demands covered by the task analysis by McKNIGHT & ADAMS (1970a), the task proposal elaborated by TÜV Rheinland in 1977 (HAMPEL & STURZBECHER, 2010) and the observation categories of the present reform proposal for optimisation of the practical driving test in Germany

From a comparison of the observation categories developed for the present reform proposal against the situation-independent action demands of the task proposal elaborated by TÜV Rheinland in 1977 (HAMPEL & STURZBECHER, 2010), it is similarly evident that the five categories suggested here correspond quite precisely – in terms of their content – with the observation categories of the TÜV Rheinland proposal[68] (see Table 7): "This brings historical lines of methodical development to light, and illustrates, moreover, that it is less a fundamental upheaval in respect of content, but rather methodical reformulation and streamlining which is needed to optimise the system of the practical driving test in Germany" (HAMPEL & STURZBECHER, 2010, p. 110).

Overall, the comparisons of the present reform proposal with its scientific precursors confirm the conclusion already reached by BÖNNINGER et al. (2010) with regard to the similar optimisation recommendations of STURZBECHER, BIEDINGER et al. (2010): "The result which emerged was a streamlined category list …, which corresponds essentially to that which was already shown to be expedient over thirty years ago …, was used by a number of Technical Examination Centres in very similar form and to methodical advantage within the framework of their test reports up to 1996 …, and is above all still used almost with exception by today's methodically innovative European countries" (BÖNNINGER et al., 2010, p. 173). This latter statement, and the necessity to verify whether the future observation categories for the German licensing system are conformant with the EU stipulations, directs our attention back to international practice – as was already the case in respect of the driving tasks.

The fundamental observation standards to be applied in practical driving tests in the member states

[68] The categories "Obstructing or endangering of other road users", "Intervention by the driving instructor" and "Observance of traffic signs or traffic rules" were not taken over into the reform proposal; on the other hand, they also appear dispensable, as the comments made above with regard to a possible observation category "Observing right-of-way" apply similarly here: They can be better operationalised in the assessment criteria of the relevant driving tasks and do not constitute (separate) areas of driving competence.

of the EU are described in the EU Directive on Driving Licences 2006/126/EC of 20th December 2006, Annex II, Section 9 “Marking of the test of skills and behaviour”, under paragraph 9.3:

“9.3.1. Controlling the vehicle; taking into account: proper use of safety belts, rear-view mirrors, head restraints; seat; proper use of lights and other equipment; proper use of clutch, gearbox, accelerator, braking systems (including third braking system, if available), steering; controlling the vehicle under different circumstances, at different speeds; steadiness on the road; the weight and dimensions and characteristics of the vehicle; the weight and type of load (categories BE, C, CE, C1, C1E, DE, D1E only); the comfort of the passengers (categories D, DE, D1, D1E only) (no fast acceleration, smoothly driving and no hard braking);

9.3.2. Driving economically and in an environmentally friendly way, taking into account the revolutions per minute, changing gears, braking and accelerating (categories BE, C, CE, C1, C1E, D, DE, D1 und D1E only);

9.3.3. Observation: all-round observation; proper use of mirrors; far, middle, near-distance vision;

9.3.4. Priority/giving way: priority at crossroads, intersections and junctions; giving way at other occasions (e.g. changing direction, changing lanes, special manoeuvres);

9.3.5. Correct position on the road: proper position on the road, in lanes, on roundabouts, round bends, suitable for the type and the characteristics of the vehicle; prepositioning;

9.3.6. Keeping distance: keeping adequate distance to the front and the side; keeping adequate distance from other road users;

9.3.7. Speed: not exceeding the maximum allowed speed; adapting speed to weather/traffic conditions and where appropriate up to national speed limits; driving at such a speed that stopping within distance of the visible and free road is possible; adapting speed to general speed of same kind of road users;

9.3.8. Traffic lights, road signs and other indications: acting correctly at traffic lights; obeying instructions from traffic controllers; acting correctly at road signs (prohibitions or commands); take appropriate action at road markings;

9.3.9. Signalling: give signals where necessary, correctly and properly timed; indicating directions correctly; taking appropriate action with regard to all signals made by other road users;

9.3.10. Braking and stopping: decelerating in time, braking or stopping according to circumstances; anticipation; using the various braking systems (only for categories C, CE, D, DE); using speed reduction systems other than the brakes (only for categories C, CE, D, DE)” (EUROPEAN PARLIAMENT & EUROPEAN COUNCIL, 2006, L 403/43).

Comparing the listed observation requirements of the EU Directive on Driving Licences with the proposed observation categories for a future practical driving test in Germany, as elaborated within the framework of the present project (see Table 8), it can be ascertained that – with the sole exception of item 9.3.4 “Priority/giving way” – the reform proposal satisfies all the requirements of the EU directive: There are certain minor technical deviations and overlaps in the structural assignments of the individual categories (e.g. it is possible to view “Braking and stopping” as an aspect of both speed adaptation and vehicle control), but it is generally reasonable to assume congruence at the level of content – taking into account the aforementioned discussion regarding an independent observation category “Observing right-of-way”.

Observation demands of the EU Directive of Driving Licences (2006)		Proposal for optimisation of the observation categories (2010)
9.3.3 **9.3.8**	Observation Traffic lights, road signs and other indications[1]	Traffic observation
9.3.5 **9.3.6**	Correct position on the road Keeping distance	Vehicle positioning
9.3.7 **9.3.10**	Speed Braking and stopping	Speed adaptation
9.3.9 **9.3.8**	Signalling Traffic lights, road signs and other indications	Communication
9.3.1 **9.3.2**	Controlling the vehicle Driving economically and in an environmentally friendly way, taking into account the revolutions per minute, changing gears, braking and accelerating	Vehicle control / Environment-aware driving
9.3.4	Priority/giving way	

Tab. 8: Comparative overview of the observation categories specified in the EU Directive on Driving Licences and the present reform proposal for optimisation of the practical driving test in Germany

Additional note:

[1] Aspects of this multi-dimensional observation demand of the EU Directive on Driving Licences are to be found under two observation categories of the German reform proposal, namely “Traffic observation” (traffic lights/road signs) and “Communication” (other indications).

As already in the case of driving tasks, brief analysis of the use of observation categories in international practice, and especially in the test procedures implemented in selected GDE and GDL countries, is intended to identify innovative national concepts which go beyond the minimum demands of the EU Directive on Driving Licences.

Comparing the systems of testing in the aforementioned 36 countries, it can be noted that almost all the countries considered specify aspects of candidate driving behaviour which are to be observed and taken into particular account by the driving test examiner. In many cases, general situation-independent observation standards also exist in the sense of the observation categories recommended here. These categories are occasionally supplemented with lists of concrete indicators, which offer more or less precise and behaviour-referenced opportunities to document test and driving performance (usually driving errors, sometimes also aspects of positive performance). There are nevertheless differences – as shown in the following discourse – with regard to the manner of definition and the degree of differentiation with which the situation-independent demands are formulated. Generally speaking, there were no noteworthy observation standards to be found with content exceeding the demands of the EU Directive on Driving Licences.

Among the reform-oriented GDE countries of Northern Europe, two fundamentally different approaches to the specification of situation-independent demand standards for the practical driving test can be identified: In Finland, the Netherlands and Norway, observation categories are defined in the same way as in Germany, in other words concrete observation instructions for the examiner; from the point of view of content, these category definitions correspond essentially to the German reform proposal. In Sweden, on the other hand, the system describes four “competence categories” to which specific training objectives and the correspondingly desired driving behaviour are assigned as a basis for assessment of the candidate's performance in different traffic situations (see Table 9).

If we look at the former group of countries in more detail, the Dutch system can be seen to stand out with the relatively high number of 13 observation categories. This gives rise to the question as to whether this multitude of observation categories is able to properly fulfil the initially mentioned relief, orientation and objectivisation functions. Experience gained in psychological testing (e.g. BARTHELMESS, 1976; KROJ & PFEIFFER, 1973) and the trials conducted with the TÜVIS draft of 1978[69] both indicate that, in line with international practice, these functions are guaranteed ideally with a total of five or six categories: The necessary differentiation and situation specificity for driving performance assessment is best safeguarded not by way of an increased number of observation categories, but rather through combination of a smaller number of categories with situation-related driving tasks and assessment criteria in a multi-dimensional assessment matrix (see Chapter 3.4), especially when an electronic method of documen-

[69] HAMPEL and STURZBECHER (2010, p. 63) noted in this connection: “Overall, it can be said that especially the classifications of ‘driving situations’ or driving tasks in the VdTÜV recommendations of 1978 seem unconvincing and even a little confused, whereas the observation categories underlying the ‘driving errors’ are in their essence acceptable.”

tation is to be employed (see Chapter 4). Assuming strict application of this awareness, it would be more appropriate to operationalise category contents such as compliance with traffic rules (see Finland, the Netherlands and Sweden in Table 9) or a safe manner of driving (see the Netherlands and Sweden) in the form of assessment criteria. At the same time, the subcategories defined under speed adaptation and vehicle positioning in the Dutch test report could be combined.

Germany	Finland	Netherlands	Norway	Sweden[1]
Traffic observation	Observation	Observation Attentiveness towards other road users	Traffic observation	
Vehicle positioning	Correct vehicle positioning	Position on the road	Positioning on the road Adaptation to traffic flow	
		Safe distance		
Speed adaptation	Speed control	Speed	Speed adaptation	
		Slowing down/braking/stopping		
Communication	Interaction	Signalling and reaction to the signals of other road users	Signalling	
Vehicle control / Environment-aware driving		Preparation for driving, operation and control	Vehicle control	Vehicle knowledge and handling
		Eco-friendly driving		Eco-friendly and economical driving
	Compliance with traffic rules[2]	Reaction to traffic lights/signals given by police officers		Traffic regulations, including speed
		Reaction to other prescribed signals		
		Observing right-of-way		
	Judgement			
		Safe and convincing manner of driving		Traffic safety and behaviour, including speed

Tab. 9: Situation-independent observation categories in the reform-oriented European GDE countries in comparison to the German reform proposal

Additional notes:

[1] In Sweden, these categories are termed “competence categories”. It remains unclear, however, how exactly these categories are operationalised and which elements of driving behaviour are assessed thereby.

[2] In the German category system, the corresponding aspects of driving behaviour are to be found under “Traffic observation” (e.g. traffic signs) or “Vehicle positioning” (e.g. road markings).

A glance into selected driver training curricula (New Zealand: “Driver Training Syllabus, Learner Stage” and “Driver Training Syllabus, Restricted Stage”; USA: ADTSEA curriculum) and test materials (Canada, British Columbia: Test report; Australia, Victoria: “What you need to know about your Drive Test”) in use in reform-oriented GDL countries already confirms the German reform proposal and the findings to date (see Table 10): The situation-independent demand standards “Traffic observation”, “Vehicle positioning”, “Speed adaptation”, “Communication” and “Vehicle control” are found once more, albeit occasionally – as above in the case of the Netherlands – with slightly narrower differentiation (see “Vehicle positioning” in Victoria/Australia) or with alternative designations. In New Zealand and British Columbia/Canada, there is in both cases an independent category relating to hazard perception; the required behaviour on the part of the candidate, however, could also be viewed as the result of effective traffic observation and thus adequately operationalised under this latter category.

The analysis of test reports from 25 countries similarly indicates that the five observation categories defined for the German reform proposal mirror the essential core of the situation-independent demand standards in use on an international scale: More than half of the countries specify such categories in the sense of explicit observation instructions for the driving test examiner. In addition, many countries expect the candidate to demonstrate an environment-aware and fuel-saving manner of driving. A closer appraisal of the test reports also reveals that most reports lack clear specifications of the concrete candidate behaviour to be observed by the examiner during the driving test. On the other hand, it is preferable to tackle such specifications within the framework of correspondingly referenced assessment criteria, where they can in turn be combined with concrete traffic situations or driving tasks, rather than when defining observation categories. The lack of precise specifications is probably also attributable to the fact that a clearly organised and practicable multi-dimensional assessment matrix can hardly be realised in the form of a “paper-and-pencil” report and must instead await electronic implementation.

Germany	New Zealand	Canada (British Columbia)	Australia (Victoria)	USA[1] (ADTSEA)
Traffic observation	Observation	Observation	Observation	The candidate uses the vehicle displays in an appropriate manner.
	Use of mirrors			The candidate performs adequate visual checks (mirrors and shoulder check).
	Shoulder check			
Vehicle positioning	Positioning	Distance	Following distance	The candidate uses appropriate steering actions to drive in a straight line, to negotiate curves, to turn at junctions and to turn around.
	Gap selection		Lateral position	
			Gap selection	
Speed adaptation	Speed choice	Speed	Speed choice	The candidate uses the accelerator and brake pedals to control speed and to stop the vehicle.
Communication	Signalling	Communication	Signal use	The candidate uses the vehicle's signalling devices in an appropriate manner.
Vehicle control / Environment-aware driving	Braking[2]	Steering	Control	The candidate uses safety, monitoring and display devices in an appropriate manner.
	Hazard detection and response	Hazard perception		The candidate is able to make correct judgements, e.g. gap assessment: The candidate is able to correctly assess gaps between vehicles.[3]

Tab. 10: Situation-independent demand standards and observation categories in the test reports or framework curricula of reform-oriented GDL countries in comparison to the German reform proposal

Additional notes:

[1] The curriculum provides descriptions, but does not use specific designations.

[2] “The candidate must use the brake efficiently and smoothly, and must not drive either with the vehicle in neutral gear or with a foot on the clutch.”

[3] In our opinion, the performing of such manoeuvres belongs to the category of vehicle positioning.

To summarise, the analysis of international test procedures showed that the situation-independent demand standards or observation categories are defined with very divergent degrees of differentiation in the individual countries: In some countries, only very general situation-independent observation standards exist, and are then usually termed “competence categories” (e.g. in Sweden); in other countries, lists of concrete indicators are provided – occasionally as supplements to observation categories – with more or less precise definitions of the aspects of behaviour to be assessed (e.g. in Austria, France, Great Britain, Ireland, South Africa and Spain). This suggests that – as to date also in Germany – the methodically desirable distinction between observation categories and assessment criteria is often neglected. Furthermore, it is conspicuous that demand standards referenced to particular traffic situations (in our terminology “driving tasks”) and situation-independent demand standards (in our terminology “observation categories”) are not seldom mixed together, without there being any recognisable methodical system to explain this approach.[70]

It has already been emphasised on several occasions, referring to relevant scientific sources (e.g. McKNIGHT & ADAMS, 1970a; HAMPEL, 1977; BÖNNINGER & STURZBECHER, 2005; STURZBECHER, BÖNNINGER & RÜDEL, 2010), that the demand and assessment standards for driving licence testing must be aligned specifically to novice-specific competence deficits and accident causes, and that this is expedient for various reasons – not only as a means to enhance road safety, but also for cost control. The same applies correspondingly to the observation categories for the practical driving test: They must focus on those aspects of candidate behaviour which represent an increased risk of accident involvement for novice

[70] This applies at least with regard to the test reports analysed for the current project; whether and to what extent it applies also to national examination guidelines or to manuals describing methodologies for psychological testing, where they exist, could not be determined within the framework of the present studies.

drivers. Over a series of scientific studies[71], it has been possible to identify those hazardous behaviour patterns and competence deficits which can typically be observed at the early stages of driving competence acquisition and thus during the initial phases of participation in motorised road traffic under the supervision of a driving instructor or other accompanist. These competence deficits on the part of novice drivers can usually be assigned to the level of vehicle handling (see Chapter 2) and relate specifically to the five areas of competence or observation categories defined in the reform proposal.

The most frequent cause of fatal accidents involving novice drivers is “inappropriate speed” (STATISTISCHES BUNDESAMT, 2010). The test drive should thus enable the driving test examiner to determine whether and to what extent the candidate adapts his driving speed to the prevailing road, traffic, weather and visibility conditions, and not least also to his usually still limited scope of driving experience. This requirement is embodied in the observation category “Speed adaptation”. Further common causes of novice driver accidents are rear-end collisions with preceding vehicles, errors when overtaking and conflicts with other road users (MAYCOCK & FORSYTH, 1997; MAYHEW & SIMPSON, 1996). These accidents can be attributed above all to the failure to maintain adequate safe distances; the corresponding aspect of behaviour or driving competence is covered by the observation category “Vehicle positioning”. At the same time, this category embraces those deficits which are manifested as incorrect positioning on the road when negotiating bends or during an overtaking manoeuvre, including proper observation of the road markings.

Another remarkable characteristic of the behaviour displayed by young drivers concerns their inadequate observation of the immediate traffic environment. The results of eye-tracking studies have shown that novice drivers, due to their limited driving experience, employ less efficient scanning strategies and thus recognise potential hazards later than an experienced driver (UNDERWOOD, CRUNDALL & CHAPMAN, 2002). Furthermore, novice drivers use their rear-view mirror less often and pay less heed to the peripheral areas of their field of vision. These deficits indicate that the driving test examiner should also assess whether the candidate is at all times aware of the overall traffic situation, which could be recognised, for example, from his adequate use of the vehicle mirrors and appropriate head and eye movements; these aspects of behaviour are covered by the category “Traffic observation”.

As traffic density increases, the participation in road traffic places ever greater demands on road users in general, and on their social behaviour in particular: Without unambiguous communication and functioning coordination between the individual road users, it is inconceivable to drive a motor vehicle safely in dense road traffic. Driving instructors and driving test examiners report that novice drivers at times experience difficulties in connection with such communicative demands, because they must – to a large degree – still consciously regulate the fundamental psychomotor action sequences associated with vehicle control and manoeuvring; compared to an experienced driver, this occupies considerable mental working capacities, which are consequently no longer available for the purposes of communication. For this reason, it seems important to provide for explicit observation and assessment of the communication with other road users during the practical driving test (e.g. whether the candidate sets indicators in good time). The relevant aspects of competence are addressed by the observation category “Communication”.

The observation category “Vehicle control/environment-aware driving”, finally, relates above all to novice-typical driving errors and deficits such as late braking, irregular steering movements or driving in the wrong gear (ELLINGHAUS & STEINBRECHER, 1990). The subcategory “Environment-aware driving” here plays a special role, because a contrary manner of driving, however detrimental for the environment, does not constitute an increased accident risk for the novice driver or other road users. This subcategory is nevertheless to be deemed important, as novice drivers should from the very beginning be encouraged to acquire a feeling of responsibility with regard to environment-aware driving.

Description of the observation categories

The mutually referenced revision and description of driving tasks, observation categories and assessment criteria represents a decisive step towards optimisation of the practical driving test (BÖNNINGER, et al., 2010) and further development of the system of driving school training (including learner assessment). It seemed expedient to entrust the outstanding description of future observation categories to the same working group

[71] The following sources were analysed: CAVALLO, BRUINDEI, LAYA & NEBOIT, 1989; ELLINGHAUS & STEINBRECHER, 1990; DUNCAN, WILLIAMS & BROWN, 1991; JAMSON, 1999; BARTL & HAGER, 2006; BRAITMAN et al., 2008; GRATTENTHALER, KRÜGER & SCHOCH, 2009; MCCARTT et al., 2009; STATISTISCHES BUNDESAMT, 2010.

– comprising experts from the Technical Examination Centres, driving instructors and others (see above) – which took responsibility for elaboration of the driving tasks under the present project. The procedure followed to elaborate the driving tasks was also retained: As a first step towards situation-independent, action-referenced description of the five observation categories defined in the project support group, a central pool of scientific data was gathered from analysis of the aforementioned sources. All knowledge derived from this information base was then condensed into draft descriptions and subsequently discussed and developed further in the working group. The outcome was the following table of category descriptions, which were formulated both as dimensions of competence and as observation instructions for the driving test examiner (see Table 11).

Category	Description as dimensions of competence	Description as observer instructions
1. Traffic observation	The driving test candidate must at all times maintain awareness of the traffic conditions around the vehicle – also by making due use of the vehicle's outside and inside rear-view mirrors, and by glancing to the side as necessary to check “blind spots”.	The examiner is to observe whether, how and when the test candidate performs actions serving to gain awareness of the traffic conditions around the vehicle.
2. Vehicle positioning	The driving test candidate must position the vehicle correctly on the road and make proper use of the available road space. This includes observing road lane markings, maintaining a safe clearance or distance to stationary or moving obstacles and other road users, as well as timely and unambiguous selection of a particular road lane or position on the road. When performing turning manoeuvres at junctions, attention must be paid to ensure that the vehicle does not stray unnecessarily onto the lane used by oncoming traffic.	The examiner is to observe whether the test candidate observes the road markings and how use is made of the available road space. Furthermore, the examiner is to observe the clearances and distances maintained to other road users, road infrastructure features or obstacles.
3. Speed adaptation	The driving test candidate must adapt the speed of the vehicle to the prevailing road, traffic, weather and visibility conditions. This includes not driving too slowly without good reason, so as not to become a hindrance to other traffic, but at the same time also not exceeding the maximum speed limit of the roads used.	The examiner is to observe whether and how the test candidate adapts the speed of the vehicle to the prevailing road, traffic, weather and visibility conditions, taking into account the maximum speed limit.
4. Communication	The driving test candidate must signal potential hazards and planned driving manoeuvres to other road users unambiguously and in good time (e.g. by setting the vehicle's turn indicators), and – insofar as necessary – react accordingly to the signals of other road users.	The examiner is to observe whether the test candidate signals relevant driving intentions unambiguously and in good time. Furthermore, the examiner is to observe whether and how the test candidate coordinates driving behaviour with the behaviour of other road users.
5. Vehicle control / Environment-aware driving	The driving test candidate must demonstrate reliable handling of the technical features of the vehicle by way of defined action sequences, and at the same time pay attention to an environment-aware manner of driving. This includes above all shifting into the next higher gear as early as possible, avoiding unnecessary acceleration and braking, and switching off the engine when the vehicle is stopped for a foreseeably longer period.	The examiner is to observe the motor action sequences employed by the test candidate when operating the test vehicle, taking into account the available technical features, and furthermore the extent to which attention is paid to an environment-aware manner of driving.

Tab. 11: Description of observation categories

3.4 Assessment and decision criteria

The practical driving test is essentially a performance assessment and decision process, by which to judge the driving competence of the driving licence applicant – as an indirectly observed parameter – and to decide whether the candidate possesses the required minimum level of competence to be able to participate independently in motorised road traffic under defined protective regulations (e.g. absolute zero-alcohol rule, or designated supervision in case of “Accompanied driving from age 17”). As indicators for such competence, alongside knowledge relating to technical preparation and completion of the drive, the examiner judges the behaviour displayed by the candidate when performing driving tasks assigned to particular areas of competence (or observation categories). The assessment and decision criteria are stipulated in legal provisions, but must also be founded from the professional and methodical perspective; in other words, the defined legal framework must be supplemented by methodically systematic procedures based on the principles of psychological testing.

The use of assessment and decision criteria in the German system of driver licensing

Situation-related and situation-independent test demands (in the sense of driving tasks and aspects of driving behaviour or observation categories) existed as test standards in the German system of driver licensing from the very beginning; up to 1957, however, there were no legislative regulations describing assessment criteria: "Assessment of the test drive was left solely to the discretion of the examiner; there was no list of defined criteria to determine whether the test was passed" (MÖRL, KLEUTGES & ROMPE, 2009, p. 62). It was only on 28th January 1958 that "Guidelines for the Examination of Applicants for a Licence to Drive Motor Vehicles" came into force. This established clear legal framework conditions and specifications also for the practical driving test, and included a list of assessment criteria which had been developed within the Association of Technical Inspection Agencies (VdTÜV) in cooperation with the Federal Ministry of Transport (ibid.).

Reformulation of the examination guidelines in 1963 served to describe the test demands and assessment criteria more precisely (MÖRL, KLEUTGES & ROMPE, 2009); these examination guidelines then remained applicable until 1970. It is conspicuous, however, that the methodical instruments of the practical driving test were up to this point of a purely legal nature: "In many points, therefore, the driving test examiners were forced to rely on their own interpretations of the traffic regulations, and such interpretations were naturally based on technical considerations rather than any approach with foundations in test psychology. The Technical Examination Centres attempted to overcome this lack of a methodical basis, which they evidently themselves regarded as a serious deficiency, through an increasingly specific description of the test contents, both in respect of the test demands and possible observation categories and assessment criteria, as expressed in the Examination Guidelines of 1970 (see above) and the supplementary Notice 731 issued by the Association of Technical Inspection Agencies ... in 1973" (HAMPEL & STURZBECHER, 2010, p. 48). The aforementioned Notice 731 and the subsequent VdTÜV guideline "Driving error marking for driving tests" can apparently be seen as the first catalogue of assessment criteria ("Catalogue of driving errors") aside from statutory regulations. The methodical test standards of the Technical Examination Centres were at that time thus focussed not on a concrete and systematic description of situation-related and situation-independent driving tasks and observation categories, but rather on specifications of driving errors.

Still in 1973, the Technical Examination Centres also reached agreement on a uniform layout for a test report, which was to serve primarily to document the errors made during the practical driving test, and accordingly aligned to the paragraphs of the Road Traffic Regulations (SCHNEIDER, 1977). In the period which followed, the results of practical driving tests were recorded by way of EDP-ready marking sheets by a few of the Technical Examination Centres, and also subjected to a statistical analysis in 1976 (HAMPEL et al., 2009). On the basis of this analysis, HAMPEL criticised various aspects of the assessment criteria introduced in 1973, despite his considering them a major methodical advance over the previous, essentially non-systematic recording of errors. The three points which he already brought up in 1977 are still valid today:

- Firstly, he questioned the effectiveness of the demand that good performance also be taken into account in the test assessment, as it was contained in Notice 731: "On the other hand, the Examination Guidelines demand that especially any errors made during the test drive are to be recorded. It should be evident under these circumstances, that the requirement to assess also positive performance will have only limited impact" (ibid., p. 53).
- Secondly, he regretted that the uniform "Standard Driving Error Scheme" recommended in Notice 731 was structured according to the categorisation of the Road Traffic Regulations (StVO): "Consequently, the StVO indirectly acquires the status of a description of learning objectives. This procedure is at least beneficial in terms of more specific definitions, even though it is no substitute for a genuine description of driving tasks" (ibid., p. 48). Furthermore, the reference to the Road Traffic Regulations promoted "a manner of assessing driving performance which is biased towards merely obeying rules" and required that "the catalogue of driving errors must also be amended each time the StVO is amended" (ibid., p. 133).
- Thirdly, he criticised the impracticability of the around 100 assessment criteria proposed by the VdTÜV in Notice 731, because this repeatedly gave rise to "difficulties regarding the assignment of particular errors to the error scheme", and the multitude of assessment items encouraged the tendency to "record judgements not during the drive, but only after its completion" (ibid., p. 133).

The sustained topicality of the criticism expressed almost four decades ago results from the facts that

- there is still today no obligation to record good performance, even though the Examination

Guidelines require this to be taken into account,

- the currently applicable methodical standards still lack concrete descriptions of driving-task-referenced assessment criteria, and
- no standardised, practicable and purposeful documentation tool exists to date to enable efficient recording of the candidate's driving performance during the actual test.

Taking up the latter challenge, a structurally optimised matrix of 20 “driving situations” and nine definitions of “situation-relevant behaviour" – based on the aforementioned VdTÜV catalogue of driving errors – was presented by TÜV Bayern in the mid-1970s as a means to record driving errors in tabular form. This matrix was nevertheless still too extensive to permit efficient documentation of all the desirable properties of a candidate’s driving behaviour during the test. For this reason, TÜV Rheinland took up the core idea of the Bavarian proposal in its own work to further develop test documentation, but condensed the original proposal of 20 “driving situations” into a matrix of seven “prototypical driving tasks” (see Chapter 3.2); at the same time, corresponding assessment criteria were specified.

Unfortunately, the efforts to further develop task-referenced assessment criteria and to operationalise these criteria in a test report were not maintained in the subsequent years: A number of Technical Examination Centres, among them TÜV Rheinland, abandoned EDP-ready test reports and returned instead to the use of informal written notes on test performance; others adopted the list of driving errors arranged in accordance with the paragraphs of the Road Traffic Regulations from VdTÜV Notice 731. A third group of Technical Examination Centres (e.g. TÜV Hannover/Sachsen-Anhalt and DEKRA) initially pursued further development of the observation categories and a matrix-style test report until 1994, but later ceased this promising[72] work, when it became clear that the test candidate was only to receive a written test report in case of failure. Since 1996, all Technical Examination Centres have used the test report which is currently stipulated in accordance with Annex 13 to the Examination Guidelines (Prüfungsrichtlinie, PrüfRiLi) of 2004. The assessment criteria themselves have experienced no significant further development ever since that time, aside from the fact that the group of particularly accident-relevant “serious errors” leading to termination of the test and automatic failure (e.g. straying onto the wrong side of the road when preparing to turn left, changing lanes without observing other traffic, or lack of consideration for children or disabled pedestrians), was greatly expanded in the 1987 Examination Guidelines (HAMPEL & STURZBECHER, 2010).

The assessment and decision criteria contained in the currently applicable Examination Guidelines were already presented in detail – together with their psychological foundations – and at the same time subjected to critical methodical analysis by STURZBECHER, BIEDINGER et al. (2010). Consequently, the following paragraphs provide only a brief recap of the most important points of criticism and recommendations for optimisation relating to the elements of the test[73] and to the test decision:

(1) No assessment criteria exist to date for the test elements “Technical preparation of the vehicle” and “Technical completion of the drive”; it is thus not possible to guarantee the objectivity of examiner assessments. It should be defined unambiguously, which aspects of behaviour are to be considered errors, and how they are to influence the test decision.

(2) A series of concrete task-referenced assessment criteria with very detailed and strict error definitions exist for the test element “Basic driving manoeuvres”. These minimum standards have no background in traffic science, however, and the particular safety relevance of certain errors which lead to failing of the driving test is not apparent. The scope of judgement granted to the examiner by PrüfRiLi 5.17 (“The rules are not to be interpreted pettily.”) is also insufficient to overcome this deficit, because, for the candidate, “there is a significant difference between an assessment on the basis of understandably derived, tolerant criteria and the need to place hope in the examiner’s generosity in case of failure to comply with petty requirements” (STURZBECHER, BIEDINGER et al., 2010, p. 116). Consequently, the assessment criteria for the basic driving manoeuvres should be referenced to the observation categories, revised accordingly, and – in line with their

[72] Already after the first trials with a matrix-style test report at TÜV Rheinland in 1974, it was concluded that “the recorded driving errors are distributed satisfactorily over the individual judgement categories, and the task of recording is significantly simplified for the examiner compared to the old procedure” (HAMPEL, 1977, p. 141).

[73] It is to be noted that, for capacity reasons, the present project was unable to contribute to the elaboration or revision of assessment criteria for the test elements “Technical preparation of the vehicle”, “Technical completion of the drive” and “Basic driving manoeuvres”. As a result, no solution is offered in respect of the disproportionate significance of the basic driving manoeuvres for the test decision, as criticised by STURZBECHER, BIEDINGER et al. (2010).

safety relevance – relativised in terms of their significance for the test results.

(3) The demands applicable to the test element "Test drive" are structured by way of driving tasks (see Chapter 3.2), the assessment of which is based on the determination of driving errors. In accordance with the Driving Licence Regulations (FeV), a distinction is made between so-called "serious errors", which lead to the candidate failing the practical driving test even if they are committed only once, and less serious errors, which are sometimes also designated "simple errors" and only result in test failure if they are observed repeatedly or in accumulation (Annex 7 FeV, 2.5.2). In the Examination Guidelines, the serious errors are listed in detail and conclusively (PrüfRiLi 5.17.2.1); the simple errors, on the other hand, are merely described by way of examples (PrüfRiLi 5.17.2.2). Despite the partial referencing of errors to driving tasks and observation categories, above all in Annex 10 to the Examination Guidelines, these rather non-systematic and illustrative specifications remain insufficient with regard to the demands of psychological testing. From the critical methodical perspective, it must also be noted that no grounds are given for the assignments of individual driving errors to one or other of the two categories of inappropriate behaviour (serious or simple errors); such grounds should take up the road safety relevance of the errors concerned.[74] Furthermore, certain weaknesses are revealed in the formulations: What exactly is to be understood, for example, by "gross disregard" or "excessive hesitation"? At the same time, objective fulfilment of the statutory requirement to take good performance into account (PrüfRiLi 5.17) is thwarted by the lack of a task-referenced description of good performance. Overall, the methodical challenge associated with the above criticism can be summarised as follows: In conjunction with the detailed description of driving tasks (see Chapter 3.2), systematic and differentiated assessment criteria must be described for each individual task with reference to the defined situation-independent action contexts or observation categories (see Chapter 3.3), i.e. it must be determined – and justified in terms of road safety – what is to be considered "good performance", a "simple error" or a "serious error".

(4) With regard to the test decision, critical note has already been made of the strong importance attached to correct performance of the basic driving manoeuvres, despite their limited road safety relevance; the corresponding regulations must thus be reviewed accordingly. Furthermore, concrete specifications are required to define the significance of the test elements "Technical preparation of the vehicle" and "Technical completion of the drive" for the test decision. Concerning the aforementioned "simple errors" which only lead to failing of the test when observed repeatedly or in accumulation (Annex 7 FeV, 2.5.2), it must be clarified, what exactly is meant by "repeatedly" and "accumulation". In addition, it must be specified, how good performance is to be reflected in the test decision. Last but not least, a system manual based on the methodical principles of psychological testing should provide an unambiguous decision algorithm for the practical driving test, with concrete indications of the procedure to be followed to reach an overall decision and the intended weighting of performances in the different test elements.

Following revision of the assessment and decision criteria contained – in part in different forms and with varying degrees of differentiation – in the individual licensing regulations, the next step is appropriate streamlining and systematisation of the legal basis for the licence testing system.

It is to be noted that, even after implementation of all the given optimisation proposals relating to the decision criteria, one methodical petitum would still remain unanswered: For the determination of test success in the context of a learning-objective-referenced test, it is usual to place the number of incorrectly solved tasks or test items in relationship to their overall number or to the number of correctly solved items. HAMPEL (1977) writes in this connection: "It can be assumed that the reliability of an empirical claim increases with the number of observations on which the claim is based. A trivial correlation thus exists between the validity of an observation system and the extent to which it refers to quantifiable events. This now gives rise to

[74] In this context, STURZBECHER, BIEDINGER et al. remarked: "It appears that the category of serious errors is reserved above all for those instances of inappropriate behaviour which represent a significant endangering of road safety or inconsiderateness towards "weaker" road users (e.g. failure to observe right-of-way, ignoring of children). Simple errors, on the other hand, seem to refer instead to uncertainty on the part of the test candidate (e.g. hesitation at crossroads) or deficiencies in vehicle handling. If the assignments were indeed based on the aforementioned principles, however, it still remains unexplained, why the failure to drive at an appropriate distance behind the preceding vehicle, for example, is counted a simple error: This behaviour, after all, is one of the most common causes of accidents in road traffic" (2010, p. 117).

the following complication: Correct behaviour can be taken as the rule in most driving tests and demonstrations, as is basically to be expected where the declared objective is to verify the mastering of a particular task. As a result, only errors are recorded during most demonstrations of driving ability. It should be obvious, however, that the significance of a certain number of errors can only be interpreted with knowledge of the applicable situative demands. Such interpretation, in turn, is only feasible if not only errors, but also correctly solved tasks are counted, and thus the number of errors can be related to the total number of relevant situations or events" (p. 96-97). Such procedures, which could also be termed "learning progress assessment", are found especially in observations of driving behaviour serving to visualise the learning progress of a licence applicant during the course of driver training (e.g. McKNIGHT & HUNDT, 1971a; McGLADE, 1965); in the case of learner assessment in the sense of driving tests ("status assessment"), however, they are something of an exception from the international perspective (see below). In Germany, the introduction of such a procedure would have far-reaching consequences for test organisation and test documentation, as premature termination of a test would only be permissible in exceptional cases, and all arising driving tasks would need to be recorded; we will return to this point (see Chapter 3.5 "Control concept").

Assessment and decision criteria for an optimised practical driving test

Contrary to the situations with regard to driving tasks (see Chapter 3.2) and observation categories (see Chapter 3.3), no reform proposal for optimised assessment and decision criteria already existed at the beginning of the current project. This can be attributed to the fact that, at that time, there was also no scientifically founded catalogue of driving tasks whose content validity had been confirmed by corresponding domain experts and to which assessment and subsequently decision criteria could have referred. It must be pointed out, furthermore, "that the determination of what constitutes a serious error of decisive significance for the test assessment should be based on appraisals of safety relevance by traffic experts and cannot be viewed as a task for the methodologist" (STURZBECHER, BIEDINGER et al., 2010, p. 117). It nevertheless seems expedient – as already in the previous chapters – to commence the discourse on future assessment standards for the practical driving test in Germany with a brief survey of existing international practice, in order to identify possible starting points or inspiration for the German optimisation process.

The EU Directive on Driving Licences 2006/126/EC (EUROPEAN PARLIAMENT & EUROPEAN COUNCIL, 2006) defines the following assessment standards for successful completion of the practical driving test in Annex 2, Section 9: "For each of the abovementioned driving situations, the assessment must reflect the degree of ease with which the applicant handles the vehicle controls and his demonstrated capacity to drive in traffic in complete safety. The examiner must feel safe throughout the test. Driving errors or dangerous conduct immediately endangering the safety of the test vehicle, its passengers or other road users shall be penalised by failing the test, whether or not the examiner or accompanying person has to intervene." This establishes a behaviour and interpretation framework as a minimum standard, which the individual EU member states can then concretise with situation-related (i.e. task-referenced) and situation-independent (i.e. observation-category-referenced) assessment and decision criteria for the driving test examiner in accordance with their particular traditions.

It is conspicuous that the assessment criteria anchored in the EU Directive on Driving Licences – familiarity with the vehicle controls, safe handling of the vehicle in road traffic, feeling of safety on the part of the examiner – cannot be deemed a methodically founded and practicable basis for assessment from the perspective of test psychology; it is rather the case – as HAMPEL (1977, p. 94) writes – that they represent "mere descriptions of the examiner's impression", which, without further precision, can hardly be considered acceptable either as psychometric conditions or from the viewpoint of the test candidate. To satisfy these requirements, it is necessary to formulate assessment criteria for test performance (e.g. driving errors, good performance) which are accessible to direct observation, refer to actual behaviour as specifically as possible, and demand as little additional interpretation and classification as possible on the part of the examiner (KROJ & PFEIFFER, 1973). Such assessment criteria must apparently be developed at national level, as they are not a subject of the EU stipulations.

The pass criterion "No immediate endangering of road users" contained in the EU Directive on Driving Licences also appears impracticable without supplementary specifications: The directive fails to indicate the situation characteristics by which such endangering can be recognised objectively, and it ignores the scientific tradition founded by

McKNIGHT and HUNDT (1971a), according to which the passing of a driving test should be linked to scientifically founded decision algorithms taking into account the empirically determined road safety relevance of driving errors. In view of the extraordinary diversity in the manner in which assessment and decision criteria are handled in the individual EU member states, one can readily gain the impression that, in the case of the assessment standards (and in contrast to the demand standards), the EU Directive on Driving Licences here served not harmonisation on a scientific basis, but rather description of the “least common denominator”.

Attention is thus turned to the results of the international comparative analysis of assessment and decision criteria: More or less differentiated stipulations relating to assessment of the candidate's performance in the practical driving test exist in all 36 countries whose systems of testing were analysed within the framework of the present project (see above). In many cases, the test organisations also publish more specific explanations of the provisions contained in the relevant legislation, regulations and guidelines in the form of “test handbooks”; they always refer to the driving errors committed by the candidate, but in some countries (e.g. British Columbia/Canada, Finland, Iceland and Norway) additionally to particularly good driving performance. In Finland, for example, driving behaviour which can only be expected from a candidate with an above-average level of competence is recorded in the category “Good performance”; this is nevertheless not taken into account for the test decision. Outstanding performance is similarly recorded in the test report in Iceland and Norway, but may here serve – as in Sweden, where no report is created, however – to compensate minor errors when the examiner considers his overall test decision.

The driving errors are everywhere recorded by way of standardised forms (so-called test reports, see Chapter 4) which list the demands to be met by the test candidate (formulated mainly as situation-related driving tasks, but often also mixed with situation-independent demand standards in the sense of observation categories) and typically observed errors. In approximately 75 per cent of the countries analysed, driving errors are further subdivided in accordance with their safety relevance. None of those countries, however, offered grounds for either the chosen classifications or the corresponding weighting. This is to be deemed critical from the methodical perspective: “Weightings only seem meaningful where their corresponding content can be derived convincingly and on a statistically robust basis” (HAMPEL, 1977, p. 108).

Given the lack of justification for the error classifications, it is not surprising to find such broad variation, for example,

- two-tier error classifications (e.g. “Simple errors” and “Serious errors”) in Germany, Finland, Greece, Lithuania and Victoria (Australia),
- three-tier error classifications (e.g. “Minor errors”, “Serious errors” and “Critical errors”) in Great Britain, Ireland, Latvia, Malta, Norway and Austria, and
- four-tier error classifications (e.g. “Minor errors”, “Errors”, “Serious errors” and “Very serious errors”) in Denmark and Luxembourg.

In the three and four-tier assessment systems, errors which relate to particular instances of endangerment or the hindering of other road users during the test, or else result in an accident or intervention by the driving instructor, are usually operationalised as an independent (highest) error category. Measured against the aforementioned recommendations from KROJ and PFEIFFER (1973), three and four-tier error classifications appear overly differentiated and thus difficult to handle[75]; a distinction between simple and – a few – particularly safety-relevant and thus decisive errors, on the other hand, can be seen as practicable and adequate in terms of driving safety.

Most countries specify a maximum number of errors which may be tolerated (so-called “cut-offs”, see above), but at the same time often apply different weightings to the error categories. Exceeding of the specified threshold leads to the candidate failing the test. As already noted with regard to the error classifications, however, no grounds are offered as foundation for the similarly broad variation in the applicable threshold values[76]. Im-

[75] HAMPEL (1977, p. 139) also reports – on the basis of results from trials– that “a whole series” of driving test examiners felt overtaxed by the requirement to apply a three-tier error classification.

[76] A number of contrary examples serve to illustrate this diversity: In Great Britain, the driving test is deemed failed in case of more than 15 “Driving faults” (mistakes in vehicle handling or incorrect reactions in non-dangerous situations), irrespectively of whether these faults are spread over different categories or the same fault is observed repeatedly. The test is similarly failed if the candidate commits any one “Serious fault” (errors which could potentially endanger others) or “Dangerous fault” (errors resulting in the actual endangering of others). Overall, 60 possible errors are distinguished. In Ireland, the candidate fails the driving test in case of any “Dangerous or potentially dangerous fault”, more than either three or five “More serious faults” where these faults belong to the same aspect of driving (e.g. observation when negotiating roundabouts) or under the same category heading (“Observation”), respectively, or more than eight different “More serious faults” in total. The driving test regulations in Latvia distinguish between “Minor errors” (which do not endanger traffic safety and carry one penalty point),

portant country-specific differences are found in the manner in which simple or minor errors are taken into account in the test decision: In a few countries (e.g. Ireland and Austria), minor errors have no influence on the test decision, whereas the majority of countries – including Germany – specify that the test is failed in case of the repeated occurrence of similar “minor errors” or accumulations of several different “minor errors” (e.g. British Columbia/Canada, Denmark, Ireland, Latvia, Lithuania, Great Britain, New Zealand and Queensland/Australia). The applicable definitions of “repeated occurrence” and “accumulations” of errors, however, remain unclear. Insofar as this indefinite specification is intended to enable adaptation of the assessment standards – as already demanded by HAMPEL (1977) – to the still relatively low level of the candidate's driving competence after only a short period of driving school training (see above), the discounting of minor errors when determining the test decision seems more consistent from the professional perspective and at the same time more conducive to test objectivity than diffuse assessment standards.

It is only in exceptional cases that the maximum permissible number of errors is related to the number of tasks to be handled or alternatively to the duration of the test drive – as was already practised by McKNIGHT and HUNDT (1971a) with their Driving Situations Test, and later also demanded by HAMPEL (1977). One example of a correspondingly relativised, and furthermore very complex assessment and decision system is that implemented in South Africa. If a test candidate makes three mistakes when changing lanes, for example, this earns 15 penalty points, because lane-changing errors are weighted with a factor of five (abrupt stopping without endangering others, on the other hand, is weighted with only one penalty point). If a certain cut-off value is exceeded (calculated by multiplying the test duration in minutes by eight), the candidate is deemed to have failed the test.

While most countries use a cumulative system when defining failure thresholds, there are also a few cases where the result is based on deduction: In Iceland, for example, the candidate starts the driving test with a credit of 100 points; over the course of the test drive, certain numbers of credit points (based on a weighting of n = 1, 3 or 12) are then deducted for any errors committed, depending on their severity. To pass, the novice driver must reach the end of the test with at least 80 points remaining. One credit point can be regained, however, for particularly good performance. In Luxembourg, driving errors lead to the deduction of 3, 5, 10 or 20 points. To pass, the candidate must retain at least 45 of the original credit of 60 points.

Around one-third of the countries analysed use summarised, competence-oriented judgements of the candidate's performance. This generally means that driving errors are initially assigned to certain situation-independent action contexts (in the sense of observation categories) or to the corresponding areas of competence. Subsequently, the driving errors recorded for the individual areas of competence are further categorised according to either the number of errors or their severity, as a basis for often very detailed criteria for passing of the test.[77] In countries such as Estonia, Finland, the Netherlands, Norway and Sweden, on the other hand, no such differentiated calculation rules exist: Here, the final test decision – insofar as the candidate commits no particularly safety-relevant errors which lead de facto to automatic failure – is very much dependent on the overall impression which the examiner gains of the candidate's driving competence during the drive, and less so on individual driving errors. Such a procedure appears also logical, if no effort is to be taken to elaborate precise descriptions of the test demands: “A schematic reference to numbers of errors would, on the contrary, merely result in an illusion of exactness and injustice for the test candidate, as long as the manner of error assessment is not defined precisely” (HAMPEL, 1977, p. 54).

It was already mentioned that certain countries allow driving errors which are not exceedingly relevant in terms of road safety to be compensated by good performance. Special possibilities for compensation exist in France, Croatia, New Zealand, the Netherlands and Austria, where the discussion of certain driving and traffic situations between the candidate and driving test examiner is a planned

“Medium errors” (which endanger traffic safety to only a slight degree and carry four penalty points) and “Serious errors” (which endanger traffic safety significantly and carry ten penalty points). The driving test is deemed failed if the candidate accumulates more than nine penalty points or else the same “Medium error” is observed twice. In Lithuania, the test is failed if the examiner observes nine or more “Non-critical errors” or any “Critical error” or “Repeated error”.

[77] A few examples are intended to illustrate the diverse practice: In Belgium, errors are assigned to one of eleven competence categories; if no errors are recorded for a particular category, the rating “Satisfactory” is recorded (1 error = “With reservations”, 2 errors = “Poor”, 4 errors = “Inadequate”). If the candidate receives a rating of “Inadequate” in any category, the test is deemed failed. In British Columbia, the individual errors are assigned to five competence categories, each of which possesses its own cut-off. In France, adequate performance must be demonstrated in three competence categories to pass the test; the driving test is not passed if a “Critical error” is recorded in any of the competence categories.

element of the test drive (so-called "situational questions"). In Austria, for example, the candidate is asked to speak about a traffic situation specified by the examiner; this is intended to verify the candidate's "traffic awareness". The examiner should preferably choose a situation in which he has previously observed a driving error. The candidate is then able to explain his error and driving behaviour in the given circumstances: The driving error is only recorded as such if the candidate fails to demonstrate adequate awareness and understanding for the traffic situation in this conversation; the original error can thus be compensated in a phase of reflection.

To summarise, the international comparative analysis of assessment and decision criteria for the practical driving test reveals a considerable variety of methodical approaches. Substantial differences are found between the individual countries with regard to

- the manner and degree of differentiation in error descriptions and error classifications,
- the weighting of errors in accordance with their road safety relevance,
- whether only actual endangering or also potential dangers are to be treated as errors in certain situations,
- the applicability of minor errors and specifications relating to a required score or maximum permissible number of errors,
- the relativity of demands and the complexity of the assessment and decision system,
- the holistic nature of test assessments and decisions, and last but not least
- the possibilities to compensate errors by way of good performance.

Alongside the occasionally considerable differences between the individual countries, there are also a number of common features. One example is the fact that driving errors entailing an immediate endangering of road safety lead to the test being deemed failed in all the countries analysed. Furthermore, the fundamental, content-related or psychological considerations serving as the basis for assessment and decision criteria remain universally unclear; no scientifically founded system could be found within the framework of project research.

As neither scientific foundations nor empirical verification of the assessment and decision criteria for the practical driving test could be identified, and in view of the diversity of approaches followed in international practice, only limited inspiration can be derived for further development of the German assessment system, and it is not yet possible to speak of harmonisation in Europe. It is evident that the development of professionally adequate assessment and decision criteria – given the lack of corresponding guidance from the EU – must necessarily be pursued at national level. For this reason, the assessment standards were also revised within the framework of the current project.

Description of the assessment and decision criteria

It was already mentioned (see Chapter 3.2, "Driving tasks") that the assessment criteria – together with the demand standards (driving tasks and observation categories) to which they refer – were elaborated on the basis of scientific information gathered by the experts appointed to corresponding working group. The first step in this elaboration process was to describe what constitutes correct mastering of the individual driving tasks ("normal performance"). With this done, it was subsequently possible to determine and define the behaviour deemed to represent "above-average performance" on the one hand, and a "simple error" or "serious error" on the other. In this connection, due consideration was also given to study findings relating to novice-typical driving errors, competence deficits and accident causes: The road safety relevance of the particular driving behaviour was to be reflected in the distinction between simple and serious errors.

The assessment of test performance and the subsequent test decision in a future, optimised practical driving test should be realised on three levels:

(1) Event-oriented assessment: Situation-related assessment of the candidate's performance of a driving task, with corresponding reference to the observation categories, by way of a four-level behaviourally anchored ordinal scale ("Above-average performance", "Normal performance", "Simple error", "Serious error")

(2) Competence-oriented assessment: Situation-independent assessment of the candidate's mastering of driving tasks in the sense of elements of driving competence or observation categories, similarly by way of a four-level behaviourally anchored ordinal scale ("Very good", "Good", "Sufficient", "Inadequate")

(3) Test decision: Weighing up of the results of levels (1) and (2) in order to reach a dichotomous final decision ("Passed", "Failed").

The undertaken differentiation of the assessment criteria and the proposed support for the decision process are intended to strengthen the character of the practical driving test as an instrument for competence assessment and address three central objectives: Firstly, the acquired test data are to permit more meaningful representation of the level of candidate's driving and traffic competence than is possible by the current method of recording only errors; this would also extend the possibilities for improvement-oriented performance feedback to the candidate. Secondly, the assessments obtained are to enable conclusions to be drawn on possible preparation deficits and thus on the driving training completed by the candidate. Finally, an adequately extensive database is to be established for evaluation of the practical driving test, and here in particular its validity with regard to traffic behaviour prognoses. In the following, the aforementioned assessment and decision levels are considered in more detail.

For the event-oriented assessment, the examiner judges the candidate's fulfilment of the action demands specified in the observation categories during completion of individual driving tasks, by classifying the recognised aspects of relevant behaviour under one of the four levels of an ordinal assessment scale immediately upon observation of the behaviour concerned. The individual levels are: (1) “Above-average performance”, (2) “Normal performance”, (3) “Simple error” and (4) “Serious error”. As already explained above, the objectivity of an assessment process is increased where the distinctions between assessment levels are sufficiently clear-cut and it is thus possible to assign the behaviour observations unambiguously. The elaborated classification is presented below; the corresponding assessment criteria must be tested accordingly in a subsequent project to verify their practicability:

re 1: In the opinion of the experts, “above-average performance” is understood to mean performance observed in connection with a driving task which exceeds that typically displayed by a novice driver in respect of a certain aspect of competence (see observation categories). This will often refer to behaviour in essentially unforeseeable or unusually complex hazardous situations which arise not least due to the incorrect behaviour of other road users and to which candidate reacts in a professional and calm manner (e.g. excellent performance relating to hazard anticipation, hazard avoidance and, where necessary, hazard management). Such behaviour is indicated by way of examples referring to the relevant observation categories for most items of the driving task catalogue. Where appropriate, above-average performance may be taken into account in the overall assessment as compensation for any “simple errors” observed. A failure to display above-average performance is not to be considered a deficiency.

re 2: “Normal performance” is understood to refer to the error-free completion of a driving task. This level of performance is described in the driving task catalogue as the fundamental action demand – with corresponding reference to all applicable observation categories – for each driving task.[78]

re 3: “Simple error” describes incorrect behaviour which constitutes the substantial potential endangering of road users – including the candidate and other occupants of the test vehicle – or else leads to the avoidable obstruction or irritation of other road users. These errors are taken into account in the overall assessment and serve to objectivise the assessment of competence. If “simple errors” are observed repeatedly, the driving test examiner must decide, in accordance with the demand level of the underlying traffic situation and the road safety relevance of the behaviour concerned, whether the observed errors in behaviour are to lead to failing of the test; to this end, it is still necessary to concretise both the error definitions and the corresponding decision rules (see below). “Simple errors” can be compensated by way of above-average performance; they are (to date) specified merely by way of examples in the driving task catalogue.

re 4: “Serious error” describes incorrect behaviour which constitutes particularly critical potential to endanger or injure either the occupants of the test vehicle or other road users, or else results in concrete endangering. “Serious errors” have to date resulted in immediate premature termination of the test; they are listed conclusively in the driving task catalogue.

With regard to the differentiation of “simple errors” and “serious errors”, it remains to be noted that the closely associated discussion on the content

[78] In certain respects, the terms “above-average performance” and “normal performance” are perhaps inappropriate, because they suggest a reference to a social norm. This can be taken to apply at least in the case of the term “above-average performance”, whereas “normal” is used in the current project – for reasons of better legibility – to mean “flawless in standard situations without particular aggravating demands”. The designations of these categories – and likewise their handling – is to be reviewed within the framework of further research and development work, not least because an entirely flawless performance cannot be considered typical or normal in the case of a driving test candidate (see above). Furthermore, the “normality” of behaviour or novice-typical driving performance can only be determined finally within the framework of empirical test evaluation (see Chapter 5), and not by way of experienced-based expert opinions.

meaning of the terms “concrete endangering” and “potential endangering”, as well as on use of these terms in test practice, is still in progress and must be continued under all circumstances within the framework of methodical further development of the practical driving test: Precise definition of these two terms can already be deemed imperative in the context of road traffic, because risks can never be fully excluded, i.e. “potential endangering” in this sense is necessarily conceivable, or can at least be construed, under the most diverse situative conditions.

The introduction of competence-oriented assessment adds a holistic aspect of performance judgement to the event- and error-based manner of assessment which is traditional in Germany; similar developments can also be witnessed in other countries with reform-oriented driver licensing systems (see above). The overall assessment relates to how frequently correct driving behaviour is displayed in connection with the different types of driving task, or how frequently certain partial competences are demonstrated across the completion of all driving tasks (in other words, in the different observation categories). In an optimised practical driving test, the driving tasks and observation categories should in future be subject of competence-oriented assessment by way of a four-level ordinal scale with the following classifications:

- “Very good”: The candidate acts correctly, efficiently and with foresight in (almost) all traffic situations (with reference to the type of driving task or area of competence).
- “Good”: The candidate acts correctly, efficiently and with foresight in most of a diversity of traffic situations; simple errors represent an exception.
- “Sufficient”: The candidate acts for the most part correctly, efficiently and with foresight in standard traffic situations (i.e. situations without special demands); simple errors are observed in complex or unfamiliar situations (e.g. situations which are rarely encountered during training).
- “Inadequate”: The candidate for the most part fails to act correctly, efficiently and with foresight even in standard traffic situations. Serious errors and/or multiple or repeated simple errors are observed; further improvement is necessary before solo participation in motorised road traffic.

The competence-oriented assessments should reflect the overall course of the test from the perspective of the driving test examiner; the event-oriented assessments contribute to this judgement as objectivising factors. The four differentiated levels for the assessment of driving competence as it relates to driving tasks on the one hand, and to observation categories on the other, establishes the necessary prerequisites for evaluation of the effectiveness of novice driver preparation in general, and the practical driving test in particular: To date, there has been no differentiated assessment and documentation of the test performances of candidates who pass the test; correspondingly, it has also not been possible to analyse the mastering of everyday traffic demands by novice drivers against their previous test performance. For this reason, no studies have so far been conducted to verify the criterion validity (see Chapter 5). Furthermore, competence assessments also bring significant improvements with regard to the possibilities to provide feedback on test performance to the candidate.

Let us now turn to the criteria which determine whether the test is passed or failed. According to the currently applicable legislative regulations, a practical driving test is failed if any “serious error” is documented during the test drive and/or in case of the repeated occurrence or accumulation of “simple errors”. STURZBECHER, BIEDINGER et al. (2010) already pointed to the associated methodical weaknesses relating to the decision criteria for the practical driving test, and described the non-exhaustive listing of “simple errors”, in particular, as unacceptable: Given the severity of the consequences of “simple errors” for the test decision, a complete definition would appear to be no less expedient that the exhaustive listing of “serious errors” leading to immediate termination of the test. Furthermore, no unambiguous explanations are provided for the meanings of “repeated occurrence” and “accumulation”. It is therefore an imperative prerequisite for improvement of the methodical reliability (and thus, in turn, also the validity) of the practical driving test to elaborate precise definitions conformant with the principles of test psychology for the decisive terminology used in the context of test assessment

This notwithstanding, an unambiguous and exhaustive clarification of what is to be considered a “simple error” should not result in an overly complicated multitude of error definitions; the objective should rather be to identify those forms of incorrect behaviour which represent a significant (potential) endangering of road safety, and to provide for differentiation from novice-typical conspicuities in driving behaviour which – in contrast to “simple errors” – are to have no negative influence on the outcome of the test. Current licensing legislation makes no explicit reference to aspects of behaviour which, although suboptimal, are to be considered irrelevant for the test decision, although this could be the intention of the requirement for the examiner to exercise discretion (PrüfRiLi 2.17). This may well be satisfactory from a legal viewpoint; from the test psychology perspective, however, it should better be treated merely as a framework for the elaboration of further differentiated methodical stipulations. The distinguishing of "simple errors" from essentially irrelevant conspicuous behaviour would also accommodate the widely advocated opinion that novices cannot be considered well-versed drivers, and that successful completion of a driving test should be understood not as proof of perfection, but rather as a "licence to continue learning" (HAMPEL, 1977; STURZBECHER, BIEDINGER et al., 2010). The proposed re-appraisal of driving errors will presumably lead to a shift in the weight of "simple errors" for the overall test assessment; in this case, it would then be necessary to review also the present stipulations relating to "serious errors" and their influence on the test decision.

It is generally agreed that the test decision should not be based solely on the recording of driving errors, and must instead follow on from a competence-oriented assessment, wherein event-oriented observations nevertheless contribute to objectivisation. What is still lacking as a basis for a competence-oriented test decision, however, is a decision algorithm by which event-specific assessments can be mapped unambiguously to the competence-related level and, finally, a test decision can be derived. The scope of professional judgement granted to the driving test examiner remains necessary, also in our opinion, but should be limited to the assessment level, and there to the distinction between "simple errors" and conspicuous driving behaviour which is irrelevant for the test decision; furthermore, a set of decision criteria should define the intended path to a test decision (STURZBECHER, BIEDINGER et al., 2010).

As a final remark, it is to be noted that the work on optimisation of the assessment and decision criteria within the framework of the present project does not extend significantly beyond the existing legislative framework. This was not the intention[79], however, and certain outcomes of the revision are thus felt to remain inadequate from the methodical perspective. Accordingly, revision of the assessment and decision criteria is by no means concluded with the research and development findings presented here. It should rather be considered a continuous development task, which must constantly be tackled anew in accordance with scientific progress (e.g. studies investigating novice-typical driving errors) and the expected evaluation results (see Chapter 5).

[79] The first analysis of the methodical system of the practical driving test (STURZBECHER, 2010) had shown that, with regard to content, the recommended future driving tasks and observation categories are already contained in the stipulations of today's licensing legislation. The original intention, therefore, was essentially to retain the assessment and decision criteria, in order to facilitate testing of the streamlined demand standards. With hindsight, however, it seems desirable to provide for fundamental revision of the assessment standards in the near future.

3.5 Control concept

As a method of learning-objective-referenced testing and systematic (driving) behaviour observation, the practical driving test requires a methodical control concept which

- arranges the individual modules of its methodical architecture – situation-related and situation-independent demand standards in the sense of driving tasks and observation categories, as well as corresponding assessment and decision criteria – within an overall methodical system,
- takes into account the special circumstances of the framework conditions – implementation in the lifeworld domain "road traffic", where planning and control are only feasible to a limited degree – and
- is able to serve as a basis for the determination and description of implementation standards.

One such control concept is the "Adaptive test strategy for the practical driving test"[80], which was founded in detail, illustrated in the form of a circular process model and discussed with regard to its consequences as a method of psychological testing by STURZBECHER, BIEDINGER et al. (2010). If this test strategy is applied, then the practical driving test is to be viewed not as a classic method of testing, but rather as a partially standardised, criterion-driven process of competence verification (see Chapter 3.1) and assessment.

An adaptive test strategy interconnects planning, observation, assessment, verification and decision processes. Correspondingly, the adaptive test strategy comprises five action elements, which the driving test examiner realises more or less frequently, and in part also simultaneously, during the course of a test, and which – similarly to JÜRGENS and SACHER (2008) with regard to oral examinations (see Chapter 3.1) – can be de-

[80] In the interest of better legibility, only the shortened term "adaptive test strategy" is used hereafter in the present report.

scribed as framework conditions for test implementation:

1. Planning and structuring of the test or observation situation by way of prescribed demand standards (driving tasks) and on the basis of knowledge of the test route, taking into account the performance already displayed by the test candidate, where appropriate
2. Systematic (targeted) observation of the candidate's behaviour in accordance with specified observation categories which define the aspects of behaviour and candidate competences to be examined
3. Assessment of the candidate's behaviour against specified assessment criteria, documentation of the assessment results by way of a test report, and elaboration of decision preferences for the final test result (pass or fail)
4. Verification (reflection) of the current assessment and decision basis with regard to the attained degree of certainty and its suitability as justification for a valid decision
5. Decision-making relating to the planning of further observation situations and formulation of a final test decision.

In the case of the practical driving test, the described adaptive test strategy is used not so much to facilitate finer assessment of the candidate's abilities, but instead for validation purposes (see Chapter 3.1): If the candidate completes a set task correctly, then he will not – as in a classic adaptive test – be confronted with progressively more difficult driving tasks as a basis for iterative localisation of the precise maximum of his relevant driving competence. It is rather the case that, in response to ambivalent test performances which do not permit an unambiguous assessment, a further, comparable driving task is set with the aim of minimising assessment doubts “where it is not yet possible to reach a valid test decision” (STURZBECHER, 2010, p. 103-104). If the test candidate fails to perform the driving task “Changing lanes” satisfactorily, for example because he neglects to observe the traffic and fails to maintain the required safety margins, then the examiner will attempt to give the candidate a second opportunity to demonstrate the desired behaviour in a similar situation. With the planning and realisation of this similar situation, the examiner adapts the further course of the test to the previously displayed performance.

It remains necessary – after completion of the current project – to substantiate the scientifically founded, but to date only generally described “adaptive test strategy” by way of implementation guidelines for the driving test examiner. These future implementation standards must then be presented, together with the demand standards, assessment criteria and the algorithm for determination of the test decision, in a methodical manual elaborated in accordance with test psychology principles. This methodical manual must be incorporated into the set of procedural instructions to be followed by driving test examiners, and should furthermore provide precise specifications of the driving tasks (see driving task descriptions) to be examined and the situative conditions (subclasses of situation) under which this examination is to be realised within the framework of a candidate-oriented minimum demand standard. The current status of discussion and processing in the expert working group indicates that this minimum demand standard can be assumed to comprise the 15 driving tasks (subtasks) listed in Table 12, supplemented in certain cases with corresponding prerequisite conditions.

