

# Revision zur optimierten Praktischen Fahrerlaubnisprüfung

Berichte der  
Bundesanstalt für Straßenwesen

Mensch und Sicherheit Heft M 268

**bast**

# Revision zur optimierten Praktischen Fahrerlaubnisprüfung

von

Dietmar Sturzbecher  
Universität Potsdam

Philipp Luniak  
Institut für angewandte Familien-,  
Kindheits- und Jugendforschung e.V.  
Universität Potsdam

Susann Mörl  
Institut für Prävention und Verkehrssicherheit  
Kremmen

**Berichte der  
Bundesanstalt für Straßenwesen**

Mensch und Sicherheit Heft M 268

**bast**

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 82.0529/2011:**  
Revision zu einer optimierten Praktischen Fahrerlaubnisprüfung

### 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](http://www.schuenemann-verlag.de)

ISSN 0943-9315

ISBN 978-3-95606-261-2

Bergisch Gladbach, Mai 2016

## Kurzfassung – Abstract

### Revision zur optimierten Praktischen Fahrerlaubnisprüfung

Die Praktische Fahrerlaubnisprüfung nimmt als systematische Fahrverhaltensbeobachtung aufgrund ihrer Steuerungs- und Selektionsfunktion im Gesamtsystem der Fahranfängervorbereitung eine wichtige Rolle ein. Eine Voraussetzung dafür ist der Nachweis ihrer psychometrischen Güte und Praxistauglichkeit. Das vorliegende Revisionsprojekt dient v. a. diesem Nachweis im Hinblick auf die optimierten Durchführungs-, Anforderungs- und Bewertungsstandards, um auf empirischer Grundlage über die Implementierung dieser Standards in Deutschland zu entscheiden.

Zu Beginn des Revisionsprojekts wurden alle notwendigen theoretischen und konzeptionellen Arbeiten abgeschlossen. Dazu gehörten die Fertigstellung des Fahraufgabenkatalogs für alle Fahrerlaubnisklassen, die Programmierung eines einsatzreifen elektronischen Prüfprotokolls (e-Prüfprotokoll), das Weiterbildungskonzept für Sachverständige zur Einführung in die Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung, die Ausarbeitung des Rückmelde-systems einschließlich der schriftlichen Rückmeldung für alle Fahrerlaubnisbewerber und das Evaluationskonzept. Danach wurden die Durchführungs-, Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung mit Untersuchungen zur instrumentellen Evaluation und mit ca. 9.000 realen Praktischen Fahrerlaubnisprüfungen in vier Modellregionen erprobt.

Die Beobachterübereinstimmung wurde als bedeutendster Indikator für die Objektivität und Reliabilität des Messverfahrens bei 15 simulierten und 20 realen Prüfungen der Fahrerlaubnisklasse B, 20 realen Prüfungen verschiedener Fahrerlaubnisklassen und bei drei videografierten Prüfungen untersucht. Im Ergebnis der Untersuchungen im Realverkehr kann festgehalten werden, dass bezüglich der Prüfungsentscheidungen und der Dokumentation der für die Prüfungsentscheidung besonders bedeutsamen schweren Fehler die vorab festgelegten Mindestbeobachterübereinstimmungswerte bei allen drei Untersuchungen erreicht bzw. – teilweise deutlich – übertroffen wurden. Hinsichtlich der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche sowie der Dokumentation der leichten Fehler wurden nur in den beiden erstgenannten Untersuchungen (simulierte und reale Prüfungen Klasse B) zufriedenstellende Übereinstimmungswerte erzielt. Bei der dritten Untersuchung (reale Prüfungen verschiedener Klassen)

konnte ein solcher belastbarer Nachweis (noch) nicht erbracht werden. Dies dürfte durch die erhöhten, noch ungewohnten Anforderungen eines ständigen Wechsels zwischen den zu prüfenden Fahrerlaubnisklassen und einen „Trainingsrückstand“ der Prüfer im Vergleich mit den ersten beiden Untersuchungen begründet sein. Bestätigt wird diese Annahme durch die kontinuierliche Zunahme der Beobachterübereinstimmungswerte bei der Beurteilung videografierten Prüfungen vom ersten bis zum dritten Messzeitpunkt. Demnach gelangten die an dieser Untersuchung teilnehmenden Prüfer mit zunehmender Übung im Umgang mit dem e-Prüfprotokoll zu immer objektiveren und reliableren Ergebnissen, was auf die Erreichbarkeit einer hohen Verfahrensgüte bei ausreichendem Beobachtertraining schließen lässt.

Hinsichtlich der Prüfungsvalidität fanden sich folgende Belege: (1) Die Analyse der Prüfungsergebnisse bestätigte das theoretische und methodische Konzept der optimierten Praktischen Fahrerlaubnisprüfung: Die Inhalte und die Struktur des Fahraufgabenkatalogs und der Kompetenzbereiche sowie das System der Bewertungskriterien haben sich als tragfähig erwiesen. (2) Weiterhin fanden sich bei der Auswertung der Prüfungsleistungen im Hinblick auf die Fahraufgaben und Kompetenzbereiche Ergebnisse, die aus der Verkehrsforschung und Unfallstatistik bekannt sind (z. B. fahranfängertypische Kompetenzdefizite beim Bewältigen der Fahraufgabe „Kreuzungen, Einmündungen“ und im Kompetenzbereich „Verkehrsbeobachtung“) oder zumindest fachlich plausibel erscheinen (z. B. Häufigkeiten bestimmter Bewertungskombinationen). (3) Die Bewerbermerkmale „Vorbereitung einer anderen Fahrerlaubnis“, „Fahrerfahrung“ und „Konsum straßenverkehrsbezogener Computerspiele“ wirkten sich erwartungsgemäß förderlich auf das Bestehen der Prüfung aus. (4) Es existieren akzeptable Zusammenhänge zwischen den prüfungsbasierten Fahrlehrereinschätzungen und Prüferereinschätzungen.

Die Ergebnisse der Reliabilitäts- und Validitätsanalysen belegen also weitestgehend die Verfahrensgüte und damit die Einsatztauglichkeit der optimierten Praktischen Fahrerlaubnisprüfung. Die Implementierung des vorliegenden Verfahrens kann demnach aus wissenschaftlicher Sicht klar empfohlen werden.

Schließlich sollten im Rahmen des Revisionsprojekts mögliche Umsetzungsstrategien und Verfahrensweisen für eine künftige bundesweite Bewerberbefragung und Fahrlehrerbefragung erprobt werden. Die Erprobungsergebnisse zeigen, dass die im Fünfjahresturnus geplanten Befragungen mit aussagekräftigen Ergebnissen und vertretbarem Aufwand durchführbar sind.

### Revision of the optimised practical driving test

The practical driving test, as a systematic observation of driving behavior, plays a key role in the overall system of novice driver preparation due to two functions: “control” and “selection”. As an important prerequisite for the fulfilment of these functions, evidence of its psychometric quality and practicality is required. In the revision project at hand, these issues are investigated with respect to the optimised procedural, demand and assessment standards of the practical driving test in order to create an empirical basis for deciding on the implementation of these standards in Germany.

At the beginning of the revision project, the theoretical and conceptual work necessary was completed. This included the preparation of a catalogue of driving tasks for all licence categories, the programming of a ready-for-use electronic test protocol, the development of a training concept for introducing examiners to the optimised practical driving test, the elaboration of a feedback system including a written feedback for every licence applicant, and the development of an evaluation concept. Following that, the procedural, demand and assessment standards of the optimised practical driving test were tested in a set of field and laboratory studies, including about 9.000 practical driving tests conducted in four model regions under real conditions.

Inter-rater reliability, as the most important indicator of the method’s objectivity and reliability, was examined based on 15 simulated and 20 real driving tests of licence category B (passenger cars), 20 real driving tests of various licence categories and three videotaped simulated driving tests of category B. As a result of the field tests, it can be noted that the criteria for inter-rater reliability that had been defined a priori were met or exceeded – considerably, in part – with respect to the examiners’ decisions on passing the test and with respect to the documentation of serious mistakes, being particularly important for the decision on passing. Regarding the overall assessment of driving tasks and competencies as well as the documentation of slight mistakes, only the first two of the studies mentioned (simulated and real tests of category B) yielded satisfactory reliability scores. Concerning the third study (real tests of various licence categories) sustainable evidence of reliability could not (yet) be provided. This is likely to be due to the increased demands of a steady and

not yet familiar change between the licence categories as well as to a backlog in examiner training compared to the first two studies. This interpretation is corroborated by the finding of a continuous increase in inter-rater reliability scores across the three times of measurement in the ratings of videotaped driving tests: With increasing experience in using the electronic test protocol, raters yielded more objective and reliable results. This suggests that – given sufficient observer training – it is possible to achieve a high test quality.

Concerning validity, evidence was found as follows: (1) Analyses of driving test results endorsed the theoretical and methodological concept of the optimised practical driving test: The contents and structure of the driving task catalogue proved sustainable. (2) Furthermore, in the analysis of test performance data, results replicated findings known from traffic and accident research (e. g. deficiencies typical of novice drivers with regard to the driving task “intersections” and the competence of “traffic observation”) or at least matched theoretical expectations (e. g. concerning the frequency of particular combinations in scoring). (3) With regard to applicant characteristics, “prior possession of another driving licence”, “driving experience” and “experience in playing driving-related video games”, as expected, were positively correlated with passing the test. (4) Test-based assessments made by instructors and examiners showed reasonable correlations.

Hence, the results of the reliability and validity analyses vastly prove the psychometric quality and practicality of the optimised practical driving test. The implementation of the method at hand can thus clearly be recommended from a scientific perspective.

Finally, potential strategies of implementation and procedures for a future nationwide survey of licence applicants and driving instructors were to be tested within the revision project. Results show that such surveys, intended to be conducted every five years, are feasible with reasonable effort and yield meaningful results.

## Summary

### Revision of the optimised practical driving test

#### 1 Initial situation and objectives

In 2013, as in former years, the proportion of 18- to 24-year-olds involved in road accidents was excessive in comparison with other age groups. Novice drivers also have a higher risk of being injured or killed in road traffic than experienced drivers. Against this background, advances in the system of novice driver preparation have to be considered highly desirable. Within the overall system of novice driver preparation, the practical driving test plays a key role in enhancing road safety: Firstly, it regulates access to participation in motor traffic by refusing admittance to those licence applicants who do not meet certain criteria regarding their driving and traffic competencies ("selective function"). Secondly, the demands of the practical driving test serve as a framework and define priorities with regard to the system of driving education, as the contents in education is largely oriented towards the contents tested ("control function"). Due to this particular importance, the practical driving test was brought into focus right at the beginning of the research and development work initiated by the BASt (German Federal Highway Research Institute) in 1997.

Several steps can be distinguished in the previous years' work on optimizing the practical driving test: First of all, starting in 2005, the methodological bases of the test were described and options for optimization were outlined. These options were elaborated, beginning in 2008, and finally consolidated in the form of an optimized test model for licence category B – the OPDT (optimized practical driving test). The OPDT relies on demand standards ("driving tasks") and competence standards ("observation categories") as well as assessment and decision criteria that are scientifically founded and have been developed in cooperation with practitioners from the fields of driving education and examination. These standards and criteria form the basis of an electronic test protocol ("e-test protocol") intended to serve examiners as the essential tool for documentation and assessment in conducting the test, at the same time being an indispensable

prerequisite for a future scientific evaluation of the test. Furthermore, the competence-based assessment of test performance by means of the e-test protocol makes it possible to give licence applicants educationally valuable feedback concerning strengths and weaknesses in their driving behavior right after the practical test. All of these conceptual and technological innovations were put to a first test ("feasibility study") in 2012. The results of the research and development work summarised so far were the starting point of the "revision project on an OPDT" described here.

The focus of the revision project was on empirical evidence of the OPDT's psychometric quality and practicality. From a methodological point of view, the practical driving test is a systematic observation of driving behavior, and, as such, has to meet the criteria of objectivity, reliability, and validity. Furthermore, as a service provided to licence applicants by the Technical Examination Centres on behalf of the German government, the practical driving test has to meet criteria of customer satisfaction. The trial studies of the revision project were intended to enable scientifically founded statements concerning the quality of and - where applicable - the potential for optimization in the overall concept for conducting and evaluating the OPDT, in order to create an empirical basis for deciding on the nationwide implementation of these standards in Germany.

#### 2 Methods

For the trial of the OPDT concept, the following results of previous work were taken up and refined:

- (1) The OPDT demand and assessment standards (as per catalogue of driving tasks),
- (2) the standards for documenting test performance by means of the e-test protocol,
- (3) the evaluation concept for studies on psychometric quality and everyday practicality,
- (4) the feedback concept on the way of giving the licence applicants valuable information concerning their test performance, and
- (5) the training concept for introducing examiners to the new framework of the OPDT.

The trial studies were developed and conducted based on the OPDT evaluation concept that had first been outlined in 2011 in the draft of the



“Handbook on the Driving Licence Testing System (Practice)” and later been described in detail and explained in the BAST report “OPDT” (Sturzbecher et al., 2014).

The overall evaluation concept is based on three pillars: the first one supports the evaluation of psychometric quality regarding the OPDT as an instrument of testing (so-called “instrumental evaluation”), the second and third one (“customer surveys” and “evaluation of test results”) serve the analysis of everyday quality in conducting the practical driving test.

The first set of studies in the revision project, focusing on psychometric quality, incorporated analyses of assessor agreement (providing estimates of objectivity and reliability) and various aspects of validity. The standards outlined concerning test demands, competencies, assessment, documentation, and evaluation were to be tested in a sample of licence applicants for the first time. To this end, multiple approaches were combined: field tests (simulated<sup>1</sup> and genuine practical driving test in real traffic) were conducted as well as controlled studies on the basis of videotaped driving tests. This multi-method approach was chosen in order to compensate for the (inevitable) constraints of individual approaches (e. g. poor potential for standardization and limited number of assessors inside the vehicle in real driving tests) by exploiting the strengths of other methods (e. g. larger number of assessors per test and replicability of stimulus conditions in video-based assessment) for determining the psychometric quality of the OPDT.

The second set of studies, focusing on everyday quality in the conduct of the OPDT, incorporated a field trial of 9.533 practical driving tests conducted with the e-test protocol in four model regions during a period of four and a half months. Along with information from the test protocols, data from surveys of driving test examiners, driving instructors and licence applicants were collected and analysed. In the course of several exchange of experience meetings and additional telephonic

interviews, all examiners were interviewed two times concerning their experience in conducting the optimized test. Licence applicants were interviewed by their examiners directly after the practical tests in order to collect data on personal characteristics relevant to the research questions (e. g. age, gender, education). This allowed analysing if certain characteristics exert an influence on the test results (putting certain groups at a disadvantage), which at the same time provides valuable hints regarding construct validity. A second survey of applicants (mainly concerned with the quality of the performance feedback given after the test) and a survey of driving instructors concerning customer satisfaction were conducted online.

### 3 Results

Before the results of the psychometric evaluation, everyday quality, and customer satisfaction studies are presented, a summary will be given concerning the results of the conceptual work (cf. section 2) that had to be accomplished in advance within the project to allow for subsequent empirical testing:

- (1) The catalogue of driving tasks, which lists the situationally specific demands (driving tasks) to be mastered by licence applicants and the respective criteria for assessment (description of positive and negative patterns of behaviour in carrying out a certain driving task with respect to defined areas of driving competence or observation categories), was comprehensively revised and extended. As a result, it is now available for all licence categories.
- (2) With regard to content and method, the catalogue of driving tasks served as a template for the development and implementation of the e-test protocol, the new IT-based documentation tool for the OPDT. The e-test protocol was implemented as a software application for tablet computers. It was refined based on the results of the feasibility study and made available for all licence categories. Additionally, the IT foundations for the electronic management of test data were laid.
- (3) The individual elements of the methodological system for a future continuous evaluation of the OPDT were worked out in detail and empirically tested. Thus, a well-founded and demonstrably functional inventory of evaluation

---

<sup>1</sup> A „simulated“ practical driving test in this context means that licence applicants, at the end of their driving education, are subjected to the same procedure as in a “real” test, the only difference being that the result is not decisive in terms of licence acquisition.

methods was created with a view to the potential implementation of the OPDT in the future.

- (4) A didactically valuable concept for an individual feedback, based on test performance as documented in the e-test protocol and helping applicants to promote their abilities, was developed and tested. Besides the customary oral discussion of performance after the driving test, the new feedback system includes an elaborate written feedback listing mistakes, instances of above-average performance and hints on how to advance driving competence subsequently. Such well-founded and differentiated feedback concerning their level of driving competence is envisioned to be given to every licence applicant in the future, independent of their test result.
- (5) Based on insights from adult education, a training concept for introducing examiners to the OPDT was developed and tested. The superordinate goal of the training method is to enable examiners to use the e-test protocol in their everyday practice autonomously, correctly and with minimal effort without neglecting their observation task. The concept also served as the basis for training the examiners involved in the field trials of the revision project.

### Psychometric evaluation

For the purpose of assessing the OPDT's psychometric quality, inter-rater reliability, as the most important indicator of objectivity and reliability, was examined based on 15 simulated driving tests of licence category B (passenger cars), 20 real driving tests of licence category B, 20 real driving tests of various licence categories and three videotaped simulated driving tests of category B. In each of the three studies in real traffic, agreement between independent observers (i.e. examiners) was perfect with respect to the final test decisions (passed vs. failed). In sum, as a result of the field tests, it can be noted that the criteria for inter-rater reliability that had been defined a priori were met or exceeded – considerably, in part – with respect to the examiners' decisions on passing the test and with respect to the documentation of serious mistakes, being particularly important for the decision on passing.

The videotaped tests were presented to the examiners for systematic observation and

assessment at three times of measurement: at the beginning, in the middle, and at the end of the four-and-a-half-month period during which the field trials took place. Interrater-reliability deviated from perfect agreement in this study. This could in part be due to constraints in information presentable by means of videography. However, the findings concerning assessment stability across rating times indicate reliable measurement of driving competence by means of the optimized practical driving test: The coefficients of assessment stability, i. e. test-retest-reliability, were medium or high for overall test outcome as well as for overall ratings of individual driving tasks and areas of competence.

With respect to the overall assessment of driving tasks and competencies as well as the documentation of slight mistakes, satisfactory reliability scores were also attained in the simulated and real tests of licence category B. Concerning the real tests of various licence categories, sustainable evidence of reliability could not (yet) be provided. This is likely to be due to the increased demands of a steady and not yet familiar change between the licence categories to be tested or the different parts of the driving task catalogue, respectively, as well as due to a backlog in examiner training compared to the first two studies. This interpretation is corroborated by the finding of a continuous increase in inter-rater reliability scores across the three times of measurement in the ratings of videotaped examinations: With increasing experience in using the electronic test protocol, raters yielded more objective and reliable results. This suggests that, given sufficient observer training, it is possible to achieve a high test quality.

For the purpose of estimating validity, a given examiners' ratings of driving competence for a given applicant were compared to the respective ratings made by a driving instructor, taking the instructor ratings as an external criterion of validity. The two measures showed reasonable correlations.

Factor analyses showed the contents and structure of the catalogue of driving tasks, the areas of competence and the system of assessment criteria to be sustainable: all driving tasks and areas of competence represent independent and non-redundant test demands and facets of driving competence, respectively. Furthermore, in the analysis of test performance data with regard to driving tasks and areas of competence, results



replicated findings known from traffic and accident research (e. g. deficiencies typical of novice drivers with regard to the driving task “intersections” or the competence of “traffic observation”) or at least matched theoretical expectations (e. g. concerning the frequency of particular combinations in scoring).

Evidence of validity was also found in the analyses revolving around applicant characteristics: As expected, “prior possession of another driving licence”, “driving experience” and “experience in playing driving-related video games” were positively correlated with passing the test.

Hence, the results of the reliability and validity analyses vastly prove the psychometric quality and operational suitability of the optimised practical driving test. The implementation of the method at hand can thus be clearly recommended from a scientific perspective. This is especially true if options for improvement occasionally identified in the course of the trial can be put into practice until then.

#### **Customer satisfaction (licence applicants, driving instructors)**

The survey of licence applicants and driving instructors points out high overall satisfaction with regard to the OPDT (e.g., concerning written feedback, oral debriefing, general organizational conditions). The feedback system to be tested was accepted by the majority of applicants questioned: 82% were “very satisfied” (38%) or “somewhat satisfied” (44%) with the written feedback, 76% were “very satisfied” (35%) or “somewhat satisfied” (41%) with the hints on how to advance driving competence, and 70% stated they felt the written feedback would be helpful for their subsequent learning and training in road traffic.

This represent a convincing vote for the implementation of the new feedback system. The appreciation for the system is independent of gender, age, and education of the participants questioned.

#### **4 Conclusions for practice**

Based on the development of a scientifically supported model of an OPDT, the foundations of an ambitious and uniform practical driving test procedure were laid and empirically tested. One aspect is of particular value in this context: the

chance to improve systematically the overall system of novice driver preparation (including driving education) by means of a detailed description of the driving tasks that have to be mastered by applicants and by means of a competence-based assessment and feedback with regard to test performance observed.

Via the determination of situational (driving tasks) and cross-situational (areas of driving competence / observation categories) demands on driving behavior, the contents of the driving test is sharpened and emphasis is put on essential aspects in the process of learning and teaching. Thus, a nationwide implementation of these demand standards would reinforce the control function of the practical driving test.

In laying down uniform test demands by means of an OPDT, a consistent design and conduct of the practical driving test could be granted throughout the country. Simultaneously, the explicit specification of demand standards would contribute to enhancing transparency of demands for licence applicants, driving instructors, and driving examiners. Finally, the use of uniform assessment criteria that relate to specific driving situations as well as well as to comprehensive areas of driving competence enables an unambiguous and transparent assessment of driving performance. Thus, the precision of statements concerning the applicant’s level of driving competence is enhanced, and a more differentiated feedback can be given in order to effectively support the advancement of driving experience.

In the future, an elaborate feedback is envisioned to promote focused retraining of those applicants who did not pass their test as well as the further learning process of those who passed. This will enhance the learning efficacy of the practical driving test and could also promote the subsequent acquisition of driving experience and thus the practical driving test’s efficacy for road safety. As another reason for expecting increased safety efficacy and thus a reinforcement of the practical driving test’s selective function, deficiencies and causes of accidents typical of novice drivers were particularly considered in the determination of driving tasks and assessment criteria.

Furthermore, the electronic registration of test data opens up opportunities for an evidence-based management (“empirical turning point”) in novice

driver preparation. The implementation of the OPDT enables formative and summative evaluation of the driving test, the results of which could form the basis of an empirically founded optimisation of driving education and testing. Thus, in the context of a continuous evaluation, we could assess, e.g., what driving tasks are tested in what testing locations or to what extent testing locations are suitable to assess certain aspects of driving competence. Based on scientific support, i.e. the analysis of test data and other relevant sources such as accident reports, the contents and methods of the practical driving test can be further developed in the future. In an output-oriented control process like that, we could, e.g., investigate in which driving tasks errors occur particularly often and use the findings to optimize driving education accordingly.

The current project aimed at providing empirical evidence of psychometric quality regarding the optimised implementation, demand and assessment standards for the practical driving test as a systematic observation of driving behavior.

The completion of the project lays the foundations to decide on the implementation of these standards on an empirical basis.



## Inhalt

<b>1</b>	<b>Ausgangslage und Zielsetzung</b> . . . . .	13	4.7	Bewerberbefragung zur Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung . . . . .	57
1.1	Ausgangslage . . . . .	13	4.8	Fahrlehrerbefragung . . . . .	58
1.2	Zielstellung . . . . .	14	<b>5</b>	<b>Ergebnisse</b> . . . . .	59
<b>2</b>	<b>Konzeptionelle Arbeiten und Projektvorbereitung</b> . . . . .	16	5.1	Objektivität und Reliabilität . . . . .	59
2.1	Anforderungs- und Bewertungsstandards im Fahraufgabenkatalog . . . . .	16	5.1.1	Überblick . . . . .	59
2.2	Dokumentationsstandards und e-Prüfprotokoll . . . . .	21	5.1.2	Überprüfung der Objektivität und Reliabilität anhand der Beobachterübereinstimmung . . . . .	60
2.3	Evaluationskonzept . . . . .	25	5.1.3	Überprüfung der Reliabilität anhand der Bewertungsstabilität . . . . .	71
2.4	Rückmeldesystem . . . . .	27	5.1.4	Fazit . . . . .	73
2.5	Weiterbildungskonzept . . . . .	32	5.2	Validität . . . . .	74
<b>3</b>	<b>Untersuchungsgegenstände</b> . . . . .	34	5.2.1	Überblick . . . . .	74
3.1	Überblick . . . . .	34	5.2.2	Überprüfung der Validität anhand einer Fahrlehrerbefragung . . . . .	74
3.2	Objektivität . . . . .	34	5.2.3	Überprüfung der Validität anhand der „Technik der bekannten Gruppen“ . . . . .	78
3.3	Reliabilität . . . . .	35	5.2.4	Lernverläufe im Rahmen der Beurteilung videografiertener Prüfungen . . . . .	80
3.4	Validität . . . . .	36	5.2.5	Untersuchung der populations-spezifischen Äquivalenz im Rahmen der Felderprobung . . . . .	80
3.5	Populationsspezifische Äquivalenz . . . . .	37	5.3	Populationsspezifische Äquivalenz . . . . .	81
3.6	Alltägliche Durchführungsqualität . . . . .	40	5.3.1	Überblick . . . . .	81
3.7	Kundenzufriedenheit . . . . .	42	5.3.2	Diskriminante Bewerbermerkmale . . . . .	81
<b>4</b>	<b>Anlage und Durchführung der Erprobungsuntersuchungen</b> . . . . .	44	5.3.3	Konvergente Bewerbermerkmale . . . . .	86
4.1	Überblick . . . . .	44	5.3.4	Zusammenfassung der Befunde . . . . .	89
4.2	Untersuchungen zur instrumentellen Evaluation im realen Straßenverkehr . . . . .	45	5.4	Auswertung von Prüfungsergebnissen . . . . .	89
4.3	Untersuchungen zur instrumentellen Evaluation mittels videografiertener Prüfungen . . . . .	46	5.4.1	Überblick . . . . .	89
4.4	Felderprobung . . . . .	52	5.4.2	Deskriptive Auswertung der Prüfungsanforderungen, Prüfungsbewertungen und Prüfungsentscheidungen . . . . .	90
4.5	Rückmeldungen der Fahrerlaubnisprüfer zur Praxiserprobung des e-Prüfprotokolls . . . . .	54			
4.6	Bewerberbefragung zur populations-spezifischen Äquivalenz . . . . .	55			

5.4.3	Ausgewählte Zusammenhänge zwischen Prüfungsrahmenbedingungen, Prüfungsbewertungen und Prüfungsentscheidungen.....	94
5.4.4	Prüfungsbewertungsmuster und Prüfertypen .....	98
5.5	Kundenzufriedenheit.....	105
5.5.1	Überblick.....	105
5.5.2	Bewerberbefragung .....	105
5.5.3	Fahrlehrerbefragung.....	113
<b>6</b>	<b>Implementierungs- voraussetzungen</b> .....	<b>119</b>
6.1	Schaffung der fahrerlaubnisrechtlichen Voraussetzungen .....	119
6.2	Schaffung der technischen und organisatorischen Voraussetzungen .....	122
6.3	Zeitplan für die Implementierung .....	125
<b>7</b>	<b>Zusammenfassung und Ausblick</b> ...	<b>125</b>
<b>8</b>	<b>Literatur</b> .....	<b>128</b>
	<b>Danksagung</b> .....	<b>133</b>

## Danksagung

Wir danken Herrn Willmes-Lenz und Herrn Bahr für die Fachbetreuung und Unterstützung durch die Bundesanstalt für Straßenwesen.

### Mitwirkende

Die nachstehend aufgeführten Experten haben am Projekt und an der Erarbeitung des vorliegenden Berichts mitgewirkt:

Herr Tino Friedel (TÜV|DEKRA arge tp 21) hat Recherchen und Manuskriptentwürfe für Teile der Kapitel 2, 6 und 7 bereitgestellt.

Herr Mathias Rüdell hat als Geschäftsführer der TÜV|DEKRA arge tp 21, dem Kooperationspartner des Forschungsnehmers, die Forschungs- und Entwicklungsbeiträge der technischen Prüfstellen koordiniert, den Entwurf des Gesamtmanuskripts kommentiert und seine Überarbeitung fachlich beraten.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Schließlich gebührt den technischen Prüfstellen DEKRA, TÜV NORD Mobilität, TÜV Rheinland Group, TÜV SÜD, der Bundeswehr und nicht zuletzt der Bundesvereinigung der Fahrlehrerverbände Dank für ihre vielfältige Unterstützung.