Driving tasks	Subtasks of a candidate-oriented minimum demand standard
Driving task 1	1.1 “Joining traffic” 1.2 “Leaving traffic” 1.3 “Changing lanes” The three subtasks are to be performed at least once under each of the following conditions: “Roads which can be used up to a maximum speed of 50 km/h”, “Roads which can be used up to a maximum speed between 50 and 100 km/h” and “Roads which can be used at maximum speeds in excess of 100 km/h”.
Driving task 2	2.1 “Approaching and negotiating curves” 2.2 “Driving on connecting road sections” The subtask “Approaching and negotiating curves” must be performed at least once on “Roads which can be used up to a maximum speed between 50 and 100 km/h”.
Driving task 3	3.1 “Passing obstacles” 3.2 “Overtaking” The two subtasks are to be performed at least once under each of the following conditions: “Roads which can be used up to a maximum speed of 50 km/h” and “Roads which can be used up to a maximum speed between 50 and 100 km/h”.
Driving task 4	4.1 “Passing crossroads and junctions” 4.2 “Turning right at crossroads and junctions” 4.3 “Turning left at crossroads and junctions” The three subtasks are to be performed at least once under each of the following conditions: “Priority for traffic from the right”, “With signs indicating priority” and “With light signals”.
Driving task 5	5.1 “Negotiating roundabouts”
Driving task 6	6.1 “Approaching and passing railway level crossings” **or** 6.2 “Approaching trams, and overtaking and being overtaken by trams”
Driving task 7	7.1 “Approaching and passing bus and/or tram stops” 7.2 “Approaching and passing pedestrian crossings” **or** 7.3 “Approaching and passing pedestrians”
Driving task 8	8.1 “Approaching and passing cyclists”

Tab. 12: The driving tasks and subtasks of a candidate-oriented minimum demand standard

If it is assumed that, wherever possible, performance of the driving tasks (subtasks) is to be assessed in several (i.e. at least two) independent traffic situations, so as to permit reliable analysis of the underlying driving behaviour (KANNING, 2004), each practical driving test would ideally require the planning, observation and assessment of at least 30 traffic situations. On the basis of experience gained with other methods of driving behaviour observation, this seems feasible within the framework of the prescribed test duration – insofar as the corresponding road infrastructure and traffic circumstances are actually available at the particular test location: The Road Test described by McGLADE (1965), for example, comprises 28 driving tasks, of which a varying number must be repeated; McKNIGHT and HUNDT (1971a) define 17 driving tasks for their Driving Situations Test; these tasks must similarly be repeated, and it is considered feasible to observe a total of more than 100 traffic situations during a drive lasting between 30 and 45 minutes. Despite the promising starting point, only studies of the planned trials and continuous evaluation on the basis of empirical findings will show whether the reformed catalogue of driving tasks presented above can be implemented fully at all test locations, given their individual – and, especially in rural regions, occasionally unfavourable – traffic infrastructures. At the same time, there can be other reasons, in addition to infrastructure deficiencies at a particular test location, why it may not be possible to assess all elements of the candidate-oriented minimum demand standard presented in Table 12, despite the fact that the underlying behaviour is deemed a general requirement of the test. There are also many cases in which the realisation of certain driving tasks cannot be reconciled with overall road safety aspects. Attention was already drawn to one important exceptional case, namely to the decision to forego overtaking in case of adverse weather conditions or high traffic density, in Chapter 3.2 (“Driving tasks”). It is furthermore well known from driving test practice that it is not always possible to examine certain driving tasks (e.g. changing lanes) in winter conditions, especially in mountain regions where snow-covered roads may only be partially cleared, if at all. Wet roads and autumn carpets of fallen leaves are also situations in which it may be expedient not to demand the demonstration of certain driving manoeuvres. The driving tasks of the minimum demand standard thus represent only an ideal demand framework, and cannot necessarily be examined in their entirety under all road and traffic conditions.[81] It is imperative, however, that

[81] Unfavourable conditions at the test location cannot always be compensated adequately by the driving test examiner, and should be taken as occasion to consider medium-term improvement of the test conditions by way of modifications to the corresponding test location specifications.

any deviations from the minimum demand standard are documented with corresponding explanation in the test report; this ensures that the test decision remains transparent, the evaluation of test results is not hindered by unanswered questions relating to missing data, and the test location conditions can be analysed in more detail in case of inadequate road infrastructure circumstances. Even where the minimum demand standard is not met due to the inability to assess certain driving tasks, the described procedure still represents a major advance compared to current practice, where the driving tasks actually handled by the candidate are not recorded at all.

Following the opinion of BARTHELMESS (1976), it is to be emphasised that a driving test examiner must not record the "carefully slow" performance of driving tasks – or likewise the necessity to forego driving manoeuvres which involve a significant residual risk for less experienced drivers due to special traffic circumstances, e.g. high traffic density, adverse weather conditions (see above) – as inadequate driving behaviour against a novice driver who has so far undergone only a relatively short period of basic driver training. It is even less reasonable, assuming that road safety is not endangered in any way, for the examiner to (adaptively) demand the repetition of cautiously performed manoeuvres under time pressure in order to reveal or exclude the presumed anxious or uncertain driving behaviour. On the contrary, thoughtful and vigilant driving behaviour deserves particular recognition, because the test candidate has evidently reflected his still limited driving experience, and has actively adapted his actions to take this limitation into account: The test must not concentrate so much attention on maximising the attained level of training that the aspect of the candidate's attitude to road safety is – paradoxically – neglected (ibid., p. 60-61).[82]

It was already mentioned in connection with the decision criteria (see Chapter 3.4), that the gapless recording of all driving tasks performed is a prerequisite for psychometrically correct consideration of good and error-free test performances, as would be desirable – from the psychological perspective – for a learning-objective-referenced test. This is not the case under the current control concept for the practical driving test in Germany, however, and would consequently mean considerable adjustment and an increased workload on the part of the driving test examiner. In accordance with the presently applicable control concept (in the form of examination guidelines), a practical driving test must also be terminated prematurely – irrespective of an otherwise good test performance – if any seriously incorrect behaviour is observed and it is thus determined that the candidate has failed the test (PrüfRiLi 5.18). This serves to shorten the test duration for the examiner, and in turn permits more time to be devoted to the testing of other candidates where a proper assessment of driving competence is more difficult (HAMPEL, 1977). Even though the conditions for realisation of the practical driving test have changed since the 1970s, and despite the fact that no reliable statistical data exist on premature test terminations, it would nevertheless be expedient to consider modification of the aforementioned points within the framework of a future control concept, since continuation of the current practice would falsify future evaluation results and can furthermore be deemed questionable from the didactic perspective[83]: As long as correct task performance is not recorded, but tests can be terminated prematurely in case of a decisive serious error, evaluations founded on statistical analyses necessarily give a false impression of test reality, because the partial fulfilment of the demand standards by unsuccessful candidates is ignored. "This distortion in the test statistics is a merely theoretical flaw, as long as no further consequences are drawn from the results. As soon as test results are analysed – as we consider necessary – as a means to monitor test practice, however, there is a risk of uncontrolled feedback. For these reasons, we are of the opinion that the rule stipulating premature termination of the test currently robs the candidate of an opportunity to compensate his error and in the long term leads to distortion of the test standards. Termination of the test should thus be limited to those cases in which continuation can be deemed unreasonable for the examiner for truly serious reasons" (HAMPEL, 1977, p. 89).

Furthermore, it should be demanded that both good and inadequate aspects of test performance are recorded "immediately following observation, so as to avoid memory falsification on the part of

[82] It is to be noted that this naturally applies only to the first test to obtain a class B driving licence, but not with regard to applications for a class C or D driving licence, for which paragraph 5.1 of the Examination Guidelines requires that – beyond the demands relating to class B – the candidate must display adequate driving skills in the sense of conversant, safe and deft handling of the vehicle (e.g. even acceleration, calm driving and smooth braking).

[83] The test candidate expects an opportunity to demonstrate his performance capabilities over the whole usual duration of a test. Premature termination curtails this opportunity, which the candidate has previously bought by paying the designated test fee, and although legally permissible, will thus no doubt often be viewed as unfair – especially where the candidate is not convinced of his lack of competence or the hazardous nature of his driving behaviour. This could impair learning motivation.

the examiner and influencing by other observations" (HAMPEL, 1977, p. 87). It must be added as a proviso, however, that this applies only under normal circumstances: In chaotic, turbulent, hazardous or otherwise unpredictable traffic situations, the examiner's first responsibility is to fulfil his observation, planning and not least instructional duties; accordingly, the task of documentation must then be deferred temporarily.

3.6 Summary

The starting point for the research and development work commenced by the working group TÜV ÐEKRA arge tp 21 in 2005 – and continued by way of the present BASt project – was the intention to analyse the contextual and methodical foundations of the practical driving test against the background of current progress in scientific research and international test practice, and thereby to enable optimisation of its instrumental quality as a process of driving competence assessment. The research and development work necessary for process optimisation was outlined in the previous sections of this report, together with the results obtained; in accordance with the methodical nature of this process as a means of systematic (driving) behaviour observation and partially standardised, criterion-driven adaptive assessment of competence, the work focuses on four challenges:

(1) Determination and description of the situation-related demand standards ("driving tasks") which are indispensable for the structuring of a driving test and for the acquisition of meaningful observation data.

(2) Determination and description of situation-independent demand standards ("observation categories") to concretise both the action demands to be met by the test candidate and the observation and assessment demands for the driving test examiner, with a focus being placed on those aspects of behaviour which are significant for road safety and must thus be subject of the test.

(3) Determination and description of behaviour- or event-oriented and competence-referenced assessment and decision criteria which permit instrumentally reliable and contextually valid judgement of the test candidate's driving competence.

(4) Professionally appropriate arrangement and implementation not only of the demand and observation standards, but also of the assessment and decision criteria within a psychologically adequate control concept or test strategy, which in turn governs realisation of the test and takes the relevant test conditions into due account.

re 1: For the determination and description of driving tasks, it was possible to take up the theoretically and methodically sophisticated demand analysis conducted by McKNIGHT and ADAMS (1970), which – alongside further "driver tasks" which were not concerned with the immediate process of driving a motor vehicle in road traffic – specified the typical action sequences which are elementary to the driving process and the (key) situations which facilitate driving behaviour observations. It is probably true to say that this served to identify not merely all significant tasks[84], but also – by way of criticality evaluation – those of particular relevance for road safety, together with correspondingly appropriate performance strategies and a prescribed or typical level of performance. The value of the resulting task descriptions lies above all in their completeness as a reservoir of information on relevant behaviour in road traffic, and in the inherent quantification and substantial qualification of its safety importance.

The task descriptions as they are presented in the demand analysis, however, are not yet suitable for use for training and examination purposes – as McKNIGHT and ADAMS (1970a) themselves emphasised: Where the objective is to elaborate a curriculum for driver education or a methodical manual for the practical driving test, it is necessary to subject the full set of driving demands to corresponding scientific analysis, and thereby to select those tasks which are particularly relevant with regard to safety and thus possess fundamental importance for driving competence acquisition and the development of driving experience. This step was subsequently undertaken by McKNIGHT and HUNDT (1971a) with their Driving Situations Test, which was elaborated for learner assessment purposes; leaving aside a small number of aspects which are only practicable in the context of learner assessment

[84] This assumption is supported by the multi-level process serving to verify the completeness and significance of the identified actions: The aspects of behaviour determined inductively at the first step of the demand analysis were validated deductively by way of systematic scientific dissection and description of the aggregated action sequences at the fourth step; the criticality evaluation then provided for a second validation by traffic experts.

during the actual training, the task selection for this test is quite similar to the contents of a driving task proposal inspired by TÜV Bayern in the 1970s and later presented by TÜV Rheinland. This proposal, however, was not followed up with due resolve within the framework of further development of the German system of driver licensing at that time.

It was not until the present project that the proposals were returned to the spotlight and reviewed to assess the scientific validity of their basic assumptions. The scientific robustness of the determination procedures outlined above seem to confirm such validity; on this basis, the resulting catalogue of driving tasks was then revised by a working group of domain experts. During the course of this revision, the task catalogue was updated and assessed with regard to its road safety relevance; this can be deemed equivalent to a test of content validity. Furthermore, the eight individual driving tasks were described systematically for the first time. This satisfied two necessary prerequisites for learning-objective-referenced tests (FRICKE, 1974) and answered the demands for optimisation which had been pending for several decades (HAMPEL, 1977).

re 2: As was already the case in connection with the driving tasks, the work on the determination and description of five observation categories for a future optimised practical driving test was able to build upon the scientific demand analysis by McKNIGHT and ADAMS (1970a): The category contents are there described as essential, situation-independent action demands. It was furthermore shown that the defined observation categories can be found internationally in the test specifications of numerous reform-oriented countries; they were also already used by a number of Technical Examination Centres in Germany during the period from 1973 to 1996. The road safety relevance of the driving demands described by way of the observation categories was founded not only on the basis of the criticality evaluations contributing to the demand analysis by McKNIGHT and ADAMS (1970a), but also through the findings of more recent accident research addressing novice-typical accident causes and competence deficits. These circumstances, together with the confirmation of content validity by the expert working group appointed within the framework of the present project, indicate that the aforementioned set of five observation categories shifts the most important safety-relevant situation-independent driving demands into the focus of driving behaviour observation in the context of the practical driving test.

re 3: In the course of description of the driving tasks and observation categories by the expert working group, the assessment criteria were similarly reflected from the methodical perspective and revised accordingly: Examples of conceivable "above-average performance" and typical "simple errors" were described for each individual driving task and each of the observation categories; furthermore, a exhaustive list of "serious errors" leading to test failure was compiled and presented. Finally, as a complement to the traditional, event-oriented method of assessment, a competence-oriented mode of assessment was defined, so as to permit aggregated assignment of the individual event-based assessments relating to driving tasks and situation-independent action contexts or observation categories to one of four performance levels. In this way, the performance spectrum of those candidates who pass the practical driving test could be differentiated further as a prerequisite for meaningful test system evaluation. At the same time, the conditions for performance feedback to the test candidate, and thus motivation for further learning, were improved.

Despite this important methodical progress, a number of desirable improvements relating to the proper professional assessment of driving competence must be left to future research and development work: For capacity reasons, for example, it was not possible to elaborate assessment and decision criteria for the test elements "Technical preparation of the vehicle" and "Technical completion of the drive" within the framework of the present project. It was similarly beyond the scope of the project to tackle the urgently necessary revision of assessment and decision standards for the basic driving manoeuvres. Another unanswered question refers to how the event-oriented and competence-oriented assessments can be combined in a meaningful manner and translated into a uniform, psychologically founded decision algorithm by which the examiner can reach a final test decision. Furthermore, from the methodical perspective, it seems desirable to provide also an exhaustive list of “simple errors” which – even if less so than “serious errors” – are relevant for road safety and should thus continue to influence the test decision in the future. These errors should at the same time be distinguished from driving behaviour which, although conspicuous, can be deemed “normal” on the part of a still inexperienced driving test candidate, insofar as this behaviour is irrelevant with regard to road safety and correspondingly for the test decision. It is to be emphasised at this point that immediate continuation of the work to sharpen the assessment criteria, which was commenced within the framework of the present project, is of extraordinary importance for the instrumental reliability of the practical driving test and appears urgently necessary. Finally, no solution was found for the methodical problem of how the number of driving errors observed can be placed in relation to the total number of driving tasks performed over the course of a driving test.

re 4: The systematic description of demand and assessment standards also represented successful completion of an intermediate stage towards introduction of the “adaptive test strategy” as a psychologically founded control concept for the practical driving test: The practical test must no longer be viewed as psychologically incomplete – referring to the lack of standardisation possibilities and quality principles – and can instead be realised as a scientifically founded, learning-objective-referenced process of competence assessment based on systematic, criterion-driven driving behaviour observation. This paradigm shift cannot be deemed to impair expectations relating to methodical quality in any way; nevertheless, the corresponding potential and limitations of the new methodical concept will only be revealed after the commencement of continuous evaluation. The most important prerequisite for successful implementation of the methodical progress in actual test practice, and thus exploitation of its potential, is a system of methodically oriented qualification and further training for driving test examiners. In this respect, it is important to heed a point made by McKNIGHT and HUNDT (1971a) in their explanatory remarks on the proposed Driving Situations Test: The test is only as good as the driving test examiner.

As an outcome of the present research and development work, the involved experts for the first time produced a systematic and scientifically founded description of the methodical architecture underlying the German practical driving test.[85] This represents a major advance, but must also be viewed as a new starting point rather than a conclusion for reformation of the practical driving test: Where an analysis takes into account only expert opinions, there is always “the risk that group prejudices could lead to false priorities being set. The content selection in terms of driving behaviour characteristics should thus always be confirmed by way of empirical verification” (HAMPEL, 1977, p. 91). This requirement applies not only to the specification of demand standards, i.e. to the driving tasks and observation categories, but also to the definitions of assessment and decision criteria and implementation guidelines. In future, therefore, all these standards are to be reviewed continuously within the framework of systematic evaluation (see Chapter 5) and adapted accordingly on the basis of the evaluation results. This will not only drive further development of the test procedures, but also lend new impetus to the elaboration of framework curricula and learner assessment methods for use in driver training.

One important area in which the availability of an explicitly described methodical architecture would facilitate future further development of the practical driving test concerns the question as to how driver assistance systems should be taken into account in the realisation and assessment of tests. The now concretised demand and assessment standards permit more precise investigation of the possible influences of different driver assistance systems, as a basis for corresponding modification of the standards, if deemed appropriate; some initial thoughts are presented in Chapter 6.

Finally, it is to be pointed out that the revision and description of demand, assessment and implementation standards was effected in accordance with the applicable legislative regulations. Due consideration was given, to the fullest extent possible, to the stipulations of the EU Directive on Driving Licences and to the current driver licensing legisla-

[85] It must here be added, as qualification, that the revision and description of demand, assessment and implementation standards has to date only been commenced for the system of testing leading to a class B driving licence; realisation of a similar process for the other licence classes will no doubt entail further considerable effort in the future.

tion in force in Germany. Nevertheless, implementation of the reform proposal will probably require certain modifications to the wording of the Driving Licence Regulations and Examination Guidelines when the time comes, although this is generally considered to be unproblematic (JAGOW, 2010). Before this stage is reached, however, these proposals must be proven in practical trials and further revised as shown to be necessary from the results of such trials. Alongside legal anchoring of the test standards, they must also be established in the form of a "System Manual on Driver Licensing (Practical Test)" ("Handbuch zum Fahrerlaubnisprüfungssystem (Praxis)", see Annex 2[86] to the present report) and – in even greater detail and with reference to the principles of test psychology – in a methodical manual: Despite the differences in customary styles of presentation in the fields of law and psychological methodology, the legally relevant Examination Guidelines and the methodical manual with precise specifications for implementation of a psychologically adequate test must be congruent in terms of content; in this context, the Examination Guidelines could be limited to description of the most important framework conditions, and in particular "cleared" of those detail specifications which can be expected to change over the course of evaluation-based further development of the practical driving test (see Chapter 4 and 5).

[86] The "System Manual on Driver Licensing (Practical Test)" as contained in the annex to this report is a draft for such a document which was elaborated within the framework of this project and discussed accordingly between the Federal Highway Research Institute (BASt), the working group TÜV DEKRA arge tp 21, the Association of Technical Inspection Agencies (VdTÜV) and the Technical Examination Centres. This draft reflects the state of discussions up to 23.02.2011.

4 Documentation of the optimised practical driving test

4.1 Documentation of systematic behaviour observation

4.1.1 Functions and forms of documentation relating to systematic behaviour observation

From the test psychology perspective, the practical driving test can be deemed a work sample which is assessed by way of systematic behaviour observation (STURZBECHER, BÖNNINGER & RÜDEL, 2010). As a prerequisite for methodical optimisation of the practical driving test, therefore, it is first necessary to determine the demands which are to be placed on systematic behaviour observation in general, and on the use of such observations within the framework of single-candidate tests in particular. The documentation of behaviour observations and tests plays a significant role for the answers to these questions, as it represents a decisive "setting screw" for the methodical quality of an observation procedure (KÖTTER & NORDMANN, 1987).

The term "documentation" possesses two fundamental meanings: Firstly, it may refer to a collection of data relating to a particular object; at the same time, however, it could also be used to designate the process of data acquisition and compilation which produces such a collection. The documentation process serves to gather, select, arrange and record data. As far as work organisation is concerned, the finished documentation enables the data to be made available, evaluated and archived. The test psychology functions of documentation in the case of behaviour observations lie in the contribution to perception control on the part of the observer, and in the provision of a basis for evaluation of the observation data, for which purpose the documentation should counteract possible observation or judgement errors, as well as any memory effects (AMELANG & SCHMIDT-ATZERT, 2006).

Various forms of documentation can be used for the documentation of observation data: Generally speaking, observed behaviour can be recorded by either mediated or non-mediated methods. If the data relating to an observation situation are determined, classified and recorded directly by the observer, this process can be termed an immediate or non-mediated observation. A (technically) mediated documentation process, by contrast, uses recording equipment (e.g. video) to record the observation data. Mediated documentation offers the advantage of maximally objective, complete and unmodified ("isomorphous") recordings and unlimited repeatability of the evaluation process; the subsequent evaluation of the observation data, however, must be based on a reductive assessment system with a correspondingly manageable number of categories, which may then be very complex and time-consuming (MEES, 1977). Furthermore, the evaluation of video recordings will generally not convey the "atmosphere" of the specific observation situation (e.g. affective impressions), which may lead to judgement and interpretation errors (KOCHINKA, 2010). Video recordings are nevertheless of particular importance for estimations of the reliability of a documentation method, and this point will be taken up once more in Chapter 5.

There are basically three different process designs which can be applied for the recording and assessment of observed behaviour: Firstly, the observed behaviour can be recorded and at the same time assessed within the framework of the actual observation process. Secondly, observations can be recorded continuously during the process (so-called "monitoring"), but an assessment left until after the conclusion of all observations and recordings. As a third possibility, finally, both observations and assessments can be noted from memory only at the end of the observation (FISSENI, 2004), though this "retrospective approach" will generally lead to an increased likelihood of observation and judgement errors. The most frequent observation and judgement errors which arise in conjunction with the practical driving test were already described in detail by STURZBECHER (2010).

It is additionally possible to distinguish between process and outcome records (JONAS, 2009): The difference between process and outcome records is that the former present not only an overall result, also the course of the observed behaviour over time.

To enable "systematic behaviour observation", so-called "observation systems" are used to control the observer's perceptions and the written recording of the observation data. An observation system (see also Chapter 3) is an instrument by which observation data can be controlled, structured and specified methodically, and in this way made available for systematic evaluation and interpretation. Within any such system, instructions are given to the observer regarding the selection, classification and coding of elements and patterns of behaviour (SCHNELL, HILL & ESSER, 2008).

Documentation specifications, templates and documentation forms are also to be considered components of an observation system: They generally reflect the contents and methodical particularities of the system.

By specifying behaviour indicators (e.g. descriptions and examples), observation systems are able to reduce the burden on the observer's cognitive resources, as they focus his attention on defined criteria and assessment of the extent to which they are met (SWELLER, 2006). On the other hand, the observer will consciously neglect the documentation of observations which are not listed explicitly in the prescribed observation system (HASEMANN, 1983). To summarise, the conclusion reached by FRIEDRICHS (1990) seems pertinent: Within the observation process, an observation system serves "equally to steer and record" (p. 275). The use of observation systems guarantees systematic and controlled observations and enhances their instrumental reliability (BORTZ & DÖRING, 2006).

Three forms of observation system can be distinguished in connection with systematic behaviour observation (FAßNACHT, 2007; KROHNE & HOCK, 2007; SCHNELL, HILL & ESSER, 2008), each of which involves a different type of written documentation:

1. In the case of so-called "sign systems", the specifications list (exclusively) the defined forms of behaviour which are to be recorded, where appropriate, by the observer. Unlike the "category systems" described below, a sign system permits no assumptions to be made with regard to the completeness or coherence of the specified codes (KROHNE & HOCK, 2007). Within the scope of the specifications, observations can be documented either freely or with the aid of observation sheets by way of simple checklists or tally marks, with checklists being used especially where the intention is to reveal quality deficits (so-called "discovery maximisation principle"). An observation can be assigned to exactly one, several or even none of the specified codes in the lists. To evaluate a checklist, it is counted how many of the listed elements of behaviour were displayed or not displayed by the subject of the observation (KANNING, 2004). If tally marks are used, the frequency of observation of each element of behaviour is also recorded.
2. In the case of a so-called "category system", a series of "observation categories" is defined to enable documentation, with each category serving to record certain classes of observation-relevant behaviour with defined attributes. With the aid of these categories, the overall observation-relevant behaviour displayed during the observation phase can be recorded on an observation sheet (which could thus also be termed a "behaviour report"). It must be possible to assign each relevant element of behaviour to exactly one category (FAßNACHT, 2007). The main body of such an observation sheet generally comprises a summary of the defined observation categories, together with corresponding abbreviations, and a table in which the relevant abbreviation must be entered if the associated behaviour is observed (FIEGUTH, 1977). Occasionally, an observation sheet may also include a explanatory legend with more detailed content descriptions of the individual categories (HASEMANN, 1983; MARIN & WAWRINOWSKI, 2006). In many cases where category systems are used, moreover, the records may be extended to include the chronological sequence and duration of the relevant behaviour events. Subsequent evaluation of the records is generally more differentiated and thus more complex than with checklists and tally marks.
3. With "rating systems", the observer is required not only to assign the behaviour observations to a prescribed matrix, but additionally to give a qualitative assessment of the observed behaviour (KOCHINKA, 2010). To this end, the individual form of the observed behaviour is judged with regard to its intensity or conformance with a demand standard (in the sense of required performance, for example) and plotted to a measurement scale (FAßNACHT, 2007). A rating system thus records at least two dimensions of observed behaviour: (1) Whether certain elements of behaviour were displayed or not, and (2) the degree of relevance of the displayed behaviour, where appropriate. The observer may judge the degree of relevance of a given attribute of behaviour either by way of a graduated numerical rating scale or with the aid of a graphic scale; it is furthermore common to assign additional verbal descriptions to the levels of numerical scales (see Chapter 5). When applying rating scales, therefore, the observer functions as a "measuring instrument", and maps the interesting behaviour attributes to a content-related interpretation dimension on the basis of assessment specifications and his own discretion. Compared to category systems, many rating systems can be characterised by a high level of practicability and low time require-

ments; they similarly achieve a high methodical quality.

A suitable means for the documentation of content-specific correlations between several observation dimensions, as is characteristic for a rating system, is an observation matrix, which represents an expansion and thereby the further differentiation of a simple checklist. By way of the rows and columns of a matrix, it is possible, for example, to address (1) certain classes of observation situation and (2) specified observation categories with classifications of certain elements of behaviour, and furthermore, by way of the matrix cells, to record (3) associated assessments (manifestations of the observation categories in defined observation situations). The assessments may be recorded as dichotomous judgements (e.g. "Compliant with demands" versus "Inadequate") or else in more differentiated form by way of a rating scale with three or more levels.

Irrespective of the chosen form and observation system, documentation generally incorporates an introductory section with space for the recording of pertinent information on the observation situation (e.g. date, time, occasion/purpose, conditions), the subject of the observation and the observer. In most cases, form fields are provided for the observer to enter such administrative data. Where documentation is used within the framework of single-candidate tests, it can also be referred to as a "test report", in which – alongside the administrative data – the examined demands, the candidate's behaviour, the examiner's judgements (assessment) and the test decision are recorded. Such test reports are to be deemed the focus of all following discussions.

4.1.2 Documentation of systematic behaviour observation within the framework of single-candidate tests

How can systematic behaviour observation be documented when used in the context of testing? To assist answering of this question, it is helpful to view methods applied in personnel assessment. The most widely known methods – assessment centres and work samples[87] – were already presented by STURZBECHER (2010) and there discussed in detail with regard to their relevance for driving licence testing. Therefore, in the sense of a deeper insight, the present report is to address only the possibilities for documentation offered by these two methods.

Observation sheets are used in various forms for the documentation of assessment centres and process-oriented work samples. OBERMANN (2009) points out, however, that sheets which specify the attributes to be observed but require assessments to be made on the basis of free notes, as well as the increasingly popular "polarity profiles" where certain candidate attributes are judged merely by way of two-pole scales, display deficiencies in respect of their objectivity, reliability and validity. A more suitable approach would be a combination of checklists and rating scales, wherein the demands are operationalised as behaviour criteria; the observer is thus given direct instructions as to the aspects on which attention is to be focussed. An example of a behaviour-based rating scale is shown in Figure 6 below. OBERMANN (2009) notes furthermore that the observer must not be overburdened by an excessive number of overly complex observation categories to be taken into account, and that the behaviour indicators serving to "anchor" at least a proportion of the rating scale levels ("Thurstone scale") must be specified on the observation sheets, instead of remaining "hidden" in an observer's manual: Objectivity can only be achieved – according to OBERMANN – "if the behaviour anchors actually steer the course of the assessment process" (p. 171).

[87] Different forms of systematic behaviour observation, such as assessment centres and work samples, are frequently employed methods in the field of personnel assessment. In the case of an assessment centre, several candidates work together to tackle tasks simulating corresponding job demands, and are observed and assessed with regard to their displayed behaviour by several observers (HÖFT & FUNKE, 2006; AMELANG & SCHMIDT-ATZERT, 2006). A work sample, on the other hand, serves to assess the capabilities of a candidate by way of the behaviour displayed when performing standardised and – for the given domain – representative work tasks, alongside the product of this work (SCHULER & FUNKE, 1995). In both cases, the validity of conclusions which assign observed behaviour to underlying elements of competence is safeguarded by observing the candidate's behaviour several times in different situations (KANNING, 2004, 2005).

	Notes/Observations	1	2	3	4	5
		-				+
Active structuring of meeting (steer discussion, name topics, summarise, pose questions)						
Analysis (questioning to derive background to criticism/problems)						
Unambiguous agreements on concrete activities (who, what, why)						
Structuring of next steps which could enable solution of the overall problem						
Agreement on **outcome monitoring/follow-up**						
	Overall rating					

Fig. 6: Forms of observation sheets: "Behaviour-based rating scale" (following OBERMANN, 2009, p. 184)

The Assessment Centre Working Group (AKAC, 2004) has elaborated quality standards for assessment centres. These standards establish, among others, the following demands relating to the methodical foundation of observation systems and their implementation. Excerpts from the requirements are also included here, as they provide a further, explanatory overview – from a different perspective – of the essential methodical demands placed on the practical driving test in general, and on test reports in particular:

– Demands on observation situations: "Whether certain behaviour is suitable or not depends on the general framework of task situations. Therefore, behaviour can only be observed and assessed realistically in a situational context. In order to draw up an aptitude prognosis relating to a specific target position, the task and work situations must be re-enacted as realistically as possible. … Each job requirement must be recorded in at least two exercises (principle of redundancy)" (p. 6).
– Systematic behaviour observation as the basis for aptitude diagnosis: "These documented observations serve as the fundamental basis for decisions regarding aptitude diagnosis and the identification of profile strengths and weaknesses for each participant. To ensure reliable and valid diagnoses, the application of a job-requirement-related observational system is essential" (p. 7). To this end, a requirement-exercise matrix must be created to provide an unambiguous specification of the requirements to be assessed by way of a particular exercise.
– Well-founded observer selection and observer preparation: An observer must be qualified and adequately trained for the task of assessment.
– Systematic pre-selection of candidates, including advance information: Candidates should always be provided with information on the test requirements, and should possess a realistic chance of success (which should be determined by way of a prior screening or "advance test" of their relevant abilities).
– Good preparation and transparent test realisation: The assessments of test performance must be understandable for the candidate.
– Individual feedback and derived follow-up measures: "The essential contents of the feedback conference are personal strengths

and weaknesses referring to the job requirements. If the assessment centre serves to make concrete decisions, the feedback conference should review the overall decision and concrete recommendations for personnel development" (p. 11).

- Regular quality tests and quality control: Critical methodical analyses must be performed in accordance with specified standards, in order to assess the quality of prognosis, to eliminate procedure-relevant errors and restraints, and to safeguard the fairness and acceptance of the process. The German standard DIN 33430 "Requirements for procedures and their application in job-related proficiency assessment" (DIN, 2002) demands, furthermore, that the procedures used must display the greatest possible objectivity, reliability and validity, that clear rules must exist to govern realisation and evaluation of the process and subsequently the final judgement on aptitude or fitness, and that all procedures, relevant observations, materials and decision rules must be documented in a commonly understandable manner (KERSTING, 2008).

How can (adaptive) testing now be supported and documented optimally by way of test reports? The described methodical basis can be summarised together with principles and examples of selected documentation forms to provide the following answers to this question:

- The only feasible methodical foundation for observation-based tests is an observation system which supplies theoretically founded test contents and methodically sound specifications relating to the realisation and evaluation of observations or tests and to the interpretation of test results. These contents and specifications must be described in a manual for the observer, as is customary in psychological methodology, and lead not least to a corresponding set of documentation specifications and documentation forms.
- Where an adaptive test strategy is applied in behaviour-oriented testing, assessments of the behaviour displayed serve to control and shape the further course of a test. Therefore, assessments and their documentation must be effected simultaneously with behaviour observation during the test; especially in the case of complex adaptive tests, the documentation must be available to the examiner at all times, since it serves as a memory aid with regard to the course of the test so far. This, in turn, can only be achieved with a rating scale system.
- The documentation forms for such tests must serve primarily to control the perceptions and adaptive test planning of the observer or examiner. To this end, they must comprise not only situative demand standards, but also observation categories, and must be both clearly arranged and easy to handle. These requirements can only be met by multi-dimensional or matrix-style test reports, especially where they are provided in electronic form.
- To simplify document management and for compliance with legal requirements, various administrative data must be recorded in the header and/or footer sections of a test report, (e.g. place, date, time, occasion/purpose of the test, special circumstances, name of the test candidate, and name and signature of the examiner). The prescribed input fields should here be offered in a manner which serves to reduce the amount of writing to be done on the part of the examiner. In the main section of the test report, the examiner should be able to record notes on the course and/or results of a test; at the same time, the form should already list possible tasks and exercises which reflect the demands to be met by the test candidate, so as to assist the examiner in his planning and structuring of the observation situation. In conjunction with the specification of observation categories and instructions on how to use the form (e.g. definition of anchor examples for the levels of a rating scale), this facilitates documentation, enhances the clarity of the report, and reduces the risk of mistakes. In case of uncertainty regarding the assessment of a particular candidate attribute (characteristics, competences), the examiner should make a corresponding note in the documentation so as not to "lose sight" of his reservations, and can then – insofar as possible – seek to confirm or refute his original assessment during the further course of the test.
- Finally, it is necessary for the test report to integrate notes on assessment and interpretation of the observation data, as well as possibilities to check the current basis for an assessment and decision; this could be achieved with a general checklist for observation and plausibility reviews. In addition to the recording of identified performance deficits, the examiner should also be expected to register positive aspects of behaviour – in the sense of a development-oriented assessment – and to formulate recommendations aimed at improvement of any deficiencies revealed by

the test; the test report must include suitable documentation fields for this purpose.

In the case of tests where multiple instances of a given demand situation may be encountered with varying levels of difficulty (see above, principle of redundancy), particular attention must be paid to possible inconsistencies in the observed performance. How, for example, is the examiner to assess a situation where the test candidate solves a difficult task flawlessly, but subsequently fails to master a similar, but significantly simpler task? Can errors made when performing simple tasks be compensated by behaviour demonstrating above-average competence where a similar task sets higher demands? And where, if appropriate, should the line be drawn with regard to such possibilities of compensation? To be able to answer such questions unambiguously and with a minimum burden on cognitive resources during realisation and assessment of the test, the examiner must be offered defined decision strategies in the sense of implementation rules. Such implementation rules could stipulate, for example, that even a single failure to meet a certain behaviour demand automatically leads to a negative overall test result – in other words, compensation of the error concerned would be excluded. Alternatively, the rules could indicate the allowed possibilities to compensate errors. In both cases, particular significance must be attached to differentiated documentation of the circumstances, the concrete demands and the candidate's behaviour: In the former case, it serves to justify the negative test decision; in the latter case, it enables direct comparisons of the situational conditions, from which detailed conclusions can be drawn in respect of the candidate's situation-specific competence.

How can observation data and, in particular, performance assessments gained by way of systematic behaviour observation be documented in the case of the practical driving test? And which possibilities exist to improve compliance with the described documentation quality demands through the use of an optimised test report? These questions are now to be answered in the next chapter.

4.2 Documentation of the practical driving test

4.2.1 Fundamentals and starting point

Test reports are a decisive instrument for the legally sound documentation of proper test realisation. It was also for this reason that a uniform, legally binding test report for documentation of the practical driving test was introduced in the German system of driver licensing in 1996. A specimen test report is contained in the “Guidelines for the Examination of Applicants for a Licence to Drive Motor Vehicles” (Prüfungsrichtlinie, PrüfRiLi) as Annex 13. A detailed account of the historical development can be found both in BÖNNINGER, KAMMLER and STURZBECHER (2009) and in HAMPEL and STURZBECHER (2010).

The test report has to date served a primarily administrative and legal purpose, namely to provide legally sound justification for a candidate's failing the practical driving test, in that it can be presented as legal evidence in case of action challenging the decision on failure of the test[88] (JAGOW, 2010). The diverse methodical functions of a test report which were indicated in the previous chapters (e.g. control for observation behaviour and adaptive test planning, basis for an empirically supported formative and summative evaluation of the test) were thus not the focus of test documentation in the past; the traditional test report in accordance with Annex 13 of the Examination Guidelines is not an instrument of test documentation in the methodical sense, as it has no controlling effect on the observations made by the driving test examiner and does not even seek to record all aspects of test performance.

Despite the fact that the applicable licensing regulations (PrüfRiLi 6) demand that the examiner produces a record of the test drive, and particularly of any errors made by the candidate or improper behaviour on the part of the driving instructor in the sense of PrüfRiLi 5.18 (deception or hindering of the test procedure), no specifications exist to define the required form of such records, whether they must be produced during the test drive, and whether positive performance is also to be documented; it is merely mentioned that the latter should be taken into account in the assessment. In

[88] It is to be noted in this context that, legally speaking, the negative test decision is taken not by the driving test examiner, but by the licensing authority: “The decision on failure of the driving test is not an independently challengeable administrative act. The examiner may indeed be performing a sovereign duty, but he is nevertheless acting solely as an advisory expert for the licensing authority, and his judgement on the candidate’s qualification to drive a motor vehicle serves only to prepare the authority’s rejection of the licence application. From the legal point of view, it is the rejection of the licence application on the basis of a failed driving test which constitutes the administrative act which can be legally challenged before an administrative court” (JAGOW, 2010, pp. 148-149). From the methodical perspective, this practice yields the problem that the legally relevant test decision is taken by an administrative staff member who – compared to the driving test examiner – possesses only limited information on the course of the driving test and the candidate's test performance, and will generally not hold the qualifications otherwise necessary for the proper professional assessment of driving competence.

practice, the examiner's notes are not necessarily identical with the test report, but may nevertheless serve as a basis for later creation of a test report after the test drive. It is currently the case, however, that not every candidate receives a test report; a written report is only provided to unsuccessful candidates, along with a brief summary of the most relevant errors (Annex 7 FeV, 2.6). This specification is broadened slightly in the Examination Guidelines (PrüfRiLi 6): “If the test is not passed, then the examiner is to inform the candidate accordingly with a brief indication of the significant errors and is to hand over a test report (Annex 7 FeV, 2.6) which corresponds to the specimen to be found in Annex 13.”

Annex 13 to the Examination Guidelines thus represents the minimum requirements for test documentation relating to the practical driving test. These minimum requirements were taken up and expanded in a catalogue of demands to be met by organisations operating Technical Examination Centres, which was published by the Accreditation Agency[89] for Driving Licence Services (Akkreditierungsstelle Fahrerlaubniswesen) at the Federal Highway Research Institute (BASt) in 2005. With reference to documentation of the practical driving test, it is there stipulated under point 6.7 that the records must comprise at least the test date, the names of the examiner and candidate, the class of driving licence for which the test was taken, the start and end times of the test, the number of basic driving manoeuvres performed, and information on the test environment, e.g. whether the test included driving within built-up areas, outside built-up areas and on a motorway (BASt, 2005). The driving test examiner is able to supplement the above data by way of an additional “Remarks” section; the notes must enable a justification for the test decision to be derived. Within the framework of the second revision of the specifications in 2009, it was added under point 3.8 (“Monitoring of equipment and devices used by examiners”) that, where computers are used in the context of driving licence tests, it must be guaranteed that “the suitability of hardware and software for the intended purpose in conjunction with the test environment has been proven by way of trials”[90], and that “methods and procedures to safeguard the accuracy and integrity of the data must have been introduced and applied” (BASt, 2009). These stipulations originally referred to computer-assisted realisation of the theoretical driving test, but remain unconditionally valid for the case of an optimised practical driving test thanks to the general nature of their formulation.

As already mentioned earlier, the above requirements to be met by a test report documenting the practical driving test clearly reflect administrative and legal expectations; methodical demands, on the other hand, have for the most part not been taken into account to date. The methodical demands on test documentation for the practical driving test were described systematically for the first time by STURZBECHER, BIEDINGER et al. (2010). These demands are derived from the initially described functions of documentation in connection with systematic behaviour observation in general and test reports in particular; the associated fundamental methodical prerequisites were subsequently mapped to the circumstances of the practical driving test. Accordingly, test reports serve organisational, didactic and evaluative functions (ibid.). The organisational or control function alludes to the opportunity for the examiner to record the driving tasks which have already been performed by the test candidate, together with information on the latter's performance; this helps the examiner to plan the further course of a test adaptively in accordance with the demand standards. The didactic function is realised by offering the test candidate feedback on the observed strengths and weaknesses in his driving competence, so as to enable targeted further learning. The test report is thus also to provide a basis for the verbal communication of a development-oriented assessment. The evaluative function, finally, refers firstly to self-evaluation on the part of the driving test examiner, as he is able to review his previous assessments before or even after a final test decision, and can critically reflect the judgements reached on the basis of his report notes. At the same time, it becomes possible to implement external evaluation, as the Technical Examination Centres are able to perform systematic analyses of test reports within the framework of their continuous quality assurance measures.

The three aforementioned functions are served only inadequately, if at all, by the test report currently stipulated for the practical driving test by Annex 13 of the Examination Guidelines; there is thus considerable scope for optimisation. To date, for example, the report serves merely to record errors (in the sense of a checklist), albeit with the opportunity to add written comments alongside each of the 22 potential errors. This test report was already presented in detail and subjected to critical methodical evaluation in the report “Practical Driv-

[89] This BASt office was renamed “Evaluation Agency” (“Begutachtungsstelle”) in 2010 (see Chapter 5).

[90] In the case of an electronic test report, the suitability of the relevant hardware and software is to be demonstrated by way of a feasibility study and revision project (see below).

ing Test – Foundations and Possibilities for Optimisation" (STURZBECHER, BÖNNINGER & RÜDEL, 2010). In the following, therefore, only the principal points of criticism and the ensuing recommendations for optimisation of the documentation for the practical driving test are to be summarised once more. The criticism addressed numerous limitations of the presently used test report in respect of the structuring and formulation of the listed errors, the accuracy with which driving behaviour can be recorded, and the support given to the examiner when determining a test decision. Neither the order nor the formulations of the individual errors follow the stipulations of the Examination Guidelines (PrüfRiLi 5.17.2.1 and 5.17.2.2). Furthermore, it is not clear from the report form, whether the individual errors have been classified as "simple errors" or "serious errors" by the legislator. Their designations are neither specific nor distinctive, and simple errors are in part mixed with serious errors in the list of 22 errors on the report. Since the form and course of the candidate's driving behaviour cannot be recorded, the test report is of little use to the examiner in his planning of a driving route (adaptive test strategy). The test report offers no decision aids to assist judgement of the candidate's driving competence, and the examiner is in no way encouraged to reflect his test decision. Another shortcoming is that no provisions are made for the documentation of positive performance, which makes it difficult to offer differentiated feedback to the candidate on the observed strengths and weaknesses. For the candidate himself, insofar as he fails the test and thus receives a test report, there is little in the report contents which can be taken as a basis for targeted further learning, whether independently or with the assistance of a driving instructor. With reference to the necessary evaluation of the overall manner in which driving licence testing is conducted (see Chapter 5), finally, the limited scope of data and the poor data quality must be faulted: As the present test report does not permit differentiated recording of the driving behaviour displayed by the candidate, and instead merely supplies data relating to the dichotomous test decision and the frequency of observation of indistinctly formulated individual errors, it is not possible, for example, to perform a founded methodical analysis of the correlations between driving competence documented in this way and later driving behaviour data and accident involvement figures for validation purposes.

On the basis of these critical analysis results, the aforementioned report recommended further development of the test report, and in this context revival of the methodically advantageous matrix structure which had already been used in the past, wherein "the test tasks – as the quantitatively largest group – should be assigned to the rows, the observation categories to the columns and the assessment scales to the cells" (STURZBECHER, BIEDINGER et al., 2010, p. 126). Through the implementation of this matrix concept, given the diversity and complexity of the information to be documented and the fact that documentation must presumably be realised on several levels, a corresponding test report would represent a record of "which positive aspects of performance and which errors with which safety relevance were displayed by each candidate with reference to selected elements of behaviour during each of the test tasks. In addition, the test report should naturally continue to provide space for further notes (e.g. candidate data, examiner data, driving test data, remarks concerning premature termination of the test and the behaviour of the driving instructor, signatures, general remarks)" (ibid.). These recommendations are fully in line with the fundamental methodical guidelines described in the previous chapters as a basis for the development of demanding observation systems and the associated forms of documentation.[91]

4.2.2 Test reports in international practice

In the following chapter, an initial overview of international documentation practice is to serve as a basis for more detailed description of the innovative approaches embodied in the test reports used in various other countries and of their potential as a source of inspiration for further development of the test report for the practical driving test in Germany. To this end, the documentation practice in 36 countries[92] was made the subject of deeper analysis; furthermore, survey research was conducted to investigate specifically the experience gained through the use of test reports in a number

[91] The proposals for optimisation of the test report are not new: Already in 2006, the VdTÜV working group on practical testing ("AG Praktische Prüfung, Arbeitskreis Fahrerlaubnisfragen"), whose members represented not only the Technical Examination Centres, but also the federal transport ministry, regional authorities and the driving instructors, reached agreement on the general advantages of a matrix-based report structure, and on the fact that this structure satisfies the demands of an adequate test documentation better than an error checklist (STURZBECHER, BÖNNINGER, & RÜDEL, 2010).

[92] The starting point was a research report on theoretical and practical driving tests in Europe (BÖNNINGER, KAMMLER, STURZBECHER & WAGNER, 2005); in addition, reference was made once more to the results of an international comparative study commissioned by the Federal Highway Research Institute on the systems of novice driver preparation in 44 countries (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014).

of selected countries, and the test reports from 25 countries were analysed with regard to their content and methodical design.

In twelve of the total of 36 countries considered, the examiner already produces a test record during the actual test, i.e. simultaneously with his observation of the candidate's driving behaviour and planning of the test route. A glance into the report forms used reveals considerable variety in terms of the degree of differentiation with which test performance is documented; in general, however, the collected information is standardised and – apart from certain input boxes for brief comments – not recorded in text form (see below). It was reported from five countries that the candidate's test performance is documented exclusively after completion of the test. In the other countries, either no specifications exist to govern the documentation procedure, or else no corresponding information was obtainable on this point. When interpreting these findings, however, it must be taken into account that the driving instructor does not participate in the driving test in many countries, and the examiner thus at the same time assumes the function of legally responsible driver of the test vehicle. This significantly restricts his possibilities for documentation during the test drive. Where the examiner is relieved of this function, correspondingly greater expectations can be held with regard to methodically adequate test documentation.

In almost all countries, a pen and paper are used to record the test data. In Sweden and the Netherlands, on the other hand, test results are noted on special forms with the aid of a so-called "digital pen". In this way, the entries in the corresponding boxes are immediately saved as digital data and can be sent directly to the responsible test office via a mobile phone link. The intention behind the introduction of special test forms and digital pens in Sweden is to achieve simpler and more accurate recording of the test results, alongside acceleration of the corresponding administrative processes (NILSSON, 2008). Driving test examiners in Estonia have been using an electronic test report installed on a notebook computer since 2007. The use of this electronic test report is intended to simplify the task of documentation for the examiner and also enables statistical analyses to be performed (NAGEL, 2008).

For purposes of test quality assurance, a number of countries implement further documentation measures in addition to written reports: In Latvia, for example, the actions of the examiner and candidate are recorded together with the test route by way of two video cameras installed in the test vehicle. In Estonia, alongside the cameras, the test route and driving speed are recorded via GPS (Global Positioning System) and with the aid of acceleration sensors; the brake system is fitted with visual and acoustic signalling devices. In Greece, there are always two examiners present in the test vehicle; this contributes further to assessment objectivity and documentation of the test.

Which data are documented by way of the test reports? In approx. 70 per cent of the 25 countries from which corresponding information was received, the examiner is required to provide a record of the test duration (start and end times). The test location is recorded in approx. 50 per cent of the aforementioned countries (often in the form of an assigned test location number). Information was obtained from 25 countries with regard to possible provisions for written explanations to be given on the test report in case of incorrect performance of certain test demands (driving tasks); such provisions exist in 33 per cent of the countries. As a further element of the Swedish test report, the examiner gives an assessment of the traffic density encountered during the test on a five-level scale (separately for the two categories "Within built-up areas" and "Outside built-up areas"). The weather conditions are recorded in Austria, Estonia, Finland, Great Britain, Lithuania and South Africa. The Austrian test report offers the examiner five choices to describe the weather conditions: "Dry", "Wet", "Snow", "Ice" and "Fog". In Finland, the examiner records not only the conditions "Icy" and "Rain", but also whether the test was conducted in the "Dark". The most detailed record of the weather conditions which prevailed during a driving test is kept in Great Britain. Here, a total of 10 predefined constellations are distinguished to reflect – in an apparently non-systematic manner – the dimensions "Weather" and "Road conditions: "Bright / dry roads", "Bright / wet roads", "Raining throughout test", "Showers", "Foggy / misty", "Dull / wet roads", "Dull / dry roads", "Snowing", "Icy" and "Windy". In addition, the examiner is also able to provide an accurate description of the weather conditions in a free-text box.

An overview of those countries in which a test report is produced and subsequently handed over to the candidate at the end of the practical driving test is provided by Table 13. Information on whether or not a test report is handed over to the test candidate was received from 32 countries. In 24 countries (75%), the test candidate is presented a written report at the end of the test drive, with notes on his performance of the test tasks and pointers to any driving errors observed; in some countries, such as Great Britain, Hungary, Latvia and South

Africa, the candidate is also required to sign the examiner's test report. Among the major West European countries, Germany is the only country in which the candidate only receives a report if the driving test is failed; this is similarly the case in Poland, Estonia, Luxembourg and a number of Swiss cantons. In Iceland, Romania, Slovakia and the Czech Republic, the driving test examiner never hands over a test report, irrespectively of whether the test is passed or failed. In the Australian state of Victoria, the test candidate receives not the actual test report, but instead a brief summary of the performance assessment. This summary indicates the aspects of the candidate's driving which need further improvement, distinguishing between six areas of driving competence (control, observation, signalling, positioning the car, gapselection and speed choice). It is only in Estonia and the Netherlands that the test report is made available in electronic form after the practical driving test. Information on whether or not the driving test is followed by a brief consultation with the examiner, as an opportunity to provide verbal feedback on the observed test performance to the candidate and, where appropriate, to the driving instructor, was received from 33 countries. In 23 countries (70%), such consultations are always held to enable the examiner to communicate his test decision and to point the candidate to any driving errors and possibilities for improvement, irrespectively of whether the test is passed or failed. In Luxembourg and the Czech Republic, verbal feedback is only provided if the test is failed; in eight other countries, including the major West European country France and Germany's neighbour Poland, there is never a subsequent consultation with the examiner. In a few countries, for example Belgium and the Netherlands, the consultation is held not in the vehicle, but instead in a special conference room. A German driving test examiner is only required to inform an unsuccessful candidate as to the most significant errors observed. In Austria, the candidate is entitled to know how his test performance has been assessed in the examiner's expertise. To this end, the driving test examiner must discuss the course of the test and the grounds for his decision with the candidate, and must hand a copy of the test report over to the candidate in case of a negative result or prematurely terminated test. If the candidate wishes, he may also receive a copy of the report in the case of a positive test result.

Country	Test report handed to candidate			Verbal feedback			
	No	Only if failed	Always	No	Only if failed	Always	Duration (minutes)
Belgium	-	-	X	-	-	X	~ 5
Denmark	-	-	X				
Germany	-	X	-	-	-	X	
Estonia	-	X	-	-	-	X	~ 5
Finland	-	-	X	-	-	X	~ 10
France	-	-	X	X	-	-	-
Greece				X	-	-	-
Great Britain	-	-	X	-	-	X	
Ireland	-	-	X	X	-	-	-
Iceland	X	-	-	-	-	X	~ 5
Croatia	-	-	X	-	-	X	~ 3-5
Lithuania	-	-	X	-	-	X	~ 5-10
Latvia	-	-	X	-	-	X	~ 5
Luxembourg	-	X	-	-	X	-	< 10
Malta	-	-	X	-	-	X	
New Zealand				-	-	X	
New South Wales/AUS	-	-	X	-	-	X	
Netherlands	-	-	X	-	-	X	
Norway	-	-	X	-	-	X	~ 5-10
Ontario/CAN	-	-	X	-	-	X	
Austria	-	-	X	-	-	X	
Poland	-	X	-	X	-	-	-
Portugal	-	-	X	-	-	X	
Quebec/CAN				X	-	-	-
Queensland/AUS	-	-	X	-	-	X	
Romania	X	-	-	X	-	-	-
Russia	-	-	X	-	-	-	-
Sweden	-	-	X	-	-	X	~ 5
Switzerland	-	-	X[1]	-	-	X	
Slovakia	X	-	-				
Slovenia	-	-	X	-	-	X	~ 5
Spain	-	-	X	-	-	-	-
Czech Republic	X	-	-	-	X	-	
Hungary	-	-	X	-	-	X	~ 5
Victoria/AUS	-	-	X	-	-	X	

Tab. 13[1]: Handing of a test report to the candidate and/or verbal feedback on the result of the practical driving test in international comparison

Additional remark:

[1] Legend: "X" means "applicable"; "-" means "not applicable"; grey cells indicate that no relevant information is available.

With regard to the extent to which the candidate's performance is documented during the test drive, and in this context furthermore the realised level of differentiation, significant differences can be found between the various test reports. It is initially conspicuous that, in approx. 70 per cent of the 25 countries for which the test reports were analysed, the examiner documents merely the errors made by the candidate during the test, for example by entering a mark, letter or the like alongside the corresponding item in a list of driving and vehicle handling errors. In the majority of test reports, this list contains more than 25 errors to be documented by the driving test examiner. For better clarity (and in this way also to facilitate handling), some test reports use colour and subdivisions to distinguish various types and groups of errors. The test report from Ireland, for example, uses differently coloured boxes for the individual error categories; the boxes alongside each driving task must be marked accordingly to indicate the severity of an observed error: Grade 1 errors are recorded by marking a corresponding green box, grade 2 errors by marking a blue box, and grade 3 errors by marking a red box (see Fig. 7).

FAULTS	GRADE 1	GRADE 2	GRADE 3	FAULTS	GRADE 1	GRADE 2	GRADE 3
1. RULES/CHECKS				11. MAINTAIN REASONABLE PROGRESS AND AVOID UNDUE HESITANCY WHEN			
2. POSITION VEHICLE CORRECTLY AND IN GOOD TIME							
ON THE STRAIGHT				MOVING OFF			
ON BENDS				ON THE STRAIGHT			
IN TRAFFIC LANES				OVERTAKING			
AT CROSS JUNCTIONS				AT CROSS JUNCTIONS			
AT ROUNDABOUTS				AT ROUNDABOUTS			
TURNING RIGHT				TURNING RIGHT			
TURNING LEFT				TURNING LEFT			
STOPPING				CHANGING LANES			
FOLLOWING TRAFFIC				AT TRAFFIC LIGHTS			
3. TAKE PROPER OBSERVATION				12. MAKE PROPPER USE OF VEHICLE CONTROLS			
MOVING OFF				ACCELERATOR			
OVERTAKING				CLUTCH			

Fig. 7: Excerpt from the test report used in Ireland

The Swedish test report is similarly structured with the aid of colours; this serves to enhance the practicability for the examiner, as relevant correlations can thus be recognised quickly and unambiguously. The first block is used to record general information to identify the test participants (candidate and examiner) and to indicate the class of driving licence for which the test is being taken. The second block (see Fig. 8) contains descriptions of the areas of competence to be tested, in the form of curriculum goals (partial competences, which, from the methodical point of view, function as observation categories) and driving tasks. The seven curriculum goals or observation categories are each assigned a letter (A to G) and one of four colours. Below, a total of 26 driving tasks (including basic driving manoeuvres) are listed together with a row of coloured boxes representing the seven curriculum goals or observation categories (with their assigned letters). When a driving task is performed, this is recorded by crossing off the corresponding task number (1-26). The driving tasks and curriculum goals (observation categories) thus combine to form a table or matrix in which any observed errors can be marked. In addition, the Swedish test report requires that the prevailing traffic density within and outside built-up areas be recorded on a five-level scale (see above).

The third block of the report serves to provide information on the successful completion of individual test elements (safety checks, basic driving manoeuvres) and possible interruption of the test. Space is also available for additional remarks. As is similarly the case in the Netherlands, the driving test examiner in Sweden is advised to make as few marks and notes as possible during the actual test drive; since no provisions are made to record good performance, this could otherwise unsettle the candidate by indirectly suggesting a negative assessment of his test performance so far. Accordingly, it is usual to make a brief note of at most those driving situations which are considered particularly important for the later performance assessment. At the end of the test drive, the report is handed over to both successful and unsuccessful candidates. Furthermore, the candidate and his driving instructor receive verbal feedback on the previously displayed performance during a brief subsequent consultation with the examiner.

Competence areas	Cross = Failed	Goals stated by Curriculum
Knowledge of the vehicle Manoeuvring	A	Conducting vehicle checks and identifying risks related to the function and the manoeuvring of the vehicle
	B	Manoeuvring the vehicle routinely and practising different ways of retarding the vehicle
Environ/Econ Driving	C	Practising a driving technique resulting in low fuel consumption
Traffic Rules	D	Implementing/Practising the rules valid for driving with vehicles - Speed
Traffic Safety Behaviour	E	Showing good search routines and identifying risks in different traffic situations and traffic environments
	F	Foreseeing and judging consequences and alternate courses of events and driving with appropriate safety margins
	G	Adjusting driving to actual circumstances (e.g. speed and placement), and interplay with other road users - Speed

Test content assessed	Failures according to goals in curriculum	Test content assessed	Failures according to goals in curriculum
1 Safety check/driving position	A B C D E F G	14 Passing stationary vehicles	A B C D E F G
2 Description of functions	A B C D E F G	15 Indep. driving towards destination	A B C D E F G
3 Spec. man. test/effective breaking	A B C D E F G	16 Driving in road work area	A B C D E F G
4 Parking	A B C D E F G	17 Railway/tramway crossing	A B C D E F G

Fig. 8: Excerpt from the test report used in Sweden

For maximum clarity, the Austrian test report is divided into four separate blocks ("Vehicle checks", "Low-speed manoeuvres", "Driving in traffic" and "Discussion of experienced situations"), each of which is designed differently. The "Low-speed manoeuvres", which correspond to the basic driving manoeuvres stipulated for the German driving test, are presented in the form of a schematic drawing of the course which the candidate is required to negotiate on a practice ground (see Fig. 9).

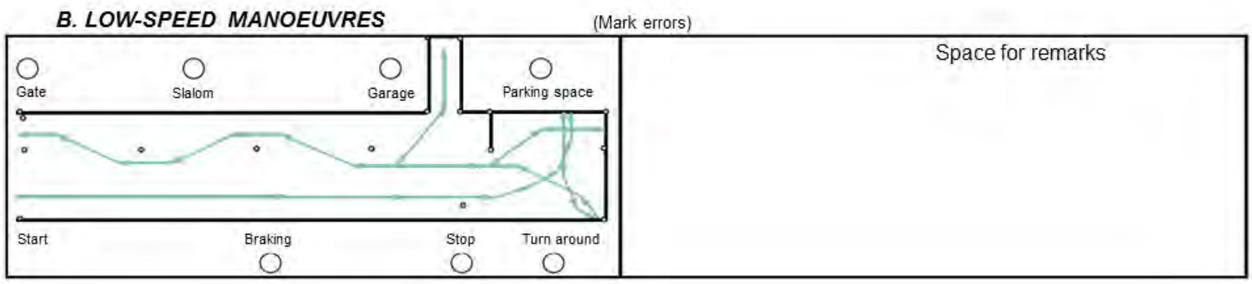

Fig. 9: Excerpt from the test report used in Austria

Alongside Sweden and the Netherlands (see below), two further countries which implement the three-dimensional documentation of candidate behaviour which, for methodical reasons, is favoured for further development of the German test report, are Finland and Estonia.

The test report used in Finland is divided into two blocks. The first block comprises general information relating to the test, alongside the candidate's self-assessment of his driving abilities, which must be given before commencing the test, but then remains unseen by the examiner until the concluding discussion after the test drive. For this self-assessment, the test candidate is asked to grade his driving competence on a five-tier scale, ranging from "poor" to "excellent", in each of eight demand categories (e.g. vehicle control, recognising and avoiding risks, social competence, economical driving). After the driving test, the driving test examiner uses the same categories to elaborate his own assessment, and then compares the results of his own observations with the candidate's initial self-assessment within the framework of a concluding discussion. In the second block (see Fig. 10), alongside basic driving manoeuvres, a series of test items relating to technical preparation and completion of the drive, and special test circumstances such as darkness, slippery roads or rain, a 21-row matrix lists also twelve driving tasks which are not necessarily performed during every test. The columns of the matrix represent six observation categories, which are identified by the letters A to F.

H Good performance
V Fault
K Conflict

A Observation
B Assessment
C Speed control
D Interaction
E Correct position of the vehicle
F Complying with traffic control

	A	B	C	D	E	F
Control of the vehicle						
Check before starting						
Speed control; starting off, crawling, stopping, changing gear, hill start						
Steering						
Dimensions of vehicle						
Reversing; reversing around a corner, parallel parking, reverse parking into a parking space						
Finishing the drive; gear, hand brake, lights, wipers						

Fig. 10: Excerpt from the test report used in Finland

The matrix is divided into four sections: “Control of the vehicle”, “Driving within built-up areas”, “Driving outside built-up areas” and “Driving in difficult circumstances”. Three categories are available to enable the examiner to assess the candidate's performance: “Fault” (abbreviated as the letter “V”; a single occurrence of an error belonging to this category does not already lead to the candidate failing the test; they thus correspond to the “simple errors” of the German system), “Conflict” (“K”; errors of this category result in immediate failure of the test and thus correspond to “serious errors” in Germany) and “Good performance” (“H”). The category “Fault” refers to behaviour which increases the general risk of an accident, but was not immediately dangerous in the actual driving situation. By contrast, an error of the category “Conflict” is recorded where the candidate is responsible for a situation in which other road users were forced to react to avert danger or else the driving test examiner was required to intervene. The appropriate letter (V, K or H) is entered in the relevant cell of the test report matrix. If neither incorrect nor outstanding behaviour was observed during the test drive, the corresponding box remains empty. In addition to such coded observations, the examiner may also add written notes to each line.

In Estonia, too, the candidate must give an assessment his own abilities before commencing the test, in this case by judging his driving to be either “good” or “satisfactory” in respect of the categories “Vehicle operation”, “Traffic observation”, “Cooperation with other road users” and “Safe style of driving”.[93] These data are noted in an electronic test report together with information on the weather and road conditions, the safety checks performed, the test duration and separate assessments of the candidate's driving performance for the categories “Driving in an urban environment” and “Driving in a rural environment” (see Fig. 11). For the assessment of driving competence, any errors observed during the performance of 18 driving tasks are related to 16 aspects of competence in the sense of observation categories and graded by way of a three-level scale (“Single simple errors”, “Repeated simple errors” and “Serious errors”). The serious errors are additionally specified more precisely.

[93] It is a subject of controversial debate in the field, whether a driving licence applicant should be required to give a self-assessment of his driving competence as part of a driving test; as yet, there has been no empirical evaluation of the true impact of capabilities or training relating to realistic self-assessment on road safety. From the methodical viewpoint, the introduction of decision-relevant self-assessments cannot be considered to raise the validity of the practical driving test as an instrument to judge driving competence, as self-assessments are not suitable for objective and error-free measurement (supposed problematic aspects are, for example, conscious distortion of the impression conveyed on the part of the candidate, and the difficulty for the examiner to reach an adequate judgement on self-assessments within the very limited duration and under the exceptional circumstances of the driving test). Consequently, the pedagogically desirable identification of discrepancies between the novice driver's self-assessment and his actual abilities is to be seen as a challenge for professional driver training, rather than a task of driving licence testing.