## Anlagen

Die Anlagen zum Bericht sind im elektronischen BAST-Archiv ELBA unter:

<http://bast.opus.hbz-nrw.de> abrufbar



# 1 Ausgangslage und Zielstellung

## 1.1 Ausgangslage

Auch im Jahr 2013 war die Altersgruppe der 18- bis 24-Jährigen wieder überproportional häufig in Straßenverkehrsunfälle verwickelt. Das Risiko, im Straßenverkehr verletzt oder getötet zu werden, fällt für Fahranfänger gegenüber erfahrenen Kraftfahrern deutlich höher aus (Statistisches Bundesamt, 2014). Vor diesem Hintergrund erscheint die Weiterentwicklung des Systems der Fahranfängervorbereitung dringend geboten.

Das System der Fahranfängervorbereitung umfasst sämtliche Maßnahmen, die dafür geeignet sind, Fahrkompetenz bei Fahrerlaubnisbewerbern<sup>1</sup> aufzubauen und zu überprüfen (GENSCHOW, STURZBECHER & WILLMES-LENZ, 2013). Seit 1997 hat die Bundesanstalt für Straßenwesen (BASt) eine Reihe von Projekten initiiert, welche die einzelnen Elemente des Systems der Fahranfängervorbereitung mit Blick auf ihre Bedeutung für die Verkehrssicherheit thematisieren und Optimierungsimpulse setzen. Im Zentrum der Weiterentwicklungsarbeiten stehen von Anfang an nicht zuletzt die theoretische und Praktische Fahrerlaubnisprüfung.

Die Praktische Fahrerlaubnisprüfung spielt im System der Fahranfängervorbereitung eine bedeutsame Rolle, da sie erstens den Zugang zur motorisierten Teilnahme am Straßenverkehr reguliert: Mit ihr wird Fahrerlaubnisbewerbern, die bestimmte Kriterien der Fahr- und Verkehrskompetenz nicht erfüllen, gezielt das Führen von Kraftfahrzeugen im öffentlichen Straßenverkehr verwehrt (Selektionsfunktion). Zweitens sind die Anforderungen der Praktischen Fahrerlaubnisprüfung strukturbildend und prioritätensetzend für das Fahranfängerausbildungssystem, da sich die Ausbildungsinhalte weitgehend an den Prüfungsinhalten orientieren (Steuerungsfunktion).

Eine notwendige Voraussetzung für die effektive Nutzung der Selektions- und Steuerungsfunktion der Praktischen Fahrerlaubnisprüfung ist die Sicherstellung ihrer instrumentellen Verfahrensgüte im Sinne testpsychologischer Kriterien. Nur wenn

die geforderten Fahrkompetenzen mit der Prüfung in objektiver, reliabler (d. h. zuverlässiger) und valider (d. h. gültiger) Weise erfasst und bewertet werden, können die gewünschten Prüfungseffekte auf die Verkehrssicherheit erzielt werden. Die Kontrolle und ggf. Erhöhung der psychometrischen Verfahrensgüte der Praktischen Fahrerlaubnisprüfung stellt daher ein besonders wichtiges verkehrspädagogisch-psychologisches Forschungs- und Entwicklungsfeld dar.

Die Verfahrensgüte und Verkehrssicherheitswirksamkeit der Praktischen Fahrerlaubnisprüfung können nur aussagekräftig evaluiert werden, wenn entsprechende Untersuchungsdaten vorliegen. In diesem Zusammenhang kommt der detaillierten Dokumentation der Prüfungsleistungen sowie der Archivierung und Evaluation darauf beruhender anonymisierter Prüfungsdaten eine entscheidende Bedeutung zu. Gegenwärtig beschränken sich die Evaluation von Prüfungsdaten und damit auch die empirisch gestützte Qualitätssicherung der Praktischen Fahrerlaubnisprüfung vorwiegend auf die Analyse der Bestehensquoten der Bewerber. Dies stellt jedoch keine ausreichende Grundlage für den Nachweis einer ausreichenden Verfahrensgüte und Sicherheitswirksamkeit der Prüfung dar. Folglich sind auch die Optimierung der Dokumentation und Evaluation von Prüfungsleistungen als wichtige verkehrspädagogisch-psychologische Herausforderungen anzusehen.

Die wissenschaftlichen Entwicklungsarbeiten zur Praktischen Fahrerlaubnisprüfung fokussieren auf die genannten Qualitätskriterien „Verfahrensgüte“ und „Sicherheitswirksamkeit“ und reichen bis in die 1970er Jahre zurück. HAMPEL (1977, 1982), der Leiter der 1974 errichteten „Studienstelle für Kraftfahrzeug-Führerprüfungen des TÜV Rheinland, begann zu dieser Zeit – teilweise im Rahmen von BASt-Projekten – unter testpsychologischen Gesichtspunkten die Praktische Fahrerlaubnisprüfung zu analysieren und wegweisende Empfehlungen zu ihrer Weiterentwicklung abzuleiten. An diese Arbeiten knüpfte die Bundesanstalt für Straßenwesen (BASt) im Jahr 1997 mit der Ausschreibung des Projektes „Optimierung der Fahrerlaubnisprüfung“ an. Wie schon in den 1970er Jahren wurden die nachfolgenden verkehrspädagogisch-psychologischen Forschungs- und Entwicklungsarbeiten in Kooperation mit den technischen Prüfstellen bzw. mit der nicht zuletzt zu diesem Zweck von ihnen gegründeten Arbeitsgemeinschaft TÜV | DEKRA arge tp 21 vorangetrieben. Seit 2005 finden sich im Hin-

<sup>1</sup> Aus Lesbarkeitsgründen gilt bei allen personenbezogenen Bezeichnungen die gewählte Form für beide Geschlechter.

blick auf die Weiterentwicklung der Praktischen Fahrerlaubnisprüfung folgende Projekte:

- das TÜV|DEKRA arge tp 21-Projekt „Methodische Grundlagen und Möglichkeiten der Weiterentwicklung der Praktischen Fahrerlaubnisprüfung“ (2005 bis 2008), in dem die theoretischen, methodischen und rechtlichen Grundlagen der Praktischen Fahrerlaubnisprüfung einschließlich ihrer Entwicklung seit den 1970er Jahren beschrieben sowie Ansätze zur Weiterentwicklung und Optimierung skizziert wurden (STURZBECHER, BÖNNINGER & RÜDEL, 2008 bzw. 2010);
- das BAST-Projekt „Optimierung der Praktischen Fahrerlaubnisprüfung“ (2008 bis 2012), in dem die methodischen Grundlagen vertieft und weitere Überlegungen zu Optimierungsmöglichkeiten mit Blick auf eine mittel- und langfristige Weiterentwicklung der Praktischen Fahrerlaubnisprüfung im System der Fahranfängervorbereitung erarbeitet wurden (STURZBECHER, MÖRL & KALTENBAEK, 2014);
- die Machbarkeitsstudie der TÜV|DEKRA arge tp 21 (2011 bis 2012) zur Praktikabilität einer elektronischen Dokumentation von Prüfungsleistungen im Prüfungsfahrzeug (FRIEDEL, MÖRL & RÜDEL, 2012) und insbesondere zur Bedienung eines elektronischen Prüfprotokolls (im Folgenden „e-Prüfprotokoll“) sowie
- das BAST-Projekt „Revision zu einer optimierten Praktischen Fahrerlaubnisprüfung“ („Revisionsprojekt“; 2013 bis 2015),<sup>2</sup> in dem ein tragfähiges methodisches Konzept für die computergestützte Durchführung sowie für die instrumentelle und prozessuale Evaluation der optimierten Praktischen Fahrerlaubnisprüfung erarbeitet und erprobt wurde. Darüber hinaus diente dieses Projekt der Schaffung der technischen Voraussetzungen für alle Verfahren und Abläufe einer computergestützten Durchführung und Weiterentwicklung der optimierten Praktischen Fahrerlaubnisprüfung.

Zusammenfassend bleibt festzuhalten, dass die Weiterentwicklung der Praktischen Fahrerlaubnis-

prüfung eine stetige wissenschaftliche Gestaltungsaufgabe darstellt, die in Deutschland seit 40 Jahren betrieben wird (HAMPEL, STURZBECHER, MÖNCH, TRAUTSCH, WAGNER & WEIßE, 2009). Insbesondere in den letzten 10 Jahren wurden die Entwicklungsarbeiten intensiviert; sie mündeten in das „Revisionsprojekt zu einer optimierten Praktischen Fahrerlaubnisprüfung“. Im Revisionsprojekt wurde an die konzeptionellen Arbeiten von STURZBECHER et al. (2014) zur konsequenten Ausgestaltung der Praktischen Fahrerlaubnisprüfung als systematischer Fahrverhaltensbeobachtung mit elektronischer Leistungsdokumentation sowie an die empirischen Ergebnisse der Machbarkeitsstudie der TÜV|DEKRA arge tp 21 (FRIEDEL et al., 2012) angeknüpft. Im Mittelpunkt stand die abschließende Erprobung des nach der Machbarkeitsstudie revidierten Entwurfs des Beobachtungs- und Dokumentationsverfahrens, um auf wissenschaftlicher Grundlage empiriegestützt über seine flächendeckende Einführung in Deutschland zu entscheiden. Die Beschreibung der Anlage und der Ergebnisse der methodenkritischen Untersuchungen zur instrumentellen Verfahrensgüte und zur alltäglichen Durchführungsqualität der optimierten Praktischen Fahrerlaubnisprüfung stellen den Inhalt des vorliegenden Forschungsberichts zum Revisionsprojekt dar.

## 1.2 Zielstellung

Das Ziel des Revisionsprojekts besteht darin, den zuvor im BAST-Projekt „Optimierung der Praktischen Fahrerlaubnisprüfung“ erarbeiteten Entwurf für ein optimiertes Prüfungskonzept einschließlich der weiterentwickelten Anforderungs-, Bewertungs-, Dokumentations- und Evaluationsstandards zu erproben und in Abhängigkeit von den jeweiligen Ergebnissen zu überarbeiten (STURZBECHER et al., 2014). Im Ergebnis soll ein erprobtes Betriebskonzept zur Durchführung und Evaluation der optimierten Praktischen Fahrerlaubnisprüfung in Deutschland einschließlich Empfehlungen für die Implementierung vorliegen. Um die Voraussetzungen für eine Erprobung des optimierten Prüfungskonzepts zu schaffen, wurden die folgenden Vorarbeiten im Revisionsprojekt aufgegriffen, weiterentwickelt und abgeschlossen:

- 1) Die Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung (gemäß Fahraufgabenkatalog),

<sup>2</sup> Dieses Projekt wurde im Auftrag der BAST als Kooperationsprojekt des Instituts für angewandte Familien-, Kindheits- und Jugendforschung an der Universität Potsdam mit der TÜV|DEKRA arge tp 21 durchgeführt.

- 2) Die Standards für die Dokumentation der Prüfungsleistungen mittels e-Prüfprotokoll,
- 3) das Evaluationskonzept bezüglich der Untersuchungen zur Verfahrensgüte („instrumentelle Evaluation“) und zur alltäglichen Durchführungsqualität (einschließlich der Studien zur Kundenzufriedenheit mit der Prüfung und zur populationspezifischen Äquivalenz) der optimierten Praktischen Fahrerlaubnisprüfung,
- 4) Das Konzept für das Rückmeldesystem zu den Prüfungsleistungen der Fahrerlaubnisbewerber sowie
- 5) Das Fortbildungskonzept für Fahrerlaubnisprüfer zur Einführung in das neue Prüfungskonzept und in die Durchführungsstandards der optimierten Praktischen Fahrerlaubnisprüfung.

Zu 1): Auf Basis fahrkompetenztheoretischer Betrachtungen, einer handlungstheoretischen Anforderungsanalyse der Kraftfahrzeugführung im öffentlichen Straßenverkehr und der methodischen Konstruktion der Praktischen Fahrerlaubnisprüfung als systematischer Fahrverhaltensbeobachtung haben STURZBECHER et al. (2014) einen Fahraufgabenkatalog entworfen, der die Anforderungs- und Bewertungsstandards der Praktischen Fahrerlaubnisprüfung beinhaltet. Er enthält insgesamt acht situationsbezogene Fahraufgaben, die teilweise in Teilfahraufgaben unterteilt werden, und fünf Beobachtungskategorien zur Beurteilung situationsübergreifender Fahrkompetenzbereiche. Weiterhin werden im Fahraufgabenkatalog für jede Fahraufgabe mit Bezug zu den Kompetenzbereichen ereignisorientierte Bewertungskriterien beschrieben: Die ereignisorientierten Kriterien bezeichnen sowohl leichte und schwere Fehler als auch Beispiele für überdurchschnittliche Leistungen. Schließlich wird die ereignisbezogene Bewertung der Prüfungsleistungen durch eine kompetenzbezogene Bewertung ergänzt. Diesbezügliche Bewertungskriterien erlauben die Einschätzung der Prüfungsleistungen auf einer vierstufigen Ratingskala. Im vorliegenden Revisionsprojekt sollen diese Anforderungs- und Bewertungsstandards geschärft und teilweise ergänzt werden. Das Ziel dieser Arbeiten besteht – neben der Erhöhung der Verkehrssicherheit – in der Objektivierung der Prüfungsdurchführung und -bewertung.

Zu 2): Der o. g. Fahraufgabenkatalog stellt die inhaltliche und methodische Grundlage für die Entwicklung und IT-gestützte Umsetzung des e-Prüfprotokolls als neuem Dokumentationsinstrument dar. Das e-Prüfprotokoll wurde als eine Software-Anwendung (Programm) für Tablet-Computer entwickelt. Dieses Programm unterstützt den Fahrerlaubnisprüfer bei der Prüfungsplanung, bei der Prüfungsdurchführung, bei der Prüfungsbewertung, bei der Prüfungsentscheidung und nicht zuletzt bei der Rückmeldung der Prüfungsleistungen an den Bewerber. Die Gebrauchstauglichkeit und Praktikabilität des e-Prüfprotokolls wurde bereits im Rahmen der Machbarkeitsstudie (FRIEDEL et al., 2012) untersucht (s. o.). Im Zuge des Revisionsprojekts soll das e-Prüfprotokoll zu einer für die Felderprobung einsatzfähigen Version weiterentwickelt werden. Darüber hinaus sollen sämtliche IT-technische Voraussetzungen für die elektronische Verwaltung der Prüfungsdaten geschaffen werden.

Zu 3): Mit der Entwicklung einer elektronischen Prüfungsdokumentation wurden auch die Grundlagen für die wissenschaftliche Evaluation der Praktischen Fahrerlaubnisprüfung wesentlich verbessert. Als Ausgangspunkt für die Erarbeitung eines Evaluationssystems wurden unter Federführung der BAST fachliche Sondierungsgespräche mit leitenden Vertretern und Qualitätsmanagementbeauftragten der vier mit der Durchführung der Fahrerlaubnisprüfung beauftragten technischen Prüfstellen und der Bundeswehr durchgeführt. In diesen Sondierungsgesprächen wurden nicht zuletzt Vorstellungen zum Evaluationskonzept der künftigen optimierten Praktischen Fahrerlaubnisprüfung erörtert; die Diskussionsergebnisse sind in das von STURZBECHER et al. (2014) skizzierte Evaluationskonzept eingeflossen. Im Revisionsprojekt sollen die einzelnen Evaluationselemente konkretisiert und empirisch erprobt werden. Das Ziel dieser Maßnahme ist – mit Blick auf eine künftige kontinuierliche Evaluation der Praktischen Fahrerlaubnisprüfung – die Schaffung eines fundierten und nachweislich funktionsfähigen Methodeninventars.

Zu 4): Für einen effektiven Lernprozess ist es von zentraler Bedeutung, dass der Bewerber nach der Praktischen Fahrerlaubnisprüfung

eine qualifizierte Rückmeldung zum erreichten Fahrkompetenzniveau erhält. Diese Rückmeldung kann der Bewerber dann sicherheitsfördernd für das anschließende Weiterlernen beim selbstständigen Fahren – auch im Rahmen des Begleiteten Fahrens mit 17 (BF17) – bzw. lerngewinnbringend für eine ggf. notwendige Prüfungswiederholung verwenden. Fachlich und verkehrspädagogisch wertvolle Rückmeldungen setzen eine prüfungsbegleitende Dokumentation der Prüfungsinhalte und Prüfungsleistungen voraus. Hierfür könnte die elektronische Prüfungsdokumentation eine ideale Voraussetzung darstellen. Im Rahmen des Revisionsprojekts soll ein prüfungsdidaktisch anspruchsvolles Konzept für eine förderorientierte, individualisierte Rückmeldung an den Bewerber erarbeitet und erprobt werden. Die – zusätzlich zum abschließenden Prüfungsgespräch zwischen Fahrerlaubnisbewerber und Fahrerlaubnisprüfer – bereitzustellende schriftliche Rückmeldung soll standardisiert und basierend auf den im e-Prüfprotokoll dokumentierten und bewerteten Prüfungsleistungen generiert werden.

Zu 5): Die Praktische Fahrerlaubnisprüfung stellt aus methodischer Sicht eine fahrkompetenzbezogene systematische (Fahr-)Verhaltensbeobachtung dar. Die Durchführungsqualität von Verhaltensbeobachtungen und somit auch die Güte der Beobachtungsergebnisse sind einerseits von der Eignung des Beobachtungsinstruments und andererseits von den Kompetenzen der Beobachter und deren Schulung abhängig (FIEGUTH, 1977; BUNGARD, HOLLING & SCHULTZ-GAMBARD, 1996). Die Beobachterschulung im Sinne einer Fortbildung der Fahrerlaubnisprüfer zum theoretischen und methodischen Konzept der optimierten Praktischen Fahrerlaubnisprüfung im Allgemeinen sowie zur Verwendung des e-Prüfprotokolls im Besonderen nimmt daher im Revisionsprojekt einen wichtigen Stellenwert ein: Nur ausreichend geschulte Fahrerlaubnisprüfer sind in der Lage, zuverlässige und gültige Beobachtungsdaten zu liefern. Im Rahmen des Revisionsprojekts soll daher – basierend auf erwachsenenpädagogischen Erkenntnissen – ein Konzept für eine Fortbildungsmaßnahme für Fahrerlaubnisprüfer zur optimierten Praktischen Fahrerlaubnisprüfung erarbeitet und

erprobt werden. Das übergeordnete Ziel der Fortbildungsmaßnahme besteht darin, dass die Fahrerlaubnisprüfer das e-Prüfprotokoll selbstständig, fachlich korrekt und mit geringem Bedienungsaufwand in ihrem Prüfungsalldag anwenden können, ohne ihre Beobachtungsaufgaben zu vernachlässigen bzw. einzuschränken.

Im vorliegenden Bericht wird dargelegt, auf welchem Wege und mit welchen Ergebnissen die skizzierten Ziele im Revisionsprojekt verfolgt wurden. Als Ausgangspunkt werden im Kapitel 2 die für das Revisionsprojekt notwendigen konzeptionellen Vorarbeiten beschrieben. Dazu gehört die Weiterentwicklung des Fahraufgabenkatalogs, in dem die Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung aufgeführt sind. Weiterhin wird die Erarbeitung und Programmierung des e-Prüfprotokolls bis hin zur Einsatzreife erläutert. Darüber hinaus werden die verschiedenen Studien für die Untersuchung der Verfahrensgüte und zur Erprobung des Evaluationskonzepts vorgestellt. Schließlich folgt die Darstellung der Grundlagen des erarbeiteten Rückmeldesystems für die Fahrerlaubnisbewerber und des Fortbildungskonzepts für Fahrerlaubnisprüfer. Im Kapitel 3 werden dann die Untersuchungsgegenstände präzisiert. Im Kapitel 4 werden die entsprechenden Untersuchungsanlagen und die Durchführung der Felderprobung beschrieben. Kapitel 5 dient der Darstellung der Erprobungsergebnisse, und im Kapitel 6 wird erläutert, wie die Implementierung der optimierten Praktischen Fahrerlaubnisprüfung umgesetzt werden könnte, welche Voraussetzungen (z. B. rechtliche Anpassungen) dafür noch notwendig sind und wie die nächsten Schritte bei der Weiterentwicklung der optimierten Praktischen Fahrerlaubnisprüfung aussehen könnten.

## 2 Konzeptionelle Arbeiten und Projektvorbereitung

### 2.1 Anforderungs- und Bewertungsstandards im Fahraufgabenkatalog

Aus methodischer Sicht stellt die Praktische Fahrerlaubnisprüfung eine Arbeitsprobe zur Ermittlung der Fahrkompetenz mittels einer systematischen (Fahr-)Verhaltensbeobachtung dar (STURZBECHER et al., 2010). Der Fahrerlaubnisbewerber



bewältigt während der Prüfungsfahrt festgelegte Anforderungen in Form von Fahraufgaben, die durch den Fahrerlaubnisprüfer anhand von Beobachtungskategorien und Bewertungskriterien dokumentiert und beurteilt werden. Für eine einheitliche Durchführung der Praktischen Fahrerlaubnisprüfung wurden in einem langjährigen Entwicklungsprozess Anforderungs- und Bewertungsstandards erarbeitet, die in einem Fahraufgabenkatalog beschrieben sind.

Den Ausgangspunkt bei der Erarbeitung der Anforderungsstandards bildete

- 1) eine empirisch gestützte handlungspsychologische Analyse der Anforderungen beim Führen eines Kraftfahrzeugs im öffentlichen Straßenverkehr sowie
- 2) eine Beschreibung und Sicherheitsbewertung der für den Straßenverkehr ausbildungsrelevanten Fahraufgaben aus den 1970er Jahren (McKNIGHT & ADAMS, 1970, 1971; McKNIGHT & HUNDT, 1971).

HAMPEL legte 1977 unter Einbeziehung dieser Arbeiten einen Vorschlag zur Vereinheitlichung und Neustrukturierung der Fahraufgaben der deutschen Praktischen Fahrerlaubnisprüfung vor. Die genannten Anforderungsanalysen und Fahraufgabenkataloge wurden von STURZBECHER et al. (2014) aufgegriffen und für die Optimierung des gegenwärtigen deutschen Prüfungssystems weiterentwickelt. Im Ergebnis entstand ein Katalog von acht situationsbezogenen Fahraufgaben, die in 17 Teilfahraufgaben unterteilt werden. Für alle (Teil-)Fahraufgaben wurden Definitionen und Handlungsalgorithmen zur Bewältigung der grundsätzlichen Handlungsanforderungen erarbeitet. Die Auswahl und Abgrenzung der Fahraufgaben erfolgte unter Berücksichtigung wissenschaftlicher Befunde zu ihrer Sicherheitsrelevanz sowie unter Einbeziehung der Ergebnisse eines internationalen Vergleichs mit den Prüfungsanforderungen anderer EU-Länder.

Die genannten Arbeiten aus den 1970er Jahren wurden ferner genutzt, um – unter Berücksichtigung des aktuellen Forschungsstandes zu fahrfähigertypischen Defiziten – situationsübergreifende Fahrkompetenzbereiche zu definieren. Die Fahrkompetenzbereiche besitzen eine Doppelfunktion: Sie stellen einerseits wichtige Komponenten der Fahrkompetenz dar, die in der Prüfung bewertet werden sollen. Andererseits strukturieren sie gleichzeitig das Beobachtungsverhalten des Prü-

fers während der Prüfungsfahrt, indem sie vorgeben, auf welche Verhaltensweisen des Bewerbers grundsätzlich zu achten ist. Zu den Kompetenzbereichen bzw. Beobachtungskategorien zählen die Bereiche bzw. Kategorien

- 1) Verkehrsbeobachtung,
- 2) Fahrzeugpositionierung,
- 3) Geschwindigkeitsanpassung,
- 4) Kommunikation und
- 5) Fahrzeugbedienung/Umweltbewusste Fahrweise (STURZBECHER et al., 2014).

Die ausführliche fachliche Beschreibung der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien wurde von einer Expertengruppe „AG Fahraufgabenbeschreibung“<sup>3</sup> vorgenommen. Mit der Einbeziehung dieses Expertengremiums sollte die Praxistauglichkeit der zunächst wissenschaftlich begründeten Entwicklungsarbeiten sichergestellt werden. Im Rahmen der „AG Fahraufgabenbeschreibung“ wurden darüber hinaus – unter Berücksichtigung der gesetzlichen Vorgaben aus der „Richtlinie für die Prüfung der Bewerber um eine Fahrerlaubnis zum Führen von Kfz“ (Prüfungsrichtlinie) – situationsbezogene Bewertungskriterien – d. h. Fehler, erwartungsgemäße Leistungen und überdurchschnittliche Leistungen – pro Fahraufgabe und Kompetenzbereich erarbeitet. Diese Festlegungen erlauben die konkrete Zuordnung einer Prüfungsleistung zu einer Fahraufgabe, einem Kompetenzbereich und einer Bewertungsstufe. Weiterhin haben STURZBECHER et al. (2014) unter Einbeziehung der „AG Fahraufgabenbeschreibung“ eine Skala für die situationsübergreifende Gesamtbewertung der Fahraufgaben und Kompetenzbereiche erarbeitet: Die dokumentierten situationsbezogenen Bewertungskriterien stellen also die Grundlage für die abschließende situationsübergreifende Gesamtbewertung der Fahraufgaben und Kompetenzbereiche dar, auf denen wiederum die Prüfungsentscheidung basiert.

Im vorliegenden Revisionsprojekt wurden die genannten Anforderungs- und Bewertungsstandards

<sup>3</sup> Der „Arbeitsgruppe Fahraufgabenbeschreibung“ gehörten Vertreter der Bundesanstalt für Straßenwesen (BASt), der Bundesvereinigung der Fahrlehrerverbände (BVF), der technischen Prüfstellen und der Bundeswehr sowie der TÜV|DEKRA arge tp 21 und verschiedener wissenschaftlicher Institutionen an.



im Rahmen der fortgeführten „AG Fahraufgabenbeschreibung“ geschärft und teilweise ergänzt. Zusätzlich wurden für einige Fahraufgaben und die dazugehörigen Bewertungskriterien Indikatoren (beispielhafte Beschreibungen von Fahrverhalten in einer bestimmten Situation) erarbeitet, die für den Fahrerlaubnisprüfer als Orientierung für die Bewertung der beobachteten Prüfungsleistungen dienen sollen. Schließlich wurden – ausgehend von den Fahraufgabenbeschreibungen für die Klasse B – Anpassungen für die weiteren Fahrerlaubnisklassen vorgenommen, sodass zu Beginn der empirischen Erprobungsphase ein Fahraufgabenkatalog für alle Fahrerlaubnisklassen vorlag.<sup>4</sup>

Im Folgenden soll auf die im Revisionsprojekt vorgenommenen Überarbeitungen der Anforderungs- und Bewertungsstandards im Vergleich zum Erstentwurf des Prüfungskonzepts (STURZBECHER et al., 2014) näher eingegangen werden. Zu den Aufgaben der o. g. „AG Fahraufgabenbeschreibung“ zählten die Überprüfung der fachlichen Richtigkeit bzw. Funktionalität jeder (Teil-)Fahraufgabe und ggf. ihre notwendige Anpassung hinsichtlich

- 1) der Definition und des Handlungsalgorithmus,
- 2) der grundsätzlichen Handlungsanforderungen zu jeder Beobachtungskategorie,
- 3) der Bezeichnung und der Inhalte der Fahraufgabe,

4) der zu unterscheidenden Situationsunterklassen sowie

5) der ereignisbezogene Bewertungskriterien.

Zu 1): Die Definitionen und Handlungsalgorithmen beschreiben, was der Bewerber grundlegend im Sinne von Mindeststandards bei der Erfüllung der jeweiligen Fahraufgabe zu tun hat; die Fahraufgabe wird also anhand ihrer typischen Situationsmerkmale sowie der typischen Handlungsziele und Handlungsstrategien der Kraftfahrzeugführer charakterisiert. Im Revisionsprojekt wurden die Definitionen und Handlungsalgorithmen hinsichtlich ihrer Vollständigkeit, Notwendigkeit, Verständlichkeit und Struktur überprüft sowie falls nötig überarbeitet.

Zu 2): Die grundsätzlichen Handlungsanforderungen beschreiben, was vom Fahrerlaubnisbewerber bei der Durchführung der einzelnen Fahraufgaben hinsichtlich der fünf Beobachtungskategorien erwartet wird. Sie wurden (analog zu den Definitionen und Handlungsalgorithmen) ebenfalls inhaltlich überprüft und falls nötig revidiert.

Zu 3): Auch hinsichtlich der Bezeichnung und der inhaltlichen Strukturierung der Fahraufgaben wurden im Vergleich mit dem Erstentwurf des Prüfungskonzepts (STURZBECHER et al., 2014) Überarbeitungen vorgenommen (s. Tabelle 1). So wurde die ehemalige Teilfahraufgabe „Verbindungsstrecken“ aus der Fahraufgabe „Kurven/Verbindungsstrecken“ herausgelöst und als eigenständige Fahrauf-

<sup>4</sup> Der Fahraufgabenkatalog – beispielhaft für die Klasse B – ist als Anlage 1 zum Bericht beigefügt.