Fig. 11: Excerpt from the test report used in Estonia

The test report used in the Netherlands contains not only administrative data, but also a performance assessment relating to technical preparation of the vehicle, and a documentation and assessment matrix in which all the relevant driving tasks and observation categories are assigned to the rows and columns. In addition, four special driving manoeuvres ("basic driving manoeuvres") are listed in a separate block. The driving test examiner is advised to complete the test report only after completion of the test drive; only the observed errors are recorded. As in Finland, candidates in the Netherlands must also give an advance self-assessment of their driving competence. This takes the form of a questionnaire applying a ten-point scale to the categories "Vehicle control", "Safety", "Traffic adaptation", "Social behaviour" and "Environmentally aware driving"; these categories are not identical with the examiner's observation categories, however, despite the fact that this would actually permit more meaningful conclusions to be drawn. The completed questionnaire is handed to the examiner in a sealed envelope at the beginning of the test, and is then opened and compared with the observations made by the driving test examiner when the final test result is announced.

In Norway, the test methodology displays a very distinctive particularity, in that the test route is already determined in detail before the test, and the examiner is then obliged to follow the specified route. For test documentation, the examiner uses a test route file with a route plan for the selected test route and an assessment table for the more than 30 driving tasks. Neither the number of errors nor the error categories are revealed to the candidate at the end of the test, as it is feared that this could devalue the targeted holistic and competence-oriented approach to driver training in favour of schematic error-prevention training. Instead, the candidate receives a test report containing the overall result of the test and a written category-independent "competence-based" assessment of his test performance. The examiner enters the test result, as well as the types and numbers of observed errors, into a special PC, whereupon, in case of a successful test, a member of the administrative staff issues a temporary driving licence to the candidate; the actual card driving licence is received by post approximately a week later. The standardised test documentation permits statistical evaluation of the error reports from all examiners and can supply pointers to possible misuse of the system or individual assessment categories, for example. Despite the very high level of detail in the documentation of test performance, it is not possible to reconstruct the specific driving task or traffic situation in which a particular assessment was made from the report: The report merely assigns the individual assessments to a certain route section, and thus serves rather as a memory aid helping the examiner to recall individual traffic situations.

From this overview of the international practice relating to test documentation, it is possible to derive various starting points for further development of the German test report in terms of the structure and design of the documentation forms, as well as expansion of the framework conditions to be documented. Particularly those test reports which specify both driving tasks and observation categories, and thus provide for multi-dimensional assessments of driving competence, confirm the potential of the path chosen for further development in Germany, which was already outlined by BÖNNINGER et al. (2010) and is to be pursued further in the following. The approach corresponds to the developments in test documentation in a whole group of countries which are considered progressive and reform-oriented in the field of novice driver preparation. It can also be recognised that a concluding discussion with the examiner on further learning potential, and the communication of a correspondingly meaningful test report to all candidates, i.e. irrespectively of whether the test is passed or failed, are seen as indispensable aspects of test design at international level.

4.3 Electronic test reports for the optimised practical driving test in Germany

Potential and limitations of electronic test documentation

One of the recommendations given by BÖNNINGER et al. (2010) to promote further development of the practical driving test in Germany was that test documentation should be realised by electronic means. Taking up this recommendation, the following chapter is now to investigate the concrete benefits which are to be expected from the introduction of an electronic test report, and how this could serve to improve test quality. The key benefits expected include:

(1) Support for the driving test examiner by fostering the professionalisation of his work (e.g. through unambiguous clarification of the demand and assessment standards)

(2) Support for the driving test candidate by way of differentiated feedback on test performance as a basis for further driving competence acquisition

(3) Simplification of test administration and the archiving of test data

(4) Optimisation of quality assurance, not least through the improved possibilities for evaluation.

re (1): On the basis of a meta-study of published research on observation methodology, KÖTTER and NORDMANN (1987) elaborated three starting points for a process to secure the methodical quality of observation results, particularly for the case of frequently performed observations: Firstly, the conception and structuring of the observation situation; secondly, documentation of the data acquired in the observation situation; and thirdly, the evaluation methodology. STURZBECHER, BIEDINGER et al. (2010) took up this analysis for their development of an adaptive test strategy, wherein five action steps address the three aforementioned approaches to quality optimisation and describe the demands to be met by the driving test examiner (see also Chapter 3). An electronic test report is able to lend significant support at each of these five steps:

– "Planning and structuring of the test or observation situation on the basis of demand standards and through determination of the test route" (first step) is simplified, as all relevant demand standards (i.e. the test demands and, in particular, the corresponding driving tasks) can be presented in a clearly arranged manner in an electronic test report. During the course of a test, the examiner can at any time refer to his progressive documentation to gain an overview of the driving tasks which have already been performed and to review his assessments of those tasks. In this way, he is easily able to plan the further course of the test on the basis of the candidate's test performance to date, and can continuously adapt the test route accordingly.

– "Systematic observation of the behaviour of the test candidate on the basis of observation categories" (second step) is supported in that the electronic test report can contain an overview of all applicable observation categories, which is then directly available to the examiner to guide his observations and regulate the focus of his attention.

– "Interpretation and assessment of the behaviour of the test candidate on the basis of assessment criteria and the documentation of performance assessments" (third step) is rendered appreciably simpler through the use of an electronic test report, as it is possible to supply flexible templates for the documentation and assessment of all applicable test demands. The amount of writing required for documentation can thus be reduced to a minimum. The availability of multimedia descriptions of concrete assessment criteria (errors, above-average performance, competence levels), together with plausibility checks, is at the same time able to compensate possible memory weaknesses and in this way serves to avoid documentation and assignment mistakes. Furthermore, a continuous test record could contribute significantly to the identification of test anxiety and its negative influence on the validity of the test performance displayed by the candidate.[94]

[94] Already at the beginning of the chapter, reference was made to the alternatives of process- and outcome-oriented methods for the recording of behaviour observations. In test situations, recording of the candidate's behaviour over a period of time may be especially relevant and desirable: In the case of the practical driving test, it appears reasonable to assume that candidates suffering from test anxiety will – as a result of the associated nervousness – make more driving errors at the beginning of the test than at a later stage after they have re-

- The "elaboration of decision preferences and appraisals of the corresponding decision certainty and justification" (fourth step) could be facilitated by an electronic test report, as the possibility to visualise a comparative overview of all observed behaviour would help the examiner to recognise and interpret indicators of competence deficits over the whole course of the test.
- "Determination of an adequate test decision" (fifth step), finally, could be promoted decisively by way of an electronic test report, as it offers a relatively straightforward means to reflect all event-related behaviour displayed by the candidate during the course of the test (errors of different types, frequency and severity; above-average performance), enabling the examiner to condense his various competence-related observations, to weigh up his corresponding competence assessments in a professionally adequate manner, and on this basis to reach a valid test decision.

re (2): Differentiated and clearly structured electronic documentation of all behaviour observed during the test would enhance the transparency of the test decision in the eyes of the candidate, and would enable the examiner to offer the candidate elaborated feedback[95] on the development status of his driving competence. The candidate could then apply this feedback to maximise the safety impact of his subsequent continued learning in the context of solo driving, or else to guide his learning for a possibly necessary repeat test.

re (3): The digital archiving of test and performance records would streamline test administration, accelerate control over test procedures, and enable practically immediate forwarding of the test results to the licensing offices and the Federal Motor Transport Authority (Kraftfahrt-Bundesamt, KBA). It can be assumed that direct electronic exchanges of existing or machine-readable data would greatly simplify the necessary documentation of administrative details, and the same would also apply with regard to the immediate electronic forwarding of test data to the Technical Examination Centres and the driver licensing authorities. In addition, necessary updates to implement amendments and improvements to the test report could be made available instantly to all Technical Examination Centres, and thus to all driving test examiners throughout the country, via the software of an electronic test report.

re (4): Another essential benefit to be derived from optimisation of the test documentation concerns the innovative opportunities for quality assurance relating to driving licence testing. An electronic test report could even contribute to validation of the overall system of novice driver preparation, within which the practical driving test represents a key control instrument. With the aid of meaningful statistical data and analyses, as the basis for formative and summative evaluations, it would become possible to draw empirically confirmed conclusions, for example on the quality of driver training or the suitability of certain test locations. This could serve to promote scientifically founded further development of the practical driving test and indeed the overall system of novice driver preparation.

gained self-confidence. Where event-oriented assessments are involved, recording of the time of a particular observation would facilitate the identification of such distortions in the demonstrated performance capabilities. It remains to be clarified, and possibly formulated in implementation regulations, how such disturbing influences are to be taken into account in the overall judgement of a candidate's test performance. With regard to the necessary decision rules, it is conceivable that simple errors could be viewed more leniently up to a certain point at the beginning of the test, whereas serious, safety-relevant driving errors remain excluded from compensation by way of subsequently positive performance, even if they are committed at the start of the test and are obviously attributable to test anxiety. The experience reported by both driving test examiners and driving instructors suggests that this essentially mirrors the current practice, as is also deemed desirable by the practicians. It is thus recommended that rules be developed to govern such cases, in order to enhance test objectivity.

[95] Development-oriented feedback on learning status is considered an important element of teaching/learning processes and provides valuable information as to whether certain learning objectives have been attained, and the extent of any discrepancy between actual and targeted learning status (CARVER & SCHEIER, 2000); this can be viewed as the diagnostic function of feedback (KULHAVY, 1977). Performance feedback permits verification of a learner's self-assessment and serves to prevent misconceptions and illusions of competence (MORY, 1996). With regard to knowledge, feedback helps to close any gaps in the learner's mental models and thus corrects imperfect knowledge (VASILYEVA, PECHENIZKIY & BRA, 2008). Furthermore, feedback influences the affective-motivational prerequisites for learning and, in the ideal case, is able to promote further learning through achievement motivation or self-efficacy, for example. Generally speaking, the precise purpose and expected benefit of feedback are highly dependent on the particular learning process and its conditions, and must thus be determined for an individual case. Possible forms of development-oriented feedback for an optimised practical driving test were not a subject of the present studies, but will be treated within the framework of a later project.

To summarise, it can be said that a methodically optimised electronic test report represents an improved basis for the implementation of an adaptive test strategy by the driving test examiner, for the proper professional assessment of driving competence, for the realisation of a concluding feedback discussion with the driving test candidate, for subsequent further learning on the part of the novice driver, and not least for test evaluation, wherein there is evidently an urgent need for further development. Consequently, the initially mentioned recommendation was taken up in the present project, leading to the development of a draft for such an electronic test report. The development was based essentially on the following four pillars:

(1) Scientific methodical knowledge relating to the documentation of systematic behaviour observation, particularly in the context of a test situation (see above)
(2) The (class-specific) legal framework governing the granting of driving licences
(3) The catalogue of driving tasks elaborated within the present project for an optimised practical driving test
(4) Practical experience gained by the Technical Examination Centres with test documentation, along with corresponding inspiration drawn from the international field.

The ergonomic challenge for software design when producing an electronic test report for the practical driving test can be illustrated by way of a brief calculation: Combining the observation categories and the various driving tasks and subtasks elaborated in the previous chapters, it can be seen that there are more than 250 possible pairings, each of which must be assessed against a four-level scale to achieve methodically persistent and complete test documentation. Using paper and pencil, this degree of detail – if at all feasible – would require a report comprising several dozen pages, and the time spent searching for individual entries would reach such proportions that reporting in this form would no longer be practicable within the framework of a normal driving test: Firstly because the working conditions for the examiner in the test vehicle would not permit this, and secondly – more importantly – because the task for the examiner is not merely documentation, but at the same time also proper observation and assessment of the candidate's performance, alongside (adaptive) planning of tasks for the further course of the test.[96] Consequently, for the development of a simple-to-use electronic test report based closely on the methodical necessities of the test process, it is imperative to pursue design solutions which minimise the mental and temporal workloads placed on the examiner and are thus conducive to a differentiated and valid reflection of the test performance displayed by the candidate.

Insofar as the aforementioned challenges are mastered, it can be assumed overall that the driving test examiners, test candidates and driving instructors, as well as the Technical Examination Centres and the administrative authorities, would benefit from the introduction of an electronic test report. To guarantee acceptance and to actually realise the full potential benefit, however, a number of important points must be taken into account when planning the transition. As a more ambitious approach to documentation compared to the present practice, an electronic test report would be accompanied by a certain increase in the amount of work involved. From the occupational psychology perspective, it must also be remembered that the examiners have become accustomed to documenting primarily selected driving errors, and have in many cases elaborated individual documentation instruments as aids to note-taking alongside the test report stipulated in Annex 13. After all, the failed driving tests, i.e. approx. one-third of all tests, are already documented in detail by way of an Annex 13 report. Compared to these familiar work processes, the use of an electronic test report would bring additional requirements and new ways of thinking. To counter possible acceptance difficulties, therefore, the chosen ergonomic and media design of both hardware and software must be implemented such that a high degree of usability is ensured in daily practice. For this same reason, it also appears imperative to flank the introduction of an electronic test report with a feasibility study and analyses of its ergonomic practicability and usability, with corresponding offers within the framework of professional qualification and further training for driving test examiners, and last but not least with the expedient supervision, work aids and computer-based training measures.

[96] Adequate implementation of the demand standards by way of adaptive test route planning, attentive observation and professional assessment of the driving behaviour displayed by the candidate, and maximally complete parallel documentation of the essential elements of test performance must be considered tasks of equal value: The neglecting of any one of these tasks would automatically devalue the treatment of the other two. Even so, proper driving competence assessment requires that the examiner invests very different amounts of time in the processing of each of the three tasks. The main share must naturally be assigned to the observation task, which should also only be interrupted for as short a time as possible, because longer interruptions – as shown by studies tracking drivers' eye movements – would entail a significant extension of the period required for re-orientation in the traffic situation, and would thus represent an additional burden on the examiner's observation resources.

It remains to be mentioned, finally, that – in contrast to a system of paper-and-pencil documentation the failure of an electronic device can never be excluded with absolute certainty. Fall-back solutions must be provided to cater for such situations. This could be achieved with a paper version of the test report comprising a similar driving task-observation category matrix, with relevant events being documented by way of a consecutive number in the appropriate cell and a corresponding note in a free-text box. Competence-oriented assessments, on the other hand, could be entered directly in the matrix.

Content and functions of the electronic test report

The content and design demands to be met by the electronic test report derive from the content structures of driving competence and the mechanisms of driving competence acquisition, from the applicable legal framework, and from the methodical principles and optimisation potential elaborated in the previous chapters. Correspondingly, the input form for the electronic test report for licence class B[97] should comprise seven sections[98]:

(1) Administrative data
(2) Technical preparation of the vehicle
(3) Basic driving manoeuvres
(4) Event-oriented assessment of the driving tasks with reference to the observation categories
(5) Technical completion of the drive
(6) Overall assessment of the driving competence displayed when performing the individual driving tasks, and overall assessment of the individual dimensions of competence (observation categories)
(7) Final test decision.

Furthermore, the electronic test report must offer possibilities to record situation-specific notes in a comment field at any given time. Such notes could refer to the circumstances of particular driving tasks or traffic situations, the behaviour of the driving instructor, or other general points of relevance for test realisation and documentation.

re (1): The first section of the electronic test report is to be used to enter necessary administrative data. This is understood to mean information about the test (date, time, licence class), the candidate (candidate ID number, surname, first name, date of birth, gender), the examiner (examiner's name, office) and the candidate's driving school. It is furthermore recommended to provide an opportunity to record the weather conditions prevailing during the test (dry, rain, snow, ice, strong winds), the traffic densities encountered (low, medium, high, congestion) and information on the test route (30 km/h zone, urban roads up to 50 km/h, roads outside built-up areas, motorway/other high-speed road), preferably with reference to particular test elements or driving tasks. Implementation examples from other countries offer corresponding inspiration for the form of operationalisation (see above).

re (2): The second section serves to record test performance relating to “technical preparation of the vehicle” before the test drive. In accordance with Annex 10 to the Examination Guidelines, the electronic test report is here to contain standardised specifications of the test demands and a list of typical errors or knowledge deficits.

re (3): The third report section “Basic driving manoeuvres” is devoted to demand and assessment specifications derived from Annex 7 to the Driving Licence Regulations and Annexes 2 to 6 to the Examination Guidelines. The corresponding report template must here be adapted to the specific requirements of a particular driving licence class (Annexes 3 to 6a to the Examination Guidelines). Furthermore, certain functions should remain accessible at all times and as standard features for all tests (e.g. comment fields to record the premature termination of a test and the corresponding reasons for such termination, such as damage to the test vehicle, an accident, adverse weather or traffic conditions, or health problems on the part of the candidate, examiner or driving instructor). It must also be possible to document doubts as to the candidate's fitness to drive in accordance with § 18 (3) FeV or Section 6 of the Examination Guidelines at any time.

[97] For licence classes C, C1, D, D1 and T, the report templates must additionally permit the recording of performance relating to “Vehicle function checks”, and for classes D and D1 furthermore the assessment of “Manual skills” (in accordance with Annex 7 to the Examination Guidelines). For licence classes BE, C1E, DE, D1E, CE and T, the report must provide a section “Coupling and uncoupling of the vehicle” (in accordance with Annexes 8 and 9 to the Examination Guidelines). It also seems expedient to automatically hide sections of the test report which are not relevant for a particular class of test (e.g. “Vehicle function checks”, “Manual skills” and “Coupling and uncoupling of the vehicle” in the case of a test for licence class B).

[98] These sections need not necessarily be arranged as separate input forms; they could also be combined in an ergonomically and functionally appropriate manner.

re (4): The fourth section of the electronic test report presents a matrix structure[99] for documentation of the behaviour displayed by the candidate when performing driving tasks in real traffic. The methodical advantage of a matrix structure can be seen in the fact that the use of rows, columns and cells enables the examiner to address the three dimensions of the required test behaviour (driving tasks, observation categories, assessment criteria) in their content-referenced relationships to each other (see above). To determine a level of driving competence, the examiner judges the driving tasks performed the candidate (represented by the rows) in the context of the observation categories (represented by the columns) and on the basis of the specified assessment criteria; his traditionally event-oriented assessments are documented in the corresponding cells of the matrix, albeit with a greater degree of differentiation than in the current test report. Normally, such event-oriented assessments should be recorded immediately after observation of the event concerned, in order to avoid memory effects. Each assessment is made according to the four-level scale described in Chapter 3 ("Above-average performance", "Normal performance", "Simple errors", "Serious errors"), wherein the behaviour of the test candidate is assessed with regard to its compliance with traffic rules and its appropriateness in the given situation. The currently applicable legislation stipulates that a "serious error" leads to immediate termination of the driving test; this termination would then also be noted in the test report (see above).

The following action sequence is implemented to document an assessment: The examiner first selects the row of the electronic test report which corresponds to the driving task (or subtask) performed by the candidate, and subsequently assigns a situation subclass to this task (e.g. driving task "Changing lanes" and situation subclass "High-density traffic"). As the third step, the examiner then searches the selected row for the column whose observation category or element of competence is most appropriate to the event to be assessed (e.g. "The candidate displays above-average performance in respect of traffic observation, by recognising the unexpected and/or hazardous driving manoeuvres of other road users."). Instances of "normal performance" are not documented explicitly; the recommendation in this case – based on the considerations discussed in Chapter 3.5 with regard to the prerequisites for proper evaluation, test location analyses and the proposed testing of demand-referenced performance assessments – is merely to mark the fact that the driving task has been performed. Driving tasks which are performed several times would also be documented with multiple marks or – in case of conspicuous behaviour – corresponding assessments.

re (5): The fifth section of the test report refers to the "technical completion of the (test) drive", and must provide standard documentation fields relating to Annex 7 to the Driving Licence Regulations and Annex 10 to the Examination Guidelines. This section could also be used to document a range of driving speeds and the types of road used during the test drive.

re (6): Following completion of the test drive the examiner should use the sixth section of the test report (bottom row and right-hand column of the fifth section, see above) to record a summarising, overall assessment of driving competence in the contexts of the eight driving tasks and five observation categories, based on the four-level scale presented in Chapter 3 (possible ratings: "Very good", "Good", "Sufficient", "Inadequate"). The event-oriented assessments documented in the cells of the matrix here serve as orientation for the competence-oriented overall assessments. Driving tasks which could not be examined due to the absence of corresponding traffic situations or infrastructure at the test location are marked accordingly with "Not applicable"; the five elements of competence described by way of the observation categories, on the other hand, can always be assessed, assuming that the test has not been terminated prematurely.

re (7): In the seventh section, the examiner can then document his final test decision. To aid this decision, a clearly structured matrix is filled with data generated automatically

[99] As already indicated, the matrix constitutes the structural, content-related and methodical "skeleton" of the electronic test report, and is to be implemented in an appropriately ergonomic manner by way of different input fields and forms. Overall, the corresponding software solution should reflect all inputs in real time, so that the most important information can be called up quickly and at any time, also in (abridged) matrix form.

from the event- and competence-oriented assessments entered in sections two to six. The examiner is thus able to reflect the overall course of the test and the entirety of the test performance displayed by the candidate. In addition, automatic plausibility checks should assist the examiner in the decision-making process.

4.4 Demands placed on the documentation instrument for creation of an electronic test report

To achieve the described documentation and data processing functions, it is necessary to make available a highly interactive technical documentation instrument or computer system comprising a visible (graphical) user interface (GUI), a not directly visible software or programming solution, and the corresponding hardware components. As is the case for any interactive technical product, this instrument must also follow the general criteria for "good design". The goal of good product design is to ensure maximum usability of the human-machine interface. Usability can be defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (DIN, 1998, p. 5). The effectiveness here determines "the accuracy and completeness with which users achieve specified goals" (ibid., p. 8), while efficiency denotes "the resources expended in relation to the accuracy and completeness with which users achieve goals" (ibid., p. 8). Effectiveness is thus a prerequisite for efficiency. Satisfaction, finally, refers to the "freedom from discomfort, and positive attitudes towards the use of the product" (ibid., p. 8). The priority issue for good product design, therefore, is to match the technical solution (hardware and software) to the intended user and the purpose of its use.

Designers of technical products see themselves confronted with a sheer endless diversity of recommendations – sometimes concurring, sometimes meaningful complements, but also occasionally contradictory (GRONER, RAESS & SURY, 2008). To identify the fundamental requirements to be met by the documentation instrument for an optimised practical driving test, a thorough analysis was started to examine recognised models of good interactive system design, enabling design criteria embedded in the models to be retrieved and condensed into five central demand categories (functionality, ergonomics, look, context of use, individual use). This analysis was based, among others, on the following approaches to interactive system design from both academic research and practical applications:

- The FURPS model (Functionality; Usability; Reliability; Performance; Supportability) elaborated by GRADY (1992), which is used by Hewlett-Packard and IBM (EELES, 2005),
- The five quality dimensions of usability (learnability; efficiency of use; memorability; errors; satisfaction) according to NIELSEN und LORANGER, 2006,
- The eight "golden rules" of interface design (consistency; universal usability; informative feedback; closed dialogues; prevent errors; reversal of actions; internal locus of control; reduce short-term memory load) according to SHNEIDERMAN, PLAISAN, COHEN and JACOBS (2009),
- The 17 demands for interface design according to STAPELKAMP (2007),
- The comparison of different software quality models according to HYATT and ROSENBERG (1996),
- The "seven principles of dialogue design" in accordance with DIN EN ISO 14915 (DIN, 2003),
- The quality criteria for software products (functionality, reliability, usability, efficiency, maintainablity, portability) in accordance with ISO/IEC 25000 (ISO/IEC, 2005),
- The foremost international standard describing guidelines for the ergonomics of human-system interaction, DIN EN ISO 9241 (DIN, 1998, 2008, 2011a, 2011b), with several parts addressing the principles which apply to dialogue design for interactive systems – e.g. in part 110: Suitability for the task, self-descriptiveness, controllability, conformity with user expectations, suitability for individualisation, error tolerance and suitability for learning (DIN, 2008; SCHNEIDER, 2008).

Criteria derived from the analysed models are assigned to each of the five extracted demand categories (functionality, ergonomics, look, context of use, individual use). The demand categories and their most important criteria are illustrated in Figure 12 below. This is followed by a brief general description of the categories and criteria, together with corresponding proposals for a documentation instrument with which to implement the proposed electronic test report.

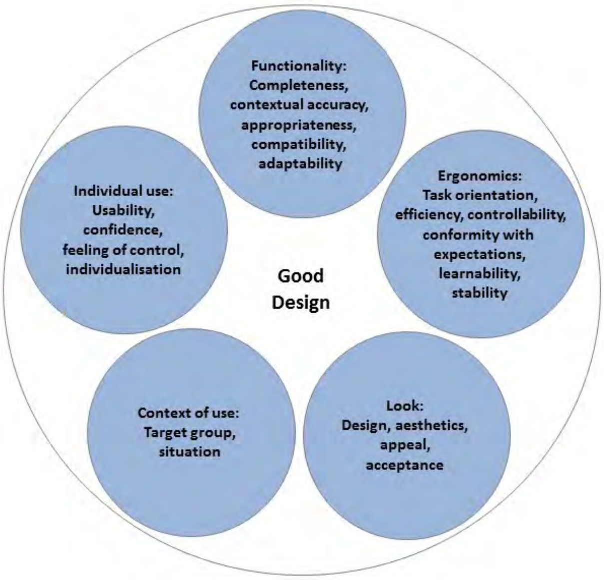

Fig. 12: Demand categories relating to the design of a documentation instrument

(1) The demand category "Functionality" is satisfied to the fullest extent if the instrument incorporates all the functions which are necessary to perform certain specified tasks. The functions and contents must furthermore be implemented in a manner which is not only professionally and methodically correct, but also appropriate for the given process of driving licence testing. Accordingly, the instrument must permit complete, contextually accurate and effective presentation and processing of all documentation requirements arising within the framework of the practical driving test (e.g. administrative data, performance details relating to the driving tasks, basic driving manoeuvres and the technical preparation and completion of the test drive, competence assessments, test decision). It is at the same time important to maintain compatibility with other systems used by the Technical Examination Centres and licensing authorities, to ensure compliance with all relevant legal provisions, including security standards relating to data privacy and data communication, and to allow for future modification of the software, so as to be able to implement improvements and amendments prompted by changes to the (legal) framework conditions as quickly and simply as possible.

(2) The demand category "Ergonomics" reflects fundamental (law-governed) demands placed on work methods, tools and equipment, and refers in general to the availability of practicable and convenient-to-use products which ensure the correct and efficient completion of pending (standard) tasks. Applied to the ergonomic design of a computer system, this concerns both the framework of possibilities dictated by the hardware and the specific design features of the software (HERCZEG, 2005). While the term "hardware ergonomics" can be used to describe the adaptation of electronic products to the systems of human locomotion and perception and to situational requirements (including, for example, battery lifetime and the weight of the mobile devices on which the software for an electronic test report is installed), "software ergonomics" (also known as "usability engineering") covers adjustment to a user's cognitive and physical competences or qualities (e.g. the ability to process complex information) by way of particular presentation forms for information and interactive elements (e.g. colours, contrast, font sizes, layouts, user inputs).

Where a computer system is intended to serve a certain work process, it must provide facilities for corresponding inputs (here, for example, the selection of a driving task) and outputs (here, for example, the marking of a driving task as "not performed correctly"). Interaction with the system is realised by way of a so-called "user interface", which is defined in the standard DIN EN ISO 9241-110 as "all components of an interactive system (software or hardware) that provide information and controls for the user to accomplish specific tasks with the interactive system" (DIN, 2008, p. 1). With regard to ergonomic design of the overall user interface, it thus appears meaningful to distinguish between screen and interaction design (STAPELKAMP, 2007): While "screen design" refers to the structuring and layout of the work area (graphical interface), "interaction design" describes the interaction and navigation features within the software (user dialogues).

The interface design for the documentation instrument must be specifically suitable to the set documentation tasks. This is the case if the instrument supports the examiner in his effective and efficient processing of the assigned responsibilities (documentation of the whole course of the test, planning of the test route by way of an adaptive test strategy, feedback to the candidate). Efficient dialogue design can be assumed if these work tasks can be processed quickly, intuitively and with the minimum possible input of cognitive resources. It is thus essential that the documentation instrument be controlled in a simple and understandable manner. The demand for simplicity requires the number of user actions necessary to be able to input data to be reduced to a minimum. This means, for example, that the individual documentation levels of the electronic test report (driving task, situation subclass, observation category, assessment) should be accessible with as few intermediate steps as possible, and standard (default) settings should be defined for the action sequences which occur most frequently in practice. The criterion of understandability is met if the individual control steps follow a logical sequence, and if all relevant texts and control buttons are immediately locatable and self-explanatory. In the case of an electronic test report, this demand could be realised by providing interactive info texts, for example. Conformity with user expectations is another way to promote simplicity and understandability: In other words, menu items, buttons, selection lists, dialogues and the layout of information must satisfy certain requirements in terms of consistency, conventionality and adaptation to the experience, expectations and competences of the driving test examiner. Simplicity, understandability and conformity with user expectations would contribute to faster familiarisation and intuitive handling of the documentation instrument or electronic test report.

Ergonomic controllability of the documentation instrument is furthermore dependent on the examiner himself being able to determine and predictably control the opening, speed, scope and reproducibility of dialogue sequences. It should not be complicated to rectify or undo operating or input errors. The instrument should instead help to avoid documentation errors, or at least enable the examiner to recognise errors (preferably by way of corresponding info texts) and, where appropriate, to make the necessary corrections as simply as possible. It also seems expedient for the examiner to be shown which dialogue is selected at any time, which step of the selected dialogue is currently active, and which further actions are available at this point. In addition, the documentation instrument should offer a general overview of the course of a driving test so far, as orientation for subsequent test planning by the examiner in the sense of an adaptive test strategy. Through such a permanently available overview of the candidate's performance to date, together with information on the driving tasks which are still outstanding and specification of the applicable observation categories and assessment criteria, the new electronic documentation instrument is intended to ease the cognitive limitations which bear on the driving test examiner when performing his complex observation task (cf. cognitive load theory according to SWELLER, 2006); these limitations result from the multiple burdens placed on the examiner by the necessity for parallel observation of the traffic situation and the candidate's driving behaviour, in combination with the requirements of immediate performance assessment and documentation. Each action performed with the instrument should be acknowledged by adequate feedback from the system. Acoustic feedback, however, should be avoided, as this could distract the candidate's attention away from the current traffic situation. It is moreover imperative that the instrument functions stably and reliably, without immediate fears of system crashes.

One of the most important ergonomic demands concerns the consistency of presentation and interaction. With regard to screen design, it is essential to ensure consistency in the positioning and grouping of content, in icon design, and in the use of colours and designations (uniform terminology). Consistent interaction design similarly requires buttons, dialogues and the achieved results to follow common principles. This also takes into account the fact that human working memory is only capable of storing between five and nine items simultaneously (MILLER, 1956), although previously familiar and consistent elements may be combined into larger units (“chunking”).

To facilitate operation by way of a touch-sensitive surface under the special situational conditions of a test drive, a minimum diagonal size of 1 cm (0.4 inch) should be observed for all action icons. For the presentation of overviews and text elements, the individual display characters should preferably not be smaller (diagonally) than 0.7 cm (0.28 inch), so as not to impair legibility, even if this at first appears difficult to realise given the complexity of the planned documentation options.

Scientifically founded knowledge also exists with regard to the ergonomically desirable design of screen buttons. Buttons with a three-dimensional appearance, for example, are immediately recognised as means to access available actions rather than mere presentations of information (HOLL, 2007). The choice and intensity of colours is equally important: Pale colours or grey backgrounds signal inactivity. The standard ISO/DIN/IEC EN 60204-1 (DIN, 2011a), for example, contains the following stipulations relating to the use of colour-coding on machines: Red is used to designate a hazardous or critical state, and thus warns of possible dangers or circumstances which call for immediate intervention, yellow points to an abnormality and the pending likelihood of a critical state, blue marks an action recommendation to be heeded by the user, and green indicates a safe (normal) state. This established use of colours should also be taken into account in the elaboration of the documentation instrument, and there reflected in the chosen design of the buttons and matrix; after all, this would effectively implement the classic and generally internalised "traffic light" colours, the meanings of which will be intuitive to the examiner. Grey would seem appropriate as the colour to represent irrelevant aspects in a given context, while blue should be used to highlight requests for input and other important features. It is here nevertheless necessary to observe a number of basic rules resulting from the physiology of the human eye. The colour blue, for example, although in itself a colour yielding high contrast, should only be chosen for texts of sufficient size, since the area of the retina which permits the sharpest vision is at the same time relatively insensitive to blue (HOLL, 2007). Unfavourable colour contrasts, such as blue on red, must be avoided, firstly because they are often perceived as disturbing and unpleasant, and secondly due to the so-called "chromostereopsis effect" (overstressing of the eye caused by combinations of colours with high levels of saturation and very different spectral wavelengths), which can easily result in headaches for the examiner (MÜLLER, 2003). Light-grey or another pastel tone would be fitting as the background colour for the user interface (HOLL, 2007). It is important to avoid unfavourable combinations (e.g. dark-coloured areas with light-coloured text), however, as light from the surroundings could here result in disturbing reflections on some screens (ibid.). As a means to enhance overall usability, it could also be beneficial to offer different combinations of appropriately contrasting colours, from which the examiner could then select the colour scheme which best matches his personal preferences. The signal colours for serious and simple errors (red and yellow) and for above-average performance (green), however, should be standard as far as possible for the aforementioned reasons.

When the examiner selects a particular control button, this selection should also be indicated graphically (e.g. by a change in the background colour or highlighting with a coloured border). For space reasons, and to aid instant recognition, control buttons with graphic symbols (icons) have been proved most effective (TIDWELL, 2009); they are thus to be given preference over text-based buttons wherever possible (SHNEIDERMAN, PLAISAN, COHEN & JACOBS, 2009). It is true that the hurdle for the use of graphics-based interaction elements is slightly higher at the beginning, but empirical studies have confirmed that the usability benefits compared to text-based elements are considerable already after a brief period of familiarisation (TIDWELL, 2009). Interaction with the documentation instrument should thus be based primarily on easily identifiable and aesthetically appealing icons.

One especially significant aspect for the development of the documentation instrument is the manner in which inputs are to be possible. For navigation design, for example, it makes a big difference whether inputs are to be made via a keyboard or by using fingers and/or a special pen in combination with either a touch-sensitive surface without display ("touchpad") or a touch-sensitive surface with display ("touchscreen"), because the form of input determines specific demands relating to minimum button size and gesture control. Touchpads and touchscreens which permit finger-based control by way of simple or multi-touch gestures offer access to a diversity of actions (HEYDEKORN, FRISCH & DACHSELT, 2010) without being tied to specific menus and control elements (BOLLHOEFER, MEYER & WITZSCHE, 2010), but are at the same time relatively expensive in terms of familiarisation and require the gestures to be memorised at the relevant points of the user interface. Irrespective of the chosen form of input, one common demand is that the required purpose should always be served by consistent actions in similar situations. The most practicable form of input, however, will only be revealed by way of the feasibility study envisaged to test the documentation instrument design (see the following chapter).

(3) The third demand category "Look" refers to the demand that the hardware and software should together possess an aesthetically ap-

pealing design. An electronic documentation instrument and test report should encourage active use by way of a tidy layout and clear structures, and should contribute to a pleasing work experience for the examiner. It also seems likely that an ergonomically and aesthetically well-designed instrument could promote much wider acceptance of the work associated with introduction of a new documentation method. In the case of a graphically less attractive screen and interaction design, on the other hand, the planned instrument could be rejected by its future users, even if it satisfies all functional requirements.

(4) The fourth demand category "Context of use" describes the environment in which the documentation instrument is to be used. The instrument must be tailored to both the intended target group, i.e. to the driving test examiners (and to their competence in working with touchpads, for example), and the observation situation, in other words to use in a moving motor vehicle during a driving test. This places particular demands on the screen design (e.g. good legibility also under constantly varying lighting conditions), on the interaction design (e.g. use of the largest feasible buttons, so as to minimise the probability of unintended inputs), and not least on the hardware. It must also be ensured that the documentation instrument and electronic test report can be adapted quickly and simply to any new circumstances which arise due to personal, situational or legal changes.

Which demands are to be met by the hardware of the documentation instrument for an optimised practical driving test? Both for economic reasons (e.g. potential cost savings through competition among potential suppliers) and in view of the typical pace of technical developments, it seems imprudent to limit developments to one specific hardware solution. On the other hand, minimum hardware requirements (possibly in conjunction with expansion options) are essential to ensure fulfilment of the described functional and ergonomic demands, and to serve as orientation for the elaboration and evaluation of a practicable and effective electronic test report. Supplementing the general demands to be met by the (ergonomic) design of the report software, as elaborated in the above, attention must be paid not least to the special place at which it is to be used (rear seat of the test vehicle) and the particular purpose it serves (support for the examiner's detailed observation, assessment and documentation of the driving competence demonstrated by the test candidate, and assistance in the sense of test planning and control on the basis of diverse function combinations). To be able to determine the precise hardware requirements, therefore, it is first necessary to consider the following crucial questions: "Which input and display devices best support the realisation, control and documentation of an optimised practical driving test?" and "How can the documentation instrument be installed in the test vehicle such that all potential endangering of the vehicle occupants (e.g. in case of heavy braking or an accident) is excluded?" As answers to these questions, essential requirements relating to the hardware of the documentation instrument are outlined below; realisation of these requirements would contribute to maximisation of the usability of the instrument:

- Display screen: To satisfy the aforementioned content-related and structural demands, the screen size should be between 17.8 cm (7 inch) and 25.6 cm (10.1 inch); the display contrast and brightness should preferably be adapted automatically to the prevailing brightness conditions. Usability must not be impaired even in case of direct sunlight from a side window. If data input is to be integrated with presentation of the course of the test to date, entries can be made directly via a touch-sensitive display (touchscreen) using fingers or a special pen. The weight of the device should not exceed 0.7 kg.
- Optional separate input device: A separate input device offers an alternative to solutions in which handling of the documentation instrument and presentation of the electronic test report are combined on common hardware. If a separate input device is used, and thus the two functions are realised independently, the hardware for presentation of the electronic test report ("display screen") can be attached directly to the head-rest of the front passenger seat. The interaction with this display device would then be implemented via a separate input device (remote control). This input device could be a smartphone, a PDA ("Personal Digital Assistant"), a specially developed remote control device without its own display, or a special paper form in combination with a special pen

("digital pen"). Such a special pen would transmit notes made by the examiner for immediate visualisation on the display screen of the electronic test report. At the same time, it would be possible to produce a paper-based report. On this point, inspiration may be drawn from the experience gained with the use of a special pen to record test data in the Netherlands and Sweden (see Chapter 4.2.2).

- Transport safety: To guarantee transport safety, an optional mounting facility could be provided for fast, variable-angle attachment to a vehicle headrest. The attachment must at the same time ensure firm, non-slip handling. This would furthermore avoid endangering the vehicle occupants in case of heavy braking or an accident.
- Optional module for mobile data links: Generally speaking, it would be desirable to enable online interfaces to the production applications installed in the Technical Examination Centres. A simple call in advance of a test would then suffice to auto-complete the required administrative data (e.g. name of the candidate, place, date) in the corresponding fields of the electronic test report. Upon completion of the test, the test result could also be communicated directly to the Technical Examination Centre. In any case, however, it is imperative that the data transmission must comply with the highest security standards.
- Optional GPS module: GPS ("Global Positioning System") could be used to save the realised test routes for evaluation purposes. Additional information could be derived from "waypoints" set by the examiner to confirm the performance of particular driving tasks. If such data are aggregated over a larger number of tests, statistics can be produced on the characteristics and the frequency of use of individual test route sections. Last but not least, new possibilities would arise for detailed performance feedback to the test candidate.
- Identity document scanner module: An interface to an identity document scanner could enable fast, legally certain identification of the test candidate, as well as automatic completion of the candidate's personal data, insofar as this has not already been realised via a data link to the Technical Examination Centre.
- Battery life: Given the typical duration of a working day for driving test examiners, at least ten hours of uninterrupted battery-powered operation should be guaranteed. Furthermore, a simple facility should be provided for recharging of the documentation instrument battery via a mains socket or in-vehicle charging cable. Alternatively, provisions could be made for fast and uncomplicated replacement with a spare battery.
- Total weight: The total weight of the documentation instrument – or the "examiner case" in which all the required components are transported – should not exceed 10 kg. This figure is based on health and safety recommendations issued for the trade and distribution sector (BGHW, 2010).[100]

(5) The fifth demand category "Individual use" covers personally specific aspects of the use of a product by a particular user. The user experience is reflected in perceptions, sentiments, impressions and reactions which are triggered by a product before use (anticipated benefits), during use and after use (identification with or distancing from the product). The documentation instrument and electronic test report must thus satisfy the (by all means heterogeneous) expectations of the driving test examiners. To this end, the benefits must be experienced as effective, efficient, transparent and reliable. This can only be achieved successfully, however, if the system is free of all content-related or technical deficiencies, and if it is simple to use. The driving test examiner must always assume the determining role in dialogue with the instrument. In other words, the dialogue design must, as far as possible,

[100] It must here be emphasised once more that the hardware design recommendations listed under this point were derived from research which referred in many cases to technical solutions from other fields of use. It is thus necessary to conduct a feasibility study to determine whether or not the individual hardware demands are transferable to the case of an electronic test report, and whether they can actually be realised. The practical feasibility is in turn dependent on further framework conditions of the practical driving test, including not least considerations relating to the safety of the vehicle occupants and the costs of realisation (the recommended identity document scanner, GPS module and online data link, for example, are significant factors which could lead to a sharp increase in the costs of testing and thus also higher test fees). Important pointers to acceptable solutions can be expected from the proposed feasibility study and revision project (see Chapter 4.4 and Chapter 5).

provide for the examiner to act rather than react, i.e. to retain (subjective and objective) control over the instrument. In cases of uncertainty, finally, the instrument should offer readily understandable guidance, and the settings and displays must be easily adaptable to the different needs of the individual examiner.

As prerequisites for the introduction of a properly functional documentation instrument and a practicable and informative electronic test report for an optimised practical driving test, it is first necessary to develop a prototype device and to subject this prototype to carefully planned trials. The purpose of such initial trials is not solely to determine whether the hardware and software of the instrument prototype satisfy the functional and technical demands described in the above chapters; it is rather that the trials must also clarify whether and, if so, to what extent the thoughts on desirable properties and support functions of an innovative documentation instrument and electronic test report are actually realistic or feasible in the specific circumstances of the practical driving test (e.g. work processes of the examiner, working conditions in the test vehicle). The necessary methodical framework for corresponding feasibility studies is to be described in the following chapter.

4.5 Demands placed on a feasibility study for initial testing of the documentation instrument

Before realisation of an innovative concept – here, for example, optimisation of the test documentation – or introduction of a new product (e.g. the documentation instrument), the organisational, technical, economic and political practicability should be investigated by way of a feasibility study (BEA, SCHEURER & HESSELMANN, 2008). Generally speaking, a feasibility study serves to judge the strengths and weaknesses of the concept or product, to test the chances for success of the current (design) ideas, and to assess whether – and where appropriate by which means and within which timeframe – these ideas can be realised. In the case of an interactive computer system, such as the planned documentation instrument, consideration must be given above all to the practical suitability (handling and acceptance) of the hardware and software. If the feasibility study yields an overall positive result, this can be taken as "proof of concept" (AMOR, 2003); subsequently, those ideas which have proven meaningful and practicable can be implemented to the extent possible. If the result of the feasibility study is negative, the (design) ideas must be modified, or perhaps even discarded altogether. SWOT analysis (Strengths, Weaknesses, Opportunities and Threats) is a particularly suitable analysis and documentation method for feasibility studies (DALCHER & BRODIE, 2007): The essence of this method is a comparative evaluation of the strengths and weaknesses, and similarly of the potential and risks, of the concept or product under development (see Fig. 13).

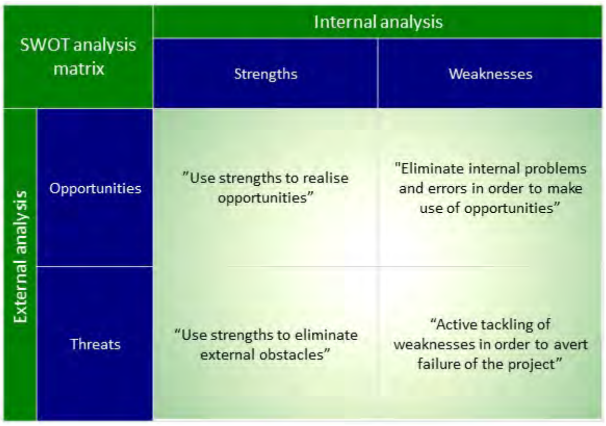

Fig. 13: SWOT analysis matrix

Whether or not a concept (in the present case: the optimised electronic test documentation) or product (here: the documentation instrument) meets the practical expectations, can be assessed in a two-stage process comprising (1) prototype tests and (2) pilot tests (LAUESEN, 2002):

(1) In prototype tests, a simplified implementation of the concept or product is put to use in an environment similar or comparable to that of the real application. At this stage, it is above all the formative aspects of evaluation which play the major role. Prototype tests can be further subdivided into "alpha tests" and "beta tests". The term "alpha testing" refers to the first test phase within the concept or product development process; it is usually performed by external experts. "Beta testing", on the other hand, is the phase at which a sample of the target group concerns itself intensively with the latest embodiment of the concept or product. Prototype tests with the documentation instrument and electronic test report represent the core of the proposed feasibility study relating to optimisation of the test documentation for the practical driving test.

(2) In pilot tests, the quality of a finished concept or product is verified in field trials (i.e. in the real application environment) by a representative circle of users chosen from the overall target group. The focus is now placed on conclusive, summative evaluation. Following the

successful completion of pilot testing, the finished concept or product is implemented for the whole target group. The pilot tests for the instrument and electronic test report for documentation of an optimised practical driving test should be incorporated into a revision project subsequent to the aforementioned prototype tests and feasibility study. In this context, the optimised contents and realisation processes for the practical driving test must be subjected to critical appraisal over the course of broad-based practical trials, as preparation for universal implementation of the documentation instrument and electronic test report.

In the following, content-related and methodical recommendations are to be presented as proposals for structuring of the pending prototype tests (referred to hereafter, in their entirety, as the "feasibility study"), and thus initial testing of the documentation instrument and electronic test report for an optimised practical driving test. These recommendations refer to questions which are still to be answered in respect of content and to the product features which are to be considered, as well as to the methodical design (target groups, survey methods, sampling) and organisational planning of the feasibility study.

With regard to the fundamental principle, the feasibility study should first of all be divided into two sections or stages: As the first stage, it is necessary to clarify or else sharpen present definitions of the scientific demands to be met by the documentation instrument and electronic test report from the perspective of the expert public (e.g. driving test examiners, driving instructors, legislator) and other involved parties (e.g. driver licensing authorities, driving licence applicants). This would serve as a basis for subsequent elaboration of a corresponding prototype. The second stage can be devoted to the actual prototype tests, enabling empirically guided further development of the instrument and electronic test report over a series of alpha tests, beta tests and several development cycles (see below). The feasibility study should be performed by the working group TÜV ÞEKRA arge tp 21, as an entity of the Technical Examination Centres mandated to conduct driving tests, and accompanied by a working group comprising representatives of the federal and state ministries responsible for traffic, the Federal Highway Research Institute (BASt), the Technical Examination Centres, the Bundeswehr, the Association of Technical Inspection Agencies (VdTÜV), the driving instructors and other scientists working in the field. This recommendation is response to the initially presented standpoint that a feasibility study should investigate various aspects – organisational, technical, economic and political – of the practicability of innovations (BEA, SCHEURER & HESSELMANN, 2008): This can best be achieved by involving all those parties who are responsible for or otherwise contribute to the implementation of the individual aspects.

Why is it deemed important for the demands placed on electronic test documentation and the associated documentation instrument, as described in the present report, to be validated from the perspective of professional experts already at the beginning of the feasibility study? The transition to an electronic test report is a caesura which holds not only unique potential, but also certain risks with regard to the administration, realisation and evaluation of the practical driving test, and it is thus imperative that it be mastered successfully: A driving licence test represents state approval for the granting of mobility entitlements to an individual citizen and in this way seeks to guarantee road traffic safety in the public interest; it is this singular context which founds the extraordinary individual and political significance of the driving test and its professional realisation. This significance must be reflected by involving as many of the participants in driver licensing as possible in the process to identify and define suitable and justified demands to be met by the test concept in general and test documentation in particular. Accordingly, it is appropriate to continue, supplement and thereby validate the present scientific demand analysis by hearing and systematically questioning representatives of the different involved parties. The content-related, methodical and technical demands and expectations are to be provided with an empirical foundation, assessed in respect of their individual importance, and discussed with regard to their technical practicability. In conclusion, criteria are to be elaborated to assess the fulfilment of those demands which are deemed meaningful and practicable, on the basis of which the practical usability of the documentation instrument and electronic test report can be judged at the end of the feasibility study.

To identify the demands and expectations of the driving test examiners, test candidates and driving instructors, one convenient starting point would be to establish a qualitative sample of the target groups in accordance with the principles of deductive sampling (MERKENS, 1997). The objective here, in the sense of variance maximisation (PATTON, 2002), is to recruit a preferably heterogeneous and – with regard to the relevant attributes – maximally contrasted and thus informative group of respondents for an analysis (in the present case an expert survey on the demands to be met by electronic test documentation). For an evaluation of the demands applicable to electronic test documentation, it seems expedient, for example, to ensure heterogeneity in the sample in respect of the categories “Computer affinity” and “Satisfaction with the presently used test report”. It is furthermore assumed that, for the two groups “Driving test examiners” and “Driving instructors”, corresponding differentiation in the category “Test experience” would be of particular value for the results. In accordance with the concept of theoretical saturation (GLASER & STRAUSS, 1967), the size of the separate samples can only be determined as the outcome of a circular process of data acquisition and immediate evaluation of that data: The required sample size in an individual case is reached as soon as no further knowledge is to be derived from the evaluation of further survey responses.

The assessments of the driving test examiners, test candidates and driving instructors can be acquired by way of semi-structured individual and group interviews built around a general framework of topics for discussion. The framework here serves primarily as guidance for the interviewer and is intended to ensure that all essential aspects are covered during the interview. From the analysis results obtained to date, and from individual and group discussions conducted with experts and others involved in the driver licensing process during the course of the current project, it has been possible to derive a number of demand criteria which will presumably be considered relevant for an electronic test report by driving test examiners, test candidates and driving instructors, respectively; these criteria could be taken as the starting point for the planning of semi-structured individual and group interviews in the context of the feasibility study:

1. As far as the driving test examiner is concerned, it can be assumed that the following five demand criteria will be deemed applicable. An electronic test report should …
   - … help to structure and control the observation situation.
   - … support the examiner in respect of uniform, efficient documentation of the candidate's test performance.
   - … facilitate professionally sound and reflected assessment of the driving competence of a test candidate, a prognosis relating to his future driving behaviour and a final test decision by presenting a clearly arranged and preprocessed overview of individual observations and assessments.
   - … provide a basis for meaningful and detailed feedback on test performance to the candidate and the driving instructor.
   - … simplify the communication of test results to the responsible offices (e.g. Technical Examination Centres, licensing authorities, KBA).

   In addition to validation of these assumed demand criteria, the interviews with driving test examiners must clarify their expectations with regard to acceptance and the benefits of an electronic test report; furthermore, it must be determined whether demands exist with regard to the precise timing of actions to record test performance and to the possibilities for correction in case of input errors.
2. It is assumed that the test candidate attaches particular importance above all to the realisation of test transparency and equality by way of unambiguous demand and assessment standards, as well as explicable documentation of the assessment: The test decision confers or withholds mobility entitlements, and the test candidate will expect to be treated fairly in this situation. Professional (verbal and written) event- and competence-oriented performance feedback is also presumed to carry particular weight, as this promotes realistic self-assessment on the part of the test candidate and targeted further development of his driving competence, as is generally also a wish of the candidate himself and furthermore in the interest of road safety.
3. For the driving instructors, it is assumed that they will see potential benefits of an electronic test report – alongside the promise of test transparency and equality – in the general contribution to improvement of their driving instruction, in the sense of quality assurance, and particularly in the basis provided for the planning of special, individualised further training offers for unsuccessful candidates.

Besides representatives of the aforementioned involved parties, the demand validation process should also address selected decision-makers and further driver licensing experts from the federal and state ministries responsible for traffic and from the Federal Highway Research Institute (BASt). Corresponding exploratory discussions were already held with high-ranking business and quality management executives from the Technical Examination Centres and the Bundeswehr within the framework of the present project; the results of these discussions have been incorporated into the following summary of demand criteria to be met by a future electronic test report from the perspective of the traffic authorities and test organisations. The conclusions drawn below are thus derived from the standpoints of the offices responsible for test administration, but are taken to apply equally for the traffic authorities; whether and, if so, to which extent this assumption is correct, should be clarified further in the course of the feasibility study, for example by way of interviews or written surveys. The expectations include:

- Simplification of test administration processes (including the preparation of tests, as well as the processing and archiving of test data) and the prescribed reporting to the licensing authorities and the Federal Motor Transport Authority
- Increased efficiency in monitoring of the observance of prescribed legal framework conditions for the practical driving test
- Enhanced professionalisation on the part of the driving test examiners within the framework of qualification training, further training measures and personnel management (e.g. by improving test transparency and revealing particularities in the assessments of individual examiners)
- Simpler updating of the demand and assessment standards contained in the test documentation thanks to more efficient coordination processes between the legislator and the Technical Examination Centres
- Optimisation of the scientific evaluation of the practical driving test and quality management (e.g. product audits and complaints management systems)
- More effective future further development of the practical driving test and the overall system of novice driver preparation on the basis of evaluation results (“output control”).

Following validation of the demand criteria, the next step for the feasibility study is to perform a SWOT analysis of the validated criteria and to elaborate a hierarchical demand catalogue; the hierarchical structure of the catalogue should reflect the relative importance of the individual demands. On this basis, it will then be possible to proceed to further development of the draft documentation instrument and electronic test report into alpha and later beta versions, to realise the corresponding alpha and beta tests, including appropriate additional development cycles, and above all to assess the results obtained from testing during and upon conclusion of the feasibility study.

As soon as a prototype or “alpha version” of the documentation instrument and electronic test report become available (together with appropriate explanatory notes on handling), a first alpha test is to be performed, followed immediately by a first development cycle to optimise the content-related documentation possibilities, software design, hardware implementation, and the test and documentation algorithms. This should then be taken as the starting point for further test series and development cycles, as the resultant “feedback loops” achieve the central objective of usability engineering – namely optimisation of the practical usability of a product – more effectively than is the case with one-off testing: Repeated testing over several development cycles serves not only to identify weaknesses, but also to verify the success of a subsequent redesign and to derive pointers for the further course of development. The reasons for this iterative approach, which is also intended to deliver the described benefits during the beta testing phase of the feasibility study, can be found in the standard DIN EN ISO 9241-210 (DIN, 2011b, p. 6). It can be described as “prototype-based development of interactive user interfaces” (GRONER, RAESS & SURY, 2008, p. 428) and is illustrated as a circular process in Figure 14 below.

Within the framework of alpha testing, a series of expert reviews (HEUER, 2003) is to be performed: Three to six driver licensing and usability experts who are not directly involved in the development work are to give heuristic appraisals of the latest status of an alpha prototype. According to NIELSEN (1994), three users of a software system are sufficient to identify approx. 65 per cent of usability problems; five users already raise this figure to more than 80 per cent, and six users even manage close to 90 per cent. From the evaluation of a diversity of projects, NIELSEN (ibid.) expresses the assumption that the proportion of usability problems discovered by a single user in one-off testing will be approx. 30 per cent. Particularly among experts, the validity of an average 30 per cent probability of identification can be assumed with a high degree of certainty (TURNER, LEWIS & NIELSEN, 2006); if the tests are conducted with lay persons, on the other hand, a lesser probability of identification should be presumed (NIELSEN, 1992).

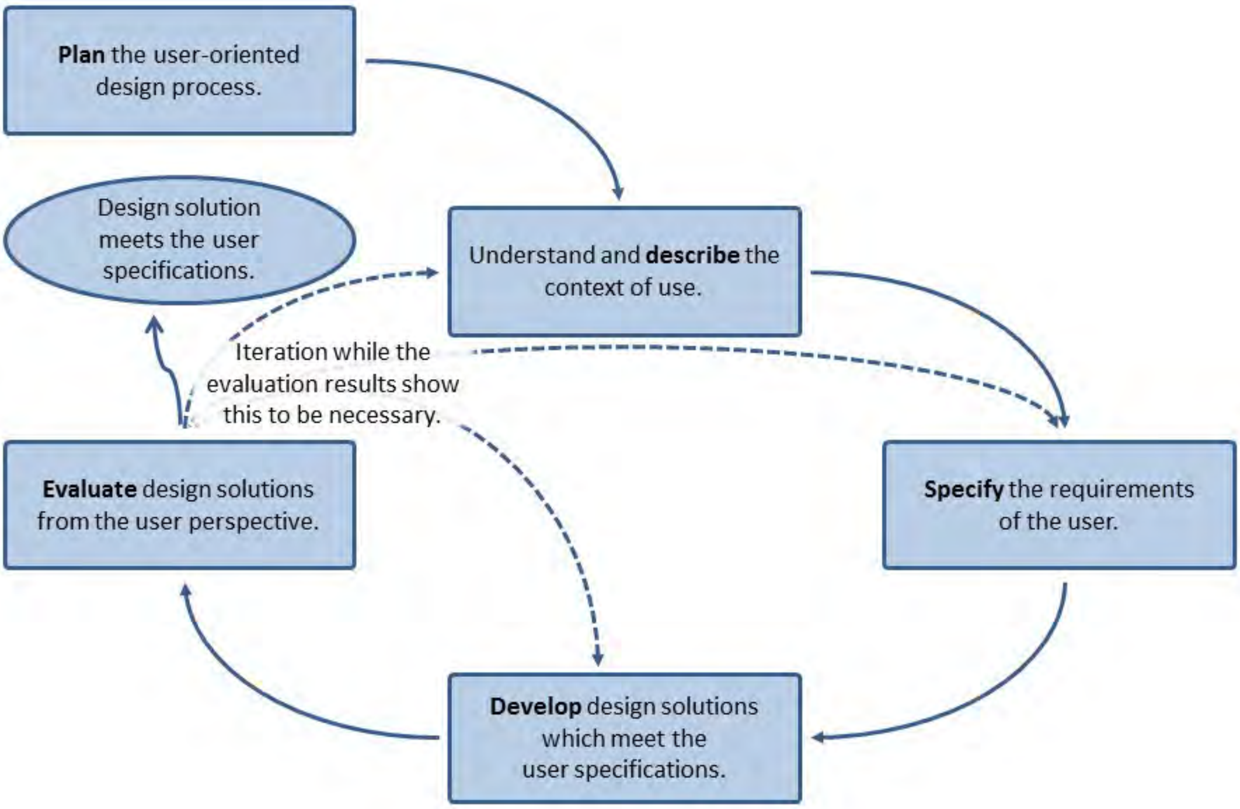

Fig. 14: Process for the designing of user-oriented interactive systems in accordance with DIN EN ISO 9241-210

The expert reviews should provide assessments of various demand aspects, in the sense of the aforementioned demand categories, for example the functionality (in respect of both user expectations and software design), ergonomics and look of the documentation instrument and electronic test report. The questions to be asked are to address possible structural weaknesses, content and programming errors, inconsistencies in design and wording, and above all experience relating to support for the task of test control (planning of the test route). In addition, so-called "cognitive walk-through methods" are to be used: The experts here perform typical user actions, and any problems revealed are collected, discussed, classified and prioritised according to their severity. Overall, as already mentioned, several feedback loops are to be realised, possibly also with parallel software versions, where appropriate (NIELSEN, 2011). By way of this iteration, the expert reviews provide an analytical basis for successive elaboration of the beta version of the documentation instrument and electronic test report. This beta version can then be made the subject of closer empirical study and improvement in a series of beta tests and further development cycles realised in cooperation with the driving test examiners during the further course of the feasibility study.

Within the framework of beta testing, the documentation instrument and electronic test report which have been elaborated and constantly improved up to this point are to undergo practical testing in simulated and – insofar as the development status of the instrument and the legal framework permit – real practical driving tests. The focus is here placed on the one hand on assessments of user experience, but at the same also on possibilities to achieve further optimisation in respect of the aforementioned functions of test documentation (see Chapter 4.2.3) and the described demand categories relating to software and hardware design (see Chapter 4.3). For realisation of the beta tests, it is suggested that five experienced examiners be selected by each Technical Examination Centre and by the Bundeswehr, and that they then each conduct at least five practical driving tests in each development cycle. This recommendation derives from the results of the aforementioned methodical research by NIELSEN (1994); at the same time, by involving not only all four Technical Examination Centres mandated to conduct driving tests on behalf of the state, but also the Bundeswehr, the implementation of this recommendation would promote comparability and acceptance, and furthermore lend the testing a certain representative character. To attain maximum effectiveness in product development, NIELSEN (1993, 2011) recommends that tests be repeated between five and ten times, i.e. at least six development cycles; with each repetition, both the testing and development steps are realised once more. If it is taken into account that alpha testing already comprises two development cycles (an initial test and one repeat test), at least four more development cycles must be planned for beta testing (an initial test and three repeat tests).

Qualified observer training is a decisive means by which to safeguard the long-term methodical quality of systematic behaviour observations such as those performed within the framework of the practical driving test (KANNING, 2004; STURZBECHER, 2010). Not least for this reason, the selected examiners must receive intensive training on handling of the documentation instrument and electronic test report – illustrated by way of practical examples – before the commencement of the prototype tests; this should also include observation, assessment and decision exercises to practise uniform application of the optimised test concept. It is similarly imperative to consider also perception and judgement errors (SCHULER, 2001).

It was already mentioned that – assuming corresponding progress in the development of the documentation instrument and electronic test report, and subject to conformance with the applicable legislation pertaining to driver licensing – the beta testing should be conducted partly within the framework of real driving tests.[101] It seems certain that the necessary prerequisites will not be met before – at the earliest – the last two development cycles of the prototype or beta testing phase. Correspondingly, the first two beta test series and their associated development cycles should concentrate on a detailed contextual analysis and general investigations of functionality and practicability, including aspects of hardware and software ergonomics, by way of simulated tests. By contrast, the two concluding beta test series, which can presumably be realised within the framework of real driving tests, should focus on full electronic documentation of a typical candidate performance, as it is necessary to guarantee the availability of a functionally robust documentation instrument for the pilot tests to be conducted after the feasibility study. To this end, it seems expedient to increase the number of tests to be handled by each examiner from five to ten for the concluding rounds where beta testing is integrated into real driving tests. Based on a conservative estimate of the probability that remaining usability problems will be identified (10 per cent), this would enable each examiner to notify at least 65 per cent of the remaining usability problems during the concluding beta tests and final development cycle. If the described methodical recommendations are followed, a total of at least 625 application trials would be realised with a new documentation instrument and electronic test report within the framework of beta testing.

It is proposed that, at least during the concluding development cycle of the feasibility study, the driving test examiners should be asked to complete semi-standardised questionnaires (with open answer options) after completing a real test, as a means to evaluate their impressions and any suggestions for optimisation from the user perspective; alternatively, semi-standardised interviews could be conducted. On the basis of knowledge from usability research, it is assumed that the majority of possibly existing inconsistencies, weaknesses and errors in the documentation instrument and electronic test report can be identified in this way. A further recommendation is to arrange a combination of individual telephone interviews and focus group discussions with the participating examiners at the end of each development cycle (GRONER, RAESS & SURY, 2008; HEGENER, 2003); in a focus group discussion, for example, a moderator could present the developer's plans for improvement of the documentation instrument and electronic test report, and then ask for the opinions of the participants based on their experiences during the previous testing.

The feasibility study should be concluded with the presentation of a final report, which should, if possible, contain at least the following results:

- A description of the tested documentation instrument and electronic test report (hardware specifications, screen and interaction design), where possible together with corresponding recommendations for closer definition in preparation for the revision project
- A SWOT analysis
- A concept for the provision of hardware and software to the driving test examiners participating in the revision project
- Thoughts on a selection of pilot test locations, taking into account the possible necessity to obtain approval for the realisation and documentation of practical driving tests in a manner which deviates from the form dictated by the current test report
- Thoughts on the installation of communication processes and interfaces for the revision project

[101] To minimise possible interference with the existing framework of licensing legislation, the written feedback to the candidate on test performance should continue to be provided by way of the currently applicable test report in accordance with Annex 13 to the Examination Guidelines; it could nevertheless be enriched with results from the electronic test report.