Fahraufgabenkatalog 2012 (Stand nach STURZBECHER et al., 2014)	Fahraufgabenkatalog 2014 (Stand im Revisionsprojekt)
1. Fahrstreifenwechsel/Ein- und Ausfädelungsstreifen	1. Ein- und Ausfädelungsstreifen, Fahrstreifenwechsel
2. Kurven/Verbindungsstrecken	2. Annähern an und Befahren von Kurven
3. Vorbeifahren/Überholen	3. Vorbeifahren an Hindernissen und Engstellen sowie Überholen
4. Kreuzungen/Einmündungen	4. Überqueren von Kreuzungen und Einmündungen sowie Rechts- oder Linksabbiegen an Kreuzungen und Einmündungen
5. Kreisverkehr	5. Befahren von Kreisverkehren
6. Bahnübergang/Straßenbahnen	6. Heranfahren an und Überqueren von Bahnübergängen, Annäherung an Straßenbahnen sowie Überholen und Überholtwerden von Straßen-
7. Haltestellen/Fußgänger	7. Annähern an und Passieren von Haltestellen für Busse und/oder Straßenbahnen, Fußgängerüberwege, Fußgängern und Radfahrern
8. Radfahrer	8. Geradeausfahren

**Tab. 1:** Vergleich der Bezeichnungen und Inhalte der Fahraufgaben im Erstentwurf des Prüfungskonzepts der optimierten Praktischen Fahrerlaubnisprüfung (2012, nach STURZBECHER et al., 2014) und im revidierten Konzept (2014)

gabe „Geradeausfahren“ festgelegt. Methodisch gesehen, stellt diese Fahraufgabe eine „Restkategorie“ dar, die eine Reihe prüfungsrelevanter – von den speziell definierten Fahraufgaben unabhängige – Bewerberleistungen abdeckt. Somit kann sie nicht als eine Teilfahraufgabe definiert werden, sondern stellt eine eigenständige Anforderungskategorie dar. Hinsichtlich der methodischen Funktion dieser Fahraufgabe ist der Ausdruck „Restkategorie“ zwar korrekt; er birgt aber das Risiko, die Fahraufgabe neben den anderen Fahraufgaben nicht gleichwertig erscheinen zu lassen. Dies wäre nicht gerechtfertigt, denn die Fahraufgabe „Geradeausfahren“ operationalisiert verkehrssicherheitsrelevantes Basisverhalten, welches unabhängig von den anderen Fahraufgaben bzw. – zeitlich und streckenbezogen – zwischen ihnen stattfindet. Die Fahraufgabe „Geradeausfahren“ findet sich auch in den Anforderungsstandards der Praktischen Fahrerlaubnisprüfungen anderer EU-Länder und hat sich dort offenbar bewährt.

Die zunächst eigenständige Fahraufgabe „Radfahrer“ wurde als Teilfahraufgabe 7.4 „Annähern an und Vorbeifahren/Überholen von Radfahrern“ in die Fahraufgabe 7 (s. u.) integriert. Ähnlich wie die Teilfahraufgaben 7.1 „Annähern und Vorbeifahren an Haltestellen mit Bussen/Straßenbahnen“, 7.2 „Annähern an und Überqueren von Fußgängerüberwegen“ und 7.3 „Annähern und Vorbeifahren an Fußgängern und anderen besonders schutzbedürftigen Verkehrsteilnehmern“ betrifft das hier beschriebene Verhalten Verkehrssituationen, in denen sich der Bewerber schwächeren Verkehrsteilnehmern nähert und diese ggf. passiert. Insofern schien eine eigenständige Fahraufgabe „Radfahrer“ inhaltlich nicht gerechtfertigt und ihre Integration in die Fahraufgabe 7 geboten.

Zu 4): Situationsunterklassen von Teilfahraufgaben sind nach STURZBECHER et al. (2014, S. 59) „häufig auftretende, noch relativ komplexe Untergruppen von Verkehrssituationen [...] [die] ein modifiziertes Verhalten vom Bewerber im Vergleich mit der Standardsituation erfordern“. Im Rahmen der Ersterprobung des e-Prüfprotokolls (FRIEDEL et al., 2012) wurde festgestellt, dass die Festlegung der Situationsunterklassen optimiert

werden kann. Gemäß obiger Definition kann eine Situationsunterklasse nur dann als sinnvoll erachtet werden, wenn sich bei ihrer Durchführung abweichende Handlungsanforderungen im Vergleich zur Standardsituation ergeben. Zudem müssen die Situationsunterklassen pro Fahraufgabe vollständig und disjunkt sein, d. h. alle denkbaren Verkehrssituationen, die zur Teilfahraufgabe gehören, müssen vorkommen und höchstens einer Situationsunterklasse zugeordnet werden können, sonst könnten zwei Fahrerlaubnisprüfer bei der Dokumentation derselben Situation zu verschiedenen Ergebnissen gelangen. Die Kontrolle dieser Kriterien führte dazu, dass Situationsunterklassen von der „AG Fahraufgabenbeschreibung“ gestrichen oder angepasst wurden.

Zu 5): Bei den ereignisbezogenen Bewertungen beurteilt der Fahrerlaubnisprüfer „die Realisierung der in den Beobachtungskategorien vorgegebenen Handlungsanforderungen bei der Bewältigung von einzelnen Fahraufgaben, indem er die wahrgenommenen relevanten Verhaltensmerkmale unmittelbar im Anschluss an ihre Beobachtung einer vierstufigen ordinalen Bewertungsskala zuordnet. Die einzelnen Stufen lauteten zunächst 1) Überdurchschnittliche Leistung, 2) Normale Leistung, 3) Einfacher Fehler und 4) Erheblicher Fehler“ (STURZBECHER et al., 2014, S. 82). Im Rahmen des Revisionsprojekts wurden diese ereignisbezogenen Bewertungskriterien auf ihre Relevanz für die Praktische Fahrerlaubnisprüfung sowie ihre Verwendbarkeit im methodischen Sinne – z. B. Disjunktheit von anderen Bewertungskriterien – hin untersucht und falls nötig überarbeitet, gestrichen oder ergänzt (s. u.). Darüber hinaus wurden im Revisionsprojekt die Benennung und der Zuschnitt der Skalenstufen teilweise revidiert bzw. präzisiert. Die Bezeichnungen „Überdurchschnittliche Leistung“ und „Normale Leistung“ blieben unverändert, wohingegen „Einfache Fehler“ in „Leichte Fehler“ umbenannt wurden und „Erhebliche Fehler“ nunmehr als „Schwere Fehler“ bezeichnet werden. Das Gefährdungspotenzial eines Fehlverhaltens stellt dabei die grundsätzliche Dimension für die Fehlerkategorisierung dar: Je nach Höhe des Gefährdungspotenzials wird ein Fehler als „Leicht“

oder „Schwer“ kategorisiert. Die „Erheblichen Fehlverhaltensweisen“ gemäß Prüfungsrichtlinie, die zu einem sofortigen Nichtbestehen der Praktischen Fahrerlaubnisprüfung führen (sog. „Todsünden“), werden in der Kategorie „Schwere Fehler“ neben anderen, aber nicht zum sofortigen Nichtbestehen führenden, „Schweren Fehlern“ aufgeführt. Die „Todsünden“ (z. B. „Grobe Missachtung der Vorfahrt- bzw. Vorrangregelung“) sind in der Prüfungsrichtlinie rechtlich festgelegt und werden im Fahraufgabenkatalog besonders gekennzeichnet.

Grundsätzlich gilt, dass unter „Normalen Leistungen“ auch suboptimale Verhaltensweisen verstanden werden, die – entsprechend des Anfängerstadiums der Bewerber – durch mangelnde Routine bedingt sind und kein relevantes Gefährdungspotenzial besitzen. Als Beispiel wäre hier ein etwas zögerlicher Fahrstreifenwechsel zu nennen.

Unter „Überdurchschnittlichen Leistungen“ werden bestimmte Leistungen in einem Kompetenzbereich bzw. einer Beobachtungskategorie verstanden, die das fahranfängertypische Leistungsvermögen deutlich übersteigen. Solche Leistungen wurden im vorliegenden Fahraufgabenkatalog des Revisionsprojekts mit Bezug zu den Beobachtungskategorien beispielhaft aufgeführt. Da „Überdurchschnittliche Leistungen“ sich meist in außergewöhnlichen Verkehrssituationen ereignen und daher in Prüfungen eher selten vorkommen, wurde auf eine vollständige Auflistung verzichtet.

Neben den oben bereits erwähnten „Todsünden“ wird in der Prüfungsrichtlinie eine Reihe von Fehlern genannt, die bei Wiederholung oder Häufung zum Nichtbestehen der Prüfung führen können. Beispiele hierfür sind „Unterlassene Bremsbereitschaft“ und „Fehlerhaftes Abstandhalten“. Solche ereignisbezogenen Bewertungskriterien (Fehler und erhebliche Fehlverhaltensweisen) wurden bei der Überarbeitung des Fahraufgabenkatalogs durch die Fachexperten der „AG Fahraufgabenbeschreibung“ für alle Fahraufgaben beschrieben. Ziel war eine möglichst vollständige Auflistung der fahranfängertypischen, häufig vorkommenden und sicherheitsrelevanten „Leichten Fehler“ und „Schweren Fehler“.

Dabei ist festzuhalten, dass konkrete Gefährdungen oder Schädigungen nicht für die grundsätzlichen Fehlerdefinitionen herangezogen wurden: Stellt der Fahrerlaubnisprüfer eine konkrete Gefährdung oder Schädigung fest, wird die Prüfung sofort beendet. Das Vorliegen einer konkreten Gefährdung oder Schädigung stellt somit nicht nur ein Bewertungskriterium dar, sondern zugleich ein Entscheidungskriterium (sog. „KO-Kriterium“).

Unterschiedliche Fahrerlaubnisprüfer, welche dieselben Prüfungsleistungen beobachten, sollten diese im Sinne einer hohen Verfahrensgüte und somit Prüfungsqualität auch übereinstimmend bewerten und dokumentieren („Objektivität der Prüfung“). Daher wurden die festgelegten ereignisbezogenen Bewertungskriterien einerseits möglichst disjunkt formuliert und andererseits mit Indikatoren versehen. Disjunkte Formulierungen ermöglichen, dass eine Prüfungsleistung eindeutig einem ereignisbezogenen Bewertungskriterium zugeordnet werden kann; mögliche Überschneidungen wurden daher von den Fachexperten diskutiert und falls nötig korrigiert. Durch die Vorgabe einer überschaubaren Zahl wichtiger Verhaltensindikatoren (z. B. anhand von Beschreibungen und Beispielen) wird bei Beobachtungsverfahren die Belastung der kognitiven Ressourcen des Beobachters reduziert, da er seine Aufmerksamkeit auf vertraute anschauliche Hinweisreize fokussieren und sich auf die Kriterienbewertung konzentrieren kann (SWELLER, 2006). Außerdem engen Indikatoren den Interpretationsspielraum bei der Verhaltensdokumentation und -bewertung ein, was zur gewünschten Objektivierung der ereignisbezogenen Bewertungskriterien beiträgt. Daher wurden prototypische Indikatoren im Sinne von Ankerbeispielen zur Erhöhung der Verfahrensgüte der optimierten Praktischen Fahrerlaubnisprüfung in den Fahraufgabenkatalog aufgenommen. Sofern sich die Festlegung von Indikatoren als schwierig erwies, wurden inhaltlich einschlägige Gesetzestexte, Gerichtsurteile und Tatbestände des polizeilichen Bußgeldkataloges recherchiert und die dort gefundenen Richtwerte bei der Erstellung der Indikatoren berücksichtigt. Beispielsweise hat das OLG Hamm entschieden (Az. 1 RBs 78/13), dass eine Ordnungswidrigkeit begehrt, wer länger

als drei Sekunden dicht auffährt. Die Fachexperten haben daher drei Sekunden als Schwellwert zwischen einer zeitnah korrigierten und einer nicht zeitnah korrigierten Abstandsunterschreitung festgelegt, natürlich nicht ohne zuvor die Sinnhaftigkeit dieser Regelung für den Kontext der Praktischen Fahrerlaubnisprüfung zu reflektieren. Hinsichtlich der Indikatoren sei abschließend noch angemerkt, dass diese den Fahrerlaubnisprüfern lediglich als Orientierungshilfen bei der Prüfungsbewertung dienen; sie entfalten keine Verbindlichkeit, wenn triftige Gründe wie eine fehlende Sinnhaftigkeit in einem speziellen Situationskontext dagegen sprechen. Ihre methodische Funktion als Orientierungshilfe können Indikatoren nur erfüllen, wenn ihre Verwendung von den Fahrerlaubnisprüfern vorher ausreichend trainiert wurde und routiniert erfolgt.

Zuweilen erscheint eine eindeutige Zuordnung beobachteten Verhaltens zu den Beobachtungskategorien nicht möglich. Dies betrifft insbesondere Überschneidungen zwischen der „Verkehrsbeobachtung“ und den anderen Beobachtungskategorien. Ursächlich hierfür ist, dass Fehler bei der „Verkehrsbeobachtung“ häufig nicht direkt beobachtbar sind, da der Fahrerlaubnisprüfer das Blickverhalten des Bewerbers nur begrenzt nachverfolgen kann. Eine mangelnde Verkehrsbeobachtung ist daher oft erst an einem fehlerhaften Folgeverhalten zu erkennen. So kann eine überfahrene Fahrbahnbegrenzungslinie in einer mangelnden Verkehrsbeobachtung oder einer fehlerhaften Fahrzeugpositionierung begründet sein. Würde man aber nicht sicher zu beurteilende Beobachtungsdefizite lediglich als fehlerhaftes Folgeverhalten registrieren, würde dies bei der Auswertung der Prüfungsleistungen zu einer systematischen Unterschätzung von Beobachtungsmängeln führen. Die Fachexperten der „AG Fahraufgabenbeschreibung“ haben sich daher auf folgendes Vorgehen verständigt: Schätzt der Fahrerlaubnisprüfer mit überwiegender Wahrscheinlichkeit ein, dass einem Fehlverhalten eine vorhergehende mangelnde Verkehrsbeobachtung zugrunde liegt, dann ist der Fehler dieser Kategorie zuzuordnen. Finden sich dagegen überwiegend Indizien, dass ein Fehlverhalten nicht auf mangelnder Verkehrsbeobach-

tung, sondern auf Einschätzungs- oder Bedienungsdefiziten beruht oder aus Rücksichtslosigkeit resultiert, dann sind die entsprechenden anderen Beobachtungskategorien (meist Fahrzeugpositionierung oder Geschwindigkeitsanpassung) auszuwählen. Um die Dokumentation solcher Fehlverhaltenssequenzen für den Fahrerlaubnisprüfer im e-Prüfprotokoll zu erleichtern, wird beim Folgefehler ein Auge-Symbol dokumentiert: Die Aktivierung dieses Symbols bedeutet, dass der Prüfer den Fehler auf eine mangelnde Verkehrsbeobachtung zurückführt. Die Handhabung dieses Symbols wird mit einer speziellen Instruktion geregelt, deren Vermittlung an die Fahrerlaubnisprüfer Bestandteil der Beobachterschulung ist.

Die beschriebenen Überarbeitungen der Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung dienen den eingangs erwähnten Zielen des Revisionsprojekts: Erhöhung der Verkehrssicherheit, Optimierung der testpsychologischen Verfahrensgüte sowie Schaffung der Voraussetzungen für eine detaillierte Leistungsrückmeldung und Prüfungsevaluation. Die überarbeiteten Anforderungs- und Bewertungsstandards wurden im Fahraufgabenkatalog festgehalten, der als inhaltliche Grundlage für das im Kapitel 2.2 vorgestellte e-Prüfprotokoll dient.

## 2.2 Dokumentationsstandards und e-Prüfprotokoll

Bereits zu Beginn der Weiterentwicklung der Praktischen Fahrerlaubnisprüfung wurde eine elektronische Dokumentation der Prüfungsleistungen vorgeschlagen: Die definierten Fahraufgaben, Beobachtungskategorien und Bewertungskriterien sollten in ein Prüfprotokoll integriert werden, das in elektronischer Form über ein mobiles Endgerät (z. B. Tablet-PC) dem Fahrerlaubnisprüfer bereitgestellt werden kann (STURZBECHER et al., 2010). Auf dieser Grundlage wurden nachfolgend Dokumentationsstandards und Möglichkeiten für die Ausgestaltung eines e-Prüfprotokolls erarbeitet, mit dem die folgenden Nutzenerwartungen verbunden sind (STURZBECHER et al., 2014):

- 1) eine Unterstützung des Fahrerlaubnisprüfers bei der Professionalisierung seiner Prüfungsdurchführung (z. B. durch die Verdeutlichung der Anforderungs- und Bewertungsstandards),



- 2) eine Unterstützung des Fahrerlaubnisbewerbers durch eine differenzierte Rückmeldung des Fahrerlaubnisprüfers zu den Prüfungsleistungen für den weiteren Erwerb von Fahrkompetenz,
- 3) eine medienbruchfreie Prüfungsverwaltung (einschließlich der Archivierung der Prüfungsdaten) und
- 4) eine Optimierung der Qualitätssicherung nicht zuletzt durch die Verbesserung der Evaluationsmöglichkeiten.

Die Arbeiten zur Umsetzung der konzeptionellen Vorstellungen in ein funktionstüchtiges Dokumentationsinstrument erfolgten im Rahmen einer Machbarkeitsstudie im Zeitraum von 2011 bis 2012 (FRIEDEL et al., 2012). Den Arbeitsauftrag stellte eine Sondierung der Nutzererwartungen der Fahrerlaubnisprüfer der technischen Prüfstellen dar. Weiterhin wurden die Erwartungen der Fahrlehrer und der Fahrerlaubnisbewerber hinsichtlich eines optimierten Rückmeldesystems für die Prüfungsleistungen eruiert. Auf dieser Grundlage wurden dann verschiedene Gestaltungsvarianten für das e-Prüfprotokoll mit unterschiedlichen Bedienfunktionalitäten erstellt. Diese Varianten wurden mit allen am Fahrerlaubnisprüfungssystem (Praxis) Beteiligten<sup>5</sup> diskutiert. Im Ergebnis dieses Diskussions- und Entwicklungsprozesses wurde von der TÜV|DEKRA arge tp 21 der erste Prototyp des e-Prüfprotokolls für die Fahrerlaubnisklasse B programmiert.

Nach den Programmierarbeiten begann die Ersterprobung des e-Prüfprotokolls. Das Ziel dieser Erprobung bestand darin, mittels des o. g. Prototyps die grundsätzlichen Möglichkeiten zur elektronischen Erfassung der Prüfungsleistungen während der Prüfungsfahrt im Prüfungsfahrzeug zu untersuchen. Der Erprobungsteil der Machbarkeitsstudie gliederte sich in vier Entwicklungszyklen, wobei jeder Zyklus aus der eigentlichen Erprobung sowie einer anschließenden teilstandardisierten telefonischen Befragung der Fahrerlaubnisprüfer und der daraus resultierenden Überarbeitung des e-Prüfprotokolls (Update) bestand. Somit wurde

das e-Prüfprotokoll im Rahmen der Machbarkeitsstudie sukzessive empirisch weiterentwickelt. Die ersten beiden Entwicklungszyklen der Erprobung bestanden aus simulierten Prüfungen, in denen die Fahrerlaubnisprüfer Prüfungsfahrten beispielsweise gemeinsam mit einem Kollegen durchführten. In den nachfolgenden zwei Entwicklungszyklen wurde das e-Prüfprotokoll in realen Prüfungen eingesetzt. Insgesamt wurde das e-Prüfprotokoll im Rahmen der Machbarkeitsstudie in 350 simulierten und 600 realen Prüfungsfahrten verwendet. Eine detaillierte Beschreibung des eingesetzten e-Prüfprotokolls sowie die ausführliche Darstellung der Entwicklungsarbeiten und Ergebnisse der Machbarkeitsstudie finden sich bei FRIEDEL et al. (2012).

Insgesamt zeigen die Ergebnisse der Machbarkeitsstudie, dass eine elektronische Dokumentation der Prüfungsleistungen während der Prüfungsfahrt praktikabel ist. Gleichwohl fanden sich Hinweise auf Verbesserungspotenziale (z. B. hinsichtlich der Bedienfunktionalitäten in Verbindung mit dem Programmlayout), die bei der fortlaufenden Weiterentwicklung berücksichtigt wurden. Dabei lag der Fokus der Weiterentwicklung des e-Prüfprotokolls primär darauf, das Programm – entsprechend der Wünsche der Fahrerlaubnisprüfer – inhaltlich zu ergänzen sowie bezüglich des Layouts selbsterklärend und übersichtlicher zu gestalten.

Ausgehend von den Ergebnissen der Machbarkeitsstudie und der Überarbeitung des Fahraufgabenkatalogs (s. o.), wurde zu Beginn des Revisionsprojekts eine erste Version des e-Prüfprotokolls – nun für alle Fahrerlaubnisklassen – programmiert. Dazu wurden Bedienkonzepte für die klassenspezifischen Prüfungsteile „Verbinden und Trennen“ und „Abfahrtskontrolle und Handfertigkeiten“ sowie die klassenspezifischen Grundfahraufgaben ausgearbeitet und in das e-Prüfprotokoll eingefügt. Weiterhin wurden allgemeine Funktionalitäten ausgestaltet (z. B. die Möglichkeit handschriftlicher Notizen). Schließlich wurden auch Plausibilitätskontrollen zur Vollständigkeit der Prüfungsdokumentation und zur Prüfungsbewertung integriert sowie Zusatzinformationen (z. B. Auszüge aus rechtlichen Grundlagen) bereitgestellt.

Im Anschluss an die Programmierung der ersten Programmversion führten Fahrerlaubnisprüfer unter wissenschaftlicher Begleitung Prototypentests durch. Die Testergebnisse wurden umgehend zur Programmoptimierung genutzt. Auf diese Weise wurde das Programm in mehreren Entwicklungs-

<sup>5</sup> Der „Ad-hoc-Arbeitsgruppe Machbarkeitsstudie“ gehörten Vertreter des BMVBS, der Landesbehörden, der BASt, der TÜV|DEKRA arge tp 21, der technischen Prüfstellen, der Fahrlehrerschaft und unabhängiger wissenschaftlicher Institutionen an.



schleifen iterativ bis zur Einsatzreife für das Revisionsprojekt verbessert. Dabei weist das e-Prüfprotokoll prüfstellenübergreifend eine einheitliche Bedienstruktur (inkl. Layout) und eine einheitliche Programmarchitektur auf. Dadurch wurden eine einheitliche Handhabung sowie ein effizientes Support- und Updatemanagement gewährleistet.

Das e-Prüfprotokoll besteht aus einer für die Dokumentation und Bewertung der Prüfungsleistungen unabdingbaren zentralen Bedienoberfläche. Mittels dieser Bedienoberfläche dokumentiert der Fahrerlaubnisprüfer die Prüfungsleistungen der Fahrerlaubnisbewerber in den jeweiligen klassenspezifischen Prüfungsteilen. Beim Prüfungselement „Prüfungsfahrt“ erfolgt die Dokumentation der Prüfungsleistungen – wie von STURZBECHER et al. (2014) empfohlen – anhand einer kombinierten Darstellung von Fahraufgaben und Beobachtungskategorien in einer Matrix. Diese Matrix stellt im e-Prüfprotokoll auch ein wesentliches Element der Grundansicht dar und bietet so einen permanenten Überblick über die bis zum jeweiligen Prüfungszeitpunkt dokumentierten Prüfungsleistungen. Damit wird den Fahrerlaubnisprüfern die Planung des weiteren Prüfungsverlaufes erleichtert (s. Bild 1).

Die Grundansicht des e-Prüfprotokolls (s. Bild 1) zeigt – nach derzeitigem Stand – am oberen Rand

allgemeine Informationen zum Bewerber (Identitätsnummer, Name, Fahrerlaubnisklasse) sowie die Uhrzeit und die bisher vergangene Prüfungszeit. Darunter lassen sich die Leistungen der Bewerber bei den Prüfungselementen „Fahrtechnische Vorbereitung“ und „Fahrtechnischer Abschluss“ dokumentieren. Daneben werden – nur wenn für die jeweilige Fahrerlaubnisklasse zutreffend – Dokumentationsflächen für die Prüfungsteile „Verbinden und Trennen“ sowie „Abfahrtskontrolle/Handfertigkeiten“ angezeigt.

Im Zentrum der Grundansicht findet sich die bereits erwähnte Matrix: In der ersten Spalte sind die acht im Fahraufgabenkatalog definierten Fahraufgaben (einschließlich der Teilfahraufgaben) auf auswählbaren Schaltflächen angeordnet. Ebenfalls in dieser Spalte befindet sich darunter die Schaltfläche für die Auswahl der klassenspezifischen Grundfahraufgaben. In der obersten Zeile sind die Schaltflächen für die Auswahl der fünf Beobachtungskategorien bzw. Kompetenzbereiche angelegt. Die Felder der sich so ergebenden Matrix dienen zur Anzeige der dokumentierten ereignisbezogenen Bewertungen. Dabei ist es nicht notwendig, in jeder Zelle – d. h. für jede Kombination aus Fahraufgabe und Beobachtungskategorie – eine ereignisbezogene Bewertung vorzunehmen: Der Fahrerlaubnisprüfer vermerkt ausschließlich die überdurchschnittlichen Leistungen und die Fehler des Bewerber-

TUV/DEKRA		Version: 1.4 (16.01.2015)	12345	Bewerber Fahrerlaubnis		T	11:41 - 05:02	100 %		
Fahrtechnische Vorbereitung		Fahrtechnischer Abschluss		Verbinden und Trennen		Abfahrtskontrolle				
Beobachtungskategorie	Fahraufgabe	Verkehrsbeobachtung	Fahrzeugpositionierung	Geschwindigkeitsanpassung	Kommunikation	Fahrzeugbedingung/umweltbewusste Fahrweise	Gesamtbewertung Fahraufgaben	Strecke		
	Ein- und Ausfädelstreifen, Fahrstreifenwechsel		A 1				Ausreichend			
	Kurve						Bewertung			
	Vorbefahren, Überholen						Gut			
	Kreuzen, Einmündung	B 1					Bewertung			
	Kreisverkehr						Bewertung			
	Schienenverkehr						Bewertung			
	Haltestelle, Fußgänger, Radfahrer			C 1			Bewertung			
	Geradeausfahren					D 1	Bewertung			
	Grundfahraufgaben	1. Rückwärtsfahren geradeaus				Leichter Fehler		nicht vollumfänglich durchgeführt		
	Gesamtbewertung Beobachtungskategorien	Bewertung	Bewertung	Sehr gut	Bewertung	Bewertung	Prüfungsentscheidung			
<b>Schnellzugriff</b>										

Bild 1: Grundansicht des e-Prüfprotokolls

bers. Leistungen, welche den im Fahraufgabenkatalog beschriebenen Mindestanforderungen (Normale Leistung) entsprechen, werden nicht explizit dokumentiert.

Die dokumentierten ereignisbezogenen Bewertungen werden in den Zellen der Matrix als farblich hinterlegte Schaltflächen (Grün für überdurchschnittliche Leistungen, Gelb für leichte Fehler und Rot für schwere Fehler) angezeigt. Die Reihenfolge der Dokumentationen wird mittels Buchstaben festgehalten. Zudem wird angegeben, wie viele (ggf. wiederholte) Ereignisse für eine bestimmte Kombination aus Fahraufgabe und Kompetenzbereich dokumentiert wurden (numerische Angabe). Sofern eine bestimmte Kombination wiederholt geprüft wird, können die Bewertungen durch den Prüfer unterschiedlich ausfallen. Entsprechend können innerhalb derselben Zelle unterschiedliche Bewertungen mittels der Farben Grün, Gelb und Rot hinterlegt werden. Auf diese Weise erhält der Fahrerlaubnisprüfer kontinuierlich einen Überblick zu den dokumentierten Leistungen des Bewerbers bezüglich der Fahraufgaben und Kompetenzbereiche.

In der rechten Spalte und in der untersten Zeile der Matrix befinden sich die Schaltflächen für die zusammenfassende Bewertung der jeweiligen Fahraufgaben und Kompetenzbereiche. Beim Anklicken dieser Schaltflächen öffnet sich jeweils eine Maske, in welcher der Fahrerlaubnisprüfer die Bewerberleistungen bei der jeweiligen Fahraufgabe bzw. beim jeweiligen Kompetenzbereich auf einer vierstufigen Skala beurteilt. Die gemeinsame Schaltfläche der rechten Spalte und der untersten Zeile enthält schließlich das Ergebnis der Prüfungsentscheidung. Damit wird der Bewertungs- und Entscheidungsprozess des Fahrerlaubnisprüfers letztlich transparent und in der Regel nachvollziehbar: Ausgehend von den ereignisbezogenen Bewertungen (Zellen innerhalb der Matrix) beurteilt der Fahrerlaubnisprüfer zeilen- bzw. spaltenweise die Fahrkompetenz des Bewerbers im Hinblick auf die verschiedenen Fahraufgaben bzw. Kompetenzbereiche und trifft auf dieser Grundlage seine Prüfungsentscheidung.