- An estimate of an average timeframe for the realisation of a practical driving test based on the optimised test concept and an electronic test report.

4.6 Summary

Optimisation of the practical driving test entails three essential changes in respect of the test documentation:

- The candidate's performance during a test drive in real traffic is to be recorded – as was already practised in the 1990s – with the aid of a matrix, which permits event-oriented assessments (errors, above-average performance) to be assigned to specific driving tasks and observation categories, and thus enables more readily understandable and more differentiated documentation of the candidate's test performance than has been possible to date.
- The event-oriented assessment and documentation of test performance is to be expanded to embrace also assessment and documentation of the underlying competence dimensions: The objective of novice driver preparation, after all, is not merely to avoid certain specific driving errors, but rather to learn to drive in a wholly error-free manner through the acquisition of driving competence. This holistic competence-oriented perspective must also apply in particular for the practical driving test, i.e. a negative test decision should not be explained by individual driving errors, but instead by an inadequate level of driving competence displayed in certain areas of competence.
- In future, the test documentation is to be realised by electronic means, because electronic data acquisition is able to minimise the work required to document test performance, and electronic data processing will furthermore greatly simplify test assessment, administration and evaluation. Optimisation of the test documentation is an important starting point for improvement of the methodical quality of the practical driving test, but by no means the only one. Aspects which appear equally important are attentive and purposeful observation of the candidate's driving behaviour, and not least continuous planning of the further course of the test drive, which requires both the test performance displayed so far and the road infrastructure conditions at the test location to be taken into account. Both demands – focussed attention during the observation of driving behaviour and planning of the further course of the test drive – occupy a large proportion of the examiner's mental capacities. The interactive screen forms of an electronic test report must therefore be designed around the latest computer technologies and in full awareness of the principles of hardware and software ergonomics, so as to achieve the intended benefits with a minimum of required effort on the part of the examiner. This also seems expedient because both driving errors and above-average test performance should preferably be documented immediately after the corresponding observation.

A fourth significant change in connection with optimisation of the practical driving test is to be seen in the improved use of the test documentation for a subsequent discussion with the candidate, in the sense of initial performance-referenced feedback, and above all for written performance feedback to all test candidates:

- Meaningful test documentation establishes a properly founded basis for a subsequent development-oriented discussion between the examiner and candidate: If errors are noted, or if inadequate driving competence is attested in the overall performance assessments, then these competence deficits must also be made the topic of a corresponding consultation, so as to indicate possibilities for effective and safety-relevant competence development. This naturally applies all the more so after a failed test, as the driving instructor must be enabled to tailor his further offers for driver training to the contents of the outcome discussion. It is no less important, however, for any above-average test performance to be mentioned in a discussion with the candidate, so as to illustrate the balanced nature of the test decision and to give the candidate incentives for further learning.
- The provision of learning-oriented written performance feedback to all candidates is deemed particularly important, because the candidate will typically only take in fragments of the verbal explanations given by the examiner immediately after the test, as he is at this time still occupied with affective processing of the test result. This applies equally to both the stress experienced if the test is failed and to the positive emotions following a successful test: Experience shows that, in this post-test situation, the interest in pointers for further learning is limited. It is known from initial studies in connection with the feasibility study on the new electronic test report (FRIEDEL,

MÖRL & RÜDEL, 2012), however, that most candidates are actually very much interested in a meaningful performance assessment and learning suggestions; it is not only possible, but indeed imperative to make use of this interest at a later time by providing corresponding feedback, whether as a print-out or in electronic form by e-mail or online. This is relevant above all because test candidates must still be viewed as inexperienced drivers at the time of the driving test, and are dependent on effective support for their further learning during the high-risk transition to solo driving. At the same time, qualified performance feedback could benefit the driving instructors and driving schools, because assessments of performance under test conditions allow conclusions to be drawn on the quality of training and indicate potential for optimisation in this respect. Such performance feedback can be generated automatically by the software of the electronic test report. Central components should be the matrix of the test report with the individual event-oriented assessments, the competence-oriented overall assessments relating to the different driving tasks and observation categories, a corresponding legend, and compact notes with recommendations for further learning.

In conclusion, it remains to be noted that the introduction of an easy-to-use electronic documentation instrument based on appropriate content-related and methodical principles would represent a decisive contribution to improvement of the quality of the practical driving test, because the realisation of tests would be simplified and controlled, and effective and efficient formative and summative evaluation of the test would be made possible. Alongside the professionalisation of driving licence testing, further aspects of improved novice driver preparation to be expected from such documentation include, not least, optimisation effects for the promotion of further learning and the quality of driver training. As a basis for successful fulfilment of these expectations, in addition to description of the scientific foundation and methodical functions of (electronic) test documentation, the preceding chapter also concretised demands relating to hardware and software design for the documentation instrument and to the testing of this instrument – and thus the electronic test report – within the framework of a feasibility study; further impetus for continued development can be derived from evaluation of the practical driving test.

5 Evaluation of the optimised practical driving test

5.1 Demands placed on the evaluation of measures

State measures which limit general action freedoms of the individual citizen are only permissible where they serve a reasonable public interest and are both necessary and suitable to achieve this objective (ALBRECHT, 2005). Proof that a measure satisfies these prerequisites can only be furnished by establishing and then monitoring the observance of implementation standards. On this basis, the effectiveness of the measure must be investigated and verified in the context of the intended purpose. The practical driving test is such a state-ordained measure: It limits the freedom of mobility of the individual in the interest of overall road safety, since – with a few legislatively stipulated exceptions – no-one is permitted to drive a motor vehicle in public road traffic without having previously passed a driving test. Consequently, it is necessary to monitor the realisation of the practical driving test and its impact in terms of road safety. This should be achieved within the framework of an evaluation, the content and procedures for which must be scientifically founded. An evaluation is here understood as a process serving to judge the quality of an item under discussion on the basis of available – or newly surveyed – and robust (i.e. objective, reliable, valid and representative) data relating to one or more questions and prescribed standards (WIDMER & BEYWL, 2009; BORTZ & DÖRING, 2006; DEGEVAL, 2008; WESTERMANN, 2002). Evaluations are thus generally based on comprehensive empirical studies (SCHUSCHKE, DAUBENSPECK & SATTELMACHER, 2008).

The subject of an evaluation may be a static product or concept, but could equally be essentially dynamic in nature, as is the case with the evaluation of a process, a project, a programme or – expressed more generally – a "measure" (DEGEVAL, 2008; WOTTAWA & THIERAU, 2003). "Measure is here the most general terminology and can be used to describe any and every form of action from the erection of a traffic sign, via the operation of a clinic, through to the unification of two states" (WESTERMANN, 2002, p. 8). The evaluation of a measure thus consists of an analysis with subsequent assessment based on scientifically recognised methods and serves in the end to verify the effectiveness of said measure. The evaluation results offer those responsible first and foremost a strategic basis for decisions regarding the continuation, modification, broader implementation (beyond a pilot phase) or termination of a measure (BORTZ & DÖRING, 2006; WOTTAWA & THIERAU, 2003).

Within any given organisation, both evaluations and quality management involve judgements of quality, but there is nevertheless an important difference between the two: The term "management" embraces "purposeful planning, control and monitoring of the business processes in organisations" (BÜLOW-SCHRAMM, 2006, p. 16); "quality management" is thus understood to refer to accompanying and coordinated activities of an organisation which are integrated into the continuous management function and serve to safeguard the standardised quality of products or services. "Evaluations", on the other hand, are special activities which are conducted by mandated experts external to the organisation (WESTERMANN, 2002); they are often realised over certain limited periods or at certain intervals, but may also be organised as an accompanying process in a similar manner to internal quality management. Summative evaluations place their focus on a particular intermediate state or else the final product of the subject under review, whereas a formative evaluation analyses and judges already the process of product development or measure implementation (BORTZ & DÖRING, 2006; DEGEVAL, 2008, WOTTAWA & THIERAU, 2003). An evaluation project will often include both formative and summative components, although different evaluation teams should then be assigned to realise the different methods (WESTERMANN, 2002).

Whenever a measure is to be evaluated, the objectives of that measure and the existing framework conditions must be determined and assessed in terms of their significance; at the same time, it is always necessary to define also the goal of the evaluation (BORTZ & DÖRING, 2006; WESTERMANN, 2002, WOTTAWA & THIERAU, 2003). Both the objectives of the measure and the objectives of the evaluation should be considered from the different perspectives of the persons involved. In line with WESTERMANN (2002), the following persons and their organisations and representations can be counted "stakeholders" with – not seldom differing – interests in the design of a measure and in the form and results of a corresponding evaluation: (1) The persons who are the subject of the evaluation or belong to an evaluated group or institution, and external persons who supply information or conduct elements of the measure; (2) the direct target group of the measure and persons who are indirectly affected, and last but not least (3) the persons or organisations who commission and bear the costs of the measure or evaluation. Within the framework of the evaluation of a measure, therefore, closer attention should also be paid to the (in)consistencies in the objectives of the individual interest groups, to the reconcilability of these different objectives, and to the compatibility of the objectives with legal provisions, scientific knowledge, and ethical and moral standards.

The evaluation of a measure can take an explorative approach, i.e. examination and description of the context of the measure, or else seek to test hypotheses, i.e. verification of the operationalised assumptions, usually with reference to the effectiveness of the measure (BORTZ & DÖRING, 2006). A further distinction can be made between instrumental, processual and result-oriented evaluation (WESTERMANN, 2002; WOTTAWA & THIERAU, 2003): While instrumental evaluation is geared to assessing the methodical quality of the instruments serving realisation of the measure, process evaluations focus on the quality of a concrete implementation; a result-oriented evaluation, finally, is able to determine the consequences and effects, and thus in turn the success of a measure.

Possible survey methods for use in connection with evaluations are written and oral questioning of experts and other involved parties, experimental studies, observations, and the collection and analysis of objective data; each of these methods has its own strengths and weaknesses (BORTZ & DÖRING, 2006; WOTTAWA & THIERAU, 2003), but these are not to be the topic of discussion at this point. To be able to combine or compensate the specific advantages and disadvantages of the individual methods with regard to the various quality aspects of the evaluation subject, particularly for the evaluation of complex measures, it is common to use systems designed for multi-method or – where different target groups are involved – even multi-perspective evaluation (STURZBECHER & MÖRL, 2008). A combination of observations and questionnaire-based surveys at the same time offers an opportunity to validate the interpretations of an observer by way of the corresponding survey responses; this is sometimes referred to as "communicative validation" (MAYRING, 2002).

For cost reasons, the evaluation of complex measures is frequently concentrated on a selection of quality attributes and a sample of the persons and groups concerned. In this case, the selected attributes must be especially relevant for the quality of the measure, and the samples should be as representative as possible of all those involved in the measure (WOTTAWA & THIERAU, 2003). The term "representative sample" is here not a clearly defined statistical concept (SCHUMANN, 2006); even so, it is usually taken to indicate the demand that the sample should be the product of a random selection process and free from systematic distortion due to confounds and the associated "confounding effects" (SCHNELL, HILL & ESSER, 2008). Accordingly, the distributions of all evaluation-relevant attributes describing the persons of a representative sample should match the corresponding distributions in the overall population, apart from coincidental deviations. Since confounding effects are nevertheless possible, they must be analysed by suitable methodical means (e.g. factor monitoring, statistical partialisation), assessed in terms of their impact, and taken into account in the evaluation of the measure (BORTZ & DÖRING, 2006; WOTTAWA & THIERAU, 2003).

Sound analysis of the evaluation data and inherently consistent interpretation of the evaluation results, finally, permit an appraisal of the quality and success of a measure. The impacts of a measure – i.e. its strengths and weaknesses – should be appropriately differentiated, both with reference to the different interest groups involved in the measure and in terms of the relevant criteria for success; furthermore, the various content-related aspects of the impact are to be taken into account. This includes the effectiveness of the measure (the intensity of its impact compared to other measures) and its efficiency (cost-benefit ratio), as well as any social significance and political dimension (BORTZ & DÖRING, 2006; DEGEVAL, 2008; WIDMER & BEYWL, 2009). For the judgement of success, it is equally imperative not to neglect any unintended or even undesirable "side effects" which may arise (WOTTAWA & THIERAU, 2003). Moreover, it must be verified whether an observed success is actually attributable – either causally or as an elemental consequence – to the evaluated measure, or whether it would have been attained independently of the evaluated measure due to other factors. Last but not least, it must be noted that, even if a measure proves effective within the framework of a pilot study, it is only by way of a robust evaluation design that it can be determined whether the observed effectiveness can be generalised for the overall target population, or whether it is essentially due to certain attribute combinations among the participants or other specific general conditions of the pilot study.

The extent and difficulty of the work required by an evaluation is dependent on the duration and complexity of the measure: "The simplest evaluation subjects are interventions which are brief and discrete, and at the same time address clearly evident objectives, e.g. a speed limit to reduce the number of accidents. The most difficult are evaluations relating to measures which are diffuse and of long duration, potentially highly variable in design from case to case, and with broad effects …" (WESTERMANN, 2002, p. 8). If the measure to be evaluated continues over a longer period and requires modifications during this time, it may be necessary to consider corresponding adjustment of the evaluation design. Should this apply, however, it must be ensured that the comparability of the results remains guaranteed despite any methodically compelling adaptation.

To summarise, there are six essential aspects which must be taken into account in the evaluation design for a measure: (1) The subject of the evaluation, (2) the objectives pursued by the measure and evaluation, (3) the groups of persons involved in the measure to be evaluated, (4) the framework conditions of the measure and evaluation, (5) the methodical instruments used within the framework of the measure and evaluation, and (6) the quality criteria applicable to judge the success of the measure.

5.2 Quality assurance for systematic behaviour observation

The measure which is placed at the centre of evaluation interest in the following is the optimised practical driving test. From the didactic and test psychology perspective, and with regard to the methodology of its realisation, this test represents a multiply repeated work sample which is recorded and assessed by way of systematic behaviour observation (STURZBECHER, 2010). Systematic behaviour observation is thus to be viewed as the methodical instrument of the practical driving test; correspondingly, the instrumental evaluation of this measure is to be based on the three classic quality criteria defined in test psychology for the case of behaviour observations, namely objectivity, reliability and validity. Alongside, it seems expedient to take into account also certain secondary quality criteria such as economy, usefulness, reasonableness, resistance to falsification and fairness (BÜHNER, 2011; DEGEVAL, 2008; KANNING, 2004; LIENERT & RAATZ, 1998; SCHNELL et al., 2008; TESTKURATORIUM, 2010).

The objectivity of a behaviour observation is defined by way of the concordance in the methodical approaches of the individual observers ("observer consensus") with regard to the gathering ("objectivity in realisation"), assessment ("objectivity in assessment") and interpretation ("objectivity in interpretation") of the observation data (SPRUNG & SPRUNG, 1984). For the context of tests, EBBINGHAUS and SCHMIDT (1999) illustrate these three aspects of objectivity as follows: Objectivity in realisation is attained if the test is conducted under common conditions for all test candidates (e.g. same test duration and tasks to be solved); this is promoted by specifying demand standards. Objectivity in assessment requires that a given test performance produces the same result even after assessment by different examiners; to this end, assessment categories (e.g. observation categories) and rules (e.g. assessment criteria) must exist as a basis on which to judge whether a task has been fulfilled correctly or incorrectly, completely or incompletely. Objectivity in interpretation, finally, means that different examiners draw identical conclusions from given assessment results; this is served by decision criteria for the passing of a test, for example. Generally speaking, a high degree of method objectivity is ensured by way of written specifications standardising contents and application modalities for all components of the method implementation and assessment as far as possible (LIENERT & RAATZ, 1998; AMELANG & SCHMIDT-ATZERT, 2006). In the case of a test like the practical driving test, which, for reasons of validity (see below), is to be conducted in a lifeworld domain (GRUBER & MANDL, 1996) on the basis of an adaptive test strategy, however, natural limitations are placed on the objectivity in realisation, because the test conditions can hardly be standardised, planned and controlled to the extent that they are identical for every single candidate.

Reliability is understood to mean that the method applied for behaviour observation functions reliably overall, i.e. that the observation result is not dependent on any random influences on the observation process (BEINER, 1982; LIENERT & RAATZ, 1998). One important aspect of reliability in this context is the so-called intra-rater reliability, which could also be interpreted as the retest reliability or stability of the observation method (HASEMANN, 1983; INGENKAMP & LISSMANN, 2008; MEES, 1977). Investigation of the intra-rater reliability determines whether an observer also records the same observations, assessments and interpretations (i.e. test decisions) when an observation or test is repeated under the same conditions. As the specific test conditions can hardly be reproduced identically in a lifeworld domain, proof of intra-rater reliability can only be furnished if individual tests are recorded on video and then presented to the examiner for renewed observation and assessment at suitable intervals ("retest interval"). When doing so, learning or memory effects on the part of the examiner must be excluded, as this would lead to overestimation of the reliability. It is not possible to provide a generally applicable rule with regard to an optimum retest interval, as the relative risk of memory effects is dependent on the individual observation contents (MOOSBRUGGER & KELAVA, 2012).

Various authors (HASEMANN, 1983; INGENKAMP & LISSMANN, 2008; MEES, 1977) point out that objectivity aspects, i.e. the observer consensus, can hardly be kept distinct from reliability aspects, in the sense of stability or intra-rater reliability, when determining the methodical quality of an observation process; consequently, the term "inter-rater reliability" is occasionally used instead of objectivity. The aspect of observer consensus is always an especially important quality criteria for systematic behaviour observation. The coefficient which is frequently used to measure observer consensus is Cohen's kappa[102] (COHEN, 1960, 1968; CONGER, 1980); the following values are found as guidelines for judgement of the quality of observer consensus: In GREWE and WENTURA (1997), and likewise in FAßNACHT (2007), a kappa value of k= .70 is quoted as acceptable; according to FLEISS and COHEN (1973), values of k= .75 and higher are taken to be "very good", a value between k= .60 and k= .75 is considered "good", and a value between k= .40 and k= .60 is still "acceptable" for complex observation systems. Following the conclusions drawn by v. KLEBELSBERG (1970), the coefficient of consensus can be enhanced almost at will by raising the intensity of training. Indeed, coefficients equivalent to the reliability measurements of many psychological tests have also been recorded for observations, for example in McGLADE (r=.88; r=.93), BARTHELMESS (r=.91) and BIEHL et al. (r=.90; r=.92). For cases where an unsatisfactory observer consensus is to be improved, the following recommendation is given by KROHNE and HOCK (2007): "If the consensus is lower than that desired or necessary for the study purpose, and if no further increase is to be expected from intensified training, the structure of the observation system should be reconsidered and, if appropriate, simplified. Such simplification can sometimes also be achieved retrospectively by combining similar categories" (p. 270). BORTZ and DÖRING (2006) recommend above all training measures for the observer as a means to improve observer consensus.

According to FISSENI (2004), the objectivity and reliability of observation data will be found to be higher, "the greater the precision with which observation units are defined, the smaller the number of such units and the more specifically observation units are formulated, in other words the less they necessitate abstraction and inference" (p. 135). Applied to observation documentation, and in particular also to test situations, this means that a high degree of objectivity and reliability can be attained by specifying unambiguous and understandably formulated disjunct observation categories, together with unambiguous assessment and decision criteria (see Chapter 3). In the case of an objective and reliable test, a candidate displaying a constant level of competence should receive the same competence assessments and the same test decision each time throughout a series of several tests.

The criterion of validity describes the degree of precision with which the method of an observation or test actually acquires the objective, content-related dimensions which it is intended to measure (LIENERT & RAATZ, 1998). This degree of precision is dependent on the diagnostic objective and the examination situation, i.e. the subject of the examination, the sample, the environmental conditions, the person conducting the examination, and the time period over which the specified validity parameter remains constant (GUTHKE, 1990). Depending on the procedure applied to gather the validity statements, a distinction is made between content validity, criterion validity and construct validity (LIENERT & RAATZ, 1998; EBBINGHAUS & SCHMIDT, 1999).

Content validity means that an observation or test is ostensibly suitable to reflect the dimensions to be recorded and is generally confirmed for a particular method by way of an expert rating; it can be

[102] Cohen's kappa is a classic method to determine the consensus of two or a very small number of observers. Where the consensus of more than two observers is to be calculated, it can be recommended to use instead Cronbach's alpha (WELLENREUTHER, 1982; CRONBACH, 1951). The procedure for the calculation of Cronbach's alpha permits a larger number of observers and dimensions to be taken into account efficiently by relative simple means; ABEDI (1996) points explicitly to the dependability and expediency of this method for the determination of inter-rater reliability. Moreover, the literature provides indications that the use of Cronbach's alpha leads to underestimation of the consensus between observers; it thus seems to be a more conservative measure of objectivity compared to Cohen's kappa. Further measures of consensus are the percentage of agreement (FLEISS, LEVIN & PAIK, 2003; WIRTZ & CASPAR, 2002), Fleiss' kappa (FLEISS, LEVIN & PAIK, 2003), Kendall's W (WIRTZ & CASPAR, 2002) and the so-called intraclass correlation coefficient ICC (McGRAW & WONG, 1996). The ICC is particularly sensitive to systematic differences between raters, but can only be applied to interval-scaled variables (WIRTZ & CASPAR, 2002): Systematic differences can be ascertained, for example, if one rater gives consistently more positive or more negative assessments than another.

taken to apply, in particular, where the demands set by the observation or test are the same as those of the real-life situation, in which case it would also be possible to speak of "ecological validity" or a "conclusion of representativeness" (AMELANG & SCHMIDT-ATZERT, 2006). The content validity of tests which take the form of systematic behaviour observation can be assumed if experts in the field have evaluated the test items, the observation categories, the assessment and decision criteria and the instructions for realisation of the test on the basis of consistent theoretical notions and robust empirical experience, and have deemed them – preferably unanimously – to be necessary and sufficient for measurement of the subject in question. Studies and discussions on the content validity of performance tests generally address the topics of learning objective orientation, the comparability of test conditions and the occurrence of judgement errors (BORTZ & DÖRING, 2006). On occasions, the "agreement coefficient" described by FRICKE is used as a measure of content validity (AMELANG & SCHMIDT-ATZERT, 2006); this specifies the extent of agreement between expert assessments relating to the content validity of relevant test components.

In the case of criterion validity, a distinction can be made between "concurrent validity" and "predictive validity", both of which refer to the degree of correlation between the dimensions supplied by a behaviour observation or test and an independently acquired external criterion: Concurrent validity considers the performance displayed in a test more or less simultaneously with the external criterion and then investigates correlations between the two, whereas predictive validity determines the extent to which an external criterion which lies in the future can be predicted on the basis of the test performance. The degree of criterion validity is generally expressed with the aid of a correlation or convergence coefficient. The "known-groups technique" is here a special variant of the methods used to assess the concurrent validity of a test: By comparing groups of persons who must evidently demonstrate different levels of competence in the field which is subject of the test, it is determined whether or not the expected performance differences are actually revealed by the test (SCHNELL et al., 2008). Through consideration of the predictive test validity, an attempt is made to correctly predict future levels of performance in the given subject field on the basis of test performance. Where the occurrence or non-occurrence of a certain event is used to operationalise the future level of performance (dichotomous prediction), the judgement of criterion validity is often based on a so-called "confusion matrix" (STEHMANN, 1997; RUBIN, 2012) (see Fig. 15): In this context, a distinction is made between the "sensitivity" of a test (understood to mean the suitability to predict "true positive cases") and its "specificity" (its suitability to predict "true negative cases"). Taking into account the significance of forecast errors, so-called "cut-off points" are defined for the desirable and acceptable degrees of sensitivity or specificity (AMELANG & SCHMIDT-ATZERT, 2006).[103]

		Predicted result	
		Positive	Negative
Actual result	Positive	**True positive** Sensitivity	**False negative** Type 2 error
	Negative	**False positive** Type 1 error	**True negative** Specificity

Fig. 15: Confusion matrix

As it is often difficult to find a single external criterion which properly covers the whole subject field of a test, BORTZ and DÖRING (2006) recommend that criterion validity be assessed against several external criteria.

The procedure to determine the construct validity for a test or similar case of systematic behaviour observation is as follows (GUTHKE, 1990; NOWAKOWSKA, 1973):

1. On the basis of theoretical knowledge of the subject of the validation, statements are collected with regard to
   - presumed relationships between the subject of the validation and other (comparable) constructs (A),
   - the presumed absence of relationships between the subject of the validation and other constructs which can be deemed disparate in terms of content or theoretical base (B), and
   - relationships between the subject of the validation and certain external criteria (C) (see criterion validity).

[103] An example can here illustrate the possible cases: A driving licence applicant may display good or poor performance in the practical driving test, and on this basis, a prognosis could be given as to the likelihood of his being involved in a road accident during later solo driving. If the test performance was good, and he later also avoids all accident involvement, this case is termed a "true positive"; if he is unexpectedly involved in an accident despite a good test performance, this represents by contrast a "false positive". If the candidate's test performance is poor and he is later indeed involved in an accident, this falls into the category "true negative", whereas "false negative" means that the predicted accident did not occur.

2. Methods are chosen or elaborated to evaluate the compared constructs (A and B) and external criteria (C) in an adequate manner and to acquire the corresponding data.
3. A nomological network of hypotheses of type A, B and C is formulated and tested against the empirical base. As the outcome of this testing, proof of construct validity is considered to have been furnished if the values relating to the subject of the validation
   – display a high degree of correlation with the compared constructs which are theoretically related to the subject of the validation ("convergent validity", A);
   – do not correlate with the value of those compared constructs which are theoretically not related to the subject of the validation ("discriminant validity", B), and
   – permit the confident prediction of criterion measurements which are theoretically connected with the subject of the validation ("criterion validity", C).

By combining the pragmatically oriented notions of criterion validity with a theoretical consideration of the subject of the validation, construct validity incorporates all other forms of validity (MICHEL & CONRAD, 1982) and continues theoretical foundation and precise definition of the test construct (GUTHKE, 1990).

Among the usual secondary quality criteria, the criteria of economy, usefulness, reasonableness, resistance to falsification and fairness appear to be of particular interest with regard to behaviour observations or tests, as in the present case. An observation or test method is economical if realisation occupies only a short period of time and can be handled routinely with a minimum of organisation; in other words, it can be realised and assessed simply, conveniently and with a minimum input of resources. A method can be deemed useful if realisation serves a practical need, while reasonableness requires that the resulting benefit of the method stands in appropriate relationship to the temporal, mental and physical loads placed on participants. The criterion of resistance to falsification is met if original performance assessments and test decisions can no longer be altered by way of later, validity-diminishing actions. Important prerequisites for fairness, finally, are population-specific equivalence and the transparency of demand and assessment standards (SCHWENKMEZGER & HANK, 1993): Population-specific equivalence means that the test results are not affected by content-independent inter-individual and population-related differences. To be able to judge this equivalence, an instrumental evaluation of the test methods must also examine how variables which characterise the candidate influence test performance, taking into account above all those attributes which seem relevant for the safeguarding of test equality (e.g. the age, gender and educational background of the candidate).

In conclusion, it should be noted that the aforementioned primary quality criteria are closely correlated: Objectivity and reliability are necessary prerequisites, but alone still insufficient evidence for the validity of a behaviour observation or test.

Before the general demands placed on the evaluation of a given measure, and here particularly on the instrumental evaluation of systematic behaviour observations, are applied to the evaluation of an optimised practical driving test in Germany, the next chapter is to investigate the possible inspiration which may be drawn from international evaluation practice.

5.3 Test evaluation in international practice

A two-stage approach was chosen for the survey of international practice relating to evaluation of the practical driving test: First, the results of the BASt project "Novice Driver Preparation – An International Comparison" (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014) were reviewed for information characterising the implemented systems of novice driver preparation in 44 countries (see Chapter 3); on this basis, a deeper analysis was started to examine the situations in 36 of those countries in more detail. The findings were nevertheless sparse, although it remains uncertain whether only few countries have to date performed scientifically founded evaluations of their practical driving tests, or whether evaluations have been performed without subsequent publication of a corresponding report. The order of the sections in this chapter reflects the elaborated systematic structure of a scientifically founded evaluation of the optimised practical driving test for Germany; the essential elements include (1) an instrumental evaluation, (2) an analysis of test results, (3) customer surveys, and (4) the realisation of so-called "product audits" (expert observations and supervisions).

Instrumental evaluation

The international comparison shows that studies of the psychometric quality of the practical driving test are apparently rare; the relatively few investigations of this kind have been conducted above all in Great Britain.[104]

As far as the reliability aspects are concerned, a study of 366 test candidates by BAUGHAN and SIMPSON (1999) found that the practical driving test did not display a particularly high degree of retest reliability with regard to the determination of driving competence in real traffic (albeit presumably under different conditions for the test and retest): While approx. 35 per cent of the candidates passed a first test, the pass rate increased to 42 per cent for a voluntary repeat test a few days later. Neither the examiners nor the candidates were aware of the results of the first test; the driving test as a whole was already deemed "passed" if one of the two tests was completed successfully. The observed increase in the pass rate was explained above all with learning effects relating to the test situation. It was conspicuous, however, that among the 36 per cent of candidates with different results in the two tests, 16 per cent failed the repeat test despite having passed the original test. Only 64 per cent of the candidates achieved the same result in both tests. Additional studies to verify observer consensus (two examiners in the same test vehicle, longer test duration) suggested that the inconsistencies in the test results were attributable primarily to the "personal form" of the candidates on the day of the test, rather than to differences in the examiners' assessments (BAUGHAN & SEXTON, 2001); the report unfortunately contained no information on differences in test conditions between the original and repeat tests, even though this can be deemed indispensable for reasonable interpretation of the findings.

KESKINEN, HATAKKA and LAAPOTTI (1988) also investigated the aspects of objectivity and reliability; with reference to the Finnish practical driving test, they calculated an observer consensus of approx. 90 per cent between the driving test examiner and the driving instructor in respect of serious errors leading to immediate termination of a driving test. For errors where the test is not automatically failed as soon as they are observed once, the consensus was 80 per cent, with an uncertainty of ±1 error.

In the Netherlands, a questionnaire ("DPA – Driver Performance Assessment") has been developed since 2007 to enable driving instructors to provide a graduated assessment of a candidate's driving competence in five categories ("Safe driving", above all with reference to speed and safety margins, "Consideration for other road users", "Facilitating traffic flow", "Vehicle control", "Environmentally responsible driving") on the basis of a four-level rating scale from "Unsatisfactory" to "Optimum" (ROELOFS, VAN ONNA & VISSER, 2010; ROELOFS, VISSER, VAN ONNA & NÄGELE, 2009). Driving instructors were given training in use of the observation instrument in a series of three three-hour workshops; alongside, they were provided with a scoring manual. To test inter-rater reliability, the instructors were presented 12 video clips showing critical aspects of the performance of driving tasks by four drivers, for which a Gower coefficient of .70 was determined. The authors specify a retest reliability between .70 and .80. With the aid of logistic regression analysis, it was possible to demonstrate a strong predictive correlation (r= .90) with the later result in the practical driving test for this instrument; it was thus proven to be predictively valid.

Validity studies relating to the practical driving test are frequently geared to the question as to whether or not the test performance displayed by a candidate permits conclusions to be drawn in the sense of a prognosis for the risk of accident involvement in later solo driving. This question addresses the predictive (criterion) validity of the driving test; statements on content validity, on the other hand, are seldom among the results of the available research, and the same generally applies to concurrent (criterion) validity or especially construct validity.[105] A few findings relating to the predictive validity of the test are given below:

– MAAG, LABERGE-NADEAU, DESJARDINS, MORIN and MESSIER (2001) found no evidence in studies conducted in the Canadian province of Quebec for a significant correlation between test performance and accident

[104] In 2005, as contribution to an investigation of the optimisation possibilities for a reform of the practical driving test, BAUGHAN, GREGERSEN, HENDRIX and KESKINEN produced a summary of all the important research results which had been published to date on the practical driving test in Great Britain.

[105] It is assumed that the content validity of the practical driving test has not been a topic of explicit discussion because the involvement of corresponding experts in development of the test contents and methodology is taken for granted and the test is thus automatically taken to be valid in this respect. If this is so, such a position should be questioned critically, because it is not the involvement of practicians per se, and not the recourse to practical experience alone, but rather the purposeful and methodically founded integration of professional expertise into the process to develop test standards which promises gains in terms of validity. This point was already made by HAMPEL (1977): "Particularly in the case of an everyday activity like driving, there is a constant danger of stereotype definitions of what constitutes a 'good driver' influencing the elaboration of objectives, if there is no monitoring on the basis of later driving mastery, in other words according to external criteria" (p. 19). Safeguarding of the content validity of the practical driving test and systematic investigation of its criterion and construct validity thus represent independent, but equally important methodical challenges to be met in further scientific research.

involvement in the first three years after the driving test.

- From a study of the correlation between driving test performance and the probability of novice driver accidents, MAYCOCK (2002) concluded that the practical driving test is not suitable for the measurement of hazard cognition.[106]
- According to CRINSON and GRAYSON (2005), young men display a higher pass rate in the practical driving test than young women; nevertheless, their later accident risk after the transition to solo driving is significantly higher.
- HATAKKA et al. (2002) found that male candidates who displayed relatively good test performance were later involved in accidents and traffic offences more frequently than male candidates whose test performance was relatively poor. The explanation given by the authors was that the focus of the practical driving test is placed on the demonstration of basic driving skills, rather than an assessment of attitudes, motivation and overall manner of driving. The necessity to apply the GDE matrix (see Chapter 2) to the driving test was also derived from this conclusion.
- In the studies conducted by WELLS et al. (2008) and EMMERSON (2008), a higher accident rate in later solo driving was determined for those candidates who described the degree of difficulty of the practical driving test as “low”. It remained unclear, however, whether these candidates did indeed possess better driving skills. The results perhaps show also an influence of self-overestimation – which can hardly be assessed during the practical driving test – on accident risk.[107]
- In other studies, an (albeit minimal) correlation was revealed between the number of minor errors (above all those in connection with traffic observation) and later accident involvement. The need to repeat tests was similarly identified as an accident predictor, especially in the case of female candidates (MAYCOCK & FORSYTH, 1997; SEXTON & GRAYSON, 2010). The accident rate among female candidates who passed the theoretical and practical driving tests at the first attempt, for example, was lower than that for female candidates who required several attempts (MAAG et al., 2001).

Viewed overall, the aforementioned studies fail to paint a consistent picture; in many cases, no significant correlation is found between test performance in the practical driving test and the candidate's later risk of accident involvement during solo driving (BAUGHAN, 2000; MAYCOCK, 2002). On the one hand, as already argued by HATAKKA et al. (2002), this is presumably due to the fact that the practical driving test assesses above all driving skills, but not attitudes to safe, defensive driving, even though the latter is likewise an aspect of driving competence and significantly influences accident risk. This limits the applicability of accident figures as an external validity criterion for the practical driving test. On the other hand, it seems feasible that a differentiated analysis of certain subsets of candidates (e.g. female candidates) or selected elements of competence (e.g. traffic observation) could nevertheless reveal clearer correlations between test performance and accident risk. The identification of such correlations would offer new opportunities to improve the control function of the practical driving test within the system of novice driver preparation and to reduce the accident risk for novice drivers.

Such opportunities are offered not only by driving tests in real traffic, but also by computer-assisted traffic perception tests, examples of which have already been introduced as an innovative form of testing in a number of countries (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014). In Australia, for instance, driving licence applicants must pass a so-called “Hazard Perception Test” (hereafter abbreviated to HPT). The Australian Council for Educational Research (ACER) conducted validity studies to determine whether candidates with low HPT scores are later involved in accidents more frequently than candidates who perform well. To answer this question, a comprehensive survey over the period from April 1996 to December 1997 acquired the HPT scores of 99,326 candidates; of this number, 2,300 drivers (2.3%) were later involved in accidents. It must be mentioned, however, that only accidents resulting in injury to persons were recorded; accidents resulting exclusively in material damage, as well as any other traffic offences, were ignored. This naturally reduces the meaningfulness of the external criterion. Subsequent analyses indicated that the

[106] It is to be noted that the demand and implementation standards for the practical driving test vary to a greater or lesser degree between the different countries, and that the specific methodical backgrounds of the individual national tests must thus be taken into account to permit differentiated interpretation of such findings. This, however, would far exceed the possible scope of the present report.

[107] The candidate's assessment of the degree of difficulty is based on a self-assessment of driving competence and thus susceptible to illusions of control, whereas passing of the test is dependent on the external assessment of the driving test examiner, which is generally objectivised by way of assessment criteria.

HPT score contributes only marginally to correct predictions of serious or fatal accidents (CONGDON, 1999); on the other hand, the predictive quality of HPT performance could possibly have been increased by broadening operationalisation of the criterion.

Finally, attention can be drawn to studies whose results identify confounding variables which diminish the validity of the practical driving test as an instrument to assess driving ability, or at least influence the test demands. FAIRCLOUGH, TATTERSALL and HOUSTON (2006), for example, discovered a significant correlation between test anxiety and failing of the driving test: Due to test anxiety, the candidate is unable to demonstrate his true driving competence. BAUGHAN et al. (2005) noted that the pass rate is three per cent lower in rainy weather compared to dry conditions.

Analysis of test results

Statistics on the results of practical driving tests are recorded in 85 per cent of the 36 countries considered by the project, at least in respect of the pass rate; in most cases, however, it remains unclear whether or to what extent these statistics are analysed for evaluation purposes, and then whether such analysis results influence quality management measures in any way. Test statistics are usually analysed with reference to a particular region or test centre, but occasionally also at the level of the individual examiner (e.g. in Iceland and Luxembourg). In the Netherlands, the pass rates achieved by individual driving schools are published. Evaluation systems which incorporate also systematic assessments and error analyses on the basis of test reports are found in Great Britain, the Netherlands, the Canadian province of Ontario, Sweden and the Czech Republic. In Great Britain, for example, the ten most common reasons for failing the driving test are inadequate use of mirrors and lack of proper traffic observation at road junctions, errors when reversing, hesitant or unnecessarily slow driving, and errors in connection with the use of signals, reverse parking, moving off safely, steering control, vehicle positioning and gear selection (BAUGHAN et al., 2005). It should be noted, however, that such findings are naturally influenced by the methods used for recording.

Surveys of customer satisfaction

Only very few research reports are available on studies to determine customer satisfaction with the practical driving test. In Northern Ireland, for example, surveys addressing satisfaction with the theoretical and practical driving tests are conducted on a regular basis. The surveys also gather certain socio-demographic data, such as gender, age, family status, disabilities, religion, political affiliation, ethnic background and native language, apparently to facilitate estimation of the population-specific equivalence. In 2003, 84 per cent of the survey respondents said that they were satisfied with the practical driving test (DVA, 2005). While men were slightly more satisfied with the practical test than women, the opposite was the case for the theory test (ibid.). In a study to assess the perceived fairness of test decisions, LAAPOTTI, KESKINEN, HATAKKA and KATILA (1998) found that successful candidates rated the test assessment as fairer (mean rating 4.4 on a five-level scale from "1 = Absolute rejection" to "5 = Full agreement") compared to candidates who failed the test (mean rating 3.8).

Product audits[108]

Various approaches exist with regard to the realisation of product audits: A number of countries choose to implement only external audits (e.g. Iceland, New South Wales) or else exclusively internal audits, with the control procedures often being prescribed by government authorities in the latter case (e.g. in Belgium, Great Britain and the Netherlands). In most countries, however, provisions are made for a combination of internal and external audits; the realisation of audits often follows official specifications and external audits are sometimes only performed where circumstances arise in which they are deemed appropriate (e.g. in Sweden in case of conspicuous statistical findings or complaints). In France, audits are combined with the completion of a survey by driving test examiners. In Lithuania, only internal product audits are performed, but then include also determination of the inter-rater reliability with regard to the test performance assessments of the auditor and the examiner.

In the Canadian province of Ontario, at least two internal audits are performed with each driving test examiner every six months. During such a "check ride", a list of defined criteria is used to judge test

[108] The term "audit" is used in general to describe inspection or monitoring methods which serve to verify that a process complies with (officially) stipulated demands. Audits are thus evaluation instruments and are frequently employed as such within the framework of quality management procedures (see above). While internal audits are performed by members of the organisation concerned, external audits are entrusted to independent experts, who are in such cases not seldom acting with a state mandate. "System audits" serve to provide a comprehensive overall assessment of the quality management system itself, whereas "product audits" focus on the fulfilment of process demands and on the compliance with quality criteria relating to the products and services which represent the outcome of the process.

realisation by the examiner (preparation, welcome, introductory briefing, vehicle check, verification of the candidate's identity, time management during the test), his knowledge and handling of the assessment criteria (correct recording of errors, legibility, accuracy, observance of a proper test route, proper realisation and sequence of the driving tasks) and the conveyed feeling of well-being and safety (checking of the safety-relevant features of the candidate's vehicle, timely instructions to the candidate, consideration of the surrounding traffic situation when giving instructions, recognition and if necessary avoidance of hazards, timely premature termination of the test in case of inadequate driving skills on the part of the candidate, intervention to avoid accidents). In addition, the test reports produced by the examiner and the auditor are here, too, compared for consistency.

Within the framework of the international project "TEST" ("Towards European Standards for Testing"), which was realised between 2003 and 2005, approx. 3,150 practical driving tests conducted in the six participating countries (Austria, France, Great Britain, the Netherlands, Sweden and Spain) were audited on the basis of a comprehensive "Auditor's Form"[109] (BAUGHAN et al., 2005). The aim was to determine whether the practical driving test offers sufficient opportunities for reliable assessment of the candidate's driving competence. This included also judgement of the suitability to assess specific aspects of the performance displayed by the candidate (e.g. behaviour at crossroads and junctions) and the influence of environmental circumstances on the assessment of test performance (e.g. traffic conditions, weather conditions). The auditing form was supplemented with an examiner questionnaire, on which the participating examiners were asked to note their experiences and any suggestions for improvement of the practical driving test. Of the 404 examiners who completed the questionnaire, a total of 84 per cent replied that they were satisfied with the test contents, but only 69 per cent expressed similar satisfaction with the test locations. Many of the test locations were indeed unable to support the requirements for the test drive contained in EU Directive 2000/56/EC, and it seemed that robust assessments of the driving and traffic competences necessary to safely complete the various driving tasks were actually impossible at many of the test locations in the six countries (above all with regard to passing and overtaking, railway crossings, trunk roads/motorways and driving on roads with gradients). The pass rate was around 25 per cent higher in rural areas compared to urban areas; both the examiners and the auditors saw better opportunities for a valid assessment of driving competence in urban areas. A significantly lower pass rate was recorded for the audited driving tests compared to tests which were not audited.

To summarise, it can be noted that the current international practice employs a diversity of (1) studies serving instrumental evaluation of the practical driving test, (2) statistical analyses of test results, (3) customer surveys, and (4) product audits at more or less frequent intervals as elements of an evaluation system. This indicates that appropriate combinations of such elements could well be incorporated into a methodically professional evaluation system. Accordingly, the scientific founding and possibilities for corresponding implementation of the individual elements in Germany are to be discussed further in the following chapter.

5.4 The evaluation system for an optimised practical driving test

5.4.1 Foundations and starting points

As already explained above, a robust evaluation concept must take into account the specific circumstances of the measure to be evaluated; these circumstances are determined by the topic addressed by the measure, its objectives, the groups of persons involved, the framework conditions, the methodical instruments and the quality criteria to be applied for the evaluation. In the present case, the measure or topic to be evaluated is the optimised practical driving test, the overarching objective of which is to reduce the accident risk faced by novice drivers. In effect, it is attainment of this principal objective which is to be investigated within the framework of the evaluation; this only seems feasible, however, if the test procedures are implemented in a methodically professional manner. Consequently, safeguarding of the methodical quality of test implementation represents a no less important evaluation objective. As far as the instrument of the practical driving test – systematic behaviour observation – is concerned, the quality criteria to be verified by way of (instrumental) evaluation have already been described in the preceding chapters of this report, along with the

[109] This form is used to record the frequencies of occurrence and observed performance for 20 driving tasks stipulated by EU Directive 2000/56/EC. At the same time, it offers diverse inspiration for the designing of test and audit reports by exemplifying the operationalisation of assessments of the test circumstances (e.g. traffic density, the test environment and road characteristics, weather and lighting conditions, road conditions).

procedures necessary to realise this verification. Attention can thus now be turned to the groups of persons involved, the framework conditions and the requirements to be met by the process and outcome evaluations.

The relevant legal and organisational conditions applicable to the practical driving test, including the institutions involved in test realisation, were already presented in detail by STURZBECHER, BÖNNINGER and RÜDEL (2010). Accordingly, the groups involved in the processes of driver licensing – beside the administrative staff in the responsible authorities – are the driving test examiners from the Technical Examination Centres, the driving instructors, and last but not least the test candidates applying for a driving licence. All these groups, with their mix of common and group-specific interests, must be taken into account by way of a multi-perspective evaluation system. Surveys conducted as part of a process evaluation are the most expedient means to determine the extent to which the practical driving test is realised in adequate quality from the perspective of the candidates and their driving instructors, whereas field studies and experimental analyses are the primary options for instrumental evaluation (see above). The results derived from the different perspectives and methods applied must be compared, weighed up and combined systematically within the framework of the evaluation.

The Technical Examination Centres are mandated to realise the practical driving test and thus bear responsibility for the process quality. Like every modern service provider, they assure the quality of the services offered by way of a comprehensive quality management system, into which both evaluative elements, such as system and product audits, and a complaint management system are integrated (STURZBECHER, BIEDINGER et al., 2010). This quality management system is organised and administered by internal experts ("quality officers"), although the compliance with standards is subject to state control through the "Evaluation Agency for Driving Licence Services" at the Federal Highway Research Institute (see below). As it follows from the aforementioned fundamentals of measure evaluation, the internal corporate quality management system must be kept distinct from the external evaluation system which is to be developed here and subsequently implemented on behalf of the legislator. That applies not only in respect of its content and methodology, but also with regard to work organisation and institutional structure: The contents and methods of the external system are geared to instrumental, result-oriented and summative evaluation; it is to be implemented by scientific institutions. The purpose of the quality management systems of the Technical Examination Centres, on the other hand, is primarily processual and formative evaluation; it is implemented either directly by the quality officers representing internal corporate units or else – as far as aspects of formative instrumental evaluation within the framework of further development of the practical driving test are concerned – by the working group TÜV | DEKRA arge tp 21 as the joint scientific institution of the Technical Examination Centres.

Which legal provisions exist to govern the contents and methods for evaluation of the practical driving test? The first pointer in this direction is to be found in the stipulations of EU Directive 2006/126/EC; these stipulations indicate – in agreement with the observed international practice – that (1) expert audits and instrumental evaluations, (2) analyses of test results and (3) customer surveys are possible evaluation elements[110]:

re 1: According to EU Directive 2006/126/EC, driving test examiners must possess adequate assessment skills for judgement of a licence applicant's competence and ability to drive a motor vehicle safely (EUROPEAN PARLIAMENT & EUROPEAN COUNCIL, 2006, Annex II, paragraph 9.1). These assessment skills include the "ability to observe accurately, monitor, and evaluate overall candidate performance, in particular correct and comprehensive recognition of dangerous situations, accurate determination of the cause and likely effect of such situations, achievement of competence and recognition of errors, uniformity and consistency in assessment…" (Annex IV, paragraph 1.4). The work of driving test examiners is to be monitored and supervised by an appropriate body authorised by the member state, so as to ensure correct and consistent application of the assessment standards (Annex IV, paragraph 4.1.5). "Moreover, the Member States must provide that each examiner is observed conducting tests once every 5 years, for a minimum period cumulatively of at least half a day, allowing the observation of several tests" (Annex IV, paragraph 4.1.3). These stipulations can be interpreted to mean that product audits are required for

[110] It is to be pointed out in this connection that the EU Directive presents only minimum requirements which are considered enforceable by all EU states; it remains possible that more demanding requirements could be deemed desirable from the professional point of view, and it is not excluded that significantly higher demands may already apply in individual EU states.

evaluation of the practical driving test; furthermore, the demand to ensure consistent test assessment suggests that an instrumental evaluation should be performed.

re 2: According to 2006/126/EC, the outcomes of the driving tests conducted by an examiner are to be reviewed periodically (Annex IV, paragraph 4.1.2). It is thus a logical consequence that test evaluation must incorporate analyses of test results.

re 3: Finally, the EU Directive also addresses the process quality of the service "Driving test": The examiner must communicate what the candidate is to expect during the test, and the content, style and language of this communication must be appropriate for the particular target group. He must respond fittingly to questions from the candidate, and must provide clear feedback to explain the test result. All candidates must be treated respectfully and without discrimination (Annex IV, paragraph 1.6). It can be derived from these requirements that a processual evaluation and customer surveys are necessary as the basis for a proper and efficient assessment of compliance.

The expectations of the German legislator with regard to quality management in the Technical Examination Centres and evaluation of the practical driving test were concretised for the first time in the document "Requirements for Operators of Technical Examination Centres" ("Anforderungen an Träger von Technischen Prüfstellen"), which was issued in 2000 by what was then the Accreditation Agency for Driving Licence Services (Akkreditierungsstelle Fahrerlaubniswesen) at the Federal Highway Research Institute (BASt) with reference to § 69 of the Driving Licence Regulations (FeV) in conjunction with §§ 10 and 14 of the Motor Vehicle Traffic Experts Act (Kraftfahrsachverständigengesetz – KfSachvG), and subsequently updated in 2003, 2005, 2007 and 2009 (BASt, 2003, 2005, 2007, 2009).[111] According to § 11 (1a) KfSachVG, the Technical Examination Centres are required to implement quality assurance systems to ensure that tests are realised properly and on the basis of consistent qualification standards, and must furnish corresponding proof to the supervisory authorities. In line with the "Requirements for Operators of Technical Examination Centres" (BASt, 2009), the Technical Examination Centres contribute to the further development and improvement of driving licence testing; in this context, they help to guarantee that driving tests are performed "professionally and in a uniform and necessary quality, taking into account the latest scientific and technical knowledge, the legislative framework and the obligations of professional ethics" (point 3.1). Prerequisites for an evaluation can also be derived from these provisions, together with the more detailed appraisals given below, and indicate once more that – as already elaborated by STURZBECHER, BIEDINGER et al. (2010) on the basis of a comparison between the "Requirements for Operators of Technical Examination Centres" and theoretical foundations for the measurement of service quality (MEFFERT & BRUHN, 2009) – it is necessary to incorporate (1) expert audits and instrumental evaluations, (2) analyses of test results and (3) customer surveys as elements of an evaluation system relating to the practical driving test:

re 1: External audits[112] providing for direct assessment of the driving tests conducted by a particular centre serve general monitoring of the work of the Technical Examination Centres and have already been performed regularly by a team from the so-called "Evaluation Agency" of the BASt for some time; this practice is a continuation of the previous accreditation procedures, which were superseded by the present evaluation system at the beginning of 2010. The scope of an external audit is dependent on the number of practical driving tests conducted by the Technical Examination Centre during the previous year (BASt, 2009, point 7.1). The records of previous internal audits are also reviewed and taken into account as an additional source of data (BASt, 2009, points 2.3.3, 2.5, 7.1). Supplementary to the use of external audits, internal appraisals of the manner in which practical driving tests are conducted are to be performed by internal auditors from within the company or organisation. It is to be noted, however, that neither the external nor the internal audits provide for verification of the consensus (inter-rater reliability) between auditors and driving test examiners, despite the extreme importance of this quality criterion in the context of systematic behaviour observation (see above) and the apparent expedi-

[111] Work on a new version of the "Requirements for Operators of Technical Examination Centres" is currently in progress.

[112] Strictly speaking, these external audits could themselves be termed "evaluations". In the interest of unambiguous terminology and to avoid possible misunderstandings, however, they are nevertheless referred to as "external audits" for the purposes of the present report.

ency of such verification within the framework of a process evaluation.

An explicit demand for verification of the psychometric quality of the practical driving test, which was still raised in the 2005 edition of the "Requirements for Operators of Technical Examination Centres" (BASt, 2005, point 7), is missing in the wording of 2009. This requirement was removed in the course of the process to anchor the specifications of the new standard DIN EN ISO/IEC 17020 ("General criteria for the operation of various types of bodies performing inspection", November 2004), which was henceforth binding for the operators of Technical Examination Centres, in the Driving Licence Regulations. This standard replaced the previously applicable standard DIN EN 45013 ("General criteria for certification bodies providing certification of personnel") and classified the operators as so-called "inspection bodies". Inspection bodies conduct their testing activities on behalf of clients (e.g. customers, authorities) with the objective of supplying those clients with information on the compliance of the inspected circumstances with regulations, standards and specifications. According to DIN EN ISO/IEC 17020, competence, independence and impartiality are essential criteria to be met by inspection bodies (KUNZ & WEINAND, 2012). The removal of the demand for verification of the psychometric quality of the practical driving test is a consequence of the formal transposition of the EU stipulations, but is not attributable to any change in scientific evaluation standards or the quality assurance expectations of the legislator. It can be assumed that the formerly very precisely formulated demand to perform instrumental evaluation is still contained – unchanged – in the aforementioned general requirement to observe relevant scientific standards.

re 2: With regard to analyses of test results, it is specified in the "Requirements for Operators of Technical Examination Centres" (BASt, 2009, point 6.9) that each Technical Examination Centre is to produce statistics which contain at least the results of the driving tests conducted, in each case differentiated according to the types and numbers of tests and the responsible examiners. This is also to include comprehensible and verifiable documentation of the elaboration algorithms and analysis programs used.

re 3: The proposal to conduct customer surveys as an element of the quality policy of the Technical Examination Centres derives from the necessity to take into account demands placed on the practical driving test by the involved customers, and the fact that multi-attributive customer surveys represent the most efficient method for the subjective, attribute-oriented measurement[113] of customer satisfaction (MEFFERT & BRUHN, 2003). Parties involved in the test process (i.e. "customers" in the present context) are, among others, the test candidates (BASt, 2009, point 3.1) and the driving schools (point 6.1). Furthermore, testing is to be scheduled and realised within a reasonable timeframe (point 6.4), and all relevant information is to be made available to the driving schools (point 6.1).

A comparison of the presented national and internal (legal) frameworks, and therein the minimum standards for quality assurance in driver licensing, indicates – in precisely the same way as the analysis of international evaluation practice – that, from a scientific point of view, a multi-perspective, multi-method evaluation system for the optimised practical driving test should comprise four elements, which are to be designed and implemented as complements to the aforementioned internal quality management systems of the Technical Examination Centres. The four elements are:

1. "Instrumental evaluation" (measurement and assessment of the psychometric quality of the systematic behaviour observations which constitute the driving test, in accordance with the specified primary and secondary quality criteria)
2. "Analysis of test results" (critical methodical and content-related statistical analysis of data pertaining to the driving tests conducted)
3. "Customer surveys" (questioning of test candidates and driving instructors to obtain their opinions relating to the process quality of the driving tests conducted)

[113] With regard to the methods used for continuous evaluation of the practical driving test, the emphasis has to date been placed on expert observations, supplemented at certain points by attribute-oriented customer surveys (STURZBECHER, BÖNNINGER & RÜDEL, 2010). For deeper analyses, more complex event-oriented (e.g. critical incident technique) and problem-oriented methods (e.g. complaint analysis) are available to acquire customer expectations and the extent of customer satisfaction (MEFFERT & BRUHN, 2009).

4. “Product audits” (measurement and assessment of the process quality of driving tests by external auditors, especially with regard to the verification of inter-rater reliability).

Insofar as the four aforementioned evaluation elements are actually used for future evaluation of the optimised practical driving test, this would satisfy the initially raised requirement that complex measures must be evaluated by way of multi-dimensional, multi-method and multi-perspective evaluation systems. The prerequisite for development of a corresponding, scientifically founded evaluation concept is an analysis of evaluation practice to date. To determine the status quo with regard to evaluation of the practical driving test and the associated quality management measures, guided interviews were arranged with management representatives and quality officers from the four Technical Examination Centres mandated to conduct driving tests and from the Bundeswehr over the period from October 2010 to February 2011; the responsible head of department and project coordinator from the BASt also took part. These exploratory meetings were used to discuss ideas for the evaluation concept for a future optimised practical driving test; the results have been incorporated into the analyses presented below. With reference to the current evaluation practice, it remains to be noted that, to date, there have been no mentionable scientific studies in the sense of instrumental evaluation, and analyses of test results have addressed merely pass rates. On the other hand, the system of quality management seems well established with regard to the provisions for internal product audits in the Technical Examination Centres, alongside external system and product audits conducted by the BASt (see Chapter 5.4.3). Finally, an inventory of common, scientifically founded and proven methods exists for company-oriented customer surveys (STURZBECHER, BÖNNINGER & RÜDEL, 2010) and offers a good methodical starting point for a nationwide summative evaluation of the quality of the practical driving test (see Chapter 5.4.5).

5.4.2 Instrumental evaluation

Fundamental remarks

Observation-based assessment methods such as the practical driving test may be subject to numerous observation and judgement errors (AMELANG & SCHMIDT-ATZERT, 2006; INGENKAMP & LISSMANN, 2008); this is especially true where a merely loosely standardised, adaptive test strategy is employed in order to raise the contextual significance of the results (STURZBECHER, BÖNNINGER & RÜDEL, 2010). It must thus be investigated whether – assuming methodically professional realisation and assessment – the observations recorded within the framework of an optimised practical driving test, and in turn the judgements and decisions derived from those observations, satisfy the initially described primary and secondary quality criteria.[114] This is to be achieved by way of instrumental evaluation: The ensuing results will indicate whether the optimised practical driving test is suitable to fulfil its control function within the system of novice driver preparation, and thus to contribute to the improvement of road safety.

An instrumental evaluation of the optimised practical driving test comprises both formative questions (i.e. those relating to instrument development) and summative aspects (i.e. consideration of the quality of the instrument in a specific intermediate or final state) (BORTZ & DÖRING, 2006; WOTTAWA & THIERAU, 2003), each of which should be handled by separate evaluation teams (WESTERMANN, 2002). The primary objective of the feasibility study described in Chapter 4 is initial testing of the documentation-relevant elements of the optimised test concept, as operationalised in the electronic test report, and empirically based further development of this concept within the framework of a multi-stage process; this study can thus be deemed a formative evaluation study, and is to be conducted by the TÜV | DEKRA arge tp 21 working group in Dresden. The subsequent revision project, on the other hand, serves to assess the modifications made to the test concept (including the electronic test report) on the basis of the feasibility study, and is thus concerned with given intermediate states of development, or else – insofar as the concept is proved to be robust and requires no further substantial revision – a final state which will presumably offer a foundation for driving licence testing for a longer time to come. The revision studies can thus be viewed as summative instrumental evaluation and should consequently be conducted by an independent scientific institution;

[114] The described studies present a broad spectrum of measures which appear necessary and desirable. A list of priorities must be agreed between the legislative bodies and the Technical Examination Centres to specify whether, to what extent and in which order these studies can actually be realised. Overall, it remains to be noted that the studies serving continuous evaluation – as the name already suggests – can only be started after implementation of the optimised practical driving test. Merely the studies relating to fundamental reliability within the framework of instrumental evaluation, and in particular the verification of inter-rater reliability, should be completed in advance of introduction of the electronic test report (i.e. as components of the revision project).

future repetition of the summative instrumental evaluation studies only appears necessary in case of essential further development of the instrument or potentially quality-relevant changes in the framework conditions for its use. The opinion of an independent scientific institution should be heard to clarify whether or not such a situation has arisen.

With regard to verification of the construct validity, it remains to be noted that significant prerequisites are still lacking; this refers above all to the absence of an elaborated model of driving competence which would permit hypotheses relating to convergent, discriminant and criterion validity to be derived and assigned within a nomological network (see above). This means that construct validation, with its desirable integration of pragmatically oriented notions of criterion validity and theoretical consideration of the subject of the validation, must initially remain a longer-term objective for quality assurance relating to an optimised practical driving test. A first step towards this objective is the description, founding and verification of appropriate external validity criteria, which is to be commenced in the present chapter (see below). The complete description and theoretical founding of driving tasks, observation categories (in the sense of dimensions of driving competence) and the corresponding assessment and decision criteria in the present project establish a good starting point for the processing of this methodical challenge.

Finally, it must be said that it has to date not been possible to conduct substantial studies relating to criterion validity for two reasons: In case of successful completion of the test, the currently applicable regulations do not require documentation of the candidate's test performance; consequently, and also because no meaningful documentation method exists, differentiated test data are and have to date never been available as a basis for validation studies. If the candidate fails the test, on the other hand, the prescribed test report in accordance with Annex 13 to the Examination Guidelines contains an (albeit not necessarily complete) list of the most important errors which led to the negative test decision; at the same time, however, these test candidates are subsequently not in a position to cause accidents or commit traffic offences, because they remain excluded from participation in motorised traffic. The question as to whether and, if so, to what extent such methodical framework conditions and, furthermore, the low degree of variance in the test data of successful candidates have also contributed to the international findings presented in Chapter 5.3, namely that there are in many cases no correlations, or at least not the expected correlations between candidate performance in the practical driving test and the risk of accident involvement or traffic offences in later solo driving (BAUGHAN, 2000; MAYCOCK, 2002), must here remain unanswered. The former hindrance of inadequate differentiation in the recorded test performance of successful candidates would not apply in future for an optimised practical driving test: The new electronic test report and the provisions for multi-level competence-oriented assessments will achieve necessary prerequisites for validity studies (see Chapter 3).

Verification of objectivity and reliability

The legal framework established by the legislator, i.e. the Driving Licence Regulations and the Examination Guidelines, is an important starting point for measures to safeguard objectivity in realisation, assessment and interpretation of the practical driving test. Its stipulations supply and substantiate rudimentary, but nevertheless fundamental methodical standards. On this basis, the individual Technical Examination Centres supply their driving test examiners with work instructions which "contain detailed descriptions of the test procedures, assistance on implementation of the Examination Guidelines and notes on how to establish a conducive test atmosphere" (STURZBECHER, BIEDINGER et al., 2010, p. 86). For all the correspondence in the content-related and methodical origins of these work instructions, however, both a synoptic appraisal of content analyses and the results of exploratory meetings with the Technical Examination Centres (see above) show that, in detail, differences still exist when it comes to implementation of the current legal stipulations in driving test practice. The reason for this phenomenon is to be sought in the fact that the legal provisions are neither able nor intended to fulfil the functions of a psychological process manual: A manual for the process of psychological testing – in contrast to legal regulations such as the Examination Guidelines – would present theoretical and methodical foundations for the process, concrete and unambiguous implementation, assessment and interpretation standards, details relating to the fulfilment of quality criteria and all materials required for realisation of the test method. No such process manual exists to date for the (optimised) practical driving test[115], even though it would be

[115] The process manual proposed here must not be confused with the already drafted "System Manual on Driver Licensing (Practical Test), which is intended to serve all institutions involved in the driver licensing system as a basis for implementation and further development of the practical driving test (operating concept). This manual describes the objectives, the involved parties and the fundamental tasks, processes and re-

desirable as a means to raise the level of standardisation and in turn the objectivity of the test. From the methodical perspective, and for reasons of practicability, it is recommended that a process manual be elaborated in electronic form[116] as replacement for the aforementioned work instructions of the individual Technical Examination Centres.

The objectivity and reliability studies within the framework of the revision project should be based on real practical driving tests rather than on simulated tests. A critical methodical analysis of test simulations would only be appropriate as an early step in method development if the optimised practical driving test were to be founded on predominantly new demand, implementation, assessment and decision standards. This is not the case, however: The driving tasks, the observation categories and the assessment and decision criteria continue to reflect the present legal framework and test practice to date; they have merely been streamlined and restructured (STURZBECHER, BÖNNINGER & RÜDEL, 2010). The only substantial changes are the introduction of an additional competence-oriented assessment of test performance and the use of an electronic test report. In this respect, the optimised practical driving test extends the current scope of test assessment and documentation[117]; on the other hand, it possesses no novelty value with regard to its fundamental contents or basic methodical architecture. Furthermore, the practical usability of the electronic test report is to be verified by way of a feasibility study before commencement of the revision project (see Chapter 4). There is thus no necessity to conduct simulated tests as a means to estimate reliability, especially as it must be feared that validity will be distorted compared to a real test (e.g. due to the reduced performance motivation on the part of the driver and the absence of test anxiety).

Which empirical findings are available for the assessment of inter-rater reliability in the context of judgements of driving behaviour? For the so-called "Vienna Driving Test" ("Wiener Fahrprobe"), which was constructed to assess driver aptitude, RISSER and BRANDSTÄTTER (1985) report an inter-rater reliability of 67 per cent; unfortunately, there is no indication of the calculation method used. BÉDARD, PARKKARI, WEAVER, RIENDEAU and DAHLQUIST (2010) investigated the consensus between two observers with regard to the numbers of errors detected in recorded driving simulator sequences; they calculated a correlation coefficient of .79 (Pearson) or .73 (ICC). The stability coefficient for an interval of approx. one month between test and retest was .83 (Pearson) or .76 (ICC). Against this background, and taking into account the typical reliability coefficients for systematic behaviour observation which were reported in Chapter 5.1.2, a minimum reliability value of .70 would represent an acceptable result for event-oriented assessments (errors and above-average performance) within the framework of the optimised practical driving test. With regard to the competence-oriented assessments for the observation categories and driving tasks, the test should strive for a minimum reliability of .80, whereas consensus in respect of the test decision should not lie below .90.

Both the consensus between different driving test examiners and the stability of the observations and decisions of one and the same examiner are to be investigated within the framework of a subsequent revision project. The implementation of the revision project is to begin with a course of experimenter training or further training[118] for at least 60 experi-

sponsibilities in connection with implementation of the practical driving test from the institutional perspective.

[116] The TÜV ĐEKRA arge tp 21 program for driving test administration and documentation, in combination with the working software for the driving test examiners (e.g. an electronic test report), already incorporates essential contents of a process manual (e.g. demand standards, observation categories, assessment and decision criteria, results of quality audits); these contents are furthermore constantly updated and brought to the attention of those involved in the process. It would be a relatively simple matter to expand this program to include the still missing contents of typical process manuals (e.g. explicit implementation instructions, assessment aids); simple access to the then uniform and holistic electronic control document for all Technical Examination Centres would be a desirable consequence of this step.

[117] In both the current Examination Guidelines and the BASt requirements, a distinction is made between a "record" of a test, i.e. the continuous notes taken by the driving test examiner, and a "test report" in accordance with Annex 13 to the guidelines. The term "test report", as it is used in the system of driving test legislation, is somewhat misleading from the scientific-methodical perspective: Annex 13 is not a true "report" or test documentation in the methodical sense, but rather a confirmatory list of selected serious errors, which is drawn up subsequently in case of a negative test decision and serves to provide legally sound justification for that decision in case of conflict with the unsuccessful candidate. The so-called "records" mentioned in the Examination Guidelines (PrüfRiLi 6), on the other hand, can be viewed as the actual methodical test documentation with all its possible beneficial functions. To date, however, such records have remained of an informal nature; as no legal stipulations exist, each examiner elaborates and makes use of such records according to his own needs and preferences.

[118] It can be assumed that, within the framework of this revision project, at least 12 examiners from each of the four mandated Technical Examination Centres and from the Bundeswehr will attend training courses lasting at least three days on realisation of the critical methodical revision studies. The training should be organised in at least four course groups with a maximum 15 participants each, and should be conducted by the same training team in each case, so as to ensure a uniform quality of

enced driving test examiners, who are then to conduct tests in accordance with the concept for the optimised practical driving test over a period of three months. The two necessary reliability studies are embedded in the experimenter training and the trials of the revision project. For the first study, the driving test examiners assess the performances displayed by candidates in real tests, whereas the second study makes use of video recordings. It was already mentioned by HAMPEL (1977) that the required observation and assessment competence for the practical driving test must be trained regularly by way of examples: "A comparison of the final judgements of examiners does not necessarily provide an adequate indication of the reliability of those judgements, and must be supplemented with a qualitative comparison to obtain information on the manner in which the judgements were effected. These comparisons should be performed regularly and discussed with the examiners within the framework of training courses" (ibid., p. 85).

In the first reliability study, the inter-rater reliability is to be determined by comparing the assessments of at least 30 examiner tandems (at least 6 tandems from each Technical Examination Centre and from the Bundeswehr) in the course of real tests conducted during the first week (first measurement) and the twelfth week (second measurement) of the trial implementation. Planning of the test drive, including the communication of driving instructions for the candidate, is placed in the hands of the examiner who is seated behind the driving instructor – in accordance with the usual practice. For the second measurements, the seating constellation of the two examiners, and thus their active or passive role for test realisation, is to be reversed for half of the tandems. The total of at least 60 report pairs from the two measurements with each tandem are then to be subjected to statistical analysis with regard to the consensus in their test decisions (evaluation level 1), competence-oriented assessments (evaluation level 2), event-oriented assessments (evaluation level 3) and notes on the framework conditions under which the test was conducted (evaluation level 4, e.g. range of speed limits, traffic density, weather conditions, lighting conditions); in addition, the objectivity could be checked by comparing the times at which individual assessments were recorded. Furthermore, the individual assessments of the necessity to terminate a test prematurely should be compared whenever this happens. Generally speaking, it seems desirable that the inter-rater reliability should improve on all evaluation levels over the course of practical implementation, i.e. between the first and second measurements, due to the increasing experience and practice of the participating examiners.

When the inter-rater reliability is determined in real test situations, the two examiners forming a tandem sit in different positions in the test vehicle and thus observe the candidate's performance from a slightly different perspective. Furthermore, the examiner who is responsible for test planning and thus specifies the immediate test demands will presumably possess more concrete performance expectations than the examiner who is merely observing the test and has no access to the background thought processes of his colleague: It could be said that the active examiner holds an information advantage, which will almost certainly influence his perception control (e.g. focus of attention). Both phenomena will probably reduce the level of consensus in the observations made during a real test. These limitations could be avoided or controlled if test drives were to be recorded on video from the perspective of the active examiner[119] and subsequently presented to driving test examiners for assessment; the correspondingly higher degree of standardisation in the observation conditions for the two observers should lead to a considerable improvement in inter-rater reliability, at least with reference to the event-oriented assessments. At the same time, this method would lift the restriction to two examiners per observer group, which is otherwise dictated by the space available in the test vehicle. On the other hand, the assessment of test videos would probably entail diminished ecological validity compared to the participation in real tests (BAUGHAN et al., 2005; STURZBECHER, BÖNNINGER & RÜDEL, 2010): The aforementioned information advantage of the active examiner and the additional opportunities for information acquisition during a real drive (e.g. a

training and a high intensity of practical exercises. Following successful completion of the courses, a pool of at least 60 driving test examiners will be available to realise the revision studies in a selected organisational unit of each Technical Examination Centre over a period of at least three months. One-day further training seminars are planned after one month and at the end of the practical study period, not only to gather information on the examiners' experiences, but also to enable assessment of the implementation quality. Given the planned course contents, the five days of training should be counted towards the legally stipulated scope of further training for driving test examiners.

[119] Inspiration for methods to record driving tests in this way can be drawn from the experience with video-based documentation of the practical driving test in Estonia and Latvia, from the system of novice driver preparation in Israel (LOTAN & TOLEDO, 2006), and from the SAF project in which vehicle control data were logged electronically and combined with video sequences of the driver and the surrounding traffic situation (SMUC, CHRIST & GATSCHA, 2002).

change of seating position), for example, are validity-enhancing framework conditions which are always available to the examiner in real-life practice.

The recourse to video recordings to support studies of the reliability and stability of observer assessments also facilitates monitoring of the demand levels of individual tests and permits systematic analysis of the influence of the level of the candidate's performance on the various aspects of reliability. It seems desirable for the assessment criteria of the optimised practical driving test to enable differentiated assessment of the driving behaviour displayed by a particular candidate against the full possible spectrum of candidate driving competence – as a prerequisite for valid performance judgements and test evaluation. To be able to investigate this aspect, the test videos used for the reliability studies should incorporate multiple instances of all relevant driving tasks, preferably with varying demands levels, and should thereby reflect different levels of performance for the completion of the tasks. These performance examples must take up the proposed four-level assessment scale with grades of "Very good", "Good", "Sufficient" and "Inadequate" for assessment of the five areas of competence dimensions (observation categories) "Traffic observation", "Vehicle positioning", "Speed adaptation", "Communication" and "Vehicle control/Environment-aware driving", i.e. at least four different test videos are required for the reliability studies within the framework of the revision project[120]:

- In the first film, the candidate should act correctly, efficiently and with foresight with reference to almost all driving tasks and areas of competence ("Very good").
- In the second film, the candidate should act correctly, efficiently and with foresight in most of a diversity of driving tasks and with reference to the overwhelming majority of the areas of competence; simple errors represent an exception ("Good").
- In the third film, the candidate should act correctly, efficiently and with foresight only in standard situations relating to the majority of the areas of competence; simple errors are observed in complex or unfamiliar situations ("Sufficient").
- In the fourth film, the candidate should fail to act correctly, efficiently and with foresight even in standard traffic situations relating to the majority of the areas of competence; serious errors and/or multiple or repeated simple errors are observed ("Inadequate").

As already implied, therefore, the second reliability study is to assess inter-rater reliability with the aid of video recordings of previously conducted tests, in other words the four "test videos" described above. Following adequate practising of the required observation and assessment techniques with the aid of corresponding exercise videos[121], the test videos are to be presented to the maximum of 15 participants at each of presumably 4 three-day experimenter training courses (first measurement); such assessments are to be repeated on two further occasions, firstly during the fifth week of the revision project trials (second measurement) and then shortly after the end of the project period (third measurement). As in the case of the real tests, the consensus between the individual observers is to be determined with reference to all four evaluation levels; here again, it is expected that the reliability coefficient will increase from one measurement to the next.

The stability of the observations and judgements recorded by a single driving test examiner, as the second aspect of reliability relating to the optimised practical driving test, can only be determined on the basis of video recordings of real tests; in practice, after all, a real driving test can never be repeated under exactly identical conditions. The aforementioned second reliability study, with series of three video-based measurements to determine inter-rater reliability (see above), therefore, is also to be used to determine the stability coefficient. Previous stability studies indicate that the chosen intervals between the three measurements (one month between first and second measurement; two months between second and third measurement) are appropriate to minimise memory effects on the part of the observers: BÉDARD et al. (2010) specified a minimum interval of one month for their driving simulator study (see above) to measure assessment stability; STURZBECHER (2004) chose an interval of eight weeks for reliability studies relating to the observation system for "Pedagogically Qualified Driving School Monitoring"

[120] It is naturally necessary to produce a diversity of further videos meeting these requirements for use as exercises during the experimenter training and in the further training of driving test examiners; here, too, further test conditions should also be varied (e.g. traffic density, weather conditions, lighting conditions).

[121] For validity reasons, both the test videos and the exercise videos should preferably be recordings of real tests. To this end, it is necessary to establish the technical and legal prerequisites (e.g. demands relating to data privacy) for the recording of test drives at a test location whose road and traffic circumstances yield a broad diversity of driving tasks. This must include also permission from the candidate and the driving instructor to record the drive and to use the resulting video.

("Pädagogisch qualifizierten Fahrschulüberwachung", PQFÜ). The stability of the assessments of test performance and test conditions is also to be determined for all four evaluation levels.

If the optimised practical driving test were to meet all the applicable demands relating to objectivity and reliability, and assuming that all tests are realised under similar framework conditions, i.e. the various driving tasks are performed in comparable demand situations, then a test candidate should always receive the same competence-oriented assessments and the same test result, irrespective of the driving test examiner and Technical Examination Centre conducting the test. Such consensus is less likely with regard to the event-oriented assessments, as the opportunities to commit errors or to display above-average performance are strongly dependent on the concrete test demands, the only conditionally controllable framework conditions (e.g. traffic density, weather conditions) and not least the uncontrollable behaviour of other road users. To determine the reliability of the optimised practical driving test, therefore, a third reliability study is to be conducted in which 30 test candidates complete two driving tests with different examiners from different test organisations on subsequent days; the relatively short interval between the tests excludes fundamental learning effects in the sense of improved driving competence, but is at the same time sufficient to avoid possible overburdening of the candidate. To eliminate training effects, the two test drives should not use the same test route (HAMPEL, 1977); the two tests or test routes should nevertheless be planned as similarly as possible with regard to the numbers and types of driving tasks, the level of the demands placed on the candidate, and the framework conditions. These necessary study conditions (in particular a diversity of test routes with similar infrastructure conditions and two Technical Examination Centres serving a single test location) are most easily met at the test location Berlin. As was already the case for the study by BAUGHAN and SIMPSON (1999), the results of the first test should remain unknown to both the examiners and the candidates until after the second test; overall, the test is still to be deemed "passed" if the candidate completes only one of the test drives successfully. An advance study should ensure that the inter-rater reliability of the examiner tandem concerned (i.e. the two examiners who conduct the test for one and the same candidate) is adequately high. Insofar as the study reveals deviations in the competence assessments and test decisions, they are to be interpreted against the background of the effect size and significance findings from the other reliability studies.

Finally, it should be analysed within the framework of the reliability studies, whether changes to the test duration and the amount of actual driving time influence inter-rater reliability and stability, above all with regard to the competence-oriented assessments and test result. Such a study would at the same time focus on aspects of the validity and economy of the optimised practical driving test: If extension of the driving time were to produce a significant gain in inter-rater reliability, this would also promise a substantial boost to the validity of the test; at the same time, however, it would lead to an increase in the costs of testing and reduced economy. The absence of an increase in consensus, but nevertheless high stability, by contrast, would favour a short duration for the test drive and could be seen as economically beneficial.

Which durations should be tested in the course of such a study? In their analysis of test implementations in 44 countries, GENSCHOW, STURZBECHER and WILLMES-LENZ (2014) found that the shortest driving time in any country was 20 minutes, and the longest 60 minutes; Germany[122] belongs to a relatively large group of countries in which the test drive lasts between 20 and 25 minutes (mostly 25 minutes), while six countries provide for driving times of 40 minutes or more. These findings suggest that two examiners should each assess the candidate's test performance and reach a (provisional) test decision after 25, 35, 45 and 60 minutes. Statistical analysis of these data must then determine whether and to what extent

- inter-rater reliability increases with the driving time,
- the competence-oriented assessments of test performance change, and
- test decisions reached after a certain driving time are revised on the basis of later observations.

Verification of content validity

The content validity of a method is generally confirmed or refuted by way of an expert rating (see

[122] The driving time specifications currently applicable in Germany are based on expert recommendations, but also take in account economic considerations (e.g. reasonableness of the test fees for the candidate); they at the same time satisfy the stipulation of EU Directive 2006/126/EC, Annex II, Section 10. Within the framework of the pending quality studies relating to an optimised practical driving test, the current specifications should be re-examined from the critical methodical perspective. The objective must be to find a meaningful balance between test validity and the test costs associated with a longer or shorter test drive.

above). With regard to the optimised practical driving test, the experts to be questioned on test quality are the driving test examiners and driving instructors, legal experts, and scientists with expertise in educational and test psychology. It seems equally expedient, in addition, for experts from the aforementioned fields and professions to be involved already in the development of a content-related and methodical concept. This was the approach taken in the case of the optimised practical driving test, and can be witnessed as a consistent practice from the elaboration of initial foundations (STURZBECHER, BÖNNINGER & RÜDEL, 2010), via the definition of demand standards in the working group "Driving tasks" (see Chapter 3), through to contributions to the present report. Following publication of the present draft concept for an optimised practical driving test, realisation of the feasibility and revision studies, and appropriate empirically based adaptation of the original draft, the revised demand and implementation standards (above all the driving tasks, the observation categories, and the assessment and decision criteria) should then be presented to a circle of experts for their appraisal; the degree of assessment consensus could be quantified with the aid of Fricke's agreement coefficient (AMELANG & SCHMIDT-ATZERT, 2006).

Verification of criterion validity

Evaluation studies on criterion validity can be deemed particularly valuable as confirmation of the methodical quality of the optimised practical driving test and as a basis for its methodical further development (HAMPEL & STURZBECHER, 2010), provided appropriate external criteria are used. Therefore, it must now be discussed whether and, if so, to what extent the factors listed below are suitable as possible criteria for meaningful validity studies:

(1) Traffic-related conspicuities (frequency of accident involvement, frequency of traffic offences, frequency of damage claims submitted to insurance companies)

(2) Driving behaviour data from in-vehicle data recorders

(3) Driving behaviour data from driving simulators

(4) Results of traffic and hazard perception tests

(5) Results of observations of driving behaviour by traffic psychologists.

re 1: One possibility to determine the predictive validity of the optimised practical driving test is to investigate the possible correlations between the test result, on the one hand, and the frequency of accident involvement or traffic offences during later solo driving[123], on the other hand, though it cannot always be rightly assumed that accidents or traffic offences are attributable to driving errors. At first sight, figures on accident involvement and traffic offences possess high face validity as external validity criteria; upon closer scrutiny, however, certain doubts arise as to their suitability for this purpose (BAUGHAN, GREGERSEN, HENDRIX & KESKINEN, 2005; BERG, KIESCHKE & SCHUBERT, 2008; MACDONALD, 1988; RISSER and BRANDSTÄTTER, 1985):

- The driving behaviour displayed during a driving test ("maximally adapted behaviour") may deviate significantly from the same person's behaviour when driving solo ("typical behaviour"). The reasons could lie in various behaviour-relevant factors, such as the motivation to comply with traffic rules, the mental and physical state of the driver or other distracting circumstances, which may differ significantly between a test situation and the "natural" situation of solo driving (MAYCOCK, 2002).
- Traffic accidents occur relatively seldom; as a result, any studies must address a very large sample of drivers in order to be able to consider an adequate number of accidents.
- The causes of traffic accidents are manifold, and the underlying effects are cumulative. Improper behaviour alone does not necessarily lead to an accident; their occurrence is to a certain extent also a product of chance situations and the behaviour of other road users, in that they either amplify or at least fail to compensate the improper behaviour on the part of the driver who is seen to have caused the accident. It is furthermore difficult to obtain valid statistics on near-accidents. It must also be mentioned, finally, that poor drivers often drive less than good drivers.
- From the methodical perspective, the recording of accident causes by the police can be deemed indistinct (KAISER, 2002)[124]; the same can also be pre-

[123] Many observers use the term "legal proving" in this context; this criminal law concept is to be considered misleading for the present purposes, however, as it implies that no further offences will be committed after serving an imposed penalty.

[124] The "Traffic Accident Notice" used by the police to record accidents is based on a fixed catalogue of causes which was introduced in 1975; this leaves practically no scope for conclu-

sumed for the recording of traffic offences. Deeper clarification of the causes of accidents could be achieved by additionally surveying and analysing the statements of those involved. Studies of this kind, however, are extremely difficult to implement (MAYCOCK, 2002). The likelihood of success can also be doubted, as such statements will often be seen to entail a risk of self-incrimination in the question of fault, which could in turn influence the probability of legal sanctions or insurance claims. The same problems would probably apply with regard to accident statistics obtained from the insurance companies.

- Alongside the lack of distinction in the recording of accident causes, it must also be taken into account that a large proportion of those traffic accidents which result in only minor damage will not be reported to the police (STATISTISCHES BUNDESAMT, 2010a), and that the true figures – as in the case of traffic offences – are considerably higher by an unknown amount. Data pertaining to offences and accidents which have not been recorded by the police can only be obtained by way of direct driver surveys. For their evaluations of the nationwide pilot implementation of the training model "Accompanied driving from age 17", for example, STIENSMEIER-PELSTER (2005) and FUNK and GRÜNINGER (2010) analysed a combination of "Traffic offences and accident involvement according to police statistics" and "Self-reported traffic offences and accident involvement", as a means to compensate the weaknesses of police statistics as a basis for validation studies. FUNK and GRÜNINGER (2010) conclude, however, that even the relatively large sample of over 3,700 survey participants was still too small to constitute "a representative reflection of the 'rare event' of accident involvement on the part of participants in the model" (p. 297). Even if the sample size for such studies were to be increased, their success would remain questionable, since CHAPMAN and UNDERWOOD (2000) were able to show that, within a period of only two weeks, drivers had already forgotten around 80 per cent of the minor (or near-miss) incidents in which they had been involved.
- Finally, the significance of relatively limited driving competence as a factor influencing the frequency of accident involvement in solo driving after the driving test is very probably dependent on how intensively the novice driver is accompanied by experienced, safety-conscious drivers and – possibly as a direct consequence thereof – the extent to which he is exposed to corresponding risks. These influencing factors, at least, need to be controlled in a validation study which uses the frequency of accident involvement and traffic offences as an external criterion.

Despite the aforementioned methodical limitations, it may be expedient – after a reasonable start-up phase of at least two years following implementation of the optimised practical driving test – to investigate whether the now differentiated test assessments correlate with figures on accident involvement and traffic offences during the initial phase of solo driving, and whether such correlations – if found – are influenced by the protective conditions applicable during this phase.[125] To this end, an appropriately large sample of driver records must be retrieved from the Central Register of Driving Licences (ZFER) held at the Federal Motor Transport Authority (KBA). To facilitate such studies, a revision of the so-called "Traffic Accident Notice" used by the police to record accidents would be desirable (see above): Ideally, the accident causes could be defined in accordance with the driving task and driv-

sions to be drawn on the specific traffic and driving competence deficits of the drivers involved in accidents. To date, a distinction has been made between general causes (e.g. road conditions, influence of the weather, obstacles), which contribute to the accident, but cannot be assigned to individual persons, and aspects of improper personal behaviour (e.g. failure to observe right of way, inappropriate speed, etc.), which can be assigned to a particular driver or pedestrian. Up to two general causes can be specified for each accident. Alongside, up to three specifications are permitted for the principally responsible person and one further involved party, which means that up to eight causes can be entered in total for each accident (STATISTISCHES BUNDESAMT, 2010a, p. 34).

[125] If it were to be possible to link such validation studies with analyses of the effects of other measures serving novice driver preparation, then even the considerable additional expense of supplementary surveys, as they were conducted for summative evaluation of the training model "Accompanied Driving from Age 17" (WILLMES-LENZ, PRÜCHER & GROSSMANN, 2010), could be justified.

ing error categories used in the proposed catalogue of driving tasks (see Annex 1 to the present report).

The external criterion of novice driver accident involvement can be considered important not only from the perspective of instrumental validation in the narrower sense, but also in its function as (politically desired) proof that optimisation of the practical driving test does indeed lead to an improvement in road safety. The furnishing of such proof appears difficult, however – as is always the case where impact analyses must be performed without randomly selected or parallelised experimental and control groups: Even if accident figures drop after introduction of the optimised practical driving test, it cannot be excluded that other factors or measures have produced or magnified this effect, either solely or in combination. For interpretation of an apparently achieved effect, therefore, it must be asked whether all other factors which could conceivably have contributed to a reduction in the number of accidents have been excluded or quantified. In practice, this will hardly be possible to a satisfactory extent, not least because not all influencing factors are known and measurable.

re 2: As an alternative or supplementary methodical approach to the – probably costly – determination of predictive validity for the optimised practical driving test, it could be useful to consider the possibilities of concurrent validity (see above). For the present case, this means that, in immediate time proximity to realisation of the practical driving test, which is to be classified an observation method, robust information on the candidate's level of driving competence must be acquired either by methodically different means (e.g. vehicle data recorder, driving simulator test, traffic perception test) or by way of a methodically similar concurrent procedure (i.e. an observation method other than psychological driving behaviour observation), and the corresponding findings then compared with the test result. Ideally, in other words where both the method selected for comparison and the method to be verified are valid for the measurement of driving competence, the performances and any differences in the performances of individual candidates must be assessed consistently by all methods. Such opportunities for concurrent validation of the optimised practical driving test are to be examined in the following; the first possibility to be considered is the recording and evaluation of competence-relevant driving data by way of a so-called "in-vehicle data recorder".

Modern in-vehicle data recorders permit the technical acquisition and scientific analysis of critical events occurring during a drive, and in doing so provide for the automatic, computer-assisted identification of different driving manoeuvres on the basis of pattern recognition algorithms and with the aid of gyroscopic sensors. Some systems also make use of interfaces to an internal communication network installed in the vehicle and can in this way enable direct online access to the recorded data. One methodical challenge for the studies is a confounding influence on the data relating to driving competence from the use of a data recorder itself: It could be shown in several studies, for example, that the driving performance of novice drivers improved significantly when an in-vehicle data recorder was used (MUSICANT, LOTAN & TOLEDO, 2007; PRATO, TOLEDO, LOTAN & TAUBMAN-BEN-ARI, 2010; SCHNEIDER, 2008). This seems by all means desirable from the perspective of road safety, and is at the same time not necessarily a disadvantage for validity studies, as additional performance incentives are present both during the driving test and later (recorded) drives, which serves to enhance the similarity of driving conditions.

With regard to data analysis, one important proviso remains to be mentioned: Even if in-vehicle data recorders – insofar as they are integrated with camera and GPS systems and can thus establish relationships between different driving tasks, driving behaviour and the prevailing traffic situation – are able to supply a diversity of information relating to vehicle control, steering response, speed adaptation, vehicle positioning and driver communication, the unambiguous assignment of these data to specific test-relevant driving tasks or manoeuvres, and meaningful referencing to behaviour standards relating to road safety are still essentially unsolved challenges.

re 3: Modern driving simulators offer diverse options and possibilities for standardised measurements of driving behaviour (ENGIN, KOCHERSCHEID, FELDMANN & RUDINGER, 2010) and can thus probably be used also in validation studies relating to the optimised practical driving test. Driving simulations permit different traffic situations and driving tasks to be presented in a realistic manner[126], and the behaviour which must be displayed to meet the demands set in the simulation is the same as that required in real traffic. Current studies also provide corresponding evidence for the equivalence of the behaviour displayed by drivers in driving simulators and in real traffic (BÉDARD et al., 2010; DE WINTER, DE GROOT, MULDER, WIERINGA, DANKELMAN & MULDER, 2009). Although studies with driving simulators have in the past served mainly to assess the driving competence of elderly and physically or mentally disabled persons, individual studies indicate that a driving simulator could well be used to determine typical elements of so-called "novice risk", for example deficits in respect of hazard perception and evaluation (i.e. traffic observation) or vehicle control (DE WINTER et al., 2009). It should be a relatively simple matter, furthermore, to operationalise the other dimensions of competence to be assessed by the optimised practical driving test (speed adaptation, vehicle positioning and communication) as test tasks in a driving simulator. Even high-risk situations could be simulated within the framework of such tasks, without exposing the candidate to a real risk of accident (e.g. overtaking manoeuvres with oncoming traffic). At first sight, therefore, an assessment of driving competence derived from a test drive in a driving simulator appears suitable as an external validity criterion for competence assessments by way of an optimised practical driving test in real traffic. As a limitation, it remains to be noted that, to date, even with use of the latest computer technologies, it has not yet been possible to overcome all problems of authenticity; certain experiential and above all perception-related differences exist between reality and simulation, which in turn detracts from the validity of the validity criterion (and the applicability of competence findings from driving simulators as an indicator of driving competence in real traffic). It is at the same time necessary to take into account the phenomenon of "simulator adaptation syndrome" (SAS) or "simulator sickness" (the physiological discomfort experienced by many users in the form of nausea, headache or disorientation) when pursuing a validation strategy which includes the use of driving simulators (HOFFMANN & BULD, 2006). Particular hazard potential can be associated with the fact that the re-adjustment to a real vehicle after use of a driving simulator may induce a renewed bout of SAS: It is for this reason, for example, that airline pilots are not permitted to fly for at least a week after taking part in simulator training.

re 4: Traffic perception tests (or "hazard perception tests" as they are often termed) also hold potential for validation of the optimised practical driving test. This innovative form of testing is currently only used in a small number of countries (e.g. in Great Britain, the Netherlands and some Australian states (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014), but nevertheless offers particularly good possibilities for assessment of the competence aspect "traffic observation" and abilities relating to timely traffic perception and hazard recognition.[127] The predominant test method is to demand a correct reaction or the correct "driving decision" in the displayed scenario; at the same time, non-verbal response is also measured (e.g. the reaction time before a computer input). The computer is thus the essential medium for task presentation and processing in traffic perception tests; the spectrum of possible test items ranges from the identification of safety-relevant hazard cues, via the observance of appropriate speeds and

[126] A high degree of realism is achieved by way of a 360-degree presentation system, a complex dynamic motion system and use of a true-to-life vehicle mock-up.

[127] A few examples here serve to illustrate the diversity of possible test contents: In Great Britain, the test comprises 14 one-minute video sequences in which a drive is presented from the driver perspective; as soon as the candidate recognises a hazardous situation, this must be indicated by clicking with the mouse. In the Netherlands, the test items are presented in the form of photographs of traffic situations depicted from the driver perspective (with information in the mirrors and with turn indicators and speedometer visible); the task for the candidate is to determine the appropriate reaction in the given situation: "Apply the brakes", "Take foot off the accelerator" or "Do nothing". In Victoria (Australia), the candidate is shown a total of 28 video sequences of traffic situations which often lead to novice-typical accidents, each presented from the driver perspective. Before each video sequence, the required driving action is specified (slow down, overtake, turn or move off). The candidate must then click with the mouse to indicate when this required action can be performed safely during the given driving scenario.

safe distances, through to the clarification, comparison and selection of available action alternatives. Traffic perception tests thus close a gap which exists in the standardised testing of (implicit) action knowledge with regard to different elements of driving competence (RÜDEL, STURZBECHER, GENSCHOW & WEISSE, 2011). For the external validation of the optimised practical driving test, this means that candidates with good test performances in respect of traffic observation, vehicle positioning and speed adaptation will probably achieve similarly good results in an adequately constructed traffic perception test (DEBUS, LEUTNER, BRÜNKEN, SKOTTKE & BIERMANN, 2008; STURZBECHER & KALTENBAEK, 2012).

re 5: As the optimised practical driving test – methodically speaking – is an instance of (systematic) driving behaviour observation, it seems logical to direct a search for external validity criteria to other scientifically founded concurrent processes of driving behaviour observation and driving competence assessment. In this context, attention is immediately drawn to the best-known procedures for standardised driving behaviour observation in the German-speaking regions, namely the "Cologne Driver Behaviour Test" ("Kölner Fahrverhaltenstest") by KROJ and PFEIFFER (1973) and the "Vienna Driving Test" ("Wiener Fahrprobe") by RISSER and BRANDSTÄTTER (1985). In both cases, standardisation is achieved by defining a fixed driving route with precisely specified observation points and behaviour sequences; although originally devised for the cities of Cologne and Vienna, respectively, the inherent situation classifications permit both methods to be transferred also to other locations. Driving behaviour is recorded on a detailed report sheet. In addition to the aforementioned observation methods, the operators of driver assessment centres have apparently realised various other psychological behaviour observations to support their examinations of fitness to drive; however, little has ever been published to explain their implementation and assessment standards or to present the findings of studies relating to the psychometric quality of such observations (SCHUBERT, SCHNEIDER, EISENMENGER & STEPHAN, 2005).[128]

In view of the fact that the authors of the present report had no access to concrete process manuals for either the Cologne Driver Behaviour Test or the Vienna Driving Test, it can here only be considered whether the validity of these two methods – as a prerequisite for their use as validity criteria – appears to be guaranteed. KROJ and PFEIFFER (1973) tested the criterion validity of the Cologne Driver Behaviour Test on the basis of the criteria "accident index" (operationalised as the number of accidents in which a driver is involved relative to the distance driven) and "offence index" (operationalised as the number of traffic offences committed by a driver relative to the distance driven): Test persons who displayed poor performance in the Cologne Driver Behaviour Test in respect of their adaptation to traffic flow and safe distances to other road users were found to possess a higher accident index than persons who displayed better performance in these two categories. Where test persons displayed poor performance in the categories "Traffic observation" (for which the authors used the designation "Safeguarding"), "Speed selection" and "Hesitation" (hindering others through excessive hesitation at crossroads and junctions or when changing lanes), furthermore, a significantly verified correlation with the offence index was determined.

Two strategies were followed to verify the criterion validity of the Vienna Driving Test[129] (RISSER & BRANDSTÄTTER,

[128] SCHUBERT et al. (2005) commented this (from a scientific perspective) surprising lack of transparency as follows: "On the other hand, one intention of liberalisation is that the providers on the market should compete with each other through the development of new and ever better methods. The current situation of a barely manageable diversity of methods for the observation of driving behaviour, which in many cases – for by all means honourable reasons of competition – are at the same time held secret or else published only with a high degree of abstraction, is certainly not satisfactory. The task is now to find constructive and solidary ways out of this difficult situation within the framework of the obligatory meetings of operators of driver assessment centres which are held at the Federal Highway Research Institute in accordance with the Driving Licence Regulations" (p. 64).

[129] In the case of the Vienna Driving Test (RISSER & BRANDSTÄTTER, 1985; CHALOUPKA & RISSER, 1995; KAUFMANN & RISSER, 2007), driving behaviour is documented by two observers: One takes free notes, while the other completes a standardised behaviour report on the basis of approx. 80 variables. The free observer records general aspects of behaviour such as driving errors (in the sense of serious traffic offences or endangerment), interaction and communication processes

1985): Firstly, interviews were conducted with the participating test persons to obtain information on their involvement in road accidents over the past ten years and any traffic offences committed during the last five years. At the same time, insurance companies were contacted with a request to send details of the planned study to customers with particularly few or particularly numerous accidents and to ask them to participate. The correlation between the number of errors recorded during the driving test and the driver's accident history was significant, although the effect size was rather small (r=.14). Further or more detailed study results on criterion validity were not reported by the authors.

Psychological driving behaviour observations are used within the framework of medical-psychological assessments of fitness to drive to clarify possibilities for the compensation of individual performance impairments. The operator of an assessment centre specifies the cases in which such psychological driving behaviour observations are performed in a corresponding manual, which must be approved by the Federal Highway Research Institute (BASt). Participation in a psychological driving behaviour observation, which lasts approximately one hour, is voluntary; the observation session is conducted by a traffic psychologist, in the presence of a driving instructor, and takes place in real traffic in a driving school vehicle. “The driving instructions, the behaviour of the driving instructor and observer during the observation, and the assessment of the displayed performance are to be standardised as far as possible” (BASt, 2009, p. 20). According to SCHUBERT and WAGNER (2003), psychological driving behaviour observation is a suitable means to assess both the rule-based (e.g. overtaking, scanning for road signs, acceleration) and skill-based action control levels (e.g. gear-changing, use of the marked road lanes, turning at junctions, traffic observation, response to a speed limit). To this end, the driver's performance is recorded for each of the four content dimensions (observation categories) “Orientation”, “Concentration and alertness”, “Risk-related self-control” and “Action reliability” at a minimum of 10 observation points; driver performance in respect of a fifth dimension “Resilience” is then calculated from the observation data.

SCHUBERT and WAGNER (2003) provide definitions and specify overt behaviour indicators for methodical “anchoring” of the dimensions to be observed. “Concentration and alertness”, for example, are defined as the ability to scan purposefully for relevant information (possible behaviour indicators are the recognising of changed conditions, such as the lifting of a speed limit, or the correct realisation of monitoring tasks involving the use of mirrors and shoulder checks); “Action reliability”, on the other hand, is manifested in the ability to handle a vehicle competently (e.g. driving within a marked lane, timely reaction/braking). Placed alongside the methodical architecture of the optimised practical driving test, it is conspicuous that the latter possesses certain dimensions or observation categories which are similar in content, namely “Traffic observation” and “Vehicle control”, but that the content design also deviates to a certain extent elsewhere: The aspects of vehicle positioning and speed adaptation, for example, are assigned to the dimension of “Action reliability” in combination with elements of vehicle control in the case of psychological driving behaviour observation (see above), whereas they are operationalised with greater differentiation as independent dimensions of driving competence in the optimised practical driving test. Furthermore, the psychological driving behaviour observation also assesses elements of driving competence which can only be observed indirectly, such as “Resilience”.

Greater similarity to the practical driving test – compared to SCHUBERT and WAGNER (2003) – is to be found in the concept for psychological driving behaviour observation presented by BRENNER-HARTMANN (2002). Even though it is emphasised that the focus of psychological driving behaviour observations is placed on possibilities for the compensation of

(neutral, friendly, unfriendly), behaviour towards weaker road users (positive, negative) and any traffic conflicts (defined as situations with a possible pending collision in less than one second). The observer using the predefined report concentrates on standardised behaviour dimensions, for example behaviour relating to the correct use of lanes, safe distances, overtaking, speed selection and traffic signals, adaptation when approaching obstacles and junctions, use of turning indicators, and behaviour towards pedestrians and in situations with stipulated right of way. The observation contents thus appear to correspond to the test contents of the optimised practical driving test.

very specific areas of personal impairment (such as concentration or attention deficits), and they thus lack the character of a test, the author acknowledges recourse to the demand standards which were defined by HAMPEL and KÜPPERS (1982) for the test locations to be used for the practical driving test and are essentially still applicable today (see Chapter 3). During the drive, the psychologist is expected to record any conspicuous behaviour under six headings (speed-related behaviour, behaviour relating to safe distances, proper use of the road lanes, defensive behaviour, endangering behaviour, communication); three of the five observation categories defined for the optimised practical driving test are thus also specified here. The conspicuous behaviour is classified according to its safety relevance as either an "uncertainty" or an "error"; the decision as to what constitutes an uncertainty or error, however, is to a large extent dependent on the subjective opinion of the observer (KAUSSNER, 2007).

It can be concluded that the formal demand and implementation standards for psychological driving behaviour observations apparently possess a greater or lesser similarity to the (optimised) practical driving test – depending on the organisation operating the assessment centre.[130] This applies to the specified driving tasks and parts of the observation categories, the setting in which the observation is conducted, and the formal procedures[131] followed by the psychologist. Nevertheless, as already indicated, no uniform catalogue of criteria exists for psychological driving behaviour observations: The individual operators of the driver assessment centres each compile their own criteria for the particular methods used; process manuals are often subject to a high degree of abstraction, insofar as they are published at all (SCHUBERT et al., 2005). In contrast to the practical driving test, and despite undeniable overlaps in the areas of traffic observation and vehicle control, the assessment of psychological driving behaviour observations appears to focus strongly on fundamental mental resources affecting the driver's performance (orientation capabilities, resilience) rather than on situation-related and rule-compliant psychomotor driving behaviour. This suggests that psychological driving behaviour observations – especially in the form described by SCHUBERT and WAGNER (2003) – are not suitable as concurrent methods for (convergent) validation of the optimised practical driving test due to their different diagnostic subject. Generally speaking, however, the question as to whether and with which dimensions psychological driving behaviour observation could serve validation must remain open as long as neither implementation and demand standards, nor quality findings have been published.

The above discourse suggests that both the recording of driving behaviour data using in-vehicle data recorders or driving simulators and the assessment of relevant components of driving competence by way of traffic perception tests could offer promising opportunities for external validation of the optimised practical driving test. At the same time, however, it is shown that these possibilities must be analysed further and elaborated accordingly before they can be used. It also seems clear that, taken separately, none of these possibilities can be considered a perfect validity criterion: The special methodical character of the (optimised) practical driving test, as a holistic test of driving competence under the scarcely standardisable conditions of real traffic, cannot be replicated in full with concurrent methods; given the complex system of conditions applicable in road traffic, the consequences of a certain level of driving competence in terms of future driving performance cannot be predicted reliably in the sense of proof of predictive validity. It is thus likely that the various external criteria can only be used to validate individual components of the driving competence assessed by way of the optimised practical driving test; the aspect of traffic observation, for example, could be validated by way of traffic perception tests, whereas data recorders could provide for validation relating to vehicle control, vehicle positioning and speed adaptation. This places the aforementioned demand for medium-term construct validation of the optimised practical driving test on the research agenda for the traffic sciences, alongside the necessity to elaborate a no-

[130] On this point, the authors of the present report contacted a number of operators of centres for the assessment of fitness to drive by telephone.

[131] The similarities include also the fact that observers conducting psychological driving behaviour observations are intended to record not only errors, but also "strengths or particularly good performance and the conditions relating to incorrect behaviour" (SCHUBERT et al., 2005, p. 62), that similar termination criteria apply (SCHUBERT & WAGNER, 2003), and that some operators record the driving conditions (e.g. traffic density and weather conditions) in addition to driving behaviour.

mological network founded in competence theory to bring together the different external validity criteria: Even if individual elements of such a validation network fail to convince when viewed in isolation, the entirety of findings could perhaps nevertheless furnish unequivocal proof of validity.

One question which arises concerns the additional steps which can be taken in the short term, for example in the course of the pending revision project, to investigate the validity of the optimised practical driving test. Given the relative simplicity, it here seems expedient to consider the previously mentioned "known-groups technique" (SCHNELL et al., 2008; see above): By way of a series of drives conducted under test-equivalent conditions ("evaluation drives"), the driving performances of different groups of drivers who must evidently demonstrate different levels of driving competence could be compared to determine whether the expected performance differences are actually revealed. To be able to verify the validity of the optimised practical driving test, therefore, it is necessary to form several groups of drivers whose driving competence will in all probability differ both significantly and systematically.

For an initial validation study relating to the optimised practical driving test, it is proposed to form four groups of drivers. The first group should comprise persons who, in the opinion of their driving instructor, have successfully completed driver training in a driving school and are thus ready to attend the (optimised) practical driving test ("Driving test candidates"). According to FUNK, SCHNEIDER, ZIMMERMANN and GRÜNINGER (2010), such candidates will have taken an average of 30 driving lessons.[132] Correspondingly, learners who have so far attended at most 12 driving lessons, and have thus completed less than half of the usual scope of driving training, should display significantly less driving competence; they are to form the second group of drivers for the validation study ("Learner drivers"). A third group should consist of drivers who have participated in the training model "Accompanied Driving from Age 17" ("BF17") and have thereby recorded an average of 2,400 km of driving over an accompanied driving phase of seven to eight months (FUNK et al., 2009); thanks to this extended period of practical driving experience, the members of this third group ("BF17 participants") should be characterised by a higher level of driving competence compared to the learner drivers and driving test candidates in the first two groups. For the fourth group, finally, it is suggested to recruit drivers with at least three years of driving experience[133] ("Experienced drivers"). Evidence for the validity of the optimised practical driving test, and at the same time for its differentiation capabilities, would be obtained if the event- and competence-oriented performance assessments recorded during the evaluation drives improve from group to group from the learner drivers, via the driving test candidates and the BF17 participants, through to the group of experienced drivers.

How should the evaluation drives be designed? Each group of drivers should comprise at least 20 persons, and the actual driving time for each person should be 25 minutes, in accordance with the duration specified for the real practical driving test. The demands of the evaluation drives should correspond to the test demands and should be standardised as far as possible; to this end, a study route must be planned and used consistently for all evaluation drives. It seems particularly important that the evaluation drives should be realised as blind tests, i.e. that the designated assessor and the participating driving instructors must be unaware of the group to which an individual driver belongs (it should also not be possible to draw conclusions as to group assignment or prior experience from other attributes, e.g. the age of the test person). The evaluation drives should be conducted by experienced driving test examiners and driving instructors, and their number should be kept as small as possible in the interest of maximum standardisation. The test location chosen for the evaluation drives should be a major city which offers a broad diversity of demanding road environments for the driving tasks to be assessed, as well as reliably predictable traffic situations. The latter requirement would also be promoted by ensuring that evaluation drives always take place at the same time of the day and in similar weather conditions.

Verification of secondary quality criteria

As a final step within the framework of instrumental evaluation of the optimised practical driving test, it is necessary to consider also the secondary quality

[132] This matches information from the German Federation of Driving Instructor Associations (BVF), according to which an average of 23 practice lessons are taken within the framework of basic practical driver training. Together with the legally prescribed 12 special drives, therefore, this indicates that, on average, novice drivers complete a total of 35 driving lessons during their practical driving training. Nevertheless, there are some candidates who already take the driving test successfully before their 20th driving lesson.

[133] Studies suggest that comprehensive controlled automation of the psychomotor actions necessary to drive a motor vehicle must be viewed as a process which, depending on the amount of driving done, may require up to three years (MAYCOCK & FORSYTH, 1997).

criteria economy, usefulness, reasonableness, resistance to falsification and fairness.

To meet the criterion of economy, it must be possible to realise and assess the test method routinely and conveniently; the realisation should furthermore occupy only a short period of time and require only a minimum input of resources. The verification of this criterion demands fundamental economic studies, for which

- an initially tested and revised observation instrument, including the associated electronic test report, and
- a team of adequately qualified driving test examiners who have received observer training with appropriately intensive elements of practical exercises and have since gained experience with the new test method in their day-to-day work

are, in turn, imperative prerequisites. These prerequisites will be met at the earliest after the new test instrument has been proved sound and practicable in the planned revision project and there is presumably no further need for modifications with regard to the test methodology, the hardware and software ergonomics, or the technical features. The driving test examiners chosen for the studies of test economy should have conducted at least 100 practical driving tests with such a finished and unmodified instrument under daily test conditions, so as to permit meaningful judgements of the actual effort and resources involved, and thus of the arising costs. This procedure was also agreed between the representatives of the Federal Highway Research Institute (BASt), the Technical Examination Centres, the Association of Technical Inspection Agencies (VdTÜV) and the working group TÜV | DEKRA arge tp 21 when elaborating a draft for the “System Manual on Driver Licensing (Practical Test)” (see Annex 2 to the present report): “The cost implications of further development of the practical driving test must be estimated in good time in respect of the additional input required and the potential savings. This assessment should be made at the earliest possible time after corresponding agreement with the legal regulator and on the basis of experience gained from the revision project” (TÜV | DEKRA arge tp 21, 2011, p. 5ff.).

It must be noted here that an answer to the question of a reasonable cost-benefit ratio and an acceptable outlay for testing will also be dependent on whether or not the overall economic costs of road accidents (STRAUBE, 2011) are taken into account, and on the demands stipulated by the legislator for documentation of the practical driving test. The demand applicable to an optimised practical driving test will probably differ from the objectives pursued to date with documentation in accordance with Annex 13 to the Examination Guidelines (see footnote 119 on the differences between methodically acceptable test documentation and documentation serving to support administrative processes in the sense of Annex 13). It can thus be assumed that the optimised practical driving test will automatically mean an increase in the scope of documentation tasks and the time spent on documentation. In this connection, it should be pointed out that, from the legal point of view, no substantial demands are placed on the documentation of the practical driving test. It is merely specified that the examiner is to produce a record of the test drive; no stipulations exist with regard to either the documentation instrument to be used or the scope of documentation (Annex 7 FeV; PrüfRiLi 6). It is likewise to be assumed, however, that the legislator holds certain expectations relating to meaningful test documentation and, on this basis, to development-oriented feedback to the candidates on their test performance (e.g. all candidates should receive a meaningful test report to support their further learning); such expectations are shared by the Technical Examination Centres (STURZBECHER, BÖNNINGER & RÜDEL, 2010). Clarification of how and to what extent these expectations can be met, is a process which has only just begun with the elaboration of theoretical and methodical foundations and with the realisation of a feasibility study, and must be continued within the framework of the revision project.

The question posed with regard to the meaningfulness of the test documentation and its suitability as a basis for development-oriented feedback on test performance to the candidate is also relevant for the criterion of usefulness. This criterion would be met if the optimised practical driving test is proven to satisfy practical needs. This applies firstly with reference to the principal objective of the measure, namely enhanced road safety, and would be achieved if improved competence diagnoses and professional competence feedback were to strengthen and broaden the selection and control functions of the optimised practical driving test within the overall system of novice driver preparation. This is by no means a fallacious expectation, not least because novice-specific competence deficits and accident causes were taken into account in the revision of the test standards. Further points are similarly indicative of the improved usefulness of the optimised practical driving test; this includes above all the electronic acquisition, transmission and processing of test data as a prerequisite for effective test evaluation.

The optimised practical driving test would meet the criterion of reasonableness if its benefits stand in appropriate relationship to the temporal, financial mental and physical loads placed on driving licence applicants.[134] Despite the fact that the road safety impact of the (optimised) practical driving test has not been proven empirically, it is justified to assume that, without the statutory obligation to pass a driving test and the associated requirement to complete appropriate training (here not referring solely to driving school training), the already high rate of accident involvement among novice drivers (see above) would be much higher still. Compared to the potential time losses and financial costs of road accidents, alongside their mental and physical consequences, the costs of testing and the loads placed on test candidates appear reasonable.

For the optimised practical driving test to meet the criterion of resistance to falsification, it is necessary for the event- and competence-oriented performance assessments relating to the individual driving tasks and observation categories to be documented as soon as possible; this serves to avoid observation and judgement errors. Subsequent alterations, as will no doubt be necessary on occasion to correct input mistakes, must be transparent and accompanied by corresponding explanations. Fulfilment of these demands is facilitated if all inputs made by the driving test examiner are documented in the electronic test report (see Chapter 4).

The criterion of fairness refers above all to the transparency of the demand and assessment standards, and to the results of equivalence studies to measure test equality. SCHWENKMEZGER and HANK (1993) use the style of empirical determination to distinguish three forms of test method equivalence which can be applied in the case of the practical driving test: "Psychometric equivalence", "Experience-related equivalence" and the already mentioned dimension of "Population-specific equivalence". Psychometric equivalence means that tests display a similar degree of fulfilment in respect of the classic quality criteria of objectivity, reliability and validity; whether or not this is true, is to be investigated within the framework of the instrumental evaluation and the product audits (see Chapter 5.4.5). Experience-related equivalence requires that similar subjective opinions on test realisation and assessment are expressed by experts, candidates and driving instructors across all tests, i.e. no tests are conducted which are felt to have been particularly simple or else particularly difficult and unfair; meaningful findings in this respect can be supplied by customer surveys (see Chapter 5.4.4) and again by product audits (see Chapter 5.4.5). Population-specific equivalence, finally, means that the test results are not affected by content-independent inter-individual or population-related differences. To be able to judge this, the continuous evaluation relating to analyses of test results (see Chapter 5.4.3) and customer surveys (see Chapter 5.4.4) is to measure and assess also the influence of candidate attributes (e.g. age, gender, education, migration background) and general test conditions (e.g. traffic density, road conditions, weather conditions) on test performance.

5.4.3 Analysis of test results

While the "instrumental evaluation" described in the previous chapter targets the psychometric quality of the methods employed by the optimised practical driving test, the objective of the three evaluation elements to be presented in the following is to analyse the implementation quality of the test in daily use. These evaluation elements thus serve a processual evaluation and are intended to provide methodically sound proof of a uniformly high quality of test design and performance assessment across the whole country; this includes – alongside "Customer surveys" (see Chapter 5.4.4) and "Product audits" (see Chapter 5.4.5) – also the "Analysis of test results".

With their analyses of test results, the Technical Examination Centres comply, on the one hand, with the stipulations of EU Directive 2006/126/EC, according to which the outcomes of the driving tests conducted are to be reviewed at regular intervals (Annex IV, paragraph 4.1.2), but at the same time also with the "Requirements for Operators of Technical Examination Centres" (BASt, 2009, point 6.9), in which it is specified that each Technical Examination Centre is to produce statistics which contain at least the results of the driving tests conducted, in each case differentiated according to the types and numbers of tests and the responsible examiners (see above). These statistics are based on the data which the examiner is required to record as documentation of the test drive. The BASt requirements stipulate that the following details are to be recorded (BASt 2009, point 6.7):

[134] This can be assumed for the driving test examiners, driving instructors and authority staff involved in the system of driving licence testing, because they have been trained accordingly and receive appropriate remuneration for their work.

- The date of the test
- The names of the examiner and candidate
- The class(es) of driving licence for which the test was taken
- The start and end times of the test
- The number of basic driving manoeuvres performed
- The test environment (proportions of driving within and outside built-up areas)
- Information on the use of motorways
- The test result.

To date, neither the above records nor the results of practical driving tests have been made the subject of studies based on scientific standards in the sense of an external summative evaluation. Current analyses in the Technical Examination Centres are performed by the respective quality officers within the framework of internal quality management and are concerned above all with the determination of pass rates.[135] Accordingly, the results of the practical driving test have in the past played no significant role for the optimisation of novice driver preparation, despite the fact that evaluation of the aforementioned test data would already today enable more specific description of factors such as the typical local test conditions ("Within built-up areas", "Outside built-up areas", driving on "Motorways or similarly constructed roads") and analysis of their influence on test assessments and decisions. As a further example, analysis of the dates and times of tests, and comparison with the corresponding test results, would permit conclusions to be drawn as to whether test performances are generally poorer in the winter months – characterised by early darkness and thus potentially reduced visibility conditions – than during the summer. On the other hand, the possibilities for such analyses are at present still rather narrow, as the test report used to date (see Chapter 4) does not provide for the systematic and differentiated documentation of driving tasks, test assessments and framework conditions (e.g. traffic density, weather conditions, visibility conditions).

Which design should be chosen for a future scientific analysis of the results of the optimised practical driving test – as a complement to the collection of general statistics for the BASt, the supervisory authorities and the KBA? From the methodical perspective, it seems expedient and practicable to distinguish different levels of evaluation, on which mutually supplemental and constitutive statements can be derived with regard to the conditions under which tests are realised, the performance displayed by candidates and test quality:

1. The first level serves initially the descriptive presentation and analysis of test performances with reference to test-relevant components of driving competence, observation categories and test decisions. Both event-oriented and competence-oriented assessments are to be taken into account.
2. On the second level, correlations between the test demands, correspondingly referenced (event- and competence-oriented) performance assessments and the resulting test decisions are to be investigated. This includes, in particular, analysis of the possibilities for prediction (e.g. for prediction of a test decision from competence assessments) as a basis for the implementation of plausibility checks.
3. The objective of the third level should be to identify and analyse possible patterns in the data with regard to test behaviour (e.g. prototypical demand patterns and planning strategies), the competence-oriented performance assessments and the test decisions.

re 1: On the first evaluation level, analyses of the frequency parameters and frequency distributions are performed for the specified analysis contents. These analyses serve, among other things, to furnish proof for the realisation of an ideal-typical test, including implementation of the stipulated requirements in the sense of driving tasks. If certain driving tasks are not or only seldom performed at a particular test location, this may indicate infrastructural circumstances which hinder test planning in this respect; such pointers must then be followed up with deeper analyses of the test

[135] The data on passed tests are used to calculate mean pass rates for a particular test location, region or branch office, which subsequently serve as guideline values for purposes of internal quality management. If the mean pass rate of an individual examiner is found to lie significantly outside a correspondingly defined corridor without plausible reasons being evident, this will generally be discussed in a personal meeting with the examiner concerned; it is also conceivable that a special product audit could be ordered (see Chapter 5.4.5). The obvious problem with this procedure is that uncontrollable normative influences on the test behaviour of examiners cannot be excluded, as they may consciously seek to avoid conspicuous results. If such influences were to apply, this would reduce the validity of the practical driving test and lead to artificial stabilisation of the pass rates, which in turn undermines the control function of the driving test within the system of novice driver preparation. It thus seems necessary to develop more sophisticated forms of quality management which are based on the identification and comparison of examiner-referenced test behaviour and assessment patterns and promote self-reflection on the quality of the conducted driving tests on the part of the examiner. Given the similarity of the methodical foundations and procedures, the development of external evaluation methods should also inspire the further development of internal quality management.

location conditions.[136] To be able to assess the quality of a test location, analyses of the general infrastructural situation on the basis of relevant documents and location inspections must be augmented with full documentation of all the driving tasks performed during test drives over the period of corresponding study. Insofar as the legal and technical prerequisites can be met, it is also feasible to record the test routes used as GPS tracks. If such analyses were to be performed over the whole country, this would enable the empirical elaboration of different test location profiles in respect of the availability of opportunities to test the required driving tasks; such profiles could then be taken into account appropriately for the purposes of both test planning and test evaluation.[137] If the analyses of test results are referred to groups of test locations with a similar profile, for example, this would produce correspondingly differentiated analysis results, and thus significantly enhance the informative quality of the test evaluation. The prerequisite for such a procedure is assignment and recording of a code number for each test location.

Evaluations of demand equality are just one example for the information which can be derived from descriptive analyses of test results. As a further possibility, analysis of the driving error distributions and the times at which errors are recorded relative to the overall duration of the test drive could reveal whether and, if so, to what extent examination stress on the part of the candidates impairs the validity of the optimised practical driving test: It is assumed that driving errors attributable to the stress of the test situation will decrease over the course of the test, whereas errors resulting from driving competence deficits should at the same time increase in line with the raising of test demands.[138] Finally, descriptive analysis of the test results also permits feedback to the training system as to those demands or driving tasks which pose particular difficulties for the test candidates; such feedback would be a valuable contribution to the optimisation of driver training and novice driver preparation overall.

re 2: Insofar as behaviour observation, performance assessment and performance documentation are realised consistently and in conformity with the specified standards over the course of the test drive, and provided a proper test decision is then recorded, certain correlations can be expected between the data of the different assessment levels. By way of explorative correlation and prediction analyses, it can be investigated whether these expected correlations are actually found in the real data: The functions of correlation and regression analysis, for example, can be used to test the hypothesis that a lower level of driving competence on the part of a test candidate will be reflected in more frequent or more serious driving errors, and accordingly in a reduced probability of passing the test. Regression analysis would also permit clarification of the extent to which the assessments relating to particular components of driving competence influence the test result, and whether test results are affected by changes in the situational conditions (e.g. through the use of driver assistance systems). From such results, in turn, it would be possible to derive proposals for optimisation of the training of novice drivers; insofar as the test documentation provides for recording of a pseudonymised ID code representing the candidate's driving school, feedback could even be given directly to the driving schools.

Correlation analyses serve furthermore to investigate objective aspects of the fair-

[136] It was already reported that, within the framework of the EU project “TEST”, 31 per cent of the driving test examiners said that they were dissatisfied with the test locations, and that many test locations lacked opportunities to test all the demands stipulated for the test drive by EU Directive 2000/56/EC. To date, no equivalent and methodically robust findings exist for Germany.

[137] The suitability of test locations also stands at the focus of local government and traffic policy discussions, as a result of which it seems important to use the scientific and technical instruments available today to elaborate founded suitability criteria and, in a next step, methodically sound evaluation strategies for test locations. When doing so, pressures to achieve standardisation should not be placed in the foreground; after all, driving test requirements play no role in the concrete planning of road infrastructures and the differences between various test environments can thus be seen as an unavoidable necessity. The developments should instead concentrate on the flexible, safety-reflected use of different methodically controlled test location profiles. In this way, new stimulus could be lent to the broader test location discussion, which has rather stagnated since the end of the 1980s.

[138] The driving test examiner is intended to organise the practical driving test according to the principle “from simple to more difficult” and is to allow the candidate a corresponding “familiarisation phase”. This means that route sections with enhanced demands should, where possible, be avoided at the beginning of the test, so as to help the test candidate to gradually reduce and overcome his uncertainty and stress during the test (STURZBECHER, BÖNNINGER & RÜDEL, 2010).

ness criterion ("test equality") and the population-specific equivalence (see Chapter 5.2) of the optimised practical driving test (the subjective aspect is addressed within the framework of the customer surveys). To this end, it is analysed whether correlations exist between the test demands, test assessments and test decisions on the one hand, and the socio-demographic data characterising the test candidates (e.g. age, gender, migration background) on the other hand.

re 3: A further objective for analyses of the test results is to search for patterns in the data relating to test behaviour and assessments. A whole range of possibilities exists for combination of the competence-oriented assessments of driving tasks, the applicable dimensions of driving competence and a test decision. Taking the five competence dimensions for the optimised practical driving test (traffic observation, vehicle positioning, speed adaptation, communication, vehicle control), together with the four assessment levels which can be assigned in each case and the dichotomous test decision, this gives a theoretical total of 2048 assessment possibilities. The interesting question is then whether these combinations display a random distribution, or whether certain "assessment patterns" are revealed. Recommended methods for the identification of test behaviour and assessment patterns are configural frequency analysis, log-linear analysis and cluster analysis.[139] If the data analysis indicates certain test behaviour and assessment patterns, these findings can be used to improve the quality of test planning and assessment by the driving test examiner. Ideally, all assessment patterns determined for a particular examiner should correspond in frequency to the distribution probability determined across all tests conducted, insofar as the tested candidates display no systematic particularities in terms of driving competence and no exceptional test location conditions apply. Accordingly, if differences are found in the distribution of assessment patterns for an individual examiner compared to the overall pattern distribution for all examiners, this is a sign of possible assessment distortion and can be investigated further within the framework of internal quality management in the Technical Examination Centre concerned. Such distortions may be the result of observer and assessor effects, which are not seldom in the context of systematic behaviour observations and also cannot be excluded in connection with the practical driving test (STURZBECHER, BÖNNINGER & RÜDEL, 2010). For such analyses, it would be necessary to include a pseudonymised ID code for the driving test examiner in the documentation of the practical driving test. Finally, it would be expedient for deeper analysis of the population-specific equivalence to verify whether certain test behaviour and assessment patterns only occur in conjunction with certain groups of test candidates (e.g. candidates of a particular age, male or female candidates, candidates with a migration background).

The analysis procedures described above should be implemented recurrently as elements of a continuous summative process evaluation of the optimised practical driving test, in order to reflect and safeguard test quality in accordance with scientific principles. In addition, it is important to expose correlations between the conditions under which the optimised practical driving test is realised (e.g. traffic density, test environment and road characteristics, weather conditions, lighting conditions, road conditions; see also the above remarks pertaining to the TEST project) and the test results, not least in order to identify those test conditions with the potential to impair the validity of the test results,[140] and to determine how such impairment could be avoided, for example by way of corresponding qualification and training offers for the

[139] All three methods serve to group (numerous) objects according to particular attributes, whereby it is possible to identify certain prototypical combinations of attribute values. Configural frequency and log-linear analysis can be used for direct investigation of whether certain attribute patterns ("configurations") occur significantly more frequently than expected. With the aid of cluster analysis, on the other hand, it is possible to categorise objects according to their attribute-related differences; subsequently, the objects within a given category should be as homogeneous as possible, whereas objects in different categories should be maximally distinct.

[140] Attention is here drawn once more to the aforementioned findings of BAUGHAN et al. (2005) relating to the dependence of test results on the weather conditions; it remains to be verified, however, whether these findings are equally applicable to the German practical driving test. Already at the beginning of the development work on optimisation of the practical driving test, namely in 2005, expert workshops with groups of examiners from the Technical Examination Centres elaborated a catalogue of factors which were presumed to influence test realisation and the test decision (STURZBECHER, BÖNNINGER & RÜDEL, 2010). These factors included, among others, the traffic and weather conditions, as well as the road characteristics at test locations.

driving test examiners. It then remains to be decided, depending on the results of these analyses, which test conditions must be recorded and analysed continuously for purposes of interpretation and assessment of the validity of the test results, and where a single scientific study with an appropriate sample is sufficient to provide the examiner with a basis for criterion-driven consideration of the prevailing test conditions in the processes of performance assessment and test decision. The first such studies should be devoted to analysis of the influence of traffic density and weather conditions on test realisation[141] and test assessment.

5.4.4 Customer surveys

In the service sector – and in this respect the practical driving test can be viewed as a service provided by the Technical Examination Centres (STURZBECHER & MÖRL, 2008) – customer surveys permit the targeted and structured acquisition of information on the expectations of current and potential users ("customers") with regard to the design and quality of the service; at the same time, information can be gathered on corresponding user experiences and customer satisfaction (SCHNEIDER & KORNMEIER, 2006). In contrast to open expert observations or quality audits, which will usually reflect the maximum performance capabilities of the service provider because the persons involved – in knowledge of the observation and assessment – will automatically "give their best", customer surveys mirror the typical level of performance provided in everyday situations without such exceptional incentives (SACKETT, ZEDECK & FOGLI, 1988; SCHULER, 2001). It is last but not least for this reason that customer surveys are planned as an independent element of the evaluation system presented here. After all, the customer perspective is to be seen as an important source of information for the professional planning of necessary innovation processes, especially with regard to the intended introduction of new (optimised) services (PILLER, 2006); this is a further argument confirming the necessity to conduct customer surveys on the optimised practical driving test.

The question of user satisfaction with a product or service (or measure) is generally the core element of a customer survey. According to v. HOLTZ (1998), "satisfaction" is to be understood as "an individual psychological phenomenon comprising emotional, cognitive and intentional dimensions. Satisfaction arises from comparison of a target component with an actual component. The target component, which is described using varying terms such as 'needs', 'expectations' or 'demand level', is an individual reference system in the sense of a comparative scale against which the actual component is judged. The role of the actual component is fulfilled by 'reality' as perceived subjectively by the individual" (p. 21). From the discrepancy between the planned (target) state and the surveyed (actual) state, it is possible to derive specific potential and measures for optimisation. The "confirmation/disconfirmation paradigm" (HOMBURG, 2008) can be considered a basic theory of customer satisfaction research and also served as the starting point for elaboration of a methodical system to assess satisfaction with driving licence tests (STURZBECHER & MÖRL, 2008)[142]. If the customer deems the actual state to match the target state, this represents satisfaction or "confirmation". A case in which the customer's assessment of the actual state exceeds the target state is termed "positive disconfirmation". If, by contrast, the actual state falls short of the target state, i.e. the customer is dissatisfied, it is customary to speak of "negative disconfirmation".