Am rechten Rand der Grundansicht des e-Prüfprotokolls sind Schaltflächen für zusätzliche Funktionen bzw. Dokumentationsmöglichkeiten platziert. Hier werden Angaben zur Streckencharakteristik, zur Prüfungsdauer oder zu Gründen eines Abbruchs der Prüfung aufgrund unvorhersehbarer besonderer Umstände (z. B. Gesundheit des Be-

werbers oder nicht selbst verschuldeter Unfall) dokumentiert. Darüber hinaus können die Fahrerlaubnisprüfer individuelle Schnellzugriffe für einzelne Masken des e-Prüfprotokolls anlegen, was einer effizienten Nutzung des e-Prüfprotokolls zuträglich ist.

Das e-Prüfprotokoll beinhaltet auch zusätzliche, vor allem zur Bewerberverwaltung nutzbare Masken. Diese ermöglichen beispielsweise die Erstellung einer Bewerberliste (zu absolvierende Prüfungen pro Tag) oder die Dokumentation der formalen Voraussetzungen für das Absolvieren einer Praktischen Fahrerlaubnisprüfung (z. B. Identitätskontrolle, Kontrolle der Ausbildungsbescheinigung oder Prüfen behördlicher Auflagen). Die zusätzlichen administrativen Programmooptionen werden aufgrund unterschiedlicher Prozessabläufe, die in Verbindung mit den jeweiligen Verwaltungssystemen der technischen Prüfstellen bestehen (Schnittstellen), prüfstellenspezifisch genutzt.

Eingesetzt wird das e-Prüfprotokoll – den Empfehlungen aus der Machbarkeitsstudie folgend – auf mindestens 10 Zoll großen Tablet-Computern bzw. vergleichbaren Geräten, welche eine Touch-Bedienung ermöglichen. Die Bedienung erfolgt, je nach Gerätemöglichkeiten, mittels Finger oder Stift durch Anklicken der entsprechenden Schaltflächen. Das e-Prüfprotokoll bietet dem Fahrerlaubnisprüfer durch ein spezielles Farbkonzept (farbliche Kennzeichnung der zum aktuellen Zeitpunkt sinnvoll bedienbaren Schaltflächen) und relativ große Schaltflächen die Möglichkeit einer intuitiven und fehlerverzeihenden Bedienung.

Im Zuge der Felderprobung des Revisionsprojekts (s. Kapitel 4.5) wurde das e-Prüfprotokoll sukzessive weiterentwickelt. Ausgehend von den Rückmeldungen der an der Erprobung teilnehmenden Fahrerlaubnisprüfer wurden die kurzfristig änderbaren Optimierungswünsche – primär Aspekte des Programmlayouts und einzelner Funktionalitäten – im Rahmen von drei Programm-Updates umgesetzt.

Für den reibungslosen Einsatz des e-Prüfprotokolls war es nicht zuletzt erforderlich, die bestehende IT-Infrastruktur der technischen Prüfstellen in den Modellregionen aufzubauen bzw. zu erweitern. Dazu wurden Schnittstellen zwischen dem e-Prüfprotokoll und – sofern bereits existierend – den bestehenden Verwaltungssystemen (Bewerberdaten, Ergebnisdaten, Rekonstruktionsdaten, Evaluations-

daten) sowie zwischen den Verwaltungs- und IT-Systemen der technischen Prüfstellen und der TÜV|DEKRA arge tp 21 (zur Übermittlung anonymisierter Evaluationsdaten) geschaffen. Die Herausforderung bestand darin, das für alle technischen Prüfstellen einheitliche e-Prüfprotokoll in die verschiedenen hausinternen IT-Systeme zur Auftragsverwaltung einzubinden.

Unmittelbar mit der Programmierung der eigentlichen Anwendungssoftware des e-Prüfprotokolls verbunden war der Beginn der Entwicklungsarbeiten an einer – parallel zur zentralen Datenbank der theoretischen Fahrerlaubnisprüfung existierenden – zentralen Datenbank für die Praktische Fahrerlaubnisprüfung. In dieser Datenbank werden, wie bereits im Entwurf des „Handbuchs zum Fahrerlaubnisprüfungssystem (Praxis)“ vorgesehen, alle Prüfungselemente und Fahraufgaben einschließlich der mit ihnen verbundenen Bewertungskriterien hinterlegt. Durch die zentrale Speicherung ist sichergestellt, dass alle Prüfungscomputer in Deutschland stets mit dem gleichen Datenbestand arbeiten.

In den an der Erprobung der optimierten Praktischen Fahrerlaubnisprüfung teilnehmenden Modellregionen der technischen Prüfstellen wurden Teile der notwendigen IT-Infrastruktur der optimierten Prüfung lokal begrenzt implementiert. Die so gewonnenen Erfahrungen dienen nach Abschluss des Revisionsprojektes zur Weiterentwicklung der technischen Voraussetzungen einer bundesweiten Implementierung der optimierten Praktischen Fahrerlaubnisprüfung. Hierzu zählen die Vorbereitung des e-Prüfprotokolls, des Rückmeldesystems (ggf. Web-Portal), des Auswertesystems zur wissenschaftlichen Evaluation der Prüfung, der zentralen Datenbank nebst zugehörigem Autorensystem wie auch die Anpassung der bestehenden IT-Systeme der technischen Prüfstellen.

### 2.3 Evaluationskonzept

Das Ergebnis der Praktischen Fahrerlaubnisprüfung entscheidet über den Zugang zum Straßenverkehr und beeinflusst somit die allgemeine Handlungsfreiheit des Einzelnen. Handlungseinschränkende staatliche Maßnahmen sind nach ALBRECHT (2005) nur zulässig, wenn sie einem vernünftigen Gemeinwohlinteresse dienen und für die Wahrung dieses Interesses erforderlich und geeignet sind. Der Eignungsnachweis für eine Maß-

nahme – im Sinne eines Beweises ihrer gewünschten Wirksamkeit – ist im Rahmen einer wissenschaftlich fundierten Evaluation zu erbringen.

Bei der Evaluation von Maßnahmen – wie der Praktischen Fahrerlaubnisprüfung – kann nach WESTERMANN (2002) sowie WOTTAWA und THIERAU (2003) zwischen einer prozessualen, einer instrumentellen und einer ergebnisorientierten Evaluation unterschieden werden: Im Rahmen einer prozessualen Evaluation wird die Qualität bei der Durchführung einer Maßnahme untersucht, während die Ergebnisevaluation überprüft, ob die Maßnahme im Ergebnis den gewünschten Effekt erzielt. Die instrumentelle Evaluation hingegen hat die Güte der verwendeten methodischen Instrumente zum Gegenstand.

Weiterhin kann zwischen formativer und summativer Evaluation unterschieden werden (BORTZ & DÖRING, 2006). Mit einer formativen Evaluation wird der Prozess der Entwicklung eines Evaluationsgegenstandes untersucht. Sie hat begleitenden Charakter und erkundet den Entstehungsprozess der Maßnahme mit dem Ziel, diese zu optimieren. Die summative Evaluation hat hingegen zum Ziel, den Evaluationsgegenstand in einem bestimmten Zwischen- oder Endzustand zu untersuchen, um Aussagen über die erreichte Qualität zu tätigen, beispielsweise im Vorfeld der Einführung einer Maßnahme.

Die optimierte Praktische Fahrerlaubnisprüfung stellt ein psychometrisches Instrument zur Messung von Fahrkompetenz dar. Die Qualität psychometrischer Messverfahren kann (summativ) anhand der Gütekriterien der klassischen Testtheorie bestimmt werden (LIENERT & RAATZ, 1998). Zu diesen Gütekriterien zählen in erster Linie die Validität, die Reliabilität und die Objektivität des Verfahrens.

Die Notwendigkeit, die Praktische Fahrerlaubnisprüfung zu evaluieren, manifestiert sich in den Evaluationsanforderungen durch den Gesetzgeber. Sowohl die Richtlinie 2006/126/EG der Europäischen Union zum Führerschein als auch die Anforderungen der Bundesanstalt für Straßenwesen an das Qualitätsmanagement der technischen Prüfstellen (2003, 2005, 2007, 2009) beinhalten Bestimmungen zur Evaluation der Praktischen Fahrerlaubnisprüfung. Wie STURZBECHER et al. (2014) dargelegt haben, lassen sich sowohl aus der EU-Richtlinie als auch aus den Bestimmungen der Bundesanstalt für Straßenwesen verschiedene Elemente

ableiten, die eine Evaluation der optimierten Praktischen Fahrerlaubnisprüfung aus fachwissenschaftlicher Sicht umfassen sollte. Dazu zählen

- 1) die instrumentelle Evaluation,
- 2) die Auswertung von Prüfungsergebnissen und
- 3) Kundenbefragungen.

Um diesen Anforderungen gerecht zu werden, haben STURZBECHER et al. (2014) ein Konzept für die Evaluation der optimierten Praktischen Fahrerlaubnisprüfung skizziert, das diese Elemente enthält. Im Revisionsprojekt soll das methodische Instrumentarium zur Durchführung dieser Evaluationsvorhaben konkretisiert und erprobt werden.

Zu 1): Bei der instrumentellen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung lassen sich ein formativer und ein summativer Teil unterscheiden. Die formative Evaluation bezieht sich auf den Prozess der Instrumentenentwicklung – in diesem Fall auf das e-Prüfprotokoll – und wurde vor allem im Rahmen der Machbarkeitsstudie (FRIEDEL et al., 2012) durchgeführt. Die Ergebnisse dieser Studie belegen die praktische Umsetzbarkeit einer elektronischen Prüfungsdokumentation mittels e-Prüfprotokoll sowie die Tauglichkeit der verwendeten Hard- und Software. Optimierungsempfehlungen der an der Machbarkeitsstudie beteiligten Fahrerlaubnisprüfer wurden – nach sorgfältiger Abwägung durch Fachexperten – größtenteils umgesetzt (s. o.). Die summative Evaluation erfolgte anhand der Gütekriterien der klassischen Testtheorie, die im Zuge des Revisionsprojekts im Rahmen wissenschaftlicher Untersuchungen (s. u.) überprüft wurden. Eine instrumentelle Evaluation der Praktischen Fahrerlaubnisprüfung hat es bis zu diesem Zeitpunkt noch nicht in Deutschland gegeben.

Neben der Analyse der drei Hauptgütekriterien Validität, Reliabilität und Objektivität kann die sogenannte populationsspezifische Äquivalenz wichtige Hinweise auf die instrumentelle Güte einer Maßnahme liefern. Das Vorliegen populationsspezifischer Äquivalenz bedeutet, dass die Ergebnisse der Prüfung von inhaltlich irrelevanten Bewerbermerkmalen unabhängig sind (SCHWENKMEZGER & HANK, 1993). Eine fehlende

populationsspezifische Äquivalenz könnte beispielsweise ein Hinweis auf eine mangelhafte Objektivität bei der Prüfungsdurchführung und -bewertung sein und sollte genauer untersucht werden. Wie die oben genannten Hauptgütekriterien stellt auch die populationsspezifische Äquivalenz einen Untersuchungsgegenstand des vorliegenden Revisionsprojekts dar.

Zu 2): Während sich die im vorangegangenen Abschnitt erläuterte instrumentelle Evaluation auf die methodische Güte der optimierten Praktischen Fahrerlaubnisprüfung bezieht, richtet sich die Auswertung von Prüfungsergebnissen auf die Kontrolle der (alltäglichen) Durchführungsqualität der Prüfung. Sie trägt den Charakter einer prozessualen Evaluation und soll „das bundesweit einheitliche hohe Qualitätsniveau der Prüfungsgestaltung und Leistungsbewertung methodisch belastbar nachweisen“ (STURZBECHER et al., 2014, S. 161). Im Fokus der Prüfungsergebnisauswertung stehen erstens die Analyse der Prüfungsanforderungen, zweitens die Analyse der Prüfungsbewertungen und Prüfungsentscheidungen sowie drittens die Analyse ausgewählter Zusammenhänge zwischen diesen Ebenen. Der vierte Schritt bezieht sich schließlich auf die Analyse häufig auftretender Bewertungsmuster und Prüfertypen

Bislang erfolgte in Deutschland keine Auswertung der Prüfungsergebnisse der Praktischen Fahrerlaubnisprüfungen nach wissenschaftlichen Standards. Die in den technischen Prüfstellen durchgeführten Analysen beschränkten sich im Wesentlichen auf die Ermittlung von (regional- bzw. fahrerlaubnisprüferspezifischen) Bestehensquoten. Dies war nicht zuletzt dem Umstand geschuldet, dass die Dokumentation der Prüfungsanforderungen und -leistungen im bisherigen Prüfprotokoll – im Gegensatz zum e-Prüfprotokoll – nicht systematisch und differenziert erfolgte.

Zu 3): Bei der Praktischen Fahrerlaubnisprüfung handelt es sich um eine Dienstleistung, die durch die technischen Prüfstellen im staatlichen Auftrag erbracht wird. Für Anbieter von Dienstleistungen spielt die Wahrnehmung der Kunden eine bedeutende Rolle, da diese die Leistung bezahlen müssen. Rückmel-



dungen und Verbesserungsvorschläge der Kunden sollten daher – soweit möglich und sinnvoll – bei der Ausgestaltung bzw. Optimierung von Dienstleistungen berücksichtigt werden. Darüber hinaus bieten Kundenbefragungen den Vorteil, das typische Leistungsverhalten von Leistungserbringern im Alltag zu erfassen. Dies gilt vor allem dann, wenn sich die Leistungserbringer nicht bewusst sind, dass ihr Verhalten einer Beurteilung unterzogen wird. Kundenbefragungen können folglich für die prozessuale Evaluation von Maßnahmen von großer Wichtigkeit sein.

Als Kunden der technischen Prüfstellen können bei der Praktischen Fahrerlaubnisprüfung in erster Linie die Fahrerlaubnisbewerber und im erweiterten Sinne – über die Forderung einer kundenfreundlichen Gestaltung der administrativen Prozesse und der Prüfungsgestaltung – die Fahrlehrer betrachtet werden (STURZBECHER et al., 2014). Die Erfassung der Nutzerzufriedenheit vor allem der Fahrerlaubnisbewerber erscheint sehr wichtig, da die Prüfung staatlich vorgeschrieben ist und nur eine bewerberzentrierte Prüfungsdurchführung valide Prüfungsergebnisse ermöglicht. Außerdem ist eine Erhöhung des Bewerbernutzens, beispielsweise durch verbesserte Leistungsrückmeldungen und darauf aufbauende Lernhinweise, ein ausdrückliches Ziel der Optimierung der Praktischen Fahrerlaubnisprüfung. Insofern erlaubt eine Bewerberbefragung auch eine Beurteilung, ob die Reformziele mit der Optimierungsmaßnahme erreicht wurden und die optimierte Prüfung daher deutschlandweit implementiert werden soll.