The realisation of meaningful customer surveys places particularly high demands on the validity and economy of the method; substantial and robust results can be expected above all from multiperspective customer survey systems, which mirror heterogeneous customer demands from the viewpoints of different customer groups. Possible survey instruments are telephone contact and direct personal interviews, as well as written question-

[141] Certain weather conditions, for example, could mean that individual driving tasks cannot be tested, because this is not possible without endangering road safety. The examiner's decision to forego such demands or driving manoeuvres is in this case not to be deemed a test deficit; it is rather a sign of well developed professional competence. Similarly, it is an indicator of substantial driving competence if a candidate declines to perform a driving task which involves undue risk. Generally speaking, this possibility to deviate from the catalogue of driving tasks enhances the validity of the test, because a competent driver would also forego dangerous driving manoeuvres outside the test situation where this is necessitated by adverse weather conditions. Consideration of whether such decisions actually served road safety, however, requires recording of the weather conditions as part of the test documentation.

[142] More detailed information on the background and on the processes underlying elaboration of this methodical system are to be found in the project report "Optimisation of the Practical Driving Test" by the working group TÜV ĐEKRA arge tp 21 (STURZBECHER, BÖNNINGER & RÜDEL, 2008). The methodical system was developed and tested by the Technical Examination Centres over the period from 2004 to 2008, in cooperation with representatives of the federal ministry responsible for traffic, the transport ministries of the federal states of Brandenburg and Rhineland-Palatinate, the Bundeswehr, the Association of Technical Inspection Agencies (VdTÜV) and the Federation of Driving Instructor Associations (BVF). The elements of this methodical system include questionnaires to be answered by the test candidates and their driving instructors, as well as interview guidelines for the licensing authorities and the responsible supreme authorities at federal state level.

naires distributed either online or by post; the methodical advantages and disadvantages of the individual options – in the present case with reference to the practical driving test – must be weighed up accordingly (STURZBECHER & MÖRL, 2008). Where data are acquired by way of an ex-post survey, i.e. with a certain time delay after performance of the service, methodical research assumes that the customer will supply a relatively reflected assessment. The methodical challenge is to choose a time between customer contact and customer survey which is neither too short nor too long: If the survey is conducted immediately after the service is provided, the customer's user experience is still very present, but it is not yet possible to give an adequate assessment of any short-, medium- and longer-term consequences. If too much time passes between the service and the survey, there is an increased likelihood of memory gaps and (incorrect) mental reconstruction.

The general methodical potential of customer surveys was already mentioned above; but where does the particular significance lie with regard to optimisation of the practical driving test? Through its selection function (see Chapter 1), the practical driving test is an element of a safety system operating in the public interest to guarantee specifically the safety of road traffic. It is thus imperative to record and react with due earnestness to the comments received from all involved parties, in order to be able to optimise system content and methodology for the safety system in general and the practical driving test in particular (STURZBECHER, BIEDINGER et al., 2010). The state demands that the quality management policy of the operators of Technical Examination Centres must take into account customer expectations and needs (BASt, 2009, point 3.1). From the scientific and economic perspectives, these expectations are best illuminated by way of customer surveys (see above). Given that these test-focussed customer surveys are to be conducted parallel to the quality management measures of the Technical Examination Centres, a repetition cycle of at most five years is recommended for a continuous, summative process evaluation. This recommendation results not least from consideration of the personal certification standard (DIN EN ISO/IEC 17020) anchored in paragraph 72 of the Driving Licence Regulations (FeV), which contains specifications relating to control of the content-oriented topicality and target-group-specific adequacy of certification measures (STURZBECHER, BIEDINGER et al., 2010).

As not only the driving test candidates, but also in the broader sense – namely via the demand for the customer-friendly realisation of testing procedures in which the driving schools are involved (ibid.) – the driving instructors are to be viewed as customers of the Technical Examination Centres, it is considered necessary to elaborate a survey instrument for both candidates and instructors for use within the framework of a multi-perspective evaluation system. The elaboration of such an element of evaluation can be based on the "Methodical System for Determination of Satisfaction with the Driving Test" (STURZBECHER & MÖRL, 2008), which was presented by the Technical Examination Centres in 2008 after several years of research and development work (STURZBECHER, BÖNNINGER & RÜDEL, 2010).

Which survey method and which survey instruments are appropriate for the evaluation of the practical driving test? In accordance with experience gained during development of the "Methodical System for Determination of Satisfaction with the Driving Test" (STURZBECHER & MÖRL, 2008), a computer-assisted telephone survey based on standardised interview guidelines and with supplementary open questions is proposed as the survey method for the test candidates and driving instructors. The advantages of a telephone survey lie above all in the relatively high sample response, especially if the respondents are informed of the forthcoming telephone survey in advance. The comparatively low costs and time requirements also support the choice of a telephone survey. For the actual interviews, it is recommended to use the computer-assisted survey method CATI ("Computer-Assisted Telephone Interview"), which is the system favoured by many opinion polling institutes; it is furthermore capable of significantly reducing the expenditure and work load for survey realisation. The interviews are conducted at specially configured computer terminals, with both the individual questions and the overall course of the interview being guided and controlled directly by pre-programmed computer processes; this extensive standardisation and control of the survey situation contributes to high data quality. The acquired data are available for further applications immediately upon completion of the interview. The telephone survey is to be conducted by appropriately trained interviewers under the permanent supervision of a scientifically qualified supervisor familiar with the circumstances of the practical driving test.

The survey instruments to be developed to obtain the opinions of candidates and driving instructors within the framework of an evaluation system for the optimised practical driving test can be based on the existing scientifically founded and tested questionnaire (STURZBECHER & MÖRL, 2008) and could take over certain elements of its approach; simple re-use, however, is out of the question, because different objectives apply: While the use of the existing survey instruments serves the internal quality management in the Technical Examination Centres and thus takes into account all the general conditions surrounding test realisation, the instruments to be elaborated for the evaluation system must focus on quality criteria which the legislator has deemed relevant for the measure "Optimised practical driving test". Consequently, a series of adaptations were necessary when developing the following proposal for a questionnaire; it was attempted, for example, to acquire a more differentiated picture of the quality of test assessments and performance feedback to the candidates by way of new indicators and questions. The instrument proposals presented here must be subjected to critical methodical appraisal – and reviewed as necessary – as part of a revision project, and can then be deployed for purposes of customer monitoring with trend and cross-section comparisons of the extent to which driving licence applicants and driving instructors are satisfied with the optimised practical driving test in Germany.

Starting points from which to derive survey indicators for the quality criteria which the legislator expects to be observed in the context of the practical driving test are to be found above all in the Examination Guidelines: According to these guidelines, the examiner must explain how driving instructions will be given to the candidate and may also comment on the required driving behaviour, for example in respect of speed (PrüfRiLi 5.12). He is subsequently intended to take into account the psychic stress bearing on the candidate; it is thus deemed unreasonable, for example, to reproach the candidate for mistakes or to ask the meaning of traffic signs during the actual driving (PrüfRiLi 5.14). When assessing the driving tasks and formulating a test decision, rules are not to be interpreted pettily; at the same time, positive aspects of performance are to be honoured (PrüfRiLi 5.17). The test drive is to be terminated as soon as it becomes clear that the candidate is not able to satisfy the demands of the test (PrüfRiLi 5.19), and if the test is failed, the examiner is to inform the unsuccessful candidate accordingly, giving a brief account of the relevant errors, and is to hand over a test report conformant with Annex 13 to the Examination Guidelines (Annex 7 FeV, 2.6). Furthermore, the BASt evaluation requirements stipulate that efficient order and appointment tracking must be guaranteed (BASt, 2009, point 6.4); the meeting place for the test drive must be chosen such that the candidate can reach this point without undue difficulties (PrüfRiLi 5.10).

How can the aforementioned quality criteria – and others which arise from the methodical demands of professional test realisation – be operationalised in the survey instruments for test candidates and driving instructors? The survey instruments should each be divided into three thematic complexes: After a first complex comprising questions on test preparation and prior experience with the practical driving test in the case of the candidates, or else questions on the organisational framework for practical driving tests in general in the case of the driving instructors, the second section of both survey instruments is to be devoted to questions relating to a particular practical driving test. For the required assessments of satisfaction, both instruments should use the proven survey scale with the levels "Very satisfied", "Largely satisfied", "Largely dissatisfied" and "Very dissatisfied", the appropriateness of which was founded by STURZBECHER and MÖRL (2008) by way of reference to the necessary degree of differentiation in the respondents' assessments. The third complex, finally, should serve to acquire socio-demographic data to describe the candidate, or else corresponding information about the driving school and driving instructor. Both survey instruments should also incorporate open questions which permit the test candidates and driving instructors to add further detail to their replies and to submit proposals for improvement. The following tables show the envisaged content sections with their corresponding operationalisations, and at the same time indicate the survey instrument(s) in which a particular indicator is to be acquired.

The survey instrument for the test candidates begins with a section entitled "Prior experience and preparation for the test"; the recommended survey items for this section are specified in Table 14 below. The data collected from the test candidate on his possession of driving licences for other vehicle classes, on whether the current driving test is a first or repeat test, on the number of driving lessons taken in preparation for the test and on the possible incorporation of simulated tests into driver training permit an estimation of the candidate's prior experience with driving tests as such. The candidate is also asked to indicate the test-related training offers which were used in the driving school, and the extent to which he was satisfied with such offers. These answers serve to reveal correlations between the quality of training and test quality, and are furthermore a source of pointers for the further development of novice driver preparation, especially where additional open questions are used to obtain suggestions for improvement from the candidate.

Survey section	Items	Candidates	Instructors
"Prior experience and preparation for the test"	Driving licence classes	X	--
	First/repeat driving test	X	--
	Number of driving lessons taken in preparation for the test	X	--
	Test simulations	X	--
	Satisfaction with preparations for the test by the driving school	X	--
	Satisfaction with the information provided by the driving school on test procedures	X	--

Tab. 14: Survey section "Prior experience and preparation for the test"

The driving instructor survey should not be limited to driving school owners, but should instead also permit the questioning of dependently employed driving instructors, because the driving school owner may not necessarily participate in driving tests himself where further staff are employed as driving instructors. It thus cannot be excluded that, in some cases, only the employed instructors will possess the specific test experience which is of interest for the analysis of customer satisfaction. As the assigned tasks and thus the work experience of the employed driving instructors will probably vary considerably from one driving school to another, the driving school owners must be allowed to decide on whether or not their driving instructors are to take part in the survey. This approach is also necessitated by the provisions of applicable employment legislation.

The first section of the survey instrument for driving instructors is devoted to the "General organisational framework" (see Table 15). This section offers particular opportunity to obtain significant information on the quality of test organisation: The driving school generally functions as a mediator between the driving licence applicant and the Technical Examination Centre, and thus takes care of all organisational matters on behalf of the test candidate. Furthermore, earlier studies have shown that satisfaction with the organisation of the practical driving test and satisfaction with the administration of test appointments are the factors with the greatest influence on the overall satisfaction of driving instructors with the services offered by the Technical Examination Centres (STURZBECHER & MÖRL, 2008).

Survey section	Items	Candidates	Instructors
"General organisational framework"	Technical Examination Centre	--	X
	Satisfaction with deadlines for bookings	--	X
	Satisfaction with processing times	--	X
	Satisfaction with allocation of appointments	--	X
	Satisfaction with feedback information	--	X
	Use of an Internet-based booking system	--	X
	Use of and satisfaction with possibilities for cancellation	--	X
	Overall satisfaction with the cooperation with the Technical Examination Centre	--	X

Tab. 15: Survey section "General organisational framework"

The essential purpose of this section is to gather the opinions of driving instructors relating to their satisfaction with test administration (above all booking deadlines, use of an Internet-based booking system, processing times, feedback information) and appointment management (above all the allocation of test appointments, possibilities for cancellation); these assessments are to refer to practical driving tests in which the instructor concerned was involved over the past 12 months.

The central element of both survey instruments is the section "Practical driving test" (see Table 16), which should be practically identical for both instruments. In this way, it is guaranteed that the customer satisfaction of the two different target groups can be presented in comparison. The instructions for this survey section are to ask both the candidates (insofar as they have already taken two or more practical driving tests) and the driving instructors to recall and assess specifically their last practical driving test. The reason for this stipulation is that the alternative possibility of a general assessment of driving tests and driving test examshould also be asked to indicate the prevailing traffic density, the local visibility conditions and whether or not the test began punctually, as well as any sense of stress induced by the aforementioned factors. The candidate should furthermore give an assessment of his overall test anxiety, alongside the measure of his satisfaction with the measures taken by the examiner to optimise the social “atmosphere” of the test and to reduce possible test anxiety. Following these assessments of satisfaction relating to the start phase of the test drive, further indicators should follow to judge the level of satisfaction with the examiner in respect of different aspects of the planning and realisation of the test.

iners could lead to significant distortion in the replies given. Such distortion effects result from the structure-seeking characteristics of human perception (e.g. selection, organisation, accentuation and fixation), which – especially in connection with aggregated assessments of numerous events or persons – may in turn produce judgement errors (STURZBECHER & MÖRL, 2008).

Survey section	Items	Candidates	Instructors
"Practical driving test"	Federal state, city/test location	X	X
	Date of the test	X	X
	Satisfaction with the meeting place for the test with regard to accessibility	X	X
	Satisfaction with the meeting place for the test with regard to the prevailing traffic density and the associated sense of stress	X	--
	Visibility conditions and the associated sense of stress	X	--
	Punctuality of the test	X	X
	Stress due to delayed commencement of the test	X	--
	Satisfaction with punctuality	X	X
	Test anxiety and satisfaction with the measures taken to reduce test stress and test anxiety	X	X
	Satisfaction with the friendliness of the driving test examiner	X	X
	Satisfaction with the prior explanation of the test procedures	X	X
	Satisfaction with the clarity of driving instructions	X	X
	Satisfaction with the timely communication of driving instructions	X	X
	Satisfaction with structure and course of the test	X	X
	Satisfaction with the realisation of the basic driving manoeuvres	X	X
	Satisfaction with the assessment of the basic driving manoeuvres	X	X
	Satisfaction with the assessment of the test drive	X	X
	Overall satisfaction with the test assessment	X	X
	Intermediate questions on the test drive	X	--
	Satisfaction with the explanations of observed errors	X	X
	Satisfaction with the remarks on possibilities for further improvement	X	X
	Satisfaction with the mentioning of good performance	X	X
	Satisfaction with the answering of own questions	X	X
	Satisfaction with the comprehensibility of the test report	X	X
	Overall satisfaction with the feedback from the driving test examiner	X	X
	Satisfaction with the measures taken to establish a relaxed atmosphere for the test	X	X
	(Self-)assessment of the candidate's test performance	X	X
	Test result	X	X
	Complaints relating to the test	X	--
	Overall satisfaction with the practical driving test	X	--
	Overall satisfaction with the driving test examiner	X	X

Tab. 16: Survey section "Practical driving test"

The complex relating to the quality of realisation of a particular practical driving test should begin for both candidates and driving instructors with questions addressing the place at which the test drive began. The point of departure for the test drive is specified by the driving test examiner on his sole responsibility; the reasonableness of the chosen point of departure can be deemed an important quality attribute of the driving test, as this choice determines the subsequent test conditions for the candidate: If the traffic conditions at the point of departure are relatively uncomplicated, this can help the candidate to gradually reduce and overcome any driving uncertainties and test stress by facilitating a "familiarisation phase" (see above). In addition, the candidate should also be asked to indicate the prevailing traffic density, the local visibility conditions and whether or not the test began punctually, as well as any sense of stress induced by the aforementioned factors. The candidate should furthermore give an assessment of his overall test anxiety, alongside the measure of his satisfaction with the measures taken by the examiner to optimise the social "atmosphere" of the test and to reduce possible test anxiety. Following these assessments of satisfaction relating to the start phase of the test drive, further indicators should follow to judge the level of satisfaction with the examiner in respect of different aspects of the planning and realisation of the test.

To assist correct categorisation and interpretation of the given assessments of satisfaction, the candidate and his driving instructor should each be required to provide their own judgement on the test performance displayed by the candidate. The test result is another important survey item, as it supplies a valuable control variable: It is to be expected that test success will have a positive influence on the candidate's assessments of satisfaction, and that failure will result in a correspondingly negative effect (STURZBECHER & MÖRL, 2008); if the test result is known, however, it is possible to control for such superimposition effects in the statistical analysis.

When designing the survey questionnaire, the important questions addressing overall satisfaction with a particular aspect of the test realisation should always be placed at the end of the survey section concerned. While it is true that the assessments of overall satisfaction may, under certain circumstances, be influenced by the order of questions and by inter-individual differences relating to the willingness and depth of the survey respondent's reflection, it is in this way possible to obtain a more deliberately reflected and balanced judgement, rather than a "gut feeling" (STURZBECHER & MÖRL, 2008). For the analysis of the survey results, acquisition of the candidate's and driving instructor's overall satisfaction with test realisation and test assessment permits weighting of the individual aspects of satisfaction in accordance with their relative importance for overall satisfaction with the practical driving test.

Survey section	Items	Candi dates	Instruc-tors
"Personal and other details"	Age	X	X
	Gender	X	X
	School education	X	--
	Migration background	X	--
	Post code of driving school location	--	X
	Size of driving school (number of pupils per year)	--	X
	Employment status (driving school owner/staff member)	--	X

Tab. 17: Survey section "Personal and other details"

The third survey section headed "Personal and other details" (see Table 17), finally, serves to record socio-demographic data characterising the candidates and driving instructors, as well as structural data on the driving schools (e.g. location and size of a driving school). With the aid of such data, it is possible to differentiate according to social groups when analysing the customer satisfaction surveys, and in this way to estimate any (undesired) influences on the quality of realisation of the practical driving test. At the same time, knowledge of the candidates' age, gender, school education and migration background enable verification of the population-specific equivalence of the practical driving test.

The survey of test candidate and driving instructor satisfaction should be implemented within the framework of a nationwide representative study. It can be assumed that more than 1.3 million practical driving tests for driving licence class B are conducted each year in Germany (KRAFTFAHRT-BUNDESAMT, 2012a); around 12,800 driving schools (IFO INSTITUT, 2012) and approx. 48,000 persons with a licence to work as a driving instructor[143] (KRAFTFAHRT-BUNDESAMT, 2012b) are involved in those tests. The totalities of all candidates and driving instructors participating in these practical driving tests form the so-called "parent populations"[144] for the two customer surveys to be conducted. A full survey of the parent populations would naturally constitute a representative study, though it seems hardly possible – already for reasons of economy and practicability – to question every test candidate and every driving instructor on the quality of the practical driving test. Consequently, two random samples of test candidates and driving instructors must be drawn from the respective parent populations. A random sample, after all, offers a similar guarantee that the study participants contacted will be representative in respect of attributes relevant for the study. Beyond this, the selection of a random sample brings further benefits: Despite the fact that there will always be deviations between a sample and the parent population, irrespective of the selection method, randomisation enables determination of the size of the selection error, and certain specifications relating to the quality of the study results can already be defined when planning the sampling procedure. Random sampling of the participants for the test candidate and driving instructor surveys, therefore, means that those customer opinions which actually exist in the parent population will be reflected in the findings of the customer survey with a determined degree of certainty. Where it is decided to use random samples for a planned customer survey, a recruitment strategy enabling access to the

[143] It must be taken into account, however, that an appreciable proportion of this number are not active in driver training.

[144] The term "parent population" here designates the totality of all "potential study participants" about whom certain statements are to be made within the framework of a study (BORTZ & DÖRING, 2006).

field and a sample concept (sampling method, sample size) must be elaborated.

The following strategy seems expedient for the recruitment of test candidates and driving instructors: The Technical Examination Centres are to send data on all practical driving tests taken in Germany during a defined evaluation period (e.g. a given week), including details of the candidate, the driving school responsible for previous driver training and the test result, to the scientific institute commissioned to conduct the survey.[145] Random samples of candidates and driving instructors who participated in a practical driving test during the relevant period can then be drawn from this data set. The advantage of this strategy compared to all other conceivable recruitment procedures lies in the certainty that the sampled test candidates and driving instructors have actually participated in a practical driving test during the chosen evaluation period; this serves to minimise the duration and cost of the survey. Furthermore, it is thus possible – as mentioned earlier in this chapter – to place the survey at an optimum point in time. Finally, it can in this way be ensured that assessments from both the candidates and the driving instructors can be compared for a large number of the practical driving tests to be judged.

The selected potential survey participants are initially to be contacted in writing, informed about the planned survey and asked to participate. This allows the candidates to decide at leisure whether or not they wish to participate in the survey, and removes all influence which may otherwise arise in the exceptional situation of a theoretical or practical driving test. If the contacted persons agree to participate in the survey, they can either enter their telephone number and a preferred time for contact via a correspondingly prepared online portal, or else complete and return the enclosed declaration of consent to the commissioned institute (similarly with specification of their telephone number) using the reply envelope provided for this purpose; alternatively, they can notify the required contact details via a specified telephone number or e-mail address of the scientific institute. The actual surveys for test candidates and driving instructors are subsequently to be conducted by telephone by staff of the institute. The proposed procedure would enable the sample to be recruited independently of the authorities, driving schools and those companies and organisations whose services are to be assessed within the framework of the survey.

How many test candidates and driving instructors need to be questioned within the framework of a nationwide study in order to permit representative statements on satisfaction with the practical driving test? According to FRIEDRICHS (1990), the sample size for a simple random sample (= n) can be calculated on the basis of proportional values with:

$$n = \frac{t^2 \times p \times q}{e^2} .$$

In this formula, “t” represents the Z score of the standard normal distribution and describes the desired degree of certainty for the planned statements. If “t” is assigned a value of 1.96, the degree of certainty is 95 per cent. This equates to a 95% confidence interval, which can be understood to mean that, within a large random series, only five per cent of the samples are likely to return values which lie outside the sought value range. The “e” value in the formula defines the maximum permissible deviation between the values of the sample and those of the parent population. In social research, a degree of certainty between 95 and 99 per cent and a margin of error of six per cent (e = 0.03) are considered acceptable. The values for “p” and “q”, finally, characterise the assumed distribution of unknown parameters in the parent population (as percentages); it thus applies that p + q = 1.0 (i.e. 100 per cent). Taking the (from the statistical perspective) least favourable case, in other words equal distribution of the parameters in the parent population, “p” and “q” must each be entered into the above formula with a value of 0.5; this constitutes the least possible risk with regard to the prerequisites for calculation.

If the required sample size “n” is calculated as above, it is found that at least 1,067 test candidates and 1,067 driving instructors must reply to the survey in order to satisfy the defined minimum standards. These sample sizes apply only for single-stage random selections and for a large parent population (for the derivation of the aforementioned formula and example calculations, see also LOHSE et al., 1982, p. 50ff.); given the particular practical relevance of the study results and the complexity of the parameter structures in the parent populations, it is imperative to observe the calculated minimum sizes. These minimum sample sizes are also required to enable differentiated analysis of the customer survey results at the level of an individual Technical Examination Centre, taking into account test success and membership in particular subpopulations (test region, gender,

[145] The communication of personal data in the interest of road safety is covered by the German Road Traffic Act (Straßenverkehrsgesetz). In addition, it is naturally necessary to take into account all relevant data privacy regulations.

age, education, experience); for methodical reasons, a minimum group size of 30 persons should be observed for all combinations of parameters to be evaluated (BORTZ & SCHUSTER, 2010). To achieve net samples of 1,067 test candidates and driving instructors, taking into account an expected systematic dropout rate of 30 per cent (e.g. failure to establish contact, declined or prematurely terminated interview) and a quality-neutral dropout rate of similarly 30 per cent (e.g. incorrect telephone number, maximum number of contacts reached), a gross starting point of n = 2,000 is to be specified in each case.

With regard to the recruitment and motivation of survey participants, it is to be weighed up, finally, whether the test candidates and driving instructors who take the trouble to participate in the survey should be rewarded in the form of a prize draw or other appropriate incentives. The survey conducted in conjunction with process evaluation for the nationwide pilot implementation of the model "Accompanied driving from age 17" (FUNK & GRÜNINGER, 2010), for example, was combined with a prize draw offering a new car to the winner, alongside an assortment of vouchers for petrol and other goods and benefits.

5.4.5 Product audits

As was already mentioned briefly above, the term "audit" is understood to refer to a study method which is used to assess (service) processes in respect of their compliance with demands and guidelines; such audits are performed, for example, within the framework of corporate quality management. By way of an audit, the current situation is analysed and compared with a desired target situation (e.g. with regard to quality specifications or expectations), in order to identify and thus be able to eliminate any need for optimisation. Audits are performed by specially trained experts in the field concerned ("auditors"); if these experts are staff of the (service) company itself, the assessment process is described as an "internal audit", otherwise as an "external audit". As a general rule for internal audits, the auditors must not themselves be directly responsible for realisation of the activities to be audited or for the staff assigned to perform those activities, so as to avoid all prejudice and conflict of interest and to guarantee an independent objective assessment. If the auditors view the services provided from the perspective of the customers and their expectations, their work is termed a "product audit". A "system audit", on the other hand, serves to assess the suitability and effectiveness of the structures and stipulations of a quality management system (MEFFERT & BRUHN, 2009).

In the case of the practical driving test, external system and product audits ("evaluations" in the officially used terminology) are performed by the so-called Evaluation Agency of the BASt (see Chapter 5.4.1); at the same time, the Technical Examination Centres themselves perform supplementary internal product audits in accordance with the applicable regulations (STURZBECHER, BIEDINGER et al., 2010). The conditions for realisation of these internal product audits were discussed with leading representatives and quality officers of the Technical Examination Centres in the aforementioned exploratory meetings at the BASt. As one outcome of these meetings, it can be noted that the internal auditors in the Technical Examination Centres follow procedures described in auditing manuals specific to the individual test organisation. These manuals contain information on the responsibilities for auditing, on the training and appointment of auditors, and on the planning, preparation, realisation and evaluation of audits. In addition, report forms and data sheets are provided in the manuals to enable documentation and analysis of the results of internal product audits. According to the results of both the exploratory meetings and a document analysis of the various manuals, the product audits performed in the different Technical Examination Centres are found to use similar methods and examine similar elements of content; even so, the desirable uniformity of audit design is yet to be achieved in respect of their core content and methods.

All Technical Examination Centres distinguish between central and decentralised internal audits. Organisation of the central internal audits is entrusted to the quality management officer of an individual Technical Examination Centre. He selects the auditors, elaborates plans for their audit work, and issues the corresponding assignments for regular and special ad hoc audits. The auditors are driving test examiners with appropriately long professional experience in the conducting of practical driving tests. The intervals at which the different Technical Examination Centres perform internal audits with each of their driving test examiners vary between one and four years.

In accordance with the audit manuals and annual plans issued by the quality management officers, the auditors agree appointments for the product audits with the regional administrative offices of the Technical Examination Centres. The regional administrative offices subsequently inform the driving schools of the forthcoming product audits; this step is necessary because the test candidate is entitled to refuse consent for his driving test to be subject of a product audit. For these pre-announced internal audits, the work rosters of the driving test examiners concerned are arranged such that preparatory and follow-up meetings can be held before and after the practical driving test. The special ad hoc internal audits which are performed in case of presumed deficits in respect of quality-compliant test realisation, and in particular pursuant to customer complaints, by contrast, are not usually announced in advance and may also be performed by the direct superior of the driving test examiner concerned. Such decentralised internal audits are also performed as a regular complement to the central audits by some Technical Examination Centres.

The results of internal product audits are documented by way of organisation-specific audit reports. The report forms used to assess the behaviour of the examiner during the practical driving test differ in certain aspects of their content criteria and recording methods (e.g. two- or three-level rating scales) from one Technical Examination Centre to another. After the test, an evaluation meeting takes place between the auditor and the audited driving test examiner; the auditor uses this meeting to explain the audit results recorded in his audit form. If any deviations from legislative provisions or internal guidelines are determined in the course of realisation or subsequent assessment of the audited test, the auditor suggests appropriate development measures to optimise the testing competence of the examiner concerned; the proposed measures are similarly documented in the audit report, which is then signed by the driving test examiner and the auditor upon completion of the audit. The audit reports are sent to the quality management officer for central evaluation and statistical analysis. There is no linking of the audit results for a driving test examiner to other personal data related to the quality of his work as an examiner (e.g. data from a systematic analysis of the test results for practical driving tests conducted by the particular examiner; see above). Generally, the ad hoc decentralised audits are only evaluated at regional level.

Against the background of the described practice for internal product audits, it is now important to discuss the fundamental, methodical and content-related requirements to be met by audits, and to outline corresponding potential for optimisation. The following basic positions can be taken as a starting point:

1. For the driving test examiner, product audits represent a process-oriented work sample, in the same way that the practical driving test can be viewed as a process-oriented work sample from the perspective of the test candidate. Both work samples are judged by way of systematic behaviour observation. Accordingly, all methodical demands which STURZBECHER, BÖNNINGER and RÜDEL (2010) derive from the methodical nature of the practical driving test – as a work sample in combination with systematic behaviour observation – as prerequisites for the safeguarding of test and result quality apply equally to product audits. This includes not least the requirement that content-related and methodical standards must exist as a basis for the realisation and evaluation of product audits: Without standards, there is no logical foundation for quality assurance measures or management decisions (ZOLLONDZ, 2002). These standards include:
   - Demand standards (specifications relating to the behaviour expected of the audited driving test examiner when conducting and assessing the test; these specifications correspond closely to the demand standards for the practical driving test and are furthermore founded in the adaptive, circular strategy for test implementation)
   - Observation categories (specifications relating to those aspects of testing to which the auditor must pay particular attention when observing the behaviour of the driving test examiner)
   - Assessment criteria (content-related and methodical specifications to indicate how the behaviour of the driving test examiner is to be judged, e.g. specification of the aspects of behaviour which the examiner must display, the required quality of behaviour, the scale levels to be used by the auditor to assess the examiner's behaviour, and the conditions under which each assessment level is to be recorded
   - Decision criteria (content-related and methodical specifications on how the assessments are to be summarised and interpreted, for example with regard to the necessity of development measures to improve the testing competence of the driving test examiner).

   In addition, in exactly the same way as the practical driving test, the product audits must also fulfil the classic primary and secondary quality criteria (above all objectivity, reliability and validity).

2. As described above, product audits assess the quality of service processes from the perspective of customers and on the basis of the latter's expectations. The customers for the practical driving test – in the figurative sense – are the legislator and the test candidate. The legislative provisions (e.g. Examination Guidelines, BASt evaluation requirements) thus represent a first starting point for the deliberations to derive content-related quality criteria for the product audit. At the same time, however, it is possible to take up the content-related quality criteria defined for the customer survey to be completed by test candidates (see Chapter 5.4.4), as these criteria reflect customer expectations and were elaborated by way of a scientifically supported discursive process involving the authorities, driving test examiners and driving instructors (STURZBECHER & MÖRL, 2008).
3. At the beginning of the present chapter, it was explained that an evaluation system for the optimised practical driving test must follow a multi-perspective approach, i.e. it must be ensured that the judgements of test and assessment quality given by the different groups (e.g. auditors, test candidates, driving instructors) are comparable and complementary. This is only possible if the content-related quality criteria underlying these judgements are identical in essence (though this does not exclude the consideration of further perspective-specific criteria) and the same methodical scales (e.g. rating scales) are used. The resulting strategy for elaboration of an optimised methodical instrument for the product audits thus begins with examination of the quality criteria defined for the customer survey to determine their usability as quality criteria for a product audit. The criteria and indicators which are deemed meaningful in the given context are then to be supplemented with further criteria (e.g. professionally adequate design of the test elements) whose assessment only seems relevant from the perspective of field experts and the test organisations. A corresponding proposal with quality criteria for the product audits is to be found below.
4. The overarching objective of the product audits is to investigate whether the driving test examiners observe the formulated demand standards, observation categories and assessment and decision criteria in their
   - test planning (e.g. preferably multiple consideration of all driving tasks, adaptive test strategy) and
   - observations, assessments and interpretation (in the sense of a test decision) of test performance.

Insofar as the auditor masters these standards – which must be ensured under all circumstances by way of corresponding professional experience and demanding further training – and the standards are likewise fulfilled by the driving test examiner in an audited driving test, both the auditor and the examiner should record the same test assessments (ideally on both the event- and competence-oriented levels) and the same test decision. In other words: When an auditor and a driving test examiner observe and assess the same practical driving test, the consensus in their assessments, and thus the inter-rater reliability, should be high. One proviso which must nevertheless be taken into account here is the fact that the auditor is performing a “double observation”: On the one hand, he must pay attention to the behaviour of the examiner, as the primary subject of the audit. At the same time, however, he must also observe and assess the behaviour and driving performance of the test candidate, because an own assessment of the candidate's performance forms the basis for judgement of the professional adequacy of the examiner's decisions. Compared to the examiner, this double burden limits the cognitive resources available to the auditor to observe and assess candidate behaviour, and could thus influence the results of this assessment above all on the event-oriented level: It cannot be excluded that certain driving errors or instances of above-average performance could escape the attention of the auditor. Furthermore, the auditor is required to sit behind the test candidate and thus observes the latter's behaviour from a different (possibly unfavourable) perspective; in some cases, this could also reduce the level of consensus between examiner and auditor in their event-oriented assessments of test performance. It can be assumed, however, that the effects of such indistinctness will be reduced with increasing abstraction of the assessment level, and that the inter-rater reliability will improve accordingly: If a test is conducted in compliance with the specified standards, a high degree of consensus should thus be found in the competence-oriented assessments (observation categories), and all the more so with regard to the test result. Within the future framework of internal product audits for the optimised practical driving test, therefore, the degree of consensus between the auditor and the audited driving test examiner should at least be determined and analysed in respect of the test decision and the assessments of the five elements of competence represented by the observation categories “Traffic observation”, “Speed adaptation”, “Vehicle positioning”, “Communication” and “Vehicle control/Environment-aware driving”; the applicable measures of inter-rater reliability were already described in the section “Instrumental evaluation” (see Chapter 5.4.2).

5. As already described, both the practical driving test and the product audits are – from the methodical perspective – instances of work samples and systematic behaviour observation which take place in a test vehicle; they are thus subject to the same documentation requirements and conditions. It is thus recommended to develop an “electronic audit report” – equivalent to the new electronic test report and on the basis of identical methodical advantages for the documentation and further processing of observation and assessment data – and to test its usability in the same manner as the aforementioned feasibility study. All the hardware and software demands to be met by the electronic test report for the optimised practical driving test are equally applicable to the electronic audit report. Such an electronic audit report could automatically integrate relevant data from the electronic test report, and the inter-rater reliabilities could also be calculated automatically. For reasons of economy and practicability, the electronic test report and the electronic audit report should share a common hardware and software base (see Chapter 4).

Independently of the described methodical circumstances and the resulting proposals for the future content-related and methodical design of internal product audits for evaluation of the optimised practical driving test, it seems expedient – given the expectation of a nationally uniform auditing procedure and in the interest of audit equality – to seek essential standardisation of the partially divergent implementation and assessment standards for product audits in the individual Technical Examination Centres (including the procedural instructions and documentation specifications in corresponding manuals), at least in the medium term. This common core, which concerns above all the fundamental quality standards prescribed through legislation, could then be expanded to include further rules specific to a particular test organisation (e.g. requirements relating to the appearance and manner of the driving test examiner, as they arise from the corporate image of each test organisation). The content architecture of the common core must reflect the contents of a common electronic audit report; these contents, in turn, should be derived from the underlying legislative provisions (Examination Guidelines, BASt evaluation requirements) and the desirable content-related and methodical parallels between the product audits and customer surveys (see above). If these requirements and the proposed procedure are accepted, then the following content-related quality criteria should be covered by future internal product audits (the criteria in italics possess content equivalents in the customer surveys):

- Verification of the proper condition of the test vehicle
- Satisfaction with the outward appearance of the examiner
- Satisfaction with the welcome and introduction
- Verification of the candidate's identity
- Check of the candidate's training certificates
- *Safety checks, vehicle function checks, general skills*
- *Satisfaction with the meeting place for the test with regard to accessibility*
- *Satisfaction with the meeting place for the test with regard to traffic density*
- *Punctuality of the test*
- *Satisfaction with the measures taken to reduce test stress and test anxiety*
- Interaction between examiner and driving instructor (e.g. distracting)
- *Satisfaction with the friendliness of the examiner*
- *Satisfaction with the prior explanation of test procedures*
- *Satisfaction with the clarity of driving instructions*
- *Satisfaction with the timely communication of driving instructions*
- *Satisfaction with the structure and course of the test*
- *Satisfaction with the realisation of the basic driving manoeuvres*
- *Satisfaction with the assessment of the basic driving manoeuvres*
- *Satisfaction with the assessment of the test drive*
- *Satisfaction with the explanations of observed errors*

- *Satisfaction with the remarks on possibilities for further improvement*
- *Satisfaction with the mentioning of good performance*
- *Satisfaction with the answering of questions from the test candidate*
- *Overall satisfaction with the feedback from the examiner*
- *Satisfaction with the measures taken to establish a relaxed atmosphere for the test*
- *Test result*
- *Overall satisfaction with the examiner*
- Overall assessment of test realisation in accordance with the statutory requirements
- Company-internal specifications, and development needs of the examiner (possibly with recommendations for appropriate measures).

To enable differentiated analysis and interpretation, and as a basis for efficient reporting on the results of the internal product audits, it seems wise to acquire also the following additional information (insofar as the documentation of internal product audits is to be passed on for central evaluation, certain data will need to be pseudonymised):

- Date and place of the test
- (Pseudonymised) data to distinguish the Technical Examination Centre, the driving test examiner and his organisational unit
- (Pseudonymised) data to distinguish the test candidate's driving school
- Driving licence class for which the test is conducted
- First/repeat driving test
- Age and gender of the driving test examiner
- Gender of the test candidate
- Assessment of the language competence of the test candidate
- Assessment of the test anxiety of the test candidate
- Assessment of the friendliness of the test candidate (or his displayed aggressiveness and arrogance).

Following the proposals presented for optimisation of the internal product audits, two possibilities emerge for evaluation and use of the audit results; these possibilities are associated on the one hand with scientific evaluation of the practical driving test, and on the other hand with company-internal quality management in the Technical Examination Centres:

1. For evaluation purposes, the anonymised audit reports should be sent to a central office (e.g. to the TÜV ĐEKRA arge tp 21 working group or to a scientific institution appointed by the latter) for corresponding analysis. These scientific analyses should be realised at the same regular intervals as the customer surveys; the analysis results relating to the product audits and customer surveys are then to be combined and compared within the framework of scientific evaluation of the optimised practical driving test and the reporting on test quality. This would also comply with the objective of a multi-perspective and multi-method evaluation system (see above): The quality of the optimised practical driving test would be assessed both by professional experts (the auditors) and customers (test candidates and driving instructors); at the same time, the use of different observation and survey methods with their individual methodical strengths and weaknesses would augment the significance of their respective results.
2. Within the framework of internal quality management in the company or test organisation, individual driving test examiners could be provided with pertinent information on their test behaviour, firstly from evaluation of the product audits, and secondly from analysis of the results of all practical driving tests conducted by the examiner concerned (see Chapter 5.4.3). Comparison of their different personal results from the multi-method evaluation – particularly when viewed over time and in relation to the findings for reference groups – enables driving test examiners to assess the status and development of their own test competence on the basis of robust empirical data and to draw appropriate conclusions with regard to optimisation of that competence. This could lend significant impetus to the process of quality development, especially if the responsible managers focus less on their supervisory duties, and more on promoting the development motivation and self-evaluation potential of their staff: It is true that management can be seen as the “motor” of quality development (DEMING, 1982), but the most important quality resource of a company is its staff (PAGE, 2000; ZOLLONDZ, 2002).

Once the content-related and methodical standards have been reviewed, tested and implemented successfully for internal product audits, it seems logical for the aforementioned Evaluation Agency at the BASt to examine these standards with regard to their suitability as a basis for external product audits in the context of the optimised practical driving test. In addition, observance of any changed specifications for the internal product audits should be monitored by way of the external system audits.

5.5 Summary

The methodically demanding evaluation system outlined for the optimised practical driving test in the present chapter of this report is the combined outcome of two fundamental lines of development, both of which are currently still in progress: Firstly, the theoretical description of a content-oriented and methodical test architecture, a task which can only be accomplished jointly by professional experts and scientists from the associated fields, and secondly – building upon the former – the elaboration and introduction of a computer-assisted means for the differentiated assessment and documentation of test performances. The importance of the first step can hardly be overestimated: “Experience, without theory, teaches management nothing about what to do to improve quality ...” (DEMING, 1986, p. 19). The second step is equally imperative if the practical driving test is to fulfil its control function within the system of novice driver preparation and is long overdue, but has only become feasible thanks to the advances in computer technology.

The described and recommended evaluation system comprises four elements, namely instrumental evaluation, analyses of test results, customer surveys and product audits; it is thus fully in line with the applicable national and international legal frameworks, and takes up practical approaches which are emerging in Germany and other countries in progressive fashion. The proposed implementations of the constituent elements are supported by fundamental methodical and methodological standards for diagnosis and evaluation from the humanities and business science. By way of these elements and their interactions, the contents, processes, conditions and results of the practical driving test experience constant critical methodical reflection, can be adapted to the practical needs of novice driver preparation and driving licence testing, and are overall the subject of a process of continuous improvement in the sense of DIN 9001 (DIN, 2008b). The outlined evaluation system is thus at the same time a dynamic system which – in addition to the benefits of its multi-perspective and multi-method approach – can adapt flexibly to changes in the test conditions (e.g. advances in driver assistance systems or electromobility). To enhance this flexibility with regard to further development of the practical driving test, the test standards and evaluation findings should be made available in the form of an electronic psychological process manual (see Chapter 5.4.2); furthermore, the work and responsibility structures of all institutions involved in driving licence testing are to be reformed, alongside optimisation of the related processes and coordination procedures. To this end, proposals were gathered within the framework of the present project, discussed with the professional community and set in writing in the draft for a “System Manual on Driver Licensing (Practical Test)”. It remains a task for further research and development projects to demonstrate the soundness of these reform proposals: A subsequent revision project of the BASt, for example, is to verify the feasibility of the organisational and technical optimisation plans in selected model regions; at the same time, the present evaluation concept and the underlying electronic test report are to be subjected to critical methodical analysis. Both projects build upon the reformed theoretical and methodical architecture of the optimised practical driving test.

6 Driver assistance systems and the optimised practical driving test

6.1 Overview of driver assistance systems

The human factor can be considered the principal source of risk in motorised road traffic: According to the GIDAS database (German In-Depth Accident Study), around 90 per cent of traffic accidents can be attributed to “human error”; this includes inattentiveness, misinterpretation of the traffic situation and incorrect reaction on the part of the driver responsible for causing the accident. Technical defects, on the other hand, are only determined as the accident cause in a little under one per cent of cases (STATISTISCHES BUNDESAMT, 2010). The use of driver assistance systems could help to avoid certain human driving errors or else reduce the severity of their consequences.

The growing prevalence and diversity of driver assistance systems is also reflected in the equipment features of the vehicles used for lessons and practice within the framework of novice driver training and in the test vehicles used for the practical driving test. This gives rise to the question as to how the use of modern driver assistance systems impacts the acquisition and testing of necessary driving competence, since driver assistance systems are increasingly taking over (partial) driving tasks which were in the past performed by the driver himself. Before seeking to answer this question in the following, a number of selected driver assistance systems are to be presented and described briefly in terms of their benefits and limitations. This is to be followed by an overview of the essential legal issues connected with the use of driver assistance systems in motor vehicles in general and in test vehicles in particular. Finally, it is to be discussed how driver assistance systems influence driving competence acquisition, and what this means in the context of the practical driving test.

Driver assistance systems are supplementary electronic systems which are installed in motor vehicles and there offer various forms of support to the driver in his handling of driving tasks. One important objective is to enhance driver safety (safety systems), and some systems also intervene directly in vehicle behaviour (intervention systems). Others serve to relieve the driver (comfort systems) and provide information on the condition of the vehicle or the surrounding traffic situation (information systems). Depending on its individual functions, the purpose of a driver assistance system is thus to inform, warn, recommend or intervene. Despite these different functions, however, all driver assistance systems have one thing in common: They are intended to offer the driver the required assistance without taking away his fundamental responsibilities in road traffic. This assistance refers above all to widening of the performance limits of human perception – as the basis for hazard recognition and hazard avoidance – and to support in hazardous situations (WINNER, HAKULI & WOLF, 2009).

The aims of driver assistance systems are thus both to improve road safety and to raise the level of driver comfort. To illustrate this, “Comfort” and “Safety” are often depicted as the two poles of a benefit dimension (see Fig. 16): Safety systems such as emergency brake assist, for example, can be expected to contribute more noticeably to a reduction of the numbers of persons killed or injured in road traffic than a parking assist system, which places the comfort aspect in the foreground. In addition, driver assistance systems are frequently assigned to a dimension with the two poles "Active" and “Passive” to represent the level of assistance provided: Active systems are in part able to perform certain driving tasks independently, whereas passive systems merely make information available to the driver, who must then himself translate this information into corresponding actions.

Differing opinions can be found in literature as to the extents to which certain driver assistance systems serve driver comfort and road safety, and whether they are essentially active or passive (DE MOLINA, 2008; KNOLL, 2009). It is generally agreed, however, that the two dimensions are inseparable, and that they apply in combination, albeit in varying proportions, to practically every driver assistance system (BELZ, HÖVER, MÜHLENBERG, NITSCHE & SEUBERT 2004). The development trends in driver assistance systems show that the distinction between active and passive safety is becoming increasing blurred, and that the systems blend ever more seamlessly into each other; correspondingly, technology developments can be seen to pursue a “concept of integrated safety” (DAIMLER, 2009).

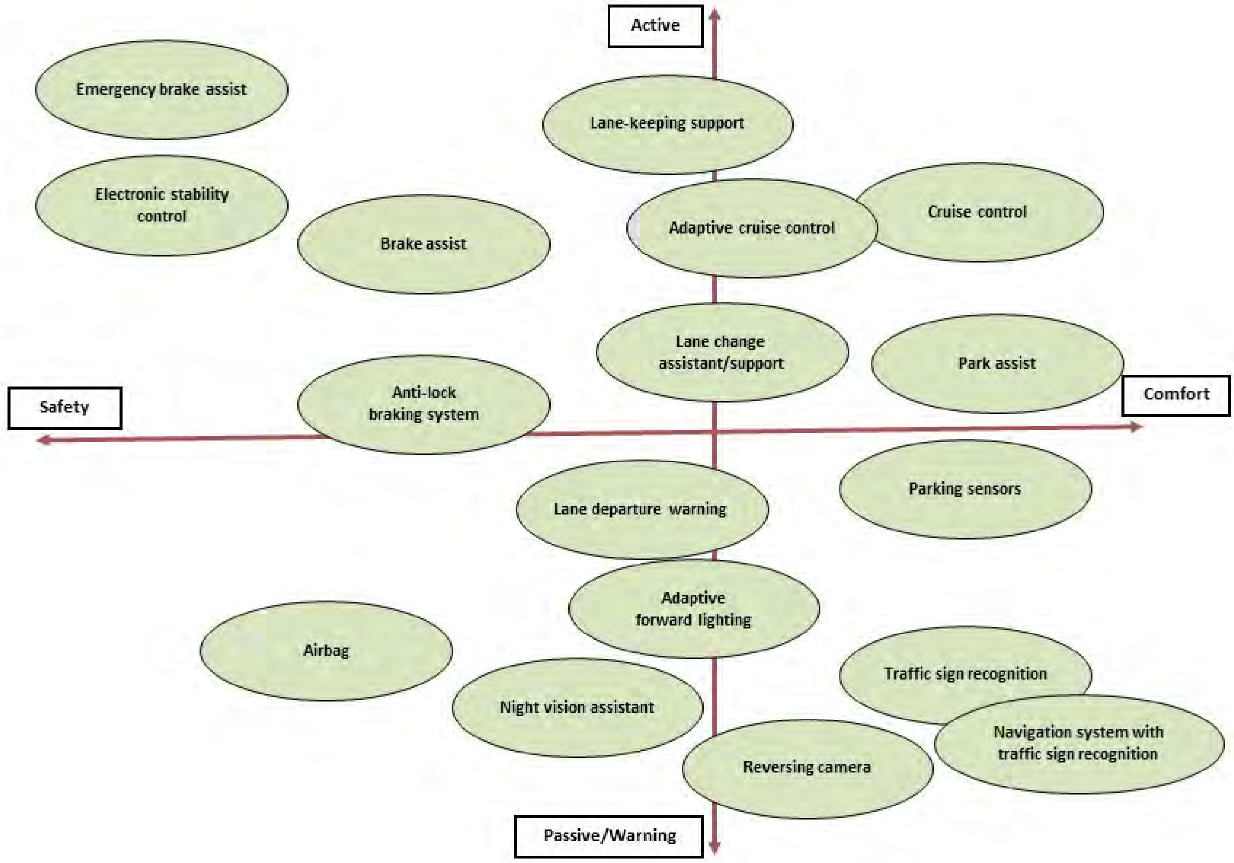

Fig. 16: Functions and degree of support of driver assistance systems (based on BANDMANN, 2008)

Driver assistance systems can be categorised according to various criteria. The following overview illustrates a selection of the most frequently used categorisation models for driver assistance systems:

- Safety systems versus comfort systems
- Passive versus active systems
- Warning and information systems versus intervention systems
- Overridable versus non-overridable systems
- Assistance on the navigation, manoeuvring or stabilisation level.[146]

The most appropriate model for the categorisation of driver assistance systems in a given case depends on the particular study focus. For the following analysis of the influence of driver assistance systems on realisation and assessment of the practical driving test, for example, the important aspects are the possibilities to override the individual driver assistance systems and the content-related action level on which assistance is provided to the driver. Clear assignment of every existing driver assistance system to a certain category, however, is not possible for many category models, i.e. some assistance systems belong to several categories at once (GELAU, GASSER & SEECK, 2009).

6.2 Function principles of selected driver assistance systems

This chapter is to present brief descriptions of the function principles of selected driver assistance systems which are of particular relevance in the context of the practical driving test. At the same time, each of the driver assistance systems is to be categorised according to the aforementioned descriptive dimensions.

As general characterisation, it can be said that driver assistance systems become effective in conjunction with the handling, control (e.g. acceleration, braking) or signalling functions of a motor vehicle, or else warn the driver in critical situations by way of suitable human-machine interfaces. The functions of driver assistance systems are activated either by the driver himself, by sensor data from the vehicle systems (e.g. wheel speeds, yaw

[146] This categorisation of driver assistance systems according to the content-related action level on which assistance is provided is based on the driving behaviour classification of DONGES (1982), which was already referenced in Chapter 2 of the present report and distinguishes between three action levels placing different demands on driving behaviour.

rate, lateral acceleration, steering angle) or by information from the environment of the vehicle. Such environment data are collected with various different kinds of sensors, including ultrasonic sensors (e.g. parking sensors), radar sensors (e.g. safe distance alert) and cameras (e.g. lane departure warning, blind spot monitor). In the following, a few of the driver assistance systems with functions considered relevant for the practical driving test are presented in more detail (STURZBECHER, BÖNNINGER, RÜDEL & MÖRL, 2011).

Adaptive cruise control (ACC):
Adaptive cruise control serves to maintain a preselected desired speed (as in the case of a speed limiter) and, by way of automatic acceleration and braking, also the following distance to a preceding vehicle set by the driver (automatic distance control). If more substantial intervention becomes necessary, e.g. because the preceding vehicle brakes suddenly, visual and acoustic signals warn the driver accordingly. ACC is a driver assistance system which supports the driver on the manoeuvring level. The driver can override ACC and influence speed himself at any time.

Adaptive forward lighting (AFL):
The purpose of adaptive forward lighting is to increase road safety by improving illumination of the driver's field of vision when turning at junctions or driving through bends at night. A distinction is made between static and dynamic systems. Static forward-lighting control is achieved by activating a separate lighting function for the negotiation of corners. Dynamic forward-lighting control, on the other hand, follows the steering actions of the driver. AFL is a driver assistance system which makes new information available to the driver and provides support on the manoeuvring level.

Anti-lock braking system (ABS):
ABS is intended to guarantee tracking stability and maintain the steerability of the vehicle under heavy braking and during abrupt braking manoeuvres. The system was developed because drivers are generally hesitant to brake with maximum force, but then to release the brake in order to enable a controlled evasive swerve around an obstacle should the wheels lock. ABS is a non-overridable driver assistance system which supports the driver on the stabilisation level.

Brake assist system (BAS):
When the system detects a necessity for emergency braking, the brake assist system ensures that the maximum braking force is applied instantly if the brake pedal is depressed quickly, but with insufficient pressure. This achieves a significant shortening of the braking distance. Brake assist is a non-overridable driver assistance system which supports the driver on the manoeuvring level.

Electronic stability control (ESC):
Electronic stability control helps the driver to retain control over the vehicle if it threatens to skid (e.g. when swerving to avoid an obstacle). The system detects the danger of skidding and reacts accordingly to compensate the loss of traction; it functions independently within the applicable physical limits and corrects the engine power and braking forces as necessary to hold the vehicle stable on its driving line. ESC is a driver assistance system which supports the driver on the stabilisation level and cannot be overridden.

High-beam assistant:
The high-beam assistant improves visual orientation when driving in the dark and relieves the driver of the need to dip the headlights himself. The sensors already detect sources of glare, other road users and built-up areas from a long distance. The high-beam headlights are switched on whenever the traffic situation permits or requires; dipping is also accomplished automatically in accordance with the specified prerequisites. The high-beam assistant is a driver assistance system which makes new information available to the driver and provides support on the manoeuvring level. It can be overridden at any time.

Night vision assistant:
The night vision assistant enables the driver to see further into the distance when driving in the dark. An active night vision system sends out infrared light and processes the reflected light by way of a special camera to produce a black-and-white image. A passive system, by contrast, has no infrared light source of its own and instead captures the infrared radiation emitted by objects in its path; the incoming signals are likewise processed into a black-and-white image (cf. thermographic cameras). The night vision assistant is a warning and information system which can be assigned primarily to the comfort-oriented category and supports the driver on the manoeuvring level.

Park assist systems:
Park assist systems facilitate parallel and/or perpendicular parking, and support the driver when leaving a parking space. They can generally be divided into manual and semi-automatic systems. The first level of system (distance alert) provides for environment and obstacle detection with acoustic or visual collision warnings (“parking sensors"). At the second level, the system not only warns of obstacles, but also informs the driver as to the suitability of a particular space for parking. The most advanced variants are park assist systems

which offer the driver support for the actual parking manoeuvre, and in the case of a semi-automatic system even guide the driver into a parking space along a calculated parking trajectory ("intelligent park assist"). The driver retains responsibility for the vehicle during the whole manoeuvre and must pay corresponding attention to the surrounding environment. Park assist systems are comfort-oriented systems which support the driver on the manoeuvring level. They are available as warning and information systems, but may also support the driver by way of intervention. Intervening park assist systems can be overridden at any time if the driver deems a different steering action to be necessary. In this case, the park assist system is deactivated.

Reversing camera:
A reversing camera projects the rearward view from the vehicle onto a dashboard monitor screen. The system thus offers the driver support for all manoeuvres which involve reversing, and here especially reverse parking. The driver must nevertheless continue to observe the surroundings of the vehicle in order to avoid all risk of endangerment or accident. A reversing camera is a warning and information system which supports the driver on the manoeuvring level.

Lane departure warning (LDW):
This system warns the driver if the vehicle unintentionally departs the chosen lane on a road with lane markings. The warning can be issued as a visual (flashing of a warning lamp), acoustic (warning tone via the car speakers) and/or haptic signal (vibration of the steering wheel or driver's seat). If the driver sets the turn indicators, the warning is suppressed, as this indicates deliberate crossing of the lane markings. Lane-keeping support is an extension of the LDW system and takes active steps to influence steering as soon as it detects the likelihood of the vehicle departing its current lane unintentionally. Lane departure warning can be classified as a warning system which supports the driver on the manoeuvring level. A lane-keeping support system also intervenes in vehicle control, but can be overridden at any time.

Lane change assistant:
The purpose of the lane change assistant is to monitor the traffic situation in the neighbouring road lanes and to warn the driver of potential collisions when changing lanes. The system is activated by setting the turn indicators (in contrast to lane-keeping support, which is deactivated in such cases) and warns the driver so as to avoid collisions with (approaching) vehicles in the neighbouring lane. The warnings are communicated as visual (warning lamps, usually in the vicinity of the side mirrors), acoustic or haptic signals (vibration of the steering wheel, driver's seat or turn indicator lever).

Lane change support represents an extension of the lane change assistant and detects obstacles in the blind spot. If vehicles are present in the blind spot or else seen to be approaching from behind, a red triangle appears in the corresponding side mirror (information level). If the driver nevertheless sets the turn indicator or commences a lane-changing manoeuvre, the red triangle begins to flash and a warning sounds or else the steering wheel vibrates (warning level). The lane change assistant supports the driver on the manoeuvring level; the lane change support function also intervenes in vehicle control, but can be overridden at any time.

Traffic sign recognition:
A camera installed behind the rear-view mirror detects traffic signs along the side of the road. If the camera recognises a speed-relevant traffic sign, this is indicated on the instrument display and the driver is warned to avoid exceeding the speed limit. If the windscreen wipers are activated, the system can also react to speed limits which only apply when the road is wet. In addition to speed limits, some systems also detect the start and end of no-overtaking zones. Traffic sign recognition is a warning and information system which supports the driver on the manoeuvring level. It can be assigned primarily to the comfort-oriented category of driver assistance systems.

The different driver assistance systems have so far been presented in terms of their benefits for road safety. In the following, selected systems are to be discussed also with regard to their safety-relevant functional limitations. These functional limitations are of special significance in the context of novice driver preparation: If such driver assistance systems are active within the framework of driver training, the novice driver must also be made aware of the traffic and hazard situations in which a particular system is unable to provide the expected support, and when it may even be counterproductive to hazard avoidance. Without such knowledge, there is a risk of misinterpretations and incorrect driving behaviour, which could in turn impair road safety. Consequently, selected (and still to be defined) knowledge of the use, functions and limitations of driver assistance systems must be conveyed in appropriate form (theory classes, practical driver training, independent theory learning) as part of novice driver preparation and taken into account accordingly in subsequent testing (knowledge test, driving test).

The functional limitations of driver assistance systems are firstly a consequence of their inherent technical limitations, which prevent the reliable and gapless detection – and even more so the prediction – of all conceivable events in a system as complex as road traffic; these technical shortcomings impair the “technical reliability”. For this reason, it is also not possible to provide technical aids to support the driver in the mastering of every imaginable situation, or to perform complex driving tasks automatically. At the same time, technical defects can influence the technical reliability of driver assistance systems. Human error and incorrect behaviour on the part of the driver, furthermore, may lead to certain capabilities of driver assistance systems being left unused, and could even give rise to new safety risks; this phenomenon is discussed under the heading of “human reliability”.

Technical reliability, or in other words the functional reliability of technical systems, is understood as the “ability of an observed entity to perform a required function under given conditions for a given time interval” (VDI 4003, 2007, p. 3). Factors which could exert a negative influence on functional reliability include, for example, system errors, physical wear in the system, the removal or manipulation of system components, and improper technical modifications. In line with the stipulations of Directive 2009/40/EC[147], the precursor to today's Directive 2010/48/EU (Commission directive on regular roadworthiness tests for motor vehicles), the German legislator decided that an inspection of safety-relevant electronically controlled systems was to be incorporated into the prescribed roadworthiness test (“Hauptuntersuchung”) from 1st April 2006, so as to ensure preservation of the original safety level over the whole service lifetime of a vehicle.[148]

Technical reliability may also be affected by special technical system limits, as illustrated by the following examples. One of the particularly demanding tasks for driver assistance systems is the recognition and classification of objects. Many of the available sensors, however, are not yet able to guarantee reliable detection in all vehicle states and under all weather conditions (GRÜNDL, 2005): The radar sensors of an adaptive cruise control system (ACC), for example, are especially susceptible in this respect in unfavourable weather conditions (e.g. rain or snow). On a curved road, furthermore, the system may respond to vehicles which are not travelling in the same lane as the vehicle with ACC; this could lead to braking which is not actually necessary in the current traffic situation. Lane departure warning only functions at higher speeds outside built-up areas, and at the same time relies on clear marking of the road lanes. A lane change assistant is likewise dependent on readily detectable and constant lane markings; interruptions of the markings due to dirt on the road and apparently ambiguous situations (e.g. a normal white lane marking alongside the temporary yellow markings used in case of road works) cannot be detected and interpreted reliably with regard to their road safety relevance. In addition, the lane change assistant can only issue reliable warnings up to a certain difference in speed between the vehicles.

Similar system limits apply equally to driver assistance systems which serve to stabilise the vehicle. One such effect is the fact that the braking distance is longer for a vehicle with ABS on a loose surface (e.g. sand, snow): Without ABS, the locked wheels would cause a wedge of loose material to be formed in front of the tyres and thus further increase the braking force. In the case of ESC, high driving speeds and a relatively narrow curve radius could exceed the physical limits for dynamic stability control, i.e. the ESC system would no longer be able to prevent skidding. Moreover, the control limits for both ESC and ABS are dependent on the friction coefficient between the road surface and the vehicle tyres: This could be reduced significantly by unfavourable road surfaces (e.g. cobblestones) or weather conditions (e.g. rain or snow), leaving only a restricted opportunity for compensation by the system. The proper functioning of a brake assist system is dependent on a fast switch from “foot on the accelerator” to “foot on the brake” in order to detect the driver's intention to perform emergency braking. While even hesitant pressure on the brake pedal always results in an appropriately enhanced braking force by way of the power braking servo, it is alone insufficient to trigger full emergency braking.

“Human reliability” is understood to mean the “ability of a person to perform a task at an acceptable level under given conditions and for a given time interval” (VDI 4006, 2002, p. 5). Negative influences on human reliability could result, for example, from a reduction or even loss of awareness for the situation, which would be the case if the driver

[147] This EU directive described only minimum requirements and contained no stipulations relating to the testing of electronics in the sense of the German approach.

[148] An electronics or system data test was introduced in 2006 by way of a reform of the Road Traffic Licensing Regulations (StVZO) addressing the scope of roadworthiness tests (41st Amendment Regulations to the StVZO).

were to suffer a stroke or fall unconscious. Undifferentiated knowledge of the function principles of driver assistance systems (e.g. "ESC keeps the vehicle controllable irrespective of the road surface") and exaggerated confidence in their effectiveness (e.g. "I have no need to worry because the car has ABS") are further factors with negative effect on human reliability. Many drivers trust that a system will function in all situations, and are then completely surprised when the system does not respond (e.g. due to sensor failure). This, in turn, may lead to them no longer being able to react in time to prevent an accident.

6.3 Legal foundations for the use of driver assistance systems in motor vehicles in general and in the test vehicle in particular

6.3.1 Legal approval of driver assistance systems

Before driver assistance systems can be installed in motor vehicles in general, and in driving test vehicles in particular, they must be granted corresponding approval. The legal decision on the permissibility of a driver assistance system is above all a question of the extent to which use of the system interferes with autonomous decision-making on the part of the driver (or the driving instructor in the case of practical training and during the practical driving test should it become necessary for the instructor – as the person legally responsible for the vehicle – to intervene in the vehicle handling of the test candidate). Alongside the reliability of driver assistance systems, it is also necessary to discuss the legal conditions under which the use of driver assistance systems can be prescribed by legislation in the interest of public road safety.

Both the Vienna Convention (VC) on Road Traffic of 8th November 1968 and specifically German legislation are significant for decisions on the granting of approval to driver assistance systems. One occasionally presented interpretation of the Vienna Convention[149] concludes that it is not applicable to driver assistance systems. The underlying formal argument here draws on the fact that the formulations "Every driver shall …" and "Every driver of a vehicle shall …" in the relevant provisions refer to behaviour obligations on the part of the driver, and not to the permissibility of particular vehicles (BEWERSDORF, 2003). This standpoint must be rejected, however, as it otherwise implies that it is quite permissible to manufacture and approve the use of vehicles with which the driver is objectively unable to meet the behaviour obligations contained in the Vienna Convention. It is furthermore incompatible with the spirit and intention of the Vienna Convention, whose (minimum) technical requirements (Chapter III) must be read in context with the general behaviour-related provisions (Chapter I). It is only in its entirety that the Vienna Convention is able to achieve its goal of guaranteeing binding minimum standards for international road traffic (cf. appropriately ALBRECHT, 2005, and the opinions expressed in the overwhelming majority of papers); corresponding evidence is to be found in ALBRECHT (2005).

A further line of argumentation against the relevance of Articles 8 and 13 of the Vienna Convention for the approval of driver assistance systems is founded, according to KEMPEN (2008), on the possible translations of the English word "control" (and its French equivalent "contrôler"), though it must be noted here that the Vienna Convention was executed in the contract languages Chinese, English, French, Russian and Spanish, but not in German: The terms "control" and "contrôler" can be translated not only in the sense of "to master" or "to operate", but also with meanings relating to "supervision", "monitoring" and "verification". If the Vienna Convention is then interpreted from the perspective of its overarching aims, the contractual purpose is to provide for common traffic rules in the interest of road safety, and not to block technical advance by erecting barriers to vehicle approval (ibid.). The predominant position in the literature thus assumes that the Vienna Convention assigns an active, commanding role to the driver, rather than a merely supervisory, monitoring function (FRENZ & CASIMIR-VAN DEN BROEK, 2009; ALBRECHT, 2005).

It is generally acknowledged that, alongside purely informational driver assistance systems and those which intervene actively but can be overridden, it is also permissible to provide and use driver assistance systems "which intervene in situations in which the driver himself is no longer able to react accordingly in a timely manner, insofar as the intervention corresponds with the will of the driver." "Furthermore, an automatic emergency braking system is permissible where it is designed purely

[149] The relevant provisions are:
Art. 8. para. 1 VC: Every moving vehicle or combination of vehicles shall have a driver.
Art. 8. para. 5 VC: Every driver shall at all times be able to control his vehicle (…) .
Art. 13. para. 1 VC: Every driver of a vehicle shall in all circumstances have his vehicle under control so as to be able to exercise due and proper care and to be at all times in a position to perform all manoeuvres required of him.

as a 'collision mitigation' system. The system design, however, is here decisive with regard to the assumed will of the driver: The intervention must not be effected until such time that collision avoidance is objectively no longer possible, as inconsistency with the will of the driver is otherwise conceivable" (SEEK & GASSER, n.d.). On the other hand, if a driver assistance system intervenes in vehicle control to prevent exceeding of an applicable speed limit, and if such a system is furthermore non-overridable, then the driver has been deprived of the prescribed full control over his vehicle. Such a driver assistance system would thus not be permissible under the provisions of the Vienna Convention. Conformance with the Vienna Convention could nevertheless be attested within the framework of research projects or field tests with such driver assistance systems, for example if the German legislator were to expressly approve such use as an unambiguously declared exception. This approval would only apply for the domestic road network, however; journeys abroad would not be permissible (FRENZ & CASIMIR-VAN DEN BROEK, 2009).

For the decision as to whether or not the introduction of certain driver assistance systems should be prescribed in law, it is necessary to consider the conformity of the envisaged regulations with Article 2 of the Basic Law for the Federal Republic of Germany (Grundgesetz, GG)[150]: It is there stipulated that interference of any kind in general personal freedoms is only permissible where it is pursuant to a law. Such a law, however, must be reasonable in the given circumstances, and the consideration of this reasonableness must include verification of the following prerequisites (cf. ALBRECHT, 2005):

- The obligation to introduce and use the system concerned must serve a sound public interest.
- The measure must be suitable and necessary.
- The ruling must be fair and reasonable, which requires, in particular, that no less invasive but equally effective alternative measure exists.

To determine the reasonableness of proposed legislation, therefore, it is necessary to weigh up all circumstances which arise from and in connection with use of the driver assistance system in question; the benefits (e.g. accident reduction potential, improved traffic flows, relief for the driver) must be set against the possible disadvantages (e.g. possible new accident risks due to malfunction, manipulation or improper handling of the new system) (cf. ALBRECHT, 2005 and furthermore the decision of the Federal Constitutional Court on the mandatory use of seat belts, BUNDESVERFASSUNGSGERICHT, 1987).

In summary, it can be said that all driver assistance systems which support the driver merely by providing information and do not intervene in control of the vehicle can be deemed generally eligible for approval. The same applies for driver assistance systems which serve to optimise vehicle functions, i.e. systems which entail control intervention, but only to better reflect the will of the driver. Non-overridable systems which intervene only in a pre-crash situation and are there not inconsistent with the will of the driver are similarly permissible. It is only where a driver assistance system intervenes in "normal" vehicle operation, for example to prevent the driver exceeding an applicable speed limit, and this function cannot be overridden, that use is only permissible on the basis of legislation meeting the aforementioned criteria.

6.3.2 Liability issues relating to driver assistance systems

Questions relating to the liability for damage resulting from a road traffic accident also play a role in decisions on the use of driver assistance systems in (test) vehicles. Various laws may be applicable when answering these questions. In accordance with § 7 (1) of the German Road Traffic Act (Straßenverkehrsgesetz, StVG)[151], the owner is liable for damage arising from the operation of his motor vehicle. Where the vehicle incorporates a driver assistance system, he is also liable for all damage attributable to the malfunctioning of such driver assistance systems or to errors in the handling of correctly functioning driver assistance systems. This so-called absolute liability on the part of the owner applies independently of blame; it is sufficient to establish an adequate causal relationship between operation of the vehicle and the damage.

According to § 18 (1) StVG[152], the driver of a motor vehicle is jointly and severally liable for damage

[150] The relevant provisions are: Art. 2 GG: (1) Every person shall have the right to free development of his personality, …. (2) … Freedom of the person shall be inviolable. These rights may be interfered with only pursuant to a law.

[151] § 7 (1) StVG: If, as a result of operation of a motor vehicle, … a person is killed, injured or affected in his health or else material property is damaged, then the owner of the vehicle is obliged to pay compensation for the ensuing damage.

[152] § 18 (1) StVG: In cases covered by § 7 (1), the driver of the vehicle … is likewise obliged to pay compensation for the ensu-

alongside the owner. In this case, however, the question of blame becomes relevant. The obligation for the vehicle driver to pay compensation applies only if the damage was caused by his negligent or wilful behaviour. “Responsibility is excluded if the damage is attributable to a defect of the vehicle (including a defect in a driver assistance system), except where the driver must necessarily be deemed responsible for the arising circumstances, for example because he has ignored warnings by which he could have foreseen the damage. A general reversal of the burden of proof must also be taken into account here; initially, it is always assumed that an obligation to pay compensation exists, and the driver is only relieved of this obligation if he can provide proof that the fault lies elsewhere” (ALBRECHT, 2005, p. 190).

The most important difference between the obligation to pay compensation derived from § 823 of the German Civil Code (Bürgerliches Gesetzbuch BGB)[153] and an obligation to pay compensation founded by §§ 7 and 18 StVG is to be seen in the fact that the driver or owner is only considered liable in accordance with § 823 BGB if he has acted culpably, and if a causal relationship exists between the damage suffered and the tortious and culpable action. Furthermore, the burden of proof relating to the prerequisites for an obligation to pay compensation lies with the party which suffers the damage. If an accident is attributable to the malfunctioning of a driver assistance system or to improper handling of such a system, and if the party suffering damage is able to prove this, then blame can be assigned to the owner or driver. Improper handling could be assumed, for example, if the driver neglects to heed warnings issued by the driver assistance system and this leads to an accident. The applicability of § 823 BGB may be of great importance for the party suffering damage, as it does not place an upper limit on the scope of liability (in contrast to § 12 StVG).

In a similar manner to § 7 (1) StVG, the notion of absolute liability for the risks of operation also underlies § 1 (1) of the Product Liability Act (Produkthaftungsgesetz, ProdHaftG)[154]. It is sufficient that the presence of a technical error and the connection with the damage can be established. The question of fault on the part of the manufacturer is not relevant. Under certain circumstances, this could be significant for cases of liability for damage arising from the malfunctioning of a driver assistance system or from improper handling of the same in accordance with § 1 (3) ProdHaftG[155], namely where it is at issue whether the damage is attributable to the driver assistance system or to the vehicle in which the driver assistance system is installed. It could furthermore play a role whether the driver assistance system is a telematic system which receives its data from an external source. In case of an error in such data, the manufacturer or provider may be subject to an obligation to pay compensation if the accident was brought about by precisely these data. In practice, the Product Liability Act will only be applicable in those cases of damage attributable to defective driver assistance systems where the party suffering the damage is able to prove a causal relationship between the damage and the error in the driver assistance system. This requirement will probably be difficult to meet. To date, no court judgements are known in connection with liability issues relating to driver assistance systems.

6.3.3 Driver assistance systems in the test vehicle for the practical driving test

For the most part, as shown above, the legal issues which could possibly arise in connection with the permissibility of driver assistance systems and liability for any damage associated with their use can be clarified on the basis of existing legal provisions, even though they make no direct reference to the subject of driver assistance systems. There is similarly no specific mention of driver assistance systems, the permissibility of their use or their handling in the context of the practical driving test in the pertinent legal regulations governing the practical driving test (§ 17 FeV, including Annex 7 thereto; Examination Guidelines, including Annex 12 thereto)[156]. It seems that this lack of concrete test specifications, together with the fact that the inherent purpose of driver assistance systems is to

ing damage … . The obligation to pay compensation is excluded if the damage is not attributable to fault on the part of the driver.

[153] § 823 (1) BGB: A person who, intentionally or negligently, unlawfully injures the life, body, health, freedom, property or another right of another person is liable to make compensation to the other party for the damage arising therefrom.

[154] § 1 (1) ProdHaftG: If, as a result of a defect in a product, a person is killed, injured or affected in his health or else material property is damaged, the manufacturer of the product is obliged to compensate the affected person for the ensuing damage. …

[155] § 1 (3) ProdHaftG: The manufacturer of a component of a product is exempted from the obligation to pay compensation, if the defect is attributable to the construction of the product into which the component is incorporated or to the instructions given by the manufacturer of the product. …

[156] § 17 (1) FeV: In the practical driving test, the candidate is to demonstrate that he possesses the technical knowledge required to operate a motor vehicle safely, where appropriate together with a corresponding trailer, sufficient knowledge of an environment-aware and energy-saving manner of driving, and the ability to apply this knowledge practically. …

help or relieve burdens on the driver when performing certain driving tasks, leads to occasional uncertainty on the part of driving test examiners where a test vehicle is equipped with driver assistance systems (e.g. uncertainty as to whether the use of a driver assistance system which can be deactivated should be permitted during the practical driving test, and how the functions of the driver assistance systems should be taken into account when assessing test performance)[157], despite the express presence of the following basic rule on the use of technical equipment and systems in the Examination Guidelines (PrüfRiLi 5.7): "Subject to the provisions of Annex 12, all equipment features and systems supplied by the vehicle manufacturer are generally permissible. This applies also where these or similar products are retrofitted at a later date …"

Applying the present legal framework, therefore, test vehicles can essentially be equipped with all approved driver assistance systems, and their use cannot be prohibited by the driving test examiner through reference to legal regulations or test practice to date. Under no circumstances is the driving test examiner entitled to refuse to conduct a driving test because the test vehicle is equipped with driver assistance systems of any kind. There is thus nothing in the currently applicable legislation which precludes the use of driver assistance systems in the practical driving test; at the same time, however, no specifications are to be found with regard to the handling of such systems by the test candidate, or the manner in which the use of driver assistance systems is to be taken into account in test planning and assessment by the driving test examiner. Correspondingly, further conclusions relating to the use of driver assistance systems in the practical driving test cannot be derived from the existing regulations.

6.4 Driver assistance systems and their significance in driving competence acquisition

The results of studies conducted in the traffic sciences suggest that the acquisition of driving competence must be understood as a complex learning process which is by no means completed upon granting of a driving licence and instead continues over a period of several years (MAYCOCK, LOCKWOOD & LESTER, 1991; SCHADE, 2001; GREGERSEN & NYBERG, 2002). Correspondingly, novice drivers usually display only a minimum level of driving competence at the beginning of their solo driving career. This circumstance is reflected, for example, in characteristic novice-specific competence deficits and an increased accident risk. A number of driver assistance systems possess the potential to compensate such novice-specific competence deficits and could thus reduce the particular risk of accident involvement for novice drivers. To date, however, there has been little discussion on how the use of driver assistance systems could influence driving competence acquisition. The additional risks which could arise for novice drivers from the use of driver assistance systems are likewise seldom a topic of study. Against this background, the present chapter is to expound the significance of driver assistance systems for novice drivers in general and for the acquisition of driving competence in particular. To this end, it is first necessary to take a closer look at the special characteristics of typical novice driver behaviour.

With regard to the novice risk which is expressed in novice-specific driving competence deficits, traffic psychology research has shown that the inappropriate behaviour of novice drivers is characterised above all by typical deficits in the perception and evaluation of hazards. It is generally the case, for example, that – compared to experienced drivers – novice drivers are less well able to distinguish the truly safety-relevant elements of a traffic situation from those which are less relevant for driving safety. Furthermore, the deficits displayed by novice drivers in terms of hazard perception result from the fact that they are often not in a position to grasp traffic situations in their entirety and to condense their different sensory impressions into a coherent overall picture of the traffic situation. Instead, novice drivers tend to concentrate their attention on specific isolated details of the traffic environment. As far as visual information acquisition is concerned, they respond above all to visual cues in the centre of their field of vision, and more often neglect corresponding cues in peripheral areas, for example children playing at the side of the road. Experienced drivers, on the other hand, scan also the periphery of their field of vision for indicators of potential hazards (UNDERWOOD, CHAPMAN, BROCKLEHURST, UNDERWOOD & CRUNDALL, 2003). To do so, they apply situatively adapted search strategies which enable them – in contrast to novice drivers – to identify the significant aspects of a traffic situation with rela-

[157] Some driving test examiners fear, in particular, that certain elements of driving competence will no longer be demonstrated by the test candidate – because the corresponding psychological and motor actions are handled on his behalf by driver assistance systems – and can thus no longer be assessed, even though the assessment of such competence is prescribed.

tively few fixations. They are thus able to recognise the essential hazards very quickly and precisely. When negotiating bends, for example, experienced drivers scan a longer section of the road ahead than on straight stretches in their search for information cues (HRISTOV, 2009). Novice drivers, on the other hand, concentrate their gaze on just a short section of road in front of their vehicle. Especially on bending roads, however, it is important to look far ahead and to analyse the form of a coming bend before the curve begins (COHEN & ZWAHLEN, 1989), in order to adapt driving behaviour to the situation in good time. Eye-tracking studies have also revealed that novice drivers seek orientation by way of the rear-view mirrors much less often than experienced drivers (LEE, OLSEN & SIMONS-MORTON, 2006).

Novice drivers must learn the correct execution and coordination of a series of psychomotor actions, such as driving off, accelerating, clutch use, gear change, steering, braking and stopping, in order to be able to operate and control a motor vehicle safely. "Psychomotor" here means that the realisation of motor actions is necessarily linked to psychological processes (e.g. control and monitoring of the actions). Due to the lack of driving experience on the part of a novice driver, however, the psychomotor skills required for driving have not yet become sufficiently automated (RASMUSSEN, 1986), and actions are still controlled and coordinated consciously for the most part. In other words: Unlike their experienced counterparts, novice drivers must devote a certain proportion of their attention to controlling the correct interaction of their psychological and motor actions. The cognitive resources which are tied up in this way are no longer available for other important tasks, for example the identification of potential hazards. Studies conducted in Great Britain and Germany suggest that comprehensive automation of the corresponding psychomotor skills is a process which, depending on the amount of driving done, may take up to three years (MAYCOCK et al., 1991; SCHADE, 2001).

As novice drivers must thus still consciously control their actions when handling a motor vehicle, and thereby generally achieve only error-prone coarse coordination of the different psychological and motor action components, their driving performance will also display characteristic deficits. ELLINGHAUS and STEINBRECHER (1990), for example, were able to show that novice drivers often brake very late and that they consequently have difficulties with the correct dosing of braking power, which is then expressed in the form of excessively hard braking. When turning at junctions, it is not seldom for novice drivers to display irregular steering behaviour and, on occasions, to swing out conspicuously at the corner. Both CAVALLO, BRUINDEI, LAYA and NEBOIT (1989) and JAMESON (1999) discovered that novice drivers often have problems holding their driving line and thus cross the road lane markings more frequently than experienced drivers when approaching, negotiating and exiting bends. The limited driving experience and skills of novice drivers at the same time mean that they are much more quickly overtaxed, especially in hazardous situations, and are thus unable to react appropriately to avert hazards in an emergency. Finally, the fact that – compared to an experienced driver – extensive cognitive processing capacities are occupied to compensate the still inadequately automated process of vehicle handling means that a novice driver is often no longer in a position to realise "auxiliary tasks" associated with driving (e.g. adjustment of a car radio or use of a hands-free telephone) without endangering road safety (GRATTENTHALER, KRÜGER & SCHOCH, 2009).

A proportion of the described novice-typical competence deficits relating to psychomotor skills and to hazard recognition and avoidance can be compensated by driver assistance systems. The various systems support different aspects of vehicle stabilisation and handling (e.g. ABS, ESC, parking steering assistance, (adaptive) cruise control and lane-keeping support), facilitate the perception of safety-relevant information (e.g. adaptive forward lighting, night vision assistant, lane departure warning, traffic sign recognition), or else intervene in case of a threatened collision and thus serve to avert danger (e.g. brake assist systems). Furthermore, even simple parking sensors are able to support timely hazard recognition – both when parking and when leaving a parking space – by informing the driver on the remaining clearance to obstacles in front of, alongside and behind the vehicle.

The problems which result from an increased readiness to take risks among some young drivers ("youth risk"), on the other hand, can only be solved to a certain degree with the aid of driver assistance systems. While it may be true that driver assistance systems are unable to prevent deliberate critical behaviour (e.g. excessive speed in a given traffic situation or insufficient safe distance to a preceding vehicle), systems such as ESC and ABS nevertheless improve vehicle handling and help the driver to master critical situations which arise out of his inappropriate behaviour.

How, then, do driver assistance systems influence driving competence acquisition? When seeking to answer this question, it is necessary to distinguish two groups of driver assistance systems: Driver assistance systems which contribute to stabilisation of the vehicle (e.g. ESC) or else only intervene directly in order to prevent a possible accident in an emergency situation (e.g. brake assist systems) have little effect on the acquisition of driving competence; they nevertheless relieve and protect the driver. If these systems were not installed and could thus not intervene in the intended manner, there would be a higher probability of an accident in a corresponding hazardous situation, and the driver – insofar as he survives the accident – would then know that he had failed to recognise a hazard in good time. Where driver assistance systems are present and intervene accordingly in a hazardous situation, the driver receives corresponding information as feedback from the vehicle (e.g. by way of control lamps); it is thus probable that he will still learn from his driving error, albeit without first needing to suffer injury or damage.

The influence on driving competence acquisition is more significant, however, where driver assistance systems perform actions which are inherent aspects of vehicle handling. When such driver assistance systems are used, there is a certain risk that a novice driver will fail to acquire the competence required to perform the corresponding actions himself. It is thus important that driving instructors not only explain the proper use of available driver assistance systems, but also teach novice drivers how to perform the relevant driving tasks without assistance. This remains imperative at least until all motor vehicles are equipped with systems to handle these tasks.

Parking is a good example to illustrate this problem. Various vehicle models are already available with so-called “intelligent park assist” systems, which offer sensor-controlled functions for steering assistance and speed control. They inform the driver on the suitability of a potential parking space, and then perform all the necessary steering actions to park the vehicle in the chosen space, in some cases even together with control of the accelerator and brakes. If such a system were to be used exclusively during driver training, the novice driver would have fewer opportunities to acquire the psychomotor skills associated with parking. This could prove to be a disadvantage later, as the novice driver may be overburdened by the situation, and could even endanger other vehicles and road users, if he is required to use a motor vehicle in which no such a park assist system is available. A further risk which may emanate from the exclusive use of park assist systems during driver training is that the novice driver could begin to rely on feedback from the distance warning sensors; important visual checks may then be neglected when manoeuvring into and out of a parking space and the necessary glance strategies would not be automated. The opinion of many experts in the field, therefore, is that, given the current state of technology, parking without the aid of a park assist system must still be taught. This would also justify its (random) inclusion as a driving task in the practical driving test, despite the fact that the high test demands currently applied to parking precision and the perfection of parking manoeuvres are hardly warranted by the requirements of road safety.

It thus remains to be concluded that novice drivers must be made aware of the functional limitations of the individual systems during the course of driver training, and that they must acquire appropriate knowledge relating to the proper handling of driver assistance systems. They must also learn not to place blind trust in the functionalities of driver assistance systems. Finally, knowledge of how the functioning of certain driver assistance systems can vary between different types of motor vehicle is likewise an element of well-trained driving and traffic competence.

6.5 Driver assistance systems and the testing of driving competence

The discussion thus far shows that, in future, greater attention must be paid to the subject of driver assistance systems in the context of novice driver preparation in general and in the theoretical and practical driving tests in particular, as they influence driving demands and may thus also affect test requirements. In the case of the practical driving test, the driving test examiner must decide – on the basis of test performance – whether the candidate's driving skills are sufficiently developed to enable further driving experience to be gained independently without endangering road safety. The fundamental questions in this connection thus concern the extent to which the functions of driver assistance systems may influence the observation and assessment of test performance, and whether driver assistance systems per se already impair the testing of driving competence. Applying the methodical foundations already described in previous chapters, professionally sound and differentiated answers to these questions can be obtained by considering how individual driver assistance systems influence the fulfilment of test standards on the three driving behaviour levels (see Chapter

6.1), and secondly how they may affect the relevance of the defined driving tasks and observation categories (see Chapter 3).

Already at the beginning of the present chapter, it was mentioned that three demand levels are to be distinguished with regard to driving behaviour, and that driving competence is thus manifested – also in the context of the practical driving test – on the navigation, manoeuvring and stabilisation levels. First answers relating to the possibly “disturbing influence” of driver assistance systems on the results of the practical driving test could thus be expected from a discussion of the manner in which driver assistance systems affect test performance on the aforementioned demand levels.

Let us begin with the navigation level. In Germany, as in most European countries (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2014), the test drive is conducted on flexible routes in real traffic. The examiner usually specifies the test route in accordance with a more or less adaptive test strategy. Navigation is not (yet) operationalised as a test task in Germany; certain navigation tasks (e.g. “Please drive to the railway station”) may be set nevertheless, insofar as the candidate declares that he possesses corresponding local knowledge. It is relatively seldom, however, that navigation systems are used to specify sections of the test route. At international level, efforts to establish navigation tasks as systematic test content, and thus to make greater use of navigation systems, are to be witnessed in a number of countries under the heading of “independent driving”.

From the methodical perspective, the broader use of navigation systems for certain sections of the practical driving test would represent an improvement: The test would then become more realistic and valid, because the use of navigation systems is by all means typical in modern road traffic. The associated support for the driver can be seen as an investment in road safety, and training of the proper handling of navigation systems should thus be an imperative element of novice driver preparation. This could also be encouraged by using such systems in the practical driving test. Compared to the longer-term orientation and routing options offered by a navigation system, the binding step-by-step instructions given by the driving test examiner may lead to additional action pressures which are rather atypical for the participation in real traffic: No-one dictates a particular route to the driver in a daily traffic situation; it is often the case that several route options are available, and navigation errors can be corrected without stress. The use of navigation systems is thus appropriate to the candidate's still limited level of driving competence. It must be added, however, that the use of navigation systems – assuming corresponding practice on the part of the candidate – is probably most meaningful at the beginning of the driving test: As the candidate is often particularly tense in the early phases of the test, he should be enabled to commence the test drive under familiar, resource-sparing driving conditions. On the other hand, the targeted assessment of individual driving tasks, which may be bound, in turn, to certain driving routes or road infrastructures, can generally only be realised on the basis of specific verbal instructions from the driving test examiner, and not by way of route recommendations from a navigation system.

Turning to the stabilisation level, the two driver assistance systems which immediately spring to mind are ABS and ESC. These driver assistance systems are non-overridable in normal operation, and intervene automatically in circumstances which could potentially lead to a hazardous loss of control over the longitudinal and transverse stability of the vehicle. In other words, they generally become effective in situations in which there is an imminent threat of accident, where their purpose is to prevent accidents or at least to reduce their consequences. Insofar as they cannot be deactivated, the use of these systems during the practical driving test is either unavoidable (e.g. ABS) or expedient for safety reasons (e.g. ESC). It is common to both systems that they serve exclusively the functional optimisation of vehicle operation, and that the driver cannot influence their effects directly. Correspondingly, the driver is unable to demonstrate that he has mastered handling of the systems in the sense of variable vehicle control; correct handling thus cannot be assessed as a test demand. Viewed overall, the use of driver assistance systems which function on the navigation or stabilisation level has no particular influence on realisation of the practical driving test.

The situation on the manoeuvring level is slightly different. There are two important reasons for this: Firstly, in contrast to navigation or the here relatively unusual case of stabilisation tasks, the safe and environmentally aware handling of a motor vehicle, i.e. the appropriate realisation of driving manoeuvres to implement a selected driving route, must be seen as the core element of the observation mandate to be fulfilled by the driving test examiner (see § 17 (1) FeV); secondly, it is evident that many of the aforementioned driver assistance systems are effective above all on the manoeuvring level. Therefore, the possible influences of driver assistance systems on the assessment of driving competence on the manoeuvring level must be evaluated very carefully. The best approach is to take those concrete driving tasks of the practical driving test which can be classified as complex action sequences on the manoeuvring level, and then to consider the observation categories which provide more precise specifications for the content-related assessment of these driving tasks by the examiner.

Most of the driving tasks anchored in the task catalogue for the optimised practical driving test (see Chapter 3) can be assigned to the manoeuvring level in terms of content (see above); this includes tasks such as changing lanes, passing and overtaking, the negotiation of crossroads and junctions, driving through curves and on connecting roads, behaviour at roundabouts and railway crossings, and interaction with pedestrians and cyclists, e.g. at pedestrian crossings and on roads with cycle lanes. To enable adequate assessment of the driving competence displayed by a test candidate when performing the set driving tasks, a series of situation-independent, behaviour-related demand standards were defined in the sense of required dimensions of driving competence, namely "Traffic observation", "Vehicle positioning", "Speed adaptation", "Communication" and "Vehicle control/ Environment-aware driving". These categories, too, can be assigned primarily to the manoeuvring level. Safe handling of a motor vehicle, after all, requires the driver to observe the surrounding traffic and to adapt his speed to the circumstances of a given situation. In the following, the driving task "Changing lanes" serves an example to illustrate how driver assistance systems influence the test demands of the practical driving test on the manoeuvring level, and which elements of driving competence are affected.

Which demands are placed on the test candidate when performing a change of lane, and which of these demands can be taken over by driver assistance systems? In terms of traffic observation, the candidate must use the vehicle mirrors, among other means, to analyse whether lane changing is permitted and expedient. It must be determined, for example, whether the traffic situation currently allows a lane change (e.g. traffic density in the destination lane, signals of other road users). Particular attention must here be paid to confirmation that no other road users are present in the driver's blind spot, and that the gaps between other vehicles are sufficient for safe merging into the other lane. When moving into a lane into which lane changes are possible from both sides, it is especially important to check whether another road user may be planning to enter the chosen gap from the other side. Yet another variant is a forthcoming zip-merging situation; this must be recognised accordingly and the corresponding traffic signs and road markings must be observed. A lane change assistant could help the driver to cope with these diverse demands: It would provide additional information on the traffic situation in the destination lane, and could warn if other vehicles are approaching from behind or possibly already present in the blind spot. This would supplement the driver's own observation of the traffic to the side of his vehicle, but nothing more. Overall, only a fraction of the required observation actions is covered, because the use of such a driver assistance system does not mean that the driver can automatically forego a side glance to check the blind spot, and certainly not that he can dispense with all further observation of the traffic environment. The same applies for all other driver assistance systems which support traffic observation (e.g. reversing camera, traffic sign detection, night vision assistant, blind spot camera): Irrespective of the use of these systems, the driver remains at all times dependent on and responsible for observation of the whole traffic environment in order to avoid hazards or accidents.

As far as the aspect of vehicle positioning is concerned, the test candidate must maintain sufficient clearance to other road users, to elements of the road infrastructure and to any other obstacles, and must subsequently adopt a new driving line in the centre of the new lane after completing a lane change. Driver assistance systems which monitor the correct lateral positioning of the vehicle within the road lane (e.g. lane departure warning, lane-keeping assistant) are unable to help in this situation, as the driver is moving out of his current lane deliberately. The distance alert function of a cruise control system, which would otherwise monitor the safe clearance to preceding vehicles, is similarly of limited use for the planning of a lane change, as it receives new reference values for its measurements as soon as the vehicle moves into a different lane. The driver thus remains essentially dependent on his traditional action control mechanisms for vehicle positioning.

Speed adaptation in connection with lane changing means that the candidate must select a vehicle speed appropriate to the traffic flow, the prevailing road and weather conditions and the current traffic situation, while at the same time observing the applicable speed limit. An (adaptive) cruise control system cannot relieve the driver of this task, which demands complex appraisal of the emerging traffic situation, as it only takes into account two single aspects of that situation, namely the set target speed and the distance to a preceding vehicle.

Insofar as a vehicle's turn indicators are not themselves to be promoted to the status of a driver assistance system, it seems hardly conceivable that driver assistance systems could offer substantial support to the test candidate's communication with other road users within the framework of the driving task "Changing lanes". When performing a lane change, the candidate must set the corresponding indicator in good time, and must pay attention to possible signals from other road users. In high-density traffic or a zip-merging situation, it may even be necessary to seek more intensive communication with other road users (e.g. direct eye contact). In case of dense traffic in the destination lane, furthermore, the intention to change lanes must be signalled by setting the flashing indicators well in advance (sometimes referred to as a "begging signal"). The manoeuvre is thus dependent on special, situation-specific social behaviour of a kind which currently cannot be automated. It is similarly not possible to automate necessary (environment-aware) vehicle control actions as soon as they become subject to the particular conditions of a special driving manoeuvre.

The example of the driving task "Changing lanes" indicates that the possibilities for driver assistance systems to contribute to the successful completion of driving tasks on the manoeuvring level are actually significantly more limited than could be presumed from a first superficial glance. If the other driving tasks specified in Chapter 3 (e.g. approaching and passing crossroads and junctions, negotiating curves or passing through roundabouts) are subjected to the same differentiated analysis, this impression is strengthened: In their current development state, driver assistance systems are not able to relieve the driver – and thus also the test candidate in the practical driving test – of the control actions necessary for safe vehicle handling in traffic to any substantial extent. This suggests that novice driver preparation should give due consideration to the limitations of driver assistance systems in respect of the safe fulfilment of demands on the manoeuvring level, alongside the undisputed simplification of certain vehicle control actions and the safety gains offered by driver assistance systems above all on the stabilisation level.

At present, it is only in the case of the so-called basic driving manoeuvres that driver assistance systems could conceivably handle the required driving tasks autonomously to any larger degree. Such basic driving manoeuvres serve to test the candidate's performance of elementary driving tasks such as parking, braking with the maximum possible deceleration and reversing. Modern intelligent park assist and brake assist systems could already perform the first two of the aforementioned basic driving manoeuvres more or less independently; the driving demands associated with reversing are changed through the availability of a reversing camera. How are these possibilities significant for the testing of basic driving manoeuvres during the practical driving test?

From the methodical perspective, the basic driving manoeuvres permit evidently valid determination of the extent to which a candidate masters fundamental techniques of vehicle control. A further benefit lies in the good possibilities for standardisation. Viewed against the background of the stipulations contained in the Examination Guidelines, it is the combination of these two aspects which explains the methodical fascination of the basic driving manoeuvres for driving test examiners. On the other hand, there are very few known research results[158] which confirm any outstanding safety relevance for either the basic driving manoeuvres in general or the high test demands relating to the precision of the parking manoeuvres in particular. This gives rise to the question as to whether the significance of the basic driving manoeuvres for road safety and for an assessment of driving competence is perhaps overestimated; irrespective of the (counter)-arguments which were already presented elsewhere for the necessity to test parking manoeuvres without the support of driver assistance systems, corresponding standpoints have already been a subject of discussion among examiners and driving instructors for some time (STURZBECHER, BIEDINGER et al., 2010). Insofar as such overestimation cannot be excluded, it seems reasonable to ask why the use of intelligent park assist systems, brake assist systems or reversing cameras should not be permitted in the practical driving test, especially since the use of such systems does not relieve the candidate of all

[158] Studies have only been conducted to evaluate the safety impact of training relating to braking with the maximum possible deceleration (so-called "emergency braking"). Accident analyses indicate that, due to their lack of experience, novice drivers tend to apply the brakes hesitantly and indecisively in emergency situations, and that this frequently results in an accident; full emergency braking by the driver could prevent around two-thirds of all tail-end collisions (GUERRINI, 2011). Further studies on braking behaviour show that the practising of emergency braking has a positive influence on real braking behaviour (LANGWIEDER, 2001; PETZHOLTZ, 2002). It was for this reason that emergency braking was introduced as a basic driving manoeuvre for the practical driving test in Germany in 2003. The intention was to ensure that braking with the maximum possible deceleration would become part of regular training and would then also be tested at least on a random basis.

demands associated with the basic driving manoeuvres. It is rather the case that the demands are merely modified: The candidate must naturally know how to use the park assist system properly, and must at the same time still observe the traffic environment; in this case, the latter would become the primary subject of testing. The same applies in general to the basic driving manoeuvre "Reversing", as reversing with the aid of a reversing camera requires the driver to perform different action sequences to those required when reversing without camera support. There are thus, in principle, no convincing content-related or methodical grounds for demands that the test demands must be handled with the minimum possible level of technical assistance.

Taking the practical driving test overall, it can be concluded that the current state of technical developments would allow individual driving tasks, and here in particular the basic driving manoeuvres, to be partially automated to a relatively minor degree with the aid of certain driver assistance systems. This refers primarily to systems which offer the driver assistance relating to vehicle handling (e.g. ACC, brake assist, lane-keeping assistant). On the other hand, these driver assistance systems offer very little support to the foresighted planning and coordinated realisation of the driving tasks as such; they facilitate at most the instrumental realisation of action components serving traffic observation, vehicle positioning, speed adaptation and vehicle control. It is thus out of the question that such driver assistance systems could take over the handling of relatively complex driving tasks to any substantial extent, and thus that they impair or even prevent assessment of the underlying driving competence. The situation is rather that the driver assistance systems which may be installed in the test vehicle for the practical driving test establish additional test demands, because the test candidate must be able to use and – insofar as this is actually possible, since certain driver assistance systems cannot be deactivated or influenced directly – operate the driver assistance systems concerned correctly and in compliance with traffic regulations.

6.6 Summary

Growth in the diversity of driver assistance systems is no less dynamic than the technical advances they embody. Innovations are entering the market at ever shorter intervals, and their presence and functioning in the vehicle is often barely visible to the untrained eye. It is undisputed, however, that driver assistance systems influence the demands of driving, and thus – depending on their individual functionalities and development level – also the demands to be met by the candidate in the practical driving test. As reaction to the presumed impact of driver assistance systems on the test conditions, which is admittedly difficult to quantify at first glance, and at the same time for fear of reduced test uniformity and equality, it is occasionally demanded that all driver assistance systems must be deactivated during the practical driving test, at least as long as their use (in test vehicles) is not legally prescribed.

Apart from the fact that many driver assistance systems cannot be deactivated, implementation of this demand would be counterproductive in terms of both road safety and improved test validity: Imponderables and unforeseen situations can never be excluded in daily road traffic. In such cases, driver assistance systems help to reduce the associated risks. They provide comprehensive support relating to the navigation, manoeuvring and stabilisation of a motor vehicle, and relieve the driver of corresponding burdens. This applies both to experienced drivers, who are hence able to call upon greater cognitive resources for the mastering of unexpected (hazardous) situations, and all the more so to novice drivers and driving test candidates, to whom far fewer such resources are available from the beginning, due to the still undeveloped nature of their behaviour routines. Driver assistance systems thus represent an urgently necessary contribution to road safety, in that they serve to compensate novice-specific driving competence deficits and avert novice-typical accidents. It is therefore imperative to support the use of driver assistance systems during novice driver preparation in general, and in the practical driving test in particular, because they make the process of learning to drive safer and enable learner drivers to acquire knowledge and skills relating to the functions and handling of driver assistance systems. If novice drivers become aware of the potential to avoid or at least minimise the consequences of accidents from an early stage, the chances that they will attach importance to such systems within the framework of a later vehicle purchase are also increased; the final result is further improvement of road safety.

A further frequently encountered reaction to the growing diversity of driver assistance systems and their implications for the uniformity of test conditions is a call for differentiated (regulatory) legal stipulations to govern the use of driver assistance systems and their influence on test realisation and assessment. In principle, this demand seems understandable and justified; nevertheless, differentiated methodical standards or guidelines which serve to evaluate the relevance of driver assistance systems installed in a particular test vehicle and at the same time keep pace with the technical further development of such systems will remain illusory in the future: The sheer diversity of driver assistance systems and the rate at which new developments are introduced by the vehicle manufacturers simply preclude system-specific and constantly up-to-date regulatory provisions, the elaboration, regular amendment and implementation of which would furthermore demand a high work and time input. From the point of view of test methodology, however, this is not seriously problematic, as a driving test examiner already applies his professional competence in other areas to compensate a lack of standardisation possibilities and detailed specifications, and thus to safeguard the validity of the practical driving test: His assessment of the safety of an overtaking manoeuvre, for example, may well vary in accordance with the prevailing (non-standardisable and uncontrollable) weather conditions, without this being deemed a methodical deficiency which must be rectified by way of special assessment instructions. With regard to the technical framework conditions for the practical driving test, too, examiners have always taken into account the potentially variable characteristics of different test vehicles (e.g. some vehicles are more likely to be stalled than others) even without corresponding guidelines; in this way, they ensure that the technical differences between vehicles do not distort the assessment of test performance.

As is made clear by the points discussed above, it seems neither urgent nor indeed possible to regulate the use of driver assistance systems in the practical driving test by way of concrete stipulations on the level of legislation or regulations. As no corresponding legal provisions exist, it is currently left to the candidate to decide whether or not he wishes to use driver assistance systems during the test drive. This means that, to a certain extent, the candidate can himself determine how he wishes to demonstrate fulfilment of the test demands and which forms of technical support he considers expedient. To ensure a maximum level of uniformity in the realisation and assessment of all tests conducted in Germany, despite the aforementioned individualisation of the test conditions, it seems desirable to provide the driving test examiners in the Technical Examination Centres with a set of common test standards comprising observation recommendations and above all general performance assessment criteria for certain driver assistance systems which may significantly influence the fulfilment of test demands relating to vehicle handling. Such test standards can be deemed necessary in view of both coordinated further development of the practical driving test and the technical advances in driver assistance systems; they are still to be elaborated, however, and must then remain the subject of continuous further development. The necessary steps can be outlined as follows:

- First of all, the (new) driver assistance systems must be assessed to determine whether they actually impact test demands in any way; this will in most cases concern demands on the manoeuvring level. Only such driver assistance systems are to be classified as relevant in the context of the practical driving test, and their number can currently be expected to be quite small. Assessments and descriptions of the relevance of selected driver assistance systems could be obtained in the form of expert statements from traffic scientists.
- Subsequently, as a second step, any existing experience with the handling of relevant driver assistance systems within the framework of driver training and testing should be gathered by way of practice-oriented action research; this may require supplementary studies.
- On this basis, standards to govern test realisation and assessment where the candidate is able to make use of certain driver assistance systems could then be made available as a third step.

Given the dynamic technical progress and the consequently growing diversity of driver assistance systems, it seems hardly feasible – as already indicated – to anchor such test standards in concrete legal provisions, not least because the elaboration and execution of corresponding laws, regulations and guidelines would hardly be able to stay abreast of technical developments. It is thus necessary to decide who should be entrusted with the elaboration of the desired test standard, and how this is to be accomplished. A starting point for this decision is to be found in the provisions of the Road Traffic Licensing Regulations (Straßenverkehrs-Zulassungs-Ordnung, StVZO), where it is stipulated that the test standards to be applied for roadworthiness tests according to § 29 StVZO are to be established by a so-called “Zentrale Stelle” (literally: “Central Office”) – in this case the company FSD Fahrzeugsystemdaten GmbH Dresden. The challenges to be met are similar to those relating to driving licence testing: While the test standards for driving test examiners must ensure that the quality of the practical driving test does not suffer under the introduction of driver assistance systems in the test vehicle, as an expression of the technical progress in vehicle engineering, the aforementioned standards for the appointed vehicle inspectors and test engineers serve to maintain the efficiency and implementation quality of roadworthiness tests in the face of continued developments in vehicle systems and technologies. It is thus considered valuable to present the foundations and mechanisms for the elaboration of test standards for roadworthiness tests in further detail.

Operators of technical inspection centres in the sense of §10 of the Motor Vehicle Traffic Experts Act (Kraftfahrsachverständigengesetz, KfSachvG) and officially recognised inspection agencies in the sense of Annex VIIIb StVZO (these organisations are hereafter referred to collectively as "inspection institutions") are required to introduce, maintain and further develop quality assurance measures relating to their recurring sovereign duty to perform motor vehicle inspections (roadworthiness tests and safety inspections in accordance with § 29 StVZO, and modification approvals in accordance with § 19 (3) StVZO). This is intended to ensure the proper and uniform realisation of vehicle inspections in accordance with the provisions of road traffic legislation, and serves above all to guarantee observance of the quality objectives by all inspection institutions, despite the competition between the individual service providers. To this end, the inspection institutions must cooperate with the responsible authorities to elaborate a quality assurance system, and the associated methodical procedures, with which to monitor the quality of vehicle inspections at different locations and on different vehicles.

One core element of this development process is the elaboration of uniform quality indicators, together with standardised assessment criteria referenced to these indicators. It is stipulated by the regulatory authority that the officially recognised experts and test engineers conducting roadworthiness tests are to observe (1) the statutory regulations applicable to such inspections, (2) corresponding guidelines published in the Official Journal of the Ministry of Transport ("Verkehrsblatt") and, in the absence of the former, (3) the test standards issued by a "Zentrale Stelle" mandated to elaborate, provide and validate such standards. To this end, the inspection institutions operate such a "Zentrale Stelle"[159], which, in accordance with its statutory purpose, also conducts appropriate research projects, either itself or by commissioning external scientific institutions. Furthermore, the "Zentrale Stelle" analyses the experience gained by the inspection institutions when applying the elaborated test standards, and – upon request – communicates its findings to the Federal Ministry of Transport, to the Federal Highway Research Institute (BASt), to the state-level authorities, to the joint working group of the inspection institutions (AKE) and to the Federal Motor Transport Authority (KBA), among others, as a basis for further development of the relevant legal provisions. The individual inspection institutions are required to support this process by sending the results of their roadworthiness tests to the "Zentrale Stelle" for evaluation. In combination, the described structures and processes facilitate the necessary adaptation of periodic vehicle inspections to the results of technical development with its exceptionally short innovation cycles.

As far as the elaboration, provision and validation of test standards to accommodate the use of driver assistance systems in the practical driving test is concerned, it seems expedient to entrust the task to the TÜV |DEKRA Working Group "Technical Examination Centres in the 21st Century" (TÜV |DEKRA arge tp 21), which was founded in 1999 by the operators of the Technical Examination Centre mandated to conduct driving licence tests[160]. This working group is also responsible for contributions to quality assurance and optimisation of the theoretical and practical driving tests and for

[159] The "Zentrale Stelle" must be located in the Federal Republic of Germany and must be organised in accordance with the applicable legislation. The procedural rules of the "Zentrale Stelle" must be presented to the Federal Ministry of Transport and to the supreme authorities at state level, and must be approved by these authorities. The "Zentrale Stelle" is not permitted to do business with the objective of returning a profit; any profits which are generated may only be used for the purpose of further development of the roadworthiness test. The supreme state-level authority of the federal state in which the "Zentrale Stelle" has its offices assumes supervisory responsibilities in consultation with the authorities of the other federal states. The supervisory authority is empowered to monitor – either itself or by way of delegation to a supervisory committee comprising a representative of the Federal Ministry of Transport, the chairperson of the joint working group of the inspection institutions (AKE) and two representatives of the federal states appointed by the corresponding supreme authorities – whether the "Zentrale Stelle" satisfies all conditions stipulated in the regulations and whether it performs the tasks conferred by the legislation in a proper manner. The supervisory committee is authorised to issue instructions to the "Zentrale Stelle" within the framework of its tasks. To support the development of test standards for the roadworthiness test, and in particular the adaptation to technical progress and efficient, high-quality realisation of the testing, the work of the "Zentrale Stelle" is to be accompanied by a technical advisory committee comprising representatives of scientific institutions, the motor industry, consumer protection organisations, the craft trades and the relevant authorities.

[160] These operators are: TÜV Rheinland Kraftfahrt GmbH, TÜV SÜD Auto Service GmbH, TÜV NORD Mobilität GmbH & Co. KG and DEKRA Automobil GmbH.

the elaboration of framework conditions to guarantee uniform test realisation throughout Germany.

Furthermore, in the interest of road safety and especially to uphold the validity of the test as a reflection of real traffic demands, the regulatory authorities should provide for the presence of individual driver assistance systems in test vehicles to be made mandatory as soon as an appropriate market coverage is reached, as has already happened in the case of ABS: Corresponding stipulations are contained in Annex 7 FeV, 2.2, according to which vehicles of Classes C (2.2.6), CE (combinations of a tractor vehicle and a trailer or semi-trailer; 2.2.7), C1 (2.2.8), D (2.2.10) and D1 (2.2.12) must be equipped with an anti-lock braking system (ABS). Beyond Annex 7 FeV, 2.2.4, the private motor vehicles to be used as test vehicles are also subject to requirements stipulated in the Examination Guidelines (PrüfRiLi 5.5 and 5.7), which could be expanded to include conditions for the use of driver assistance systems.

7 Project summary

Novice drivers face a much higher risk of being injured or killed in road traffic compared to experienced drivers. Against this background, further development of the system of novice driver preparation seems imperative. Such a process necessarily includes also optimisation of the practical driving test, which serves above all to verify practical aspects of driving competence and, by way of its control and selection functions, must be considered a central element (of testing) within a system of protective novice driver preparation.

If a test of competence is to deliver reliable, valid and meaningful test results, it is first necessary to define sound foundations in accordance with the principles of educational and test psychology. For the present research report, therefore, a driving competence model was elaborated as a basis for more detailed determination of the dimensions of driving competence to be verified by way of an optimised practical driving test. Furthermore, a content-related and methodical concept for continuous maintenance, quality assurance and further development of the test was described. To this end, the methodical foundations developed for the practical driving test by STURZBECHER, BÖNNINGER & RÜDEL (2010) were taken up and developed further, extending and adding detail to the original proposals on possibilities for optimisation. At the same time, all institutional structures necessary to implement the envisaged system of testing were presented together with the associated test methods and procedures – including the requisite demand, assessment, documentation and evaluation standards – in a draft for a "System Manual on Driver Licensing (Practical Test)". It was here assumed that the practical driving test represents a competence-referenced diagnostic work sample, wherein the test performance of the driving licence applicant is observed, assessed and documented by the driving test examiner by way of systematic behaviour observation within the framework of an adaptive test strategy. On the basis of the observed test performance, the examiner decides whether the test candidate possesses a certain minimum level of driving competence and is thus in a position to drive a motor vehicle safely and independently in road traffic.

How exactly were the aforementioned project results elaborated? The starting point was the premise that valid (driving) competence diagnosis must ideally be based on competence models which reflect both the internal structure of the competence to be diagnosed and – by way of competence levels or stages of acquisition – the process of competence acquisition (KLIEME & LEUTNER, 2006). Inspiration for the development of such integrated models is offered firstly by DONGES (2009), whose proposal for a competence model merges action-oriented dimensions relating to the content-specific demand levels of the driving process – which in turn reflect those corresponding components of driving competence which can be deemed prerequisites for fulfilment of the demands – with levels of action control according to RASMUSSEN (1983). GRATTENTHALER, KRÜGER and SCHOCH (2009), on the other hand, combine notions of content-referenced competence and processual learning mechanisms in their spiral (driving) competence acquisition model. On the basis of these two models, the areas of competence to be conveyed in driver training (vehicle stabilisation, vehicle manoeuvring, navigation, and values or attitudes) were defined, and the components of driving competence to be assessed by way of the optimised practical driving test (primarily competence relating to vehicle manoeuvring) were identified. By way of the latter model, it was furthermore determined that, in accordance with the spiral, situation-driven process of driving competence acquisition, it seems difficult to establish a differentiated model for the levels of driving competence. Finally, following KLIEME et al. (2007), it was shown that the demands to be met by novice drivers must be described in the form of (minimum) training standards for novice driver preparation, and that the formulation of these training standards must satisfy a series of quality criteria (subject specificity, focus, cumulativity, binding applicability, differentiation, comprehensibility and feasibility). Corresponding test standards for the optimised practical driving test then need to be derived from such training standards in accordance with teaching/learning theory principles.

As common training standards meeting the aforementioned quality criteria have yet to be formulated for the purposes of novice driver preparation, it was necessary for the present project to develop scientifically founded test standards by other means. The basis for the chosen approach was the thought that both the proper elaboration of training standards and the methodically sound construction of work samples serving competence diagnosis commence with an analysis of the demands as they relate to action theory, whereby the activity in question (here the driving of a motor vehicle) is divided into its constituent, demand-related action steps or tasks. Those steps or tasks which are of particular relevance for the desired quality of the action (here safe and environment-aware solo driving) can then be described in training standards in terms of their competence prerequisites. These training standards, in turn, are operationalised in training curricula, implemented in training institutions and monitored by way of competence tests.

The manner in which the development of training and test standards must be organised in the context of driving licence testing – from demand analyses, via the definition of training goals and the elaboration of a driver training curriculum, through to the construction of learner assessments and driving tests – was demonstrated in exemplary form by McKNIGHT and ADAMS (1970a, 1970b) and by McKNIGHT and HUNDT (1971a, 1971b). On the basis of their methodically excellent studies and the sound results presented, demand standards for the optimised practical driving test were described in the present research report in the form of situation-related driving tasks and situation-independent observation categories or components of competence, as a means to specify criteria for event-oriented performance assessment and overall competence evaluation. This approach also revived methods which were already practised by the Federal Highway Research Institute and the Technical Examination Centres in the 1970s (HAMPEL, 1977; KROJ & PFEIFFER, 1973; SCHNEIDER, 1977), but were later not pursued consistently and translated into licensing legislation.

In addition to the scientific foundations from the perspectives of educational and test psychology, the concretised demand and assessment standards for the optimised practical driving test also take into account the stipulations of the EU Directive on Driving Licences (EUROPEAN PARLIAMENT & EUROPEAN COUNCIL, 2006, L 403/43), as well as international experience and implementation practice. A corresponding basis was provided by an international study of innovative models for the practical driving test and their integration with other forms of teaching/learning and testing within the system of novice driver preparation. Overall, the research covered the demand and implementation standards for the practical driving test in 36 countries (sovereign states or federated units of national states, e.g. federal states or provinces, in Europe, North America and Oceania); in this connection, the currently applicable test reports from 25 countries were analysed with regard to their content and methodical design, and framework curricula for driving school training from 13 countries were studied. Particular attention was here paid to the driving tasks to be performed by the test candidate, and the observation, assessment and decision criteria applied by the examiner.

The results of evaluation from the point of view of competence theory, analyses of the demands of driving a motor vehicle in public traffic in accordance with action theory, and the methodical construction of the practical driving test as an instance of systematic driving behaviour observation led finally to the formulation of a draft for a catalogue of driving tasks. This task catalogue describes the demand standards for the practical driving test to obtain a class B driving licence. It contains first a total of eight situation-related driving tasks, some of which are further divided into subtasks: (1) Joining/leaving traffic and changing lanes, (2) Passing and overtaking, (3) Crossroads and junctions, (4) Negotiating curves, (5) Roundabouts, (6) Trams and railway level crossings, (7) Bus/tram stops, pedestrians and pedestrian crossings, and (8) Behaviour towards cyclists. Secondly, the situation-independent observation categories or components of competence to be assessed – with reference to the aforementioned driving tasks – are likewise specified in the task catalogue: (1) Traffic observation, (2) Vehicle positioning, (3) Speed adaptation, (4) Communication, and (5) Vehicle control/Environment-aware driving. Thirdly and finally, event-oriented and competence-oriented assessment criteria are described for each driving task and each observation category: The event-oriented criteria define simple and serious errors, as well as examples of above-average performance; the competence-oriented criteria permit assessment of the test performance on a four-level rating scale.

The described catalogue of driving tasks provides a content-related and methodical basis for the programming and testing of an electronic means of test documentation ("electronic test report"). To support the necessary elaboration of documentation standards, recommendations were formulated with regard to hardware and software selection, ergonomic report design and the realisation of a feasibility study. These recommendations have in the meantime been taken up and developed further by the working group TÜV |DEKRA arge tp 21; the first results from testing of an electronic test report are already available. The introduction of the electronic test report is intended to assist the driving test examiner by enabling a constant overview of those driving tasks which have already been performed and the assessments of test performance recorded so far, as better orientation for his adaptive test planning and realisation (organisational function). At the same time, electronic documentation permits differentiated feedback to be given to the candidate on all significant – positive and negative – aspects of the test performance (didactic function).

With the development of electronic test documentation, the foundations for scientific evaluation of the practical driving test are also improved significantly. As a starting point for the elaboration of an evaluation system, exploratory meetings were held with leading representatives and quality management officers from the four Technical Examination Centres mandated to conduct driving tests and from the Bundeswehr. These meetings, which were coordinated by the Federal Highway Research Institute (BASt), served to discuss not least the expectations of the various stakeholders in respect of a future evaluation concept for the optimised practical driving test; the results of the discussions were incorporated into the new evaluation concept. With reference to current evaluation practice, it must be noted that, to date, there have been no mentionable scientific studies relating to instrumental evaluation, and analyses of test results have addressed merely pass rates. The future evaluation system, by contrast, is to comprise four complementary methodical elements: While the element "Instrumental evaluation" targets the psychometric quality of the methods employed by the optimised practical driving test, the objective of the remaining elements "Customer surveys", "Product audits" and "Analysis of test results" is to analyse the implementation quality of the test in daily use. The latter evaluation elements thus serve a processual evaluation and are intended to provide methodically sound proof of a uniformly high quality of test design and performance assessment across the whole country; as forms of external process evaluation, they correspond to the procedures of internal corporate quality management.

Finally, the present project paid tribute to the dynamic developments in vehicle technologies over the past few years by elaborating professional recommendations for the fundamental treatment of driver assistance and accident avoidance systems in connection with realisation and assessment of the practical driving test. As a first step, the functions and function principles of selected driver assistance systems were described as a basis for subsequent analyses of the possible influences of such systems on the acquisition and testing of driving competence. By way of examples, it was shown how the use of such systems may affect the performance of certain driving tasks and the assessment of corresponding components of competence (observation categories) within the context of the driving test. In conclusion, it can be said that it seems neither urgent nor indeed possible to regulate the use of driver assistance systems in the practical driving test by way of concrete stipulations on the level of legislation or regulations. To nevertheless ensure a maximum level of uniformity in the realisation and assessment of all tests conducted in Germany, it seems desirable to provide the driving test examiners in the Technical Examination Centres with a set of common test standards comprising observation recommendations and above all general performance assessment criteria for certain driver assistance systems which may significantly influence the fulfilment of test demands relating to vehicle handling. The report thus presents professional design recommendations for the necessary work processes. As far as the elaboration, provision and validation of test standards and guidelines to accommodate the use of driver assistance systems in the practical driving test is concerned, it seems expedient to entrust the task to the working group TÜV |DEKRA arge tp 21, which is operated as a scientific-technical research and development institution by the Technical Examination Centres mandated to conduct driving tests.

What do the aforementioned research findings and project results mean for novice driver preparation, and especially for the future system of driving licence testing? Optimisation of the test standards for the practical driving test will no doubt enhance demand transparency for the test candidates, the driving instructors and the driving test examiners; this will contribute to closer integration of the systems of driver training and testing, and thus not least to the implementation of a demanding and uniform driving test for all candidates. The unambiguous specification of situation-independent observation categories or components of competence, in particular, can be expected to permit more clear-cut definitions of the test contents and will set priorities for the topics to which candidates and examiners should pay special attention. Furthermore, it is likely that the safety impact of the practical driving test will be improved, as novice-specific driving competence deficits and accident causes were taken into account when elaborating the driving tasks and assessment criteria. The assessment criteria are in future to refer specifically to the required driving tasks and situation-independent components of driving competence. They thus permit differentiated statements on the candidate's level of driving competence, and give due consideration to both errors and aspects of good performance. Optimised feedback on test performance to all candidates is intended to lend effective support to their continued learning process and to the development of driving expertise;

this is to apply in future both to targeted additional training for candidates who fail the test, and to the further learning of successful candidates within the framework of accompanied or solo driving.

The input forms made available by way of the electronic test report will also simplify test documentation for the examiner in the future: Typical errors and instances of above-average performance are to be recorded by clicking on a PC screen, and the overall information base is expanded. It will at the same time become easier to weigh up a final test decision, as an overview of all performance displayed during the test can be visualised, and automatic plausibility checks can be implemented. The performance overview will similarly structure and facilitate feedback to the candidate in a subsequent discussion of the test performance. Last but not least, the electronic acquisition of test data will optimise the administration of testing; in this context, it will also be possible to modernise and simplify the coordination processes between the regulatory authorities and the Technical Examination Centres by introducing a common, computer-assisted authoring system.

Viewed overall, the results of the present project and their possible implementation in an optimised practical driving test identify three significant opportunities for the improvement of road safety in the future:

1. The scientific foundations for the contents and methods of the practical driving test will in future enable continuous and empirically supported further development of the demand and assessment standards on the basis of evaluations of test results and other traffic research sources such as accident analyses. Within the framework of such output control, it can be recognised, for example, whether a conspicuous proportion of the candidates fail to perform certain driving tasks correctly; corresponding analysis results can then be used to optimise driver training. In this way, monitoring of the test standards contributes to improvement of the training standards; the safety impact of novice driver preparation can be judged empirically.
2. The electronic documentation of test performance also opens up new possibilities for quality assurance and for professional, formative and summative evaluation of the practical driving test. It can be determined, for example, which driving tasks are tested at which test locations, and to what extent given test locations are suitable for the testing of particular driving tasks. Supplementary to external monitoring of the effectiveness of the test in respect of its prime objectives, as demanded by the legislator, a deeper analysis of test results (beyond mere consideration of the pass rates) could also offer driving test examiners valuable information on their individual test behaviour in the sense of self-evaluation, and thus a basis for further development of their own, professional competence.
3. Finally, it is possible to raise the pedagogical effectiveness of the practical driving test, and – based on the electronic test report – to provide differentiated, performance-oriented feedback to the test candidate on the strengths, weaknesses and inherent risks of his driving behaviour. This can be achieved in two ways: Firstly, the driving test examiner can offer the candidate explicit hints on further learning needs during a brief meeting after the test drive; in addition, as follow-up to the test, more detailed written recommendations should be made available (e.g. online) as a starting point for targeted future learning.

Before the anticipated gains for road safety can actually be realised, the demand, assessment, documentation and evaluation standards drafted for the optimised practical driving test must be verified within the framework of a revision project, and subsequently amended, where necessary, in accordance with the results of trial implementations.

Bibliography

ABEDI, J. (1996). Interrater/Test Reliabilty Systems (ITRS). Multivariate Behavioral Research, 31, 408-417.

AMOR (2003). E-Business aktuell: Edition 2004 - Trends, Prozesse und Technologien im Unternehmen. Weinheim: Wiley-VCH.

ANDERSON, J. R. (2001). Kognitive Psychologie. Heidelberg: Spektrum Akademischer Verlag.

AKAC Arbeitskreis Assessment Center e.V. (2004). Standards der Assessment Center Technik. Hamburg: Arbeitskreis Assessment Center e.V.

ALBRECHT, F. (2005). Die rechtlichen Rahmenbedingungen bei der Implementierung von Fahrerassistenzsystemen zur Geschwindigkeitsbeeinflussung. Deutsches Autorecht – DAR 4, 186-198.

AMELANG, M. & SCHMIDT-ATZERT, L. (2006). Psychologische Diagnostik und Intervention. Heidelberg: Springer.

BÄHR, W. & WEIBERT C.-D. (2010). Prüfungskompass – Eine Handreichung für Mitglieder von Prüfungsausschüssen zur Durchführung von Zwischen-, Abschluss- und Fortbildungsprüfungen bei den zuständigen Stellen für Berufsbildung. Bonn: IFA.

BARTHELMESS, W. (1976). Fahrprüfung – Fahrverhalten – Fahrsicherheit. Forschungsgemeinschaft "Der Mensch im Verkehr" e.V. Köln, Heft 2.0, 2. Aufl. Darmstadt.

BARTL, G. & HAGER, B. (2006). Unfallursachenanalyse bei PKW-Lenkern. Vienna: Institut Gute Fahrt.

BAUGHAN, C. J. (2000). Review of the practical driver test. Proceedings of the DTLR Novice Driver Conference, June 2000. Crowthorne, UK: Transport Research Laboratory.

BAUGHAN, C. J., & SIMPSON, H. (1999). Consistency of driving performance at the time of the L-test, and implications for driver testing. In G. B. Grayson (ed), Behavioural Research in Road Safety IX Transport Research Laboratory, Crowthorne.

BAUGHAN, C., GREGERSEN, N. P., HENDRIX, M. & KESKINEN, E. (2005). Towards European standards for testing (EU project TEST final report). Brussels: Commission Internationale des Examens de Conduite Automobile.

BAUMERT, J. (1993). Lernstrategien, motivationale Orientierung und Selbstwirksamkeitsüberzeugungen im Kontext schulischen Lernens. Unterrichtswissenschaft, 21, 327-354.

BAUMERT, J., LEHMANN, R., LEHRKE, M., SCHMITZ, B., CLAUSEN, M., HOSENFELD, I., KÖLLER, O. & NEUBRAND, J. (1997). TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde. Opladen: Leske + Budrich.

BAUMERT, J., KLIEME, E., NEUBRAND, M., PRENZEL, M., SCHIEFELE, U., SCHNEIDER, W., STANAT, P., TILLMANN, K.-J. & WEIß, M. (Hrsg.). (2001). PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.

BEA, F. X., SCHEURER, S. & HESSELMANN, S. (2008). Projektmanagement. Stuttgart: Lucius & Lucius.

BÉDARD, M., PARKKARI, M., WEAVER, B., RIENDEAU, J. & DAHLQUIST, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. American Journal of Occupational Therapy, 64, 336-340.

BEINER, F. (1982). Prüfungsdidaktik und Prüfungspsychologie: Leistungsmessung und Leistungsbewertung in der öffentlichen Verwaltung sowie in der beruflichen und allgemeinen Bildung. Cologne: Heymann.

BELZ, J., HÖVER, N., MÜHLENBERG, N. NITSCHE, B. & SEUBERT, T. (2004). Fahrerassistenz im Spannungsfeld zwischen Komfort- und Sicherheitsanforderungen. In VDI (Hrsg.), Integrierte Sicherheit und Fahrerassistenzsysteme. VDI-Berichte 1864. Düsseldorf: VDI-Verlag.

v. BENDA, H. (1985). Die Häufigkeit von Verkehrssituationen. Bericht zum Forschungsprojekt 7320/2 der Bundesanstalt für Straßenwesen, Bereich Unfallforschung. Bergisch Gladbach: BASt.

BERG, M., KIESCHKE, U. & SCHUBERT, W. (2008). Artifizielle Validitätsnachweise. Anmerkungen zur Validierung von Messapparaturen. Zeitschrift für Verkehrssicherheit, 54 (4), 183-187.

BERNOTAT, R. (1970). Anthropotechnik in der Fahrzeugführung. Ergonomics, 13, 353-377.

BEWERSDORF, C. (2003). Zur Vereinbarkeit von nicht-übersteuerbaren Fahrerassistenzsystemen mit dem Wiener Übereinkommen über den

Straßenverkehr vom 8. November 1968. NZV, 266-271.

BGHW – Berufsgenossenschaft Handel und Warendistribution (2010). Merkblatt M103: Heben und Tragen. Bonn: bghw.

BIBB – Bundesinstitut für Berufsbildung (2007). Empfehlung des Hauptausschusses des Bundesinstituts für Berufsbildung für die zur Musterprüfungsordnung für die Durchführung von Abschluss- und Umschulungsprüfungen. (URL: http://www.bibb.de/dokumente/pdf/empfehlung_120-musterpr_fungsordnung_abschluss-_und_umschulungspr_fungen_182.pdf; accessed: 15.07.2011).

BIEHL, B., FISCHER, G. H., HÄCKER, H., KLEBELSBERG, D. & SEYDEL, U. (1975). A comparison of the factor loading matrices of two driver behavior investigations. Acc. Anal. & Prev., 7, 161-178.

BOLLHOEFER, K. W., MEYER, K. & WITZSCHE, R. (2010). White Paper: Microsoft Surface und das Natural User Interface (NUI). Berlin: Pixelpark.

BÖNNINGER, J. & STURZBECHER, D. (2005). Optimierung der Fahrerlaubnisprüfung. Ein Reformvorschlag für die theoretische Fahrererlaubnisprüfung. Berichte der Bundesanstalt für Straßenwesen, Heft M 168. Bremerhaven: Wirtschaftsverlag NW.

BÖNNINGER, J., KAMMLER, K., STURZBECHER, D. & WAGNER, W. (2005). Theoretische und praktische Fahrerlaubnisprüfung in Europa. Recherchebericht. Dresden: TÜV DEKRA arge tp 21.

BÖNNINGER, J., KAMMLER, K. & STURZBECHER, D. (Hrsg.). (2009). Die Geschichte der Fahrerlaubnisprüfung in Deutschland. Bonn: Kirschbaum.

BÖNNINGER, J., STURZBECHER, D., BIEDINGER, J., BÖHNE, A., V. BRESSENSDORF, G., GLOWALLA, P., KAUP, M., KLEUTGES, CH., MÜLLER, G., MÜLLER, R., PETZHOLTZ, W., RADERMACHER, R., RÜDEL, M., SCHMIDT, A. & WAGNER, W. (2010). Möglichkeiten zur Optimierung der praktischen Fahrerlaubnisprüfung in Deutschland. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 170-178). Bremerhaven: Wirtschaftsverlag NW.

BORTZ, J.& DÖRING, N. (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.

BORTZ, J. & SCHUSTER, CH. (2010). Statistik für Human- und Sozialwissenschaftler. Berlin: Springer.

BRAITMAN, K. A., KIRLEY, B. B., MCCARTT, A. T., & CHAUDHARY, N. K. (2008). Crashes of novice teenage drivers: Characteristics and contributing factors. Journal of Safety Research, 39, 47-54.

BREDOW, B. (2012). Ergebnisse einer international vergleichenden Analyse von ausgewählten Fahrausbildungscurricula. Zwischenbericht zum Projekt “Ansätze zur Optimierung der Fahrschulausbildung in Deutschland”. Unpublished paper. Vehlefanz: IFK.

BRENNER-HARTMANN, J. (2002). Durchführung standardisierter Fahrverhaltensbeobachtungen im Rahmen der medizinisch-psychologischen Untersuchung (MPU). Vortrag auf dem 38. Kongress für Verkehrspsychologie in Regensburg: Stuttgart: Manuskript des TÜV Süddeutschland.

BREUER, K. (2003). Diagnostik kaufmännischer Bildungsstandards. Entwicklungslinien und Ansprüche aus wirtschaftspädagogischer Sicht. Universität Mainz. (URL: www.wipaed.uni-mainz.de/breuer/Dateien/HBB2004-Diagnostik_kfm_Standards.pdf; accessed 15.11.2012).

BÜHNER, M. (2011). Einführung in die Test- und Fragebogenkonstruktion. Munich: Pearson.

BÜLOW-SCHRAMM, M. (2006). Qualitätsmanagement in Bildungseinrichtungen. Münster: Waxmann Verlag GmbH. (URL: http://www.e-cademic.de/data/ebooks/extracts/9783830917526.pdf; accessed: 28.02.2012).

BUNDESANSTALT FÜR STRAßENWESEN (2003). Anforderungen an Träger von Technischen Prüfstellen (§69 FeV i.V. mit den §§10 und 14 des KfSachVG) der Akkreditierungsstelle Fahrerlaubniswesen der BASt vom 10.10.2003.

BUNDESANSTALT FÜR STRAßENWESEN (2005). Anforderungen an Träger von Technischen Prüfstellen (§ 69 FeV i.V. mit den §§ 10 und 14 des KfSachvG). Bergisch Gladbach: Akkreditierungsstelle Fahrerlaubniswesen.

BUNDESANSTALT FÜR STRAßENWESEN (2007). Anforderungen an Träger von Technischen Prüfstellen (§ 69 FeV i.V. mit den §§ 10 und 14 des KfSachvG). Bergisch Gladbach: Akkreditierungsstelle Fahrerlaubniswesen.

BUNDESANSTALT FÜR STRAßENWESEN (2009). Anforderungen an Träger von Technischen Prüfstellen (§ 69 FeV i.V. mit den §§ 10 und 14 des KfSachvG). Bergisch Gladbach: Akkreditierungsstelle Fahrerlaubniswesen.

BUNDESVERFASSUNGSGERICHT (BVerfG, 3. Kammer des Ersten Senats) (1987). Verfassungsmäßigkeit der Gurtanlegepflicht für Kraftfahrzeugführer, Neue Juristische Wochenschrift (NJW), 4, 180.

CAVALLO, V., BRUIN-DEI, M., LAYA, O. & NEBOIT, M. (1989). Percepts and anticipation in negotiating curves: the role of driving experience. In A. G. Gale (Ed.), Vision in vehicles - II (pp. 365-374). Amsterdam: Elsevier Science Ltd.

CHALOUPKA, C. & RISSER, R. (1995). Don´t wait for accidents – possibilities to assess risk in traffic by applying the 'Wiener Fahrprobe'. Safety Science, 19, 137-147.

CHAPMAN, M. G. & UNDERWOOD, A. J. (2000). The need for a practical scientific protocol to measure successful restoration. Wetlands (Australia) 19, 28-49.

CHAUDHARY, N., BAYER, L., LEDINGHAM, K. & CASANOVA, T. (2011). Driver Education Practices in Selected States. Final Report. Washington: National Highway Traffic Safety Administration.

COHEN, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

COHEN, J. (1968). Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213-220.

COHEN, A. S. & ZWAHLEN, H. T. (1989). Blicktechnik in Kurven. bfu-Report 13. Berne: Schweizerische Beratungsstelle für Unfallverhütung bfu.

CONGER, A. J. (1980). Integration and generalization of kappas for multiple raters. Psychological Bulletin, 88, 322-328.

CONGDON, P. (1999). VicRoads Hazard Perception Test: can it predict accidents? (Report CR-99-1). Camberwell, Victoria, Australia: Australian Council for Educational Research.

CRINSON, L. F. & GRAYSON, G. B. (2005). Profile of the British learner driver. In G. Underwood (Ed.), Traffic and transport psychology, theory and application (pp. 157-170). London: Elsevier.

CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

DAIMLER AG (Hrsg.). (2009). Meilensteine der Fahrzeugsicherheit. Die Vision vom unfallfreien Fahren. Stuttgart: Daimler AG Communications.

DALCHER, D. & BRODIE, L. (2007). Successful IT Projects. London: Cengage Learning.

DEBUS, G., LEUTNER, D., BRÜNKEN, R., SKOTTKE, E. M. & BIERMANN, A. (2008). Wirkungsanalyse und Bewertung der neuen Regelungen im Rahmen der Fahrerlaubnis auf Probe. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit, Heft M 194. Bremerhaven: Wirtschaftsverlag NW.

DeGEval – Deutsche Gesellschaft für Evaluation e.V. (Hrsg.). (2008). Standards für Evaluation. Mainz: DeGeval.

DEMING, W. E. (1982). Quality, Productivity, and Competitive Position. Massachusetts: Massachusetts Institute of Technology.

DEMING, W. E. (1986). Out of the Crisis. MIT Center for Advanced Engineering Study. Cambridge, MA.

DE WINTER, J. C. F., DE GROOT, S., MULDER, M., WIERINGA, P. A., DANKELMAN, J. & MULDER, J. A. (2009). Relationships between driving simulator performance and driving test results. Ergonomics, 52, 137-153.

DIN Deutsches Institut für Normung e.V. (1998). DIN EN ISO 9241 Ergonomie der Mensch-System-Interaktion – Teil 11: Anforderungen an die Gebrauchstauglichkeit; Leitsätze. Berlin: Beuth.

DIN Deutsches Institut für Normung e.V. (2002). DIN 33430 Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Berlin: Beuth.

DIN Deutsches Institut für Normung e.V. (2003). DIN EN ISO 14915-1 Software-Ergonomie für Multimedia-Benutzungsschnittstellen – Teil 1: Gestaltungsgrundsätze und Rahmenbedingungen. Berlin: Beuth.

DIN Deutsches Institut für Normung e.V. (2008). DIN EN ISO 9241 Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung. Berlin: Beuth.

DIN Deutsches Institut für Normung e.V. (2011a). DIN EN 60204-1 Sicherheit von Maschinen - Elektrische Ausrüstung von Maschinen – Teil 1: Allgemeine Anforderungen. Berlin: Beuth.

DIN Deutsches Institut für Normung e.V. (2011b). DIN EN ISO 9241 Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. Berlin: Beuth.

DONGES, E. (1982). Aspekte der aktiven Sicherheit bei der Führung von Personenkraftwagen. Automobil-Industrie, 2, 183-190.

DONGES, E. (2009). Fahrerverhaltensmodelle. In H. WINNER, S. HAKULI & G. WOLF (Hrsg.), Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort (pp. 15-23). Wiesbaden: Vieweg & Teubner.

DUNCAN, J., WILLIAMS, P. & BROWN, I. (1991). Components of driving skill: Experiences does not mean expertise, Ergonomics, 34, 919-937.

EBBINGHAUS, M. & SCHMIDT, J. U. (1999). Prüfungsmethoden und Aufgabenarten. Bielefeld. Bertelsmann.

EELES, P. (2005). Capturing Architectural Requirements. (URL: http://www.ibm.com/developerworks/rational/library/4706.html; accessed: 15.06.2011).

ELLINGSHAUS, D. & STEINBRECHER, J. (1990). Junge Fahrer - Besser oder schlechter als ihr Ruf? Cologne: UNIROYAL-Verkehrsuntersuchung.

EMMERSON, K. (2008). Learning to Drive: The Evidence. Road Safety Research Report No. 87. London: Department for Transport.

ENGIN, T., KOCHERSCHEID, K., FELDMANN, M. & RUDINGER, G. (2010). Entwicklung und Evaluation eines Screening-Tests zur Erfassung der Fahrkompetenz älterer Kraftfahrer (SCREEMO). Berichte der Bundesanstalt für Straßenwesen (BASt), Mensch und Sicherheit, Heft M210. Bremerhaven: Wirtschaftsverlag NW.

ERPENBECK, J. & HEYSE, V. (1999). Die Kompetenzbiographie. Münster: Waxmann.

EUROPÄISCHES PARLAMENT & EUROPÄISCHER RAT (2006). Richtlinie 2006/126/EG des Europäischen Parlaments und Rates vom 20. Dezember 2006 über den Führerschein (Neufassung). Brüssel: Amtsblatt der Europäischen Union.

FACK, D. (2000). Automobil, Verkehr und Erziehung. Motorisierung und Sozialisation zwischen Beschleunigung und Anpassung 1885-1945. Opladen: Leske + Budrich.

FAIRCLOUGH, S. H., TATTERSALL, A. J. & HOUSTON, K. (2006). Anxiety and performance in the British driving test. Transportation Research Part F: Traffic Psychology and Behaviour, 9, 43-52.

FAßNACHT, G. (2007). Systematische Verhaltensbeobachtung. Stuttgart: UTB.

FASTENMEIER, W. & GSTALTER, H. (2003). Bewertung von Fahrsicherheitstrainings für junge Fahrer. Munich: Gesamtverband der Deutschen Versicherungswirtschaft e.V. (GDV).

FIEGUTH, G. (1977). Die Entwicklung eines kategorialen Beobachtungsschemas. In U. Mees & H. Selg (Hrsg.), Verhaltensbeobachtung und Verhaltensmodifikation – Anwendungsmöglichkeiten im pädagogischen Bereich (pp. 33-42). Stuttgart: Klett.

FISCHER, W., KLIMA, J., OSTERMAIR, C., STEINBACHER, R., TSCHÖPE, A. & TSCHÖPE, P. (2005). Praktische Ausbildung gestalten – Professionelle Umsetzung der Stufenausbildung nach dem Curricularen Leitfaden. Munich: Verlag Heinrich Vogel.

FISSENI, H.-J. (2004). Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention. Göttingen: Hogrefe.

FLANAGAN, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-359.

FLEISS, J. L. & COHEN, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement. 33, 1973, 613-619.

FLEISS, J. L., LEVIN, B. & PAIK, M. C. (2003). Statistical methods for rates and proportions. New York: Wiley.

FRENZ, W. & CASIMIR-VAN DEN BROEK, E. (2009). Fahrerassistenzsysteme und Anpassungsbedarf im nationalen Recht - Rechtliche Aspekte des Forschungsprojekts "KONVOI". Deutsches Autorecht, Vol. 11, 625-631.

FREY, A. (1999). Aufbau beruflicher Handlungskompetenz – Theoretische Vorstellungen und diag-nostisches Instrumentarium. Empirische Pädagogik,13 (1), 29-56.

FREY, A. (2008). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), Testtheorie und

Fragebogenkonstruktion (pp. 261-278). Heidelberg: Springer.

FRICKE, R. (1974). Kriteriumsorientierte Leistungsmessung. Stuttgart: Kohlhammer.

FRIEDRICHS, J. (1990). Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag.

FUNK, W. & GRÜNINGER, M. (2010). Begleitetes Fahren ab 17 – Prozessevaluation des bundesweiten Modellversuchs. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit Heft M 213. Bremerhaven: Wirtschaftsverlag NW.

FUNK, W., SCHNEIDER, A. H., ZIMMERMANN, R. & GRÜNINGER, M. (2010). Mobilitätsstudie Fahranfänger. Entwicklung der Fahrleistung und Autobenutzung am Anfang der Fahrkarriere. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit Heft M 220, Bremerhaven: Wirtschaftsverlag NW.

GELAU, C., GASSER, T.M. & SEECK, A. (2009). Fahrerassistenz und Verkehrssicherheit. In H. Winner, S. Hakuli & G. Wolf (Hrsg.), Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für die aktive Sicherheit und Komfort (pp. 24-32). Wiesbaden: Vieweg & Teubner.

GENSCHOW, J., STURZBECHER, D. & WILLMES-LENZ, G. (2014). Novice driver preparation – an international comparison. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit. Bremerhaven: Wirtschaftsverlag NW.

GLASER, B. G. & STRAUSS, A. L. (1967). The discovery of grounded theory. Chicago: Aldine.

GRADY, R. (1992). Practical software metrics for project management and process improvement. Englewood Cliffs, NJ: Prentice-Hall.

GRATTENTHALER, H., KRÜGER, H.-P. & SCHOCH, S. (2009). Bedeutung der Fahrpraxis für den Kompetenzerwerb beim Fahrenlernen. (Berichte der Bundesanstalt für Straßenwesen, Reihe "Mensch und Sicherheit", M 201). Bremerhaven: Wirtschaftsverlag NW.

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. Linköping: Swedish National Road and Transport Research Institute.

GREENO, J. G., COLLINS A., & RESNICK, L. B. (1996). Cognition and learning (Ch. 2). In D. C. Berliner & R. C. Calfee (Eds.), Handbook of Educational Psychology (pp. 15–46). New York: MACMILLAN.

GREWE, W. & WENTURA, D. (1997). Wissenschaftliche Beobachtung – Eine Einführung. Weinheim: PVU.

GRONER, R., RAESS, S. & SURY, P. (2008). Usability: Systematische Gestaltung und Optimierung von Benutzerschnittstellen. In B. BATINIC & M. APPEL (Hrsg.), Medienpsychologie (pp. 426-446). Heidelberg: Springer.

GRUBER, H. & MANDL, H. (1996). Das Entstehen von Expertise. In J. Hoffmann & W. Kintsch (Hrsg.), Enzyklopädie der Psychologie, C/II/7, (pp.583-615). Göttingen: Hogrefe.

GRÜMER, K.-W. (1974). Beobachtung. Stuttgart: Teubner.

GRÜNDL, M. (2005). Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen, Universität Regensburg, Dissertation. (URL: http://www.opus-bayern.de/uni-regensburg/volltexte/2005/555/; accessed: 27.10.2011).

GUERRINI, R. (2011). Dossier Notbremsung. L Drive, Nr. 3/2011, pp. 8 & 10.

GUTHKE, J. (1990). Grundzüge der Testkonstruktion und Testauswertung. In J. Guthke, H. R. Böttcher & L. Sprung (Hrsg.), Psychodiagnostik (pp. 201-324). Berlin: Deutscher Verlag der Wissenschaften.

HÄCKER, H. & STAPF, K. (Hrsg.). (2004). Dorsch, Psychologisches Wörterbuch. Berne: Hans-Huber Verlag.

HAMPEL, B. (1977). Möglichkeiten zur Standardisierung der Praktischen Fahrprüfung: Bericht zum Forschungsauftrag 7516 der Bundesanstalt für Straßenwesen. Cologne: Technischer Überwachungs-Verein Rheinland e.V.

HAMPEL, B. & STURZBECHER, D. (2010). Methodische Entwicklung der praktischen Fahrerlaubnisprüfung in der Vergangenheit. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 44-69). Bremerhaven: Wirtschaftsverlag NW.

HAMPEL, B., STURZBECHER, D., MÖNCH, M., TRAUTSCH, J., WAGNER, W. & WEIßE, B. (2009). Die Fahrerlaubnisprüfung als wissenschaftliche Gestaltungsaufgabe. In J. Bönninger, K. Kammler & D. Sturzbecher (Hrsg.), Die Geschichte der Fahrerlaubnisprüfung in Deutschland (pp. 121-176). Bonn: Kirschbaum.

HAMPEL, B. & KÜPPERS, F. (1982). Ermittlung der an Fahrprüfungsorte zu stellenden Anforderung. Bericht zum FA 7516 der Bundesanstalt für Straßenwesen. Cologne: TÜV Rheinland.

HASEMANN, K. (1983). Verhaltensbeobachtung und Ratingverfahren. In K. J. Groffmann & L. Michel (Hrsg.), Enzyklopädie der Psychologie (Serie II: Bd. 4) (pp. 434-488). Göttingen: Hogrefe.

HATAKKA, M., KESKINEN, E., GREGERSEN, N. P. & GLAD, A. (1999). Theories and aims of educational and training measures. In S. SIEGRIST (Hrsg.), Driver training, testing and licensing – towards theory based management of young drivers' injury risk in road traffic (Results of EU Project GADGET, Work package 3, BFU-Report 40, pp. 13-48). Berne: Schweizerische Beratungsstelle für Unfallverhütung.

HEGENER, M. (2003). Methoden zur Evaluation von Software. Bonn: Informationszentrum Sozialwissenschaften.

HEINRICH-BÖLL-STIFTUNG (2004). Selbständig lernen – Bildung stärkt Zivilgesellschaft. 6 Empfehlun-gen der Bildungskommission der Heinrich-Böll-Stiftung. Beltz Verlag: Weinheim and Basle.

HERCZEG, M. (2005). Software-Ergonomie: Grundlagen der Mensch-Computer-Kommunikation. Munich: Oldenbourg.

HEUER J. (2003). Expertenevaluation. In S. HEINSEN & P. VOGT (Hrsg.), Usability praktisch umsetzen – Handbuch für Software, Web, Mobile Devices und interaktive Produkte (pp. 115-135). Munich: Carl Hanser.

HEYDEKORN, J., FRISCH, M. & DACHSELT, R. (2010). Eine Architektur zum flexiblen Einsatz von gestischer Interaktion. In J. Ziegler & A. Schmidt (Hrsg.), Mensch & Computer 2010 (pp. 321-330). Munich: Oldenburg.

HIGHWAY SAFETY CENTER, Indiana University of Pennsylvania, 2009).

HOFFMANN, S. & BULD, S. (2006). Darstellung und Evaluation eines Trainings zum Fahren in der Fahrsimulation. VDI-Berichte Nr. 1960 – Integrierte Sicherheit und Fahrerassistenzsysteme. Düsseldorf: VDI-Verlag.

HÖFT, S. & FUNKE, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), Lehrbuch der Personalpsychologie (pp. 145-187). Göttingen: Hogrefe.

HOLL, F. (2007). Software-Gestaltung: Farbe auf dem Bildschirm. Computer und Arbeit 8-9, 9-16.

HOLTZ, R.F. v. (1998). Der Zusammenhang zwischen Mitarbeiterzufriedenheit und Kundenzufriedenheit. Munich: FGM-Verlag.

HOMBURG, C. (Hrsg.). (2008). Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen. Wiesbaden: Gabler.

HRISTOV, B. (2009). (URL: http://www.qucosa.de/file-admin/data/qucosa/documents/2770/Dissertation-B.Hristov.pdf; accessed: 17.11.2011).

HYATT L. E. & ROSENBERG, L. H. (1996). A software quality model and metrics for identifying project risks and assessing software quality. Utah: 8th Annual Software Technology Conference.

IFO INSTITUT. (2012). Fahrschulen. VR Branchen special, Bericht Nr. 68, 1-4.

INGENKAMP, K.-H. & LISSMANN, U. (2008). Lehrbuch der pädagogischen Diagnostik. Weinheim: Beltz.

ISO/IEC (2005). ISO/IEC 25000 Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE. Geneva: ISO.

JAGOW, F.-J. (2010). Rechtliche Bewertung möglicher methodischer Veränderungen bei der praktischen Fahrerlaubnisprüfung. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 145-150). Bremerhaven: Wirtschaftsverlag NW.

JAMSON, H. (1999). Curve negotiation in the Leeds Driver Simulator: the role of driver experience (Chap. 40). In D. Harris (Ed.), Engineering Psychology and Cognitive Ergonomics (pp. 351-358). Aldershot: Ashgate.

JENSCH, M., SPOERER, E. & UTZELMANN, H. D. (1977). Verkehrsverhaltenslehre. BASt-Bericht FP 7515 Didaktische Verknüpfung von StVO und Gefahrenlehre zur Verbesserung des theoretischen Fahrunterrichts.

JONAS, R. (2009). Effiziente Protokolle und Berichte. Renningen: expert.

JÜRGENS, E. & SACHER, W. (2008). Leistungserziehung und pädagogische Diagnostik in der Schule – Grundlagen und Anregungen für die Praxis. Stuttgart: Kohlhammer.

KAISER, M.O. (2002). Erfolgsfaktor Kundenzufriedenheit: Dimensionen und Messmöglichkeiten. Berlin: Erich Schmidt-Verlag.

KANNING, U. P. (2004). Standards der Personaldiagnostik. Göttingen: Hogrefe.

KANNING, U. P. (2005). Soziale Kompetenzen – Entstehung, Diagnose und Förderung. Göttingen: Hogrefe.

KAUFFELD, S. & GROTE, S. (2002) Kompetenz – ein strategischer Wettbewerbsfaktor. Personal, 11, 30-32.

KAUFHOLD, M. (2006). Kompetenz und Kompetenzerfassung: Analyse und Beurteilung von Verfahren der Kompetenzerfassung. Wiesbaden: VS Verlag für Sozialwissenschaften.

KAUFMANN, C. & RISSER, R. (2009). Verkehrspsychologische Untersuchungen - Validierung mithilfe von Fahrverhaltensbeobachtungen. Zeitschrift für Verkehrssicherheit, 4, 188-193.

KAUSSNER, Y. (2007). Fahrtauglichkeit bei M. Parkinson. Würzburg: Dissertation der Julius-Maximilans-Universität Würzburg.

KELLY, A. V. (2009). The Curriculum: Theory and Practice. London: SAGE Publications Ltd.

KEMPEN, B. (2008). Fahrerassistenz und Wiener Weltabkommen. Vortrag auf dem 3. Sachverständigentag, 25. und 26. Feb. 2008. (URL:. http://www.sachverstaendigentag21.de/svt2008/downloads/2._Vertrag_-_Prof._Kempen_-_Praesentation.pdf; accessed: 17.10.2011).

KERSTING, M. (2008). DIN Screen, Version 2. Leitfaden zur Kontrolle und Optimierung der Qualität von Verfahren und deren Einsatz bei beruflichen Eignungsbeurteilungen. In: Kersting, M. (Hrsg.). Qualitätssicherung in der Diagnostik und Personalauswahl - der DIN Ansatz (pp. 141-210). Göttingen: Hogrefe.

KESKINEN, E., HATAKKA, M. & LAAPOTTI, S. (1988). Final report of the development project concerning the evaluation of driving skills (in Finish). Turku: Department of Psychology, University of Turku.

v. KLEBELSBERG, D., BIEHL, B., FUHRMANN, J. & SEYDEL, U. (1970). Fahrverhalten - 'Beschreibung, Beurteilung und diagnostische Erfassung. Kuratorium für Verkehrssicherheit, Kleine Fachbuchreihe, Sd.8, Vienna.

KLIEME, E. (2004). Der Beitrag von Bildungsstandards zur Qualitätssicherung und Qualitätsentwicklung in Schulen, KMK-Fachtagung "Implementation der Bildungsstandards" Berlin, 2. April 2004 (URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_04_02-Klieme-Bildungsstandards-Qualitaet.pdf; accessed: 16.11.2012).

KLIEME, E., AVENARIUS, H., BLUM, W., DÖBRICH, P., GRUBER, H., PRENZEL, M., REISS, K., RIQUARTS, K., ROST, J., TENORTH, H.-E. & VOLLMER, H. J. (2007). Zur Entwicklung nationaler Bildungsstandards. Eine Expertise (unveränderte Auflage). Bonn/Berlin: Bundesministerium für Bildung und Forschung. http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf; accessed: 16.11.2012)

KLIEME, E. & LEUTNER, D. (2006), Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Überarbeitete Fassung des Antrags an die DFG auf Einrichtung eines Schwerpunktprogramms. (URL: http://kompetenzmodelle.dipf.de/pdf/rahmenantrag; accessed: 26.03.2012).

KNOLL, P.M. (2009). Prädiktive Fahrerassistenzsysteme. Vorlesung an der Universität Karlsruhe im Wintersemester 2008/2009.

KOCHINKA, A. (2010). Handbuch Qualitative Forschung in der Psychologie. In G. Mey & K. Mruck (Hrsg.), Beobachtung (pp. 449-475). Wiesbaden: VS-Verlag.

KOEPPEN, K., HARTIG, J., KLIEME, E. & LEUTNER, D. (2008). Current issues in competence modeling and assessment. Journal of Psychology, 216 (2), 61-73.

KÖTTER, S. & NORDMANN, E. (1987). Diagnostische Beobachtungsverfahren. In: M. Cierpka (Hrsg.), Familiendiagnostik. Berlin: Springer.

KRAFTFAHRT-BUNDESAMT (2012a). Fahrerlaubnisprüfungen – Basistabelle im Jahr 2010. (URL: http://www.kba.de/cln_030/nn_1177708/DE/Statistik/Kraftfahrer/Fahrerlaubnisse/Fahrerlaubnispruefungen/2010/2010__fe__p__feklassen.html; accessed: 22.10.2012).

KRAFTFAHRT-BUNDESAMT (2012b). Fahrlehr-Erlaubnisse – Deutschland und seine Länder am 1. Januar 2012. (URL: http://www.kba.de/cln_030/nn_191556/DE/Statistik/Kraftfahrer/Fahrerlaubnisse/FahrlehrErlaub nisse/2012__fe__fl__dusl__lebensalter.html; accessed: 22.10.2012).

KROHNE, H. W. & HOCK, M. (2007). Psychologische Diagnostik – Grundlagen und Anwendungsfelder. Stuttgart: Kohlhammer.

KROJ, G. & PFEIFFER, G. (1973). Der Kölner-Fahrverhaltens-Test (K-V-F-T). Faktor Mensch im Verkehr, Heft 21.

KUNZ, J. & WEINAND, M. (2012). Qualitätssicherung im Fahrerlaubniswesen. In: HIMMELREICH, K. & HALM, W. (Hrsg.), Handbuch des Fachanwalts, 4. überarbeitete und erweiterte Auflage. Cologne: Luchterhand, 2061-2073.

LAAPOTTI, S., KESKINEN, E., HATAKKA, M. & KATILA, A. (1998). Adding feedback to driving test and customer satisfaction. In E. Keskinen, M. Hatakka, A. Katila, S. Laapotti, & M. Peraaho (Eds.), Psychology in driver education – experiences and visions (pp. 187-218). Turun yliopisto, psykologian laitos (Finnish original).

LAMSZUS, H. (2008). Curricularer Leitfaden – Praktische Ausbildung Pkw. Munich: Bundesvereinigung der Fahrlehrerverbände e.V.

LANGWIEDER, K. (2001). Neue Kenntnisse zum Unfallrisiko junger Fahrer. Fahrschulpraxis, Juni 2001, p. 327.

LAUESEN, S. (2002). Software requirements – styles and techniques. Harlow: Pearson Education.

LEE, S.E., OLSEN, E.C.B. & SIMONS-MORTON, B.G. (2006). Eye glance behavior of novice and experienced adult drivers. TRB 85th Annual Meeting, 22-24 January 2006. Washington, D.C.

LIENERT, G. A. & RAATZ, U. (1998). Testaufbau und Testanalyse. Weinheim: Beltz/PVU.

LOTAN, T., & TOLEDO, T. (2006). An in-vehicle data recorder for evaluation of driving behavior and safety. Preprints of the 85th Annual Transportation Research Board Annual Meeting.

MAAG, U., LABERGE-NADEAU, C., DESJARDINS, D., MESSIER, S. & MORIN I. (2001). Three year injury crash records of new licensees with suspensions or invalidity periods lasting 90 days or more. Annu Proc Assoc Adv Automot Med. 45, 387-401.

MACDONALD, W. (1988). Driver performance measures and licence tests: A literature review. Report CR57. Federal Office of Road Safety.

MARIN, E. & WAWRINOWSKI, U. (2006). Beobachtungslehre: Theorie und Praxis reflektierter Beobachtung und Beurteilung. Weinheim: Juventa.

MARSH, C. J. (2009). Key Concepts for Understanding Curricula. London: Routledge.

MAYCOCK, G. & FORSYTH, E. (1997). Cohort study of learn and novice drivers. Part 4. Novice driver accident 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. & LESTER, F. (1991). The accident liability of car drivers (TRRL RR 315, UMTRI-82404). Crowthorne: Transport and Road Research Laboratory.

MAYHEW, D.R. & SIMPSON, H.M. (1996). Effectiveness and role of driver education and training in a graduated licensing system. Ottawa, Canada: Traffic Injury Research Foundation.

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 (3), 209-219.

McGLADE, F. (1960). A new Road Test. Center of Safety Education New York University Shell Traffic Safety Center.

McGLADE, F. (1963). Testing Driving Performance: Development and Validation techniques. Highway Research Board, News Nr. 5, 13-22.

McKNIGHT, A.J. & ADAMS, B.B. (1970). Driver Education Task Analysis. Volume I: Task Desriptions. Alexandria, VA: Human Resources Research Organization.

McKNIGHT, A.J. & ADAMS, B.B. (1970). Driver Education Task Analysis. Volume II: Task Analysis Methods. Final Report. Alexandria, VA: Human Resources Research Organization.

McKNIGHT, A.J. & HUNDT, A.G. (1971). Driver Education Task Analysis. Volume III: Instructional Objectives. Final Report. Alexandria, VA: Human Resources Research Organization.

McKNIGHT, A.J. & HUNDT, A.G. (1971). Driver Education Task Analysis. Volume IV: The development of instructional objectives. Final Report. Alexandria, VA: Human Resources Research Organization.

MEFFERT, H. & BRUHN, M. (2009). Dienstleistungsmarketing. Grundlagen – Konzepte – Methoden. Mit Fallstudien. Wiesbaden: Gabler.

MERKENS, H. (1997). Stichproben bei qualitativen Studien. In B. Friebertshäuser & A. Prengel (Hrsg.), Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft (pp. 97-106). Weinheim: Juventa.

MEES, U. (1977). Einführung in die systematische Verhaltensbeobachtung. In U. Mees & H. Selg (Hrsg.), Verhaltensbeobachtung und Verhaltensmodifikation – Anwendungsmöglichkeiten im pädagogischen Bereich (pp. 14-32). Stuttgart: Klett.

MICHEL, L. & CONRAD, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K. J. Groffmann & L. Michel (Hrsg.), Enzyklopädie der Psychologie (Serie II: Bd. 1) (pp. 1-129). Göttingen: Hogrefe.

MICHON, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? In L. Evans & R. C. Schwing (Ed.), Human behavior and traffic safety. New York: Plenum Press.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63 (2), 81–97.

MOLINA, K. de (2008). Basiswissen über Fahrerassistenzsysteme. CTI Seminar in Dornach, June 2008.

MÖRL, S., KLEUTGES, C. & ROMPE, K. (2009). Die Entwicklung der Fahrerlaubnisprüfung in der Bundesrepublik 1989. In J. Bönninger, K. Kammler & D. Sturzbecher (Hrsg.), Die Geschichte der Fahrerlaubnisprüfung in Deutschland. Bonn: Kirschbaum.

MOOSBRUGGER, H. & KELAVA, A. (Hrsg.) (2012). Testtheorie und Fragebogenkonstruktion. (2., aktual. u. überarb. Auflage). Heidelberg: Springer.

MÜLLER, M. G. (2003). Grundlagen der visuellen Kommunikation. Konstanz: UVK.

MUSICANT, O., LOTAN, T., & TOLEDO, T. (2007). Safety correlation and implications of an in-vehicle data recorder on driver behavior. TRB Annual Meeting.

NAGEL, E. (2008). Recording device and E-card in the practical test (CIECA-HAK Congress: Harmonisation Saves Lives – Zagreb, Croatia 29-30 May 2008). Accessed on 02.09.2011 via http://www.cieca.be/download/Nagel.pdf.

NERDINGER, F. W., BLICKLE, G. & SCHAPER, N. (2011). Arbeits- und Organisationspsychologie. (2., überarbeitete Auflage). Heidelberg: Springer.

NIELSEN, J. (1992). Finding usability problems through heuristic evaluation. Proceedings (373-380). Gehalten auf der SIGCHI conference on human factors in computing systems. Monterey, CA.

NIELSEN, J. (1993). Iterative user interface design. IEEE Computer, 26 (11), 32-41.

NIELSEN, J. (1994). Heuristic evaluation. In J. NIELSEN & R. L. MACK (Ed.), Usability inspection methods (pp. 25-62). New York, NY: Wiley.

NIELSEN, J. (2011). Parallel & Iterative Design + Competitive Testing = High Usability. (URL: http://www.useit.com/alertbox/design-diversity-process.html; accessed: 23.05.2011)

NIELSEN, J. & LORANGER, H. (2006). Prioritizing web usability. Indianapolis: New Riders Press.

NILSSON, P. O. (2008). Pen and Smartphone fix the practical driving test (CIECA-HAK Congress: Harmonisation Saves Lives – Zagreb, Croatia 29-30 May 2008). (URL: http://www.cieca.be/download/Nilsson.pdf; accessed: 02.09.2011)

NOWAKOWSKA, M. (1973). Grundkonzeption der Theorie psychologischer Tests. Eine nichtformale Darstellung. Probleme und Ergebnisse der Psychologie, 44, 5-23 and 45, 5-18.

OBERMANN, C. (2009). Assessment Center: Entwicklung, Durchführung, Trends. Wiesbaden: Gabler.

OECD (2001). Bildung auf einen Blick. OECD-Indikatoren 2001. Zentrum für Forschung und Innovation im Bildungswesen. Paris.

OLIVA, P. (1997). The curriculum: Theoretical dimensions. New York: Longman.

PAGE, S. (2000). Achieving 100% Compliance of Policies and Producers. New York: Process Improvement Pub.

PARADIES, L., Wester, F. & Greving, J. (2005). Praxisbuch: Leistungsbemessung und –bewertung. Berlin: Cornelsen Scriptor.

PATTON, M. Q. (2002). Qualitative research and evaluation methods. Thousand Oaks, CA: Sage.

PETZHOLTZ, W. (2002). Maximales Abbremsen eines Kategorie-B-Kraftfahrzeuges. CIECA magazine forum, issue nr. 10, pp. 1-2.

PILLER, F. (2006). Kundenintegration im Innovationsprozess als Schlüssel zur Kundenzufriedenheit. In C. HOMBURG (Hrsg.), Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen (pp. 431-460). Wiesbaden: Gabler.

PRATO, C.G., TOLEDO, T., LOTAN, T. & TAUBMAN - BEN-ARI, O. (2010). Modeling the behavior of novice young drivers during the first year after licensure. Accident Analysis & Prevention, vol 42, no. 2, pp. 480-486.

RASMUSSEN, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13(3), 257-266.

RASMUSSEN, J. (1986). Information processing and human-machine interaction. An approach to cognitive engineering. New York: Academic Press.

RETTIG, K. & HORNKE, L. F. (2000). Adaptives Testen. In W. SARGES (Hrsg.), Management-Diagnostik (pp. 557-564). Göttingen: Hogrefe.

RHEINBERG, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. SCHNEIDER & M. HASSELHORN (Hrsg.), Handbuch Pädagogische Psychologie (pp. 178-186). Göttingen: Hogrefe.

RILEY, M. C. & MCBRIDE, R. S. (1974). Safe Performance Curriculum for Secondary School Driver Education: Program Development, Implementation, and Technical Findings. Final Report. Alexandria: Human Resource Research Organization.

RISSER, R. & BRANDSTÄTTER, CHR. (1985). Die Wiener Fahrprobe – Freie Beobachtung: Vienna: Kleine Fachbuchreihe des Kuratoriums für Verkehrssicherheit.

ROELOFS E., VISSERS, J., VAN ONNA, M. & NÄGELE, R. (2009). Validity of an on-road driver performance assessment within an initial Driver training context. Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design. University of Iowa.

ROELOFS, E., VAN ONNA, M. & VISSER, J. (2010). Development of the Driver Performance Assessment: Informing Learner Drivers Cognitive of their Driving Progress, in Driver Behavior and Training, vol. IV. Ed. by Dorn, L. Ashgate, 37-50

ROLOFF, S. (2002). Hochschuldidaktisches Seminar – Mündliche Prüfungen. (URL: http://www.lehrbeauftragte.net/documents_public/MuendlPruef_Roloff.pdf; accessed: 22.08.2011)

ROST, F. (2010). Lern- und Arbeitstechniken für das Studium. Wiesbaden: VS Verlag für Sozialwissenschaften.

ROTH, H. (1971). Pädagogische Anthropologie Band II. Entwicklung und Erziehung. Hannover: Hermann Schroedel Verlag.

ROTH, E. & HOLLING, H. (Hrsg.). (1999). Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis. Munich: Oldenburg.

RUBIN, A. (2012). Statistics for Evidence-Based Practice and Evaluation. Belmont, CA: Brooks Cole.

RÜDEL, M., STURZBECHER, D., GENSCHOW, J. & WEIẞE, B. (2011). Entwicklung innovativer Aufgabentypen und Prüfungsformen zur erweiterten Fahrkompetenzerfassung. In TÜV Þ DEKRA arge tp 21 (Hrsg.), Innovationsbericht zur Optimierung der Theoretischen Fahrerlaubnisprüfung – Berichtszeitraum 2009/2010, 68-75. Dresden: TÜV Þ DEKRA arge tp 21.

SACKETT, P.R., ZEDECK, S. & FOGLI, L. (1988). Relations between measures of typical and maximum job performance. Journal of Applied Psychology, 73, 482-486.

SCHADE, F.-D. (2001). Daten zur Verkehrsbewährung von Fahranfängern. Reanalyse von Rohdaten der Untersuchung von E. Hansjosten und F.D. Schade (1997, Legalbewährung von Fahranfängern. Berichte der BASt, Reihe Mensch und Sicherheit, Heft M71). Unpublished paper. Flensburg: Kraftfahrt-Bundesamt.

SCHECKER, H. & PARCHMANN, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. In: ZfDN 12, 45-66.

SCHNEIDER, W. (1977). Bestandsaufnahme und Entwicklungsaspekte der Fahrerlaubnisprüfung. In: Entwicklungen und Konzepte für die Fahrerlaubnisprüfung. Cologne: TÜV Rheinland.

SCHNEIDER, W. (2008). Ergonomische Gestaltung von Benutzungsschnittstellen - Kommentar zur Grundsatznorm DIN EN ISO 9241-110. Berlin: Beuth.

SCHNEIDER, W. & KORNMEIER, M. (2006). Kundenzufriedenheit: Konzept, Messung, Management. Berne: Haupt.

SCHNELL, R., HILL, P. B. & ESSER, E. (2008). Methoden der empirischen Sozialforschung. Munich: Oldenbourg.

SCHUBERT, & EDLER (1965). Die Auswirkung von Konstitution und Persönlichkeit auf die Unfallgefährdung im Verkehr. Methoden zur Begutachtung des Fahrverhaltens. FA des Bund.Min. f. Gesundheitswesen.

SCHUBERT, W. & WAGNER, T. (2003). Die psychologische Fahrverhaltensbeobachtung – Grundlagen, Methodik und Anwendungsmöglichkeiten. Zeitschrift für Verkehrssicherheit 49(1), 119-127.

SCHUBERT, W., SCHNEIDER, W., EISENMENGER, W. & STEPHAN, E. (2005). Begutachtungs-Leitlinien zur Kraftfahreignung. Bonn: Kirschbaum-Verlag.

SCHULER, H. (2001). Lehrbuch der Personalpsychologie. Göttingen: Hogrefe.

SCHULER, H. & FUNKE, U. (1995). Diagnose beruflicher Eignung und Leistung. In H. SCHULER (Hrsg.), Lehrbuch Organisationspsychologie (pp. 235-283). Göttingen: Hogrefe.

SCHUMANN, S. (2006). Repräsentative Umfrage – Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren. Munich: Oldenbourg.

SCHUSCHKE, W., DAUBENSPECK, H. & SATTELMACHER, P. (2008). Bericht, Gutachten und Urteil. 34., neu bearb. Aufl. Munich: Vahlen.

SCHWENKMEZGER, P. & HANK, P. (1993). Papier-Bleistift- versus computerunterstützte Darbietung von State-Trait-Fragebogen: eine Äquivalenzuntersuchung. Diagnostica, 39 (3), 189-210.

SEEK, A. & GASSER, T. M. (n.d.). Klassifizierung und Würdigung der rechtlichen Rahmenbedingungen im Zusammenhang mit der Einführung moderner FAS. Unpublished paper. Bergisch Gladbach: Bundesanstalt für Straßenwesen.

SEVERING, E. (2011). Prüfungen und Zertifikate in der beruflichen Bildung: eine Einführung. In E. Severing & R. Weiß (Hrsg.), Prüfungen und Zertifizierungen in der beruflichen Bildung: Anforderungen – Instrumente – Forschungsbedarf (pp. 15-36). Bielefeld: wbv.

SEXTON, B. & GRAYSON, G. B. (2010a). The accident history and behaviours of new drivers who pass their first practical driving test. TRL published report (PPR427). Crowthorne: Transport Research Laboratory.

SHNEIDERMAN, B., PLAISAN, C., COHEN, M. & JACOBS, S. (2009). Designing the user interface: Strategies for effective Human-Computer Interaction. Boston: Bearson

SMUC, M., CHRIST R. & GATSCHA, M. (2002). SAF (System zur Analyse des Fahrverhaltens). Regensburg: 38. BDP-Kongress für Verkehrspsychologie.

SPRUNG, L. & SPRUNG, H. (1984). Grundlagen der Methodologie und Methodik der Psychologie. Berlin: Deutscher Verlag der Wissenschaften.

STAPELKAMP, T. (2007). Screen- und Interfacedesign - Gestaltung und Usability für Hard- und Software. Berlin: Springer.

STATISTISCHES BUNDESAMT (2010). Verkehrsunfälle - Unfälle von 18- bis 24- Jährigen im Straßenverkehr. Wiesbaden: Statistisches Bundesamt.

STATISTISCHES BUNDESAMT (2010A). Verkehrsunfälle - Unfallentwicklung im Straßenverkehr. Wiesbaden: Statistisches Bundesamt.

STEHMAN, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment, 62 (1), 77–89.

STEINER (Hrsg.). (2009). Online-Assessment - Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis. Heidelberg: Springer.

STIENSMEIER-PELSTER, J. (2005). Integratives Konzept zur Senkung der Unfallrate junger Fahrerinnen und Fahrer – Evaluation des Modellversuchs im Land Niedersachsen. Berichte der BASt, Reihe Mensch & Sicherheit, Heft M 170. Bremerhaven: Wirtschaftsverlag NW.

STRAUBE, M. (2011). Volkswirtschaftliche Kosten durch Straßenverkehrsunfälle 2009. BASt Forschung kompakt, 4.

STURZBECHER, D. (2004). (Hrsg.). Manual für die pädagogisch qualifizierte Fahrschulüberwachung. Potsdam: Universität.

STURZBECHER, D. (2010). Methodische Grundlagen der praktischen Fahrerlaubnisprüfung. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung –

Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 17-38). Bremerhaven: Wirtschaftsverlag NW.

STURZBECHER, D., BIEDINGER, J., BÖHNE, A., BÖNNINGER, J., v. BRESSENSDORF, G., GLOWALLA, P., KAUP, M., KLEUTGES, C., MÜLLER, G., MÜLLER, R., PETZHOLTZ, W., RADERMACHER, R., SCHMIDT, A. & WAGNER, W. (2010). Die praktische Fahrerlaubnisprüfung in der Gegenwart. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 70-144). Wirtschaftsverlag NW.

STURZBECHER, D., BÖNNINGER, J. & RÜDEL, M. (Hrsg.). (2010). Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit. Bremerhaven: Wirtschaftsverlag NW.

STURZBECHER, D., BÖNNINGER, J., RÜDEL, M. & MÖRL, S. (2011). Arbeitshilfe Fahrerassistenzsysteme und die Prüfung von Fahrkompetenz in der Praktischen Fahrerlaubnisprüfung. Dresden: TÜV |DEKRA arge tp 21.

STURZBECHER, D. & KALTENBAEK, J. (I. D.). “Risikoerkennung bei Fahranfängern” - Was kann die Fahranfängervorbereitung zur Risikobeurteilung beitragen? Tagungsband des 7. gemeinsamen Symposiums DGVP und DGVM in Potsdam. Bonn: Kirschbaum.

STURZBECHER, D., MÖNCH, M., KISSIG, S. & MARSCHALL, M. (2009). Die Entwicklung der Fahrerlaubnisprüfung in Deutschland von den Anfängen bis 1945. In J. Bönninger, K. Kammler & D. Sturzbecher (Hrsg.), Die Geschichte der Fahrerlaubnisprüfung in Deutschland. Bonn: Kirschbaum.

STURZBECHER, D. & MÖRL, S. (2008). Methodensystem zur Erfassung der Zufriedenheit mit der Fahrerlaubnisprüfung. Dresden: TÜV |DEKRA arge tp 21.

STURZBECHER, D., MÖRL, S. & GENSCHOW, J. (2010). Die deutsche praktische Fahrerlaubnisprüfung im internationalen Vergleich. In D. Sturzbecher, J. Bönninger & M. Rüdel (Hrsg.), Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten, Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit (pp. 150-169). Bremerhaven: Wirtschaftsverlag NW.

STURZBECHER, D. & WEIßE, B. (2011). Möglichkeiten der Modellierung und Messung von Fahrkompetenz. In TÜV I DEKRA arge tp 21 (Hrsg.), Innovationsbericht zur Optimierung der Theoretischen Fahrerlaubnisprüfung – Berichtszeitraum 2009/2010. Dresden: TÜV I DEKRA arge tp 21.

SWELLER, J. (2006). How the Human Cognitive System Deals with Complexity. In J. Elen & R. E. Clark (Hrsg.), Handling Complexity in Learning Environments: Theory and Research (pp. 13-25). Amsterdam: Elsevier.

TENORTH, H.-E. & TIPPELT, R. (Hrsg.) (2007). Handlexikon Pädagogik. Weinheim: Beltz.

TESTKURATORIUM (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 09. September 2009. Psychologische Rundschau, 61, 52–56

TIDWELL, J. (2009). Designing interfaces. Sebastopol, CA, USA: O’Reilly Media.

TURNER, C. W., LEWIS, J. R. & NIELSEN, J. (2006). Determining usability test sample size. In W. KARWOWSKI (Hrsg.), International Encyclopedia of Ergonomics and Human Factors (pp. 3084-3088). Boca Raton, FL: CRC Press.

UNDERWOOD, G., CRUNDALL, D. & CHAPMAN, P. (2002). Selective searching while driving: The role of experience in hazard detection and general surveillance. Ergonomics, 45(1), 1-12.

UNDERWOOD, G., CHAPMAN, P., BROCKLEHURST, N., UNDERWOOD, J. & CRUNDALL, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics, 46(6), 629-646.

VEREIN DEUTSCHER INGENIEURE (2002). Menschliche Zuverlässigkeit; ergonomische Forderungen und Methoden. VDI 4006.

VEREIN DEUTSCHER INGENIEURE (2007). Zuverlässigkeitsmanagement. VDI 4003.

WEINERT, F. E. (1998). Neue Unterrichtskonzepte ZWISCHEN gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (Hrsg.), Wissen und Werte für die Welt von morgen (Dokumentation Bildungskongress) (pp. 101-125). Munich.

WEINERT, F. E. (1999). Konzepte der Kompetenz. Gutachten zum OECD Projekt “Definition and

Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo)”. Neuchâtel, Switzerland: Bundesamt für Statistik.

WEINERT, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In Ders. (Hrsg.), LEISTUNGSMESSUNGEN in Schulen (pp. 17-31). Weinheim/Basle.

WEISS, D. J. (1985). Adaptive testing by computer. Journal of Consulting and Clinical Psychology, 53, 774-789.

WELLENREUTHER, M. (1982). Grundkurs: Empirische Forschungsmethoden. Für Pädagogen, Psychologen, Soziologen. Königstein/Ts.: Athensäum.

WELLS, P., TONG, S., SEXTON, B., GRAYSON, G.B. & JONES, P. (2008). Cohort II: A Study of Learner and New Drivers – Findings. Department for Transport (Road Safety Research Report No. 81).

WESTERMANN, R. (2002). Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. In: Zeitschrift für Psychologie, 210 (1), 4–26.

WIDMER, T. & BEYWL, W. (2009). Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), Evaluation - Ein systematisches Handbuch (pp. 243 – 259). Wiesbaden: VS Verlag.

WIECZERKOWSKI, W. & ZUR OEVESTE, H. (1978). Zuordnungs- und Entscheidungsstrategien. In K. L. Klauer (Hrsg.), Handbuch der Pädagogischen Diagnostik, Bd. 4: Methoden der Bildungsberatung und Bildungsforschung (pp. 919-951). Düsseldorf: Schwann.

WILLMES-LENZ, G. (2008). Chancen des Begleiteten Fahrens in Brandenburg, Beitrag zur Verkehrssicherheitskonferenz Potsdam am 13.11.2008, Potsdam. (URL: http://www.aktuell. Forum-verkehrssicheheit.org/_data/Chancen_ des _Begleiteten_Fahrens.pdf; accessed: 26.03.2012).

WILLMES-LENZ, G. (2002). Internationale Erfahrungen mit neuen Ansätzen zur Absenkung des Unfallrisikos junger Fahrer und Fahranfänger. Bergisch-Gladbach: Wirtschaftsverlag NW.

WILLMES-LENZ, G. (2002), Verringerung des Fahranfängerrisikos durch fahrpraktische Vorerfahrung. Referat auf dem 38. BDP-Kongress für Verkehrspsychologie vom 12.-14. September 2002, Universität Regensburg, Regensburg. (URL: http://psydok.sulb.uni-saaland.de/volltexte/2006/752/pdf/willmes_01.pdf; accessed: 26.03.2012).

WILLMES-LENZ, G., GROßMANN, H. & BAHR, M. (2011). Weiterentwicklung der Fahranfängervorbereitung in Deutschland. Thesenpapier zur 4. Fachwerkstatt am 15. November 2011. Bundesanstalt für Straßenwesen, Bergisch Gladbach.

WILLMES-LENZ, G., GROßMANN, H. & PRÜCHER, F. (2010). Evaluation der Fahranfängermaßnahmen “Begleitetes Fahren ab 17” und “Freiwillige Fortbildungsseminare für Inhaber der Fahrerlaubnis auf Probe”. Bericht zum AP-Projekt F1100-4408016 “Evaluation Fahranfängermaßnahmen” vom 31.05.2010 (Erweiterte Fassung). Bundesanstalt für Straßenwesen, Bergisch Gladbach.

WINNER, H., HAKULI, S. & WOLF, G. (2009). Handbuch Fahrerassistenzsysteme. Wiesbaden: GWV Facherverlage GmbH.

WIRTZ, M. & CASPAR, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe.

WOTTAWA, H. & THIERAU, H. (2003). Lehrbuch Evaluation. Berne: Hans Huber.

ZOLLONDZ, H.-D. (2002). Grundlagen Qualitätsmanagement. Einführung in Geschichte, Begriffe, Systeme und Konzepte. Munich/Vienna: Oldenbourg.

Schriftenreihe

Berichte der Bundesanstalt für Straßenwesen

Unterreihe „Mensch und Sicherheit“

2009

M 199: **Maßnahmen zur Verbesserung der visuellen Orientierungsleistung bei Fahranfängern**
Müsseler, Debus, Huestegge, Anders, Skottke € 13,50

M 200: **Entwicklung der Anzahl Schwerstverletzter infolge von Straßenverkehrsunfällen in Deutschland**
Lefering € 13,50

M 201: **Bedeutung der Fahrpraxis für den Kompetenzerwerb beim Fahrenlernen**
Grattenthaler, Krüger, Schoch € 20,00

M 202: **Computergestützte Medien und Fahrsimulatoren in Fahrausbildung, Fahrerweiterbildung und Fahrerlaubnisprüfung**
Weiß, Bannert, Petzoldt, Krems € 16,00

M 203: **Testverfahren zur psychometrischen Leistungsprüfung der Fahreignung**
Poschadel, Falkenstein, Pappachan, Poll,
Willmes von Hinckeldey € 16,50

M 204: **Auswirkungen von Belastungen und Stress auf das Verkehrsverhalten von Lkw-Fahrern**
Evers € 21,00

M 205: **Das Verkehrsquiz – Evaluationsinstrumente zur Erreichung von Standards in der Verkehrs-/Mobilitätserziehung der Sekundarstufe**
Heidemann, Hufgard, Sindern, Riek, Rudinger € 16,50

2010

M 206: **Profile im Straßenverkehr verunglückter Kinder und Jugendlicher**
Holte € 18,50

M 207: **ADAC/BASt-Symposium „Sicher fahren in Europa“**
nur als CD erhältlich € 24,00

M 208: **Volkswirtschaftliche Kosten durch Straßenverkehrsunfälle in Deutschland**
Baum, Kranz, Westerkamp € 18,00

M 209: **Unfallgeschehen auf Landstraßen – Eine Auswertung der amtlichen Straßenverkehrsunfallstatistik**
Heinrich, Pöppel-Decker, Schönebeck, Ulitzsch € 17,50

M 210: **Entwicklung und Evaluation eines Screening-Tests zur Erfassung der Fahrkompetenz älterer Kraftfahrer (SCREEMO)**
Engin, Kocherscheid, Feldmann, Rudinger € 20,50

M 211: **Alkoholverbot für Fahranfänger**
Holte, Assing, Pöppel-Decker, Schönebeck € 14,50

M 212: **Verhaltensanweisungen bei Notsituationen in Straßentunneln**
Färber, Färber € 19,00

M 213: **Begleitetes Fahren ab 17 Jahre – Prozessevaluation des bundesweiten Modellversuchs**
Funk, Grüninger, Dittrich, Goßler, Hornung, Kreßner, Libal,
Limberger, Riedel, Schaller, Schilling, Svetlova € 33,00

2011

M 214: **Evaluation der Freiwilligen Fortbildungsseminare für Fahranfänger (FSF) – Wirksamkeitsuntersuchung**
Sindern, Rudinger € 15,50

M 215: **Praktische Fahrerlaubnisprüfung – Grundlagen und Optimierungsmöglichkeiten – Methodische Grundlagen und Möglichkeiten der Weiterentwicklung**
Sturzbecher, Bönninger, Rüdel et al. € 23,50

M 216: **Verkehrserziehungsprogramme in der Lehreraus-/Fortbildung und deren Umsetzung im Schulalltag – Am Beispiel der Moderatorenkurse „EVA“, „XpertTalks“, „sicherfahren“ und „RiSk“**
Neumann-Opitz, Bartz € 14,50

M 217: **Leistungen des Rettungsdienstes 2008/09 Analyse des Leistungsniveaus im Rettungsdienst für die Jahre 2008 und 2009**
Schmiedel, Behrendt € 16,50

M 218: **Sicherheitswirksamkeit des Begleiteten Fahrens ab 17. Summative Evaluation**
Schade, Heinzmann € 20,00

M 218b: **Summative Evaluation of Accompanied Driving from Age 17**
Schade, Heinzmann
Dieser Bericht liegt nur in digitaler Form vor und kann unter http://bast.opus.hbz-nrw.de/ heruntergeladen werden.

M 219: **Unterstützung der Fahrausbildung durch Lernsoftware**
Petzoldt, Weiß, Franke, Krems, Bannert € 15,50

2012

M 220: **Mobilitätsstudie Fahranfänger – Entwicklung der Fahr-leistung und Autobenutzung am Anfang der Fahrkarriere**
Funk, Schneider, Zimmermann, Grüninger € 30,00

M 221: **Maßnahmen zur Erhöhung der Verkehrssicherheit von Kleintransportern**
Roth € 15,00

M 222: **Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung**
Malone, Biermann, Brünken, Buch € 15,00

M 223: **Evaluation der bundesweiten Verkehrssicherheitskampagne „Runter vom Gas!“**
Klimmt, Maurer € 15,00

M 224: **Entwicklung der Verkehrssicherheit und ihrer Rahmenbedingungen bis 2015/2020**
Maier, Ahrens, Aurich, Bartz, Schiller, Winkler, Wittwer € 17,00

M 225: **Ablenkung durch fahrfremde Tätigkeiten – Machbarkeitsstudie**
Huemer, Vollrath € 17,50

M 226: **Rehabilitationsverlauf verkehrsauffälliger Kraftfahrer**
Glitsch, Bornewasser, Dünkel € 14,00

M 227: **Entwicklung eines methodischen Rahmenkonzeptes für Verhaltensbeobachtung im fließenden Verkehr**
Hautzinger, Pfeiffer, Schmidt € 16,00

M 228: **Profile von Senioren mit Autounfällen (PROSA)**
Pottgießer, Kleinemas, Dohmes, Spiegel,
Schädlich, Rudinger € 17,50

M 229: **Einflussfaktoren auf das Fahrverhalten und das Unfallrisiko junger Fahrerinnen und Fahrer**
Holte € 25,50

M 230: **Entwicklung, Verbreitung und Anwendung von Schulwegplänen**
Gerlach, Leven, Leven, Neumann, Jansen € 21,00

M 231: **Verkehrssicherheitsrelevante Leistungspotenziale, Defizite und Kompensationsmöglichkeiten älterer Kraftfahrer**
Poschadel, Falkenstein, Rinkenauer, Mendzheritskiy, Fimm, Worringer, Engin, Kleinemas, Rudinger € 19,00

M 232: **Kinderunfallatlas – Regionale Verteilung von Kinderverkehrsunfällen in Deutschland**
Neumann-Opitz, Bartz, Leipnitz € 18,00

2013

M 233: **8. ADAC/BASt-Symposium 2012 – Sicher fahren in Europa**
CD-ROM / kostenpflichtiger Download € 18,00

M 234: **Fahranfängervorbereitung im internationalen Vergleich**
Genschow, Sturzbecher, Willmes-Lenz € 23,00

M 235: **Ein Verfahren zur Messung der Fahrsicherheit im Realverkehr entwickelt am Begleiteten Fahren**
Glaser, Waschulewski, Glaser, Schmid € 15,00

M 236: **Unfallbeteiligung von Wohnmobilen 2000 bis 2010**
Pöppel-Decker, Langner
Dieser Bericht liegt nur in digitaler Form vor und kann unter http://bast.opus.hbz-nrw.de/ heruntergeladen werden.

M 237: **Schwer erreichbare Zielgruppen – Handlungsansätze für eine neue Verkehrssicherheitsarbeit in Deutschland**
Funk, Faßmann € 18,00

M 238: **Verkehrserziehung in Kindergärten und Grundschulen**
Funk, Hecht, Nebel, Stumpf € 24,50

M 239: **Das Fahrerlaubnisprüfungssystem und seine Entwicklungspotenziale – Innovationsbericht 2009/2010** € 16,00

M 240: **Alternative Antriebstechnologien – Marktdurchdringung und Konsequenzen – Berichtsjahr 2011 – Abschlussbericht**
Küter, Holdik, Pöppel-Decker, Ulitzsch
Dieser Bericht liegt nur in digitaler Form vor und kann unter http://bast.opus.hbz-nrw.de/ heruntergeladen werden.

M 241: **Intervention für punkteauffällige Fahrer – Konzeptgrundlagen des Fahreignungsseminars**
Glitsch, Bornewasser, Sturzbecher, Bredow, Kaltenbaek, Büttner € 25,50

M 242: **Zahlungsbereitschaft für Verkehrssicherheit – Vorstudie**
Bahamonde-Birke, Link, Kunert € 14,00

2014

M 243: **Optimierung der Praktischen Fahrerlaubnisprüfung**
Sturzbecher, Mörl, Kaltenbaek € 25,50

M 243b: **Optimisation of the practical driving test**
Sturzbecher, Mörl, Kaltenbaek
Dieser Bericht liegt nur in digitaler Form vor, ist interaktiv und kann unter http://bast.opus.hbz-nrw.de/ heruntergeladen werden.

M 244: **Innovative Konzepte zur Begleitung von Fahranfängern durch E-Kommunikation**
Funk, Lang, Held, Hallmeier € 18,50

M 245: **Psychische Folgen von Verkehrsunfällen**
Auerbach € 20,00

M 246: **Prozessevaluation der Kampagnenfortsetzung 2011-2012 „Runter vom Gas!"**
Klimmt, Maurer, Baumann € 14,50

AKTUALISIERTE NEUAUFLAGE VON:
M 115: **Begutachtungsleitlinien zur Kraftfahreignung – gültig ab 1. Mai 2014**
Gräcmann, Albrecht € 17,50

M 247: **Psychologische Aspekte des Unfallrisikos für Motorradfahrerinnen und -fahrer**
von Below, Holte € 19,50

M 248: **Erkenntnisstand zu Verkehrssicherheitsmaßnahmen für ältere Verkehrsteilnehmer**
Falkenstein, Joiko, Poschadel € 15,00

M 249: **Wirkungsvolle Risikokommunikation für junge Fahrerinnen und Fahrer**
Holte, Klimmt, Baumann, Geber € 20,00

M 250: **Ausdehnung der Kostentragungspflicht des § 25a StVG auf den fließenden Verkehr**
Müller € 15,50

M 251: **Alkohol-Interlocks für alkohohlauffällige Kraftfahrer**
Hauser, Merz, Pauls, Schnabel, Aydeniz, Blume, Bogus, Nitzsche, Stengl-Herrmann, Klipp, Buchstaller, DeVol, Laub, Müller, Veltgens, Ziegler € 15,50

M 252 **Psychologische Aspekte des Einsatzes von Lang-Lkw**
Glaser, Glaser, Schmid, Waschulewski
Dieser Bericht liegt nur in digitaler Form vor, ist interaktiv und kann unter http://bast.opus.hbz-nrw.de/ heruntergeladen werden.

2015

M 253: **Simulatorstudien zur Ablenkungswirkung fahrfremder Tätigkeiten**
Schömig, Schoch, Neukum, Schumacher, Wandtner € 18,50

M 254: **Kompensationsstrategien von älteren Verkehrsteilnehmern nach einer VZR-Auffälligkeit**
Karthaus, Willemssen, Joiko, Falkenstein € 17,00

M 255: **Demenz und Verkehrssicherheit**
Fimm, Blankenheim, Poschadel € 17,00

M 256: **Verkehrsbezogene Eckdaten und verkehrssicherheitsrelevante Gesundheitsdaten älterer Verkehrsteilnehmer**
Rudinger, Haverkamp, Mehlis, Falkenstein, Hahn, Willemssen € 20,00

M 257: **Projektgruppe MPU-Reform**
Albrecht, Evers, Klipp, Schulze € 14,00

M 258: **Marktdurchdringung von Fahrzeugsicherheitssystemen**
Follmer, Geis, Gruschwitz, Hölscher, Raudszus, Zlocki in Vorbereitung

M 259: **Alkoholkonsum und Verkehrsunfallgefahren bei Jugendlichen**
Hoppe, Tekaat in Vorbereitung

M 260: **Leistungen des Rettungsdienstes 2012/13**
Schmiedel, Behrendt in Vorbereitung

Alle Berichte sind zu beziehen im:

Carl Schünemann Verlag GmbH
Zweite Schlachtpforte 7
28195 Bremen
Tel. (0421) 3 69 03-53
Fax (0421) 3 69 03-48
www.schuenemann-verlag.de

Dort ist auch ein Komplettverzeichnis erhältlich.