Für die Erfassung der Kundenzufriedenheit – gewöhnlich das Kernelement von Kundenbefragungen – der Fahrerlaubnisbewerber und Fahrlehrer mit der optimierten Praktischen Fahrerlaubnisprüfung wurden im Revisionsprojekt zwei Befragungen vorbereitet und durchgeführt, deren Inhalte, methodische Ausgestaltung und Ergebnisse im Folgenden noch näher beschrieben werden.

Die beschriebenen Evaluationsmaßnahmen tragen im Wesentlichen instrumentellen und prozessualen Charakter. Allerdings ist auch die Ergebnisevaluation der Prüfung damit in einigen Teilen bereits ab-

gedeckt. So betrifft die Frage an die Fahrerlaubnisbewerber (Punkt 3), ob diese die optimierte mündliche und schriftliche Rückmeldung als hilfreich für den Aufbau ihrer Fahrkompetenz betrachten, die eingangs erwähnte Steuerungsfunktion der optimierten Praktischen Fahrerlaubnisprüfung, die im Ergebnis der Prüfungsoptimierung verbessert werden soll. Die instrumentelle Evaluation (Punkt 1) und die Auswertung der Prüfungsergebnisse mit der Analyse der populationsspezifischen Äquivalenz (Punkt 2) bieten hingegen Hinweise auf die Qualität der Prüfungsbewertung im Sinne der sog. „Prüfungsgerechtigkeit“. Inwiefern die Prüfungsoptimierung zu einer Erhöhung der Verkehrssicherheit führt, indem nur ausreichend befähigte Bewerber eine Fahrerlaubnis erhalten („Selektionsfunktion“), kann allerdings erst über einen längeren Zeitraum im Rahmen einer kontinuierlichen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung bestimmt werden, da sich derartige Effekte erst langfristig statistisch abbilden lassen.

## 2.4 Rückmeldesystem

Wie bereits beschrieben, wird mit der Praktischen Fahrerlaubnisprüfung die Fahrkompetenz von Fahrerlaubnisbewerbern erfasst und eingeschätzt. Für die Beurteilung des Fahrkompetenzniveaus werden die Fahrerlaubnisbewerber vor allem während der Prüfungsfahrt bei der Bewältigung verschiedener situationsspezifischer und situationsübergreifender Anforderungen (s. Kapitel 2.1) beobachtet. Die Prüfungsleistungen werden vom Fahrerlaubnisprüfer anhand vorgegebener Bewertungskriterien bewertet und mittels e-Prüfprotokoll dokumentiert (s. Kapitel 2.2). Für ein effektives Weiterlernen ist es von zentraler Bedeutung, dass die Bewerber nach der Praktischen Fahrerlaubnisprüfung eine qualifizierte Rückmeldung zu ihrem jeweiligen Leistungsstand erhalten. Nur wenn der Lernprozess hinsichtlich des Fahrkompetenzwerbs bis hin zur Überprüfung der Fahrkompetenz – und dazu gehört auch eine umfassende Rückmeldung – optimal verläuft, kann damit ein Beitrag zur Erhöhung der Verkehrssicherheit geleistet werden. Eine solche Rückmeldung (Feedback) sollte sowohl mündlich als auch schriftlich erfolgen.

Derzeit erhält nicht jeder Bewerber eine schriftliche Leistungsrückmeldung nach der Praktischen Fahrerlaubnisprüfung; nur bei nicht bestandener Prüfung ist dem Bewerber unter kurzer Benennung der wesentlichen Fehler ein schriftliches Prüfprotokoll



auszuhändigen. Die Fahrerlaubnis-Verordnung sieht hierfür eine Schwerpunktsetzung auf die während der Prüfungsfahrt gemachten Fehler vor: „Hat der Bewerber die Prüfung nicht bestanden, so hat ihm der Fahrerlaubnisprüfer bei Beendigung der Prüfung unter kurzer Benennung der wesentlichen Fehler hiervon zu unterrichten und ihm ein Prüfprotokoll auszuhändigen“ (Nr. 2.6 Anlage 7 FeV). Möglichkeiten der nachträglichen Auswertung – ggf. zusammen mit dem Fahrlehrer oder einem Elternteil – bietet das Prüfprotokoll aufgrund fehlender Erläuterungen allerdings kaum. Lediglich das unmittelbar nach der Praktischen Fahrerlaubnisprüfung stattfindende Auswertungsgespräch bietet bestenfalls einige Aufschlüsse über die erbrachten Prüfungsleistungen.

Im Zuge des Optimierungsprozesses zur Praktischen Fahrerlaubnisprüfung wurde vorgeschlagen, sowohl die Dokumentation der Prüfungsleistungen in der Praktischen Fahrerlaubnisprüfung (s. Kapitel 2.2) als auch die Rückmeldung an den Fahrerlaubnisbewerber im Anschluss an die Prüfungsfahrt weiterzuentwickeln, um die Steuerungsfunktion der Praktischen Fahrerlaubnisprüfung besser umsetzen zu können (STURZBECHER et al., 2010; STURZBECHER et al., 2014). Die Erarbeitung und Erprobung eines optimierten Rückmeldesystems ist daher Bestandteil des Revisionsprojekts. Nachfolgend wird dieses System näher beschrieben; zuvor werden jedoch die grundsätzlichen theoretischen Grundlagen zur Funktion und Gestaltung von Rückmeldungen in Lehr-Lernprozessen erläutert.

### **Pädagogische Funktion von Feedback in Lehr-Lernprozessen**

Nach WEIDENMANN (1998) werden Handlungsfähigkeiten schneller erlernt, wenn der Lernende über das Ergebnis seiner Handlungsausführung und über die ggf. aufgetretenen Fehler eine Rückmeldung erhält, um diesbezügliche Hinweise im Prozess des Weiterlernens berücksichtigen zu können. Förderorientierte Rückmeldungen zum Lernstand (Feedback) gelten daher als wichtige Bestandteile von Lehr-Lernprozessen und werden in pädagogischen und sozialen Prozessen gezielt als Intervention eingesetzt. Bleibt hingegen Feedback an Lernende aus, besteht keine Möglichkeit, den Lernprozess optimal zu steuern (BAMBERG, 2010). Derartiges Feedback dient der Korrektur von Fehlern und gibt Aufschluss darüber, ob bestimmte Lernziele erreicht wurden bzw. wie groß die Diskrepanz zwischen der erbrachten (Ist-Stand) und der ange-

strebten Leistung (Soll-Stand) ist (CARVER & SCHEIER, 2000); darin zeigt sich die diagnostische Funktion von Feedback (FISCHER & MANDL, 1988; KULHAVY, 1977).

Die mittels Feedback übermittelten Informationen über das Erreichen von Lernzielen sollen den Lernenden bei der Überprüfung seiner Selbsteinschätzung unterstützen sowie Fehlkonzepten und Kompetenzillusionen vorbeugen (MORY, 1996): Ohne Feedback ist es für den Lernenden kaum möglich, im Rahmen eines Lernprozesses etwas über das Niveau seiner Kompetenzen zu erfahren sowie im Hinblick auf bestimmte Fähigkeiten und Fertigkeiten seine Stärken und Schwächen kennenzulernen. Im Hinblick auf den Wissenserwerb kann Feedback darüber hinaus dazu beitragen, Lücken in mentalen Modellen des Lernenden zu schließen und fehlerhaftes Wissen zu korrigieren (VASILYEVA, PECHENIZKIY & BRA, 2008). Weiterhin wirkt sich Feedback auf affektiv-motivationale Lernvoraussetzungen aus und kann im günstigen Fall – beispielsweise über die Lernmotivation oder Selbstwirksamkeitserwartungen – einen lernförderlichen Einfluss entfalten.

Mit Feedback lassen sich nur dann Fördereffekte erzielen, wenn folgende Gestaltungsaspekte berücksichtigt werden: der Zeitpunkt der Feedbackerteilung, die Transparenz der Feedbackbotschaft, die formale Gestaltung und inhaltliche Darstellung des Feedbacks, die personalen Merkmale des Feedbacksenders und die Feedbackrezeption durch den Empfänger. Insgesamt gilt, dass die konkrete Funktion und der erwartete Nutzen von Feedback sehr stark vom konkreten Lernprozess und seinen Bedingungen abhängig und daher im Einzelfall zu klären sind (DECI, KOESTNER & RYAN, 2001; JACOBS, 2002; KRAUSE, STARK & MANDL, 2003). Hinsichtlich der Feedbackgestaltung sind vor allem die Komplexität bzw. der Elaborationsgrad eines Feedbacks entscheidend. Je nachdem, wie komplex ein Feedback gestaltet ist, finden sich unterschiedliche Feedbackformen. Grundsätzlich unterscheidet man einfache Feedbackformen (Richtig-Falsch-Aussagen, Angaben zur Anzahl gelöster Aufgaben, Benennung der Lösungen) und elaborierte Feedbackformen (auch „informatives Feedback“ genannt).

Betrachtet man Lernen als einen Informationsverarbeitungsprozess, gewinnt das elaborierte Feedback an Bedeutung. Diese Form von Feedback umfasst neben einer Richtig-Falsch-Aussage auch

weiterführende inhaltliche Informationen zu Fachbegriffen, spezifische Informationen zu den gemachten Fehlern und zur richtigen Bewältigung einer Aufgabe sowie weitergehende Informationen und Erläuterungen, wie die optimale Lösung aussieht und warum sie richtig ist (DITTON & MÜLLER, 2014). KULHAVY und STOCK (1989) unterscheiden folgende Formen der Elaboration: 1) aufgabenspezifische Elaboration, 2) instruktionsbasierte Elaboration und 3) extrainstruktionale Elaboration. Im ersten Fall wird im Feedback dargestellt, wie die Aufgabe zu bearbeiten gewesen wäre (meist Darstellung der korrekten Lösung). Im zweiten Fall wird die Aufgabeninstruktion näher erläutert, und im dritten Fall werden Informationen bereitgestellt, die in der Instruktion nicht enthalten waren.

In der Fachliteratur findet sich eine Reihe von Forschungsergebnissen, die auf eine höhere Lernwirksamkeit von elaboriertem Feedback gegenüber einfachem Feedback hindeutet (BANGERT-DROWNS, KULIK & MORGAN, 1991; COLLINS, CARNINE & GERSTEN, 1987; GILMAN, 1969; NARCISS, 2001; PRIDEMORE & KLEIN, 1991). Eine Studie von JACOBY, TROUTMAN, MAZURSKY und KUSS (1984) ergab, dass Feedback insbesondere dann ignoriert, also nicht weiter genutzt wird, wenn lediglich das Ergebnis an den Lernenden rückgemeldet wird. Die Überlegenheit von elaboriertem Feedback gilt insbesondere bei komplexen Aufgaben. Soll reines Faktenwissen gelernt werden, kann eine Korrektur unter Angabe der richtigen Antworten durchaus ausreichen (KULHAVY, WHITE, TOPP, CHAN & ADAMS, 1985). Im Falle der Praktischen Fahrerlaubnisprüfung, die eine sehr komplexe Aufgabe zur Überprüfung der Fahrkompetenz im Realverkehr darstellt, empfiehlt es sich daher, auf die Formen des elaborierten Feedbacks zurückzugreifen, um das weitere Lernverhalten substantiell zu fördern.

Insgesamt lassen sich für die lernfördernde Gestaltung von Feedback nach KRAUSE (2007, S. 54) folgende Merkmale benennen: „Feedback sollte

- 1) eher beschreiben als bewerten,
- 2) möglichst konkret formuliert werden anstatt allgemein,
- 3) eher einladen als zurechtweisen,
- 4) sich auf veränderbares Verhalten beziehen und nicht auf den Charakter einer Person,
- 5) eher erbeten sein als aufgezwungen,

- 6) eher sofort gegeben werden als verzögert und rekonstruierend,
- 7) eher klar formuliert werden als vage,
- 8) möglichst durch Dritte überprüfbar sein,
- 9) weiteres Lernen ermöglichen und
- 10) auch positive Aspekte enthalten; nach Möglichkeit sollte mit positiven Aspekten begonnen werden“.

Ein weiterer zentraler Faktor für die Wirkung von Feedback ist der Rezeptionsgrad durch den Empfänger: Ein Feedback muss vom Empfänger verarbeitet und aktiv zur Wissenskonstruktion genutzt werden, damit positive Wirkungen auf den Lernerfolg erzielt werden können (HANCOCK, THURMAN & HUBBARD, 1995; MORY, 1996). Jedes Feedback sollte mit einer Situationsbeschreibung eingeleitet werden, die möglichst konkrete Beobachtungen enthält und auf einen überschaubaren Bereich begrenzt ist (JUNG, 2003). Allgemeine bzw. zu wenig konkrete Behauptungen sollten vermieden werden, weil diese für den Empfänger wenig brauchbar sind (KNIEBEL, 2012). Damit Feedback einen positiven Effekt für das Weiterlernen nach der Praktischen Fahrerlaubnisprüfung haben kann, sollten bei den Erläuterungen der Fehler unbedingt Praxisbezüge – d. h. beispielsweise Bezüge zu einer bestimmten Verkehrssituation oder zu einem bestimmten Verhalten – hergestellt werden. Die Rückmeldungen zu erbrachten Leistungen sollten in einem ersten Schritt so erfolgen, dass der Lernende sofortige und genaue Informationen über den Erfolg oder Misserfolg seines Tuns erhält, also unmittelbar nach der Handlungsausführung bzw. nach dem Prüfungsende.

Neben einem sofortigen mündlichen Feedback stellt eine ergänzende schriftliche Rückmeldung ein wesentliches Element der Auswertung von Prüfungsleistungen dar. Die Bereitstellung einer lernfördernden schriftlichen Leistungsrückmeldung erscheint besonders wichtig, weil die mündlichen Erläuterungen eines Prüfers im Auswertungsgespräch aufgrund der affektiven Verarbeitung des Prüfungsergebnisses meist nur bruchstückhaft vom Lernenden aufgenommen werden. Dies gilt für eine nicht bestandene Prüfung und den damit verbundenen Stress genauso wie für eine bestandene Prüfung und die daraus resultierende Freude: Das Interesse an Lernhinweisen hält sich in dieser Situation erfahrungsgemäß in Grenzen. Für eine aktive

Verarbeitung und Nutzung der Lernhinweise ist es daher umso wichtiger, dass die Rückmeldungen auch schriftlich erfolgen.

Eine Metaanalyse von KLUGER und DeNISI (1996) zeigt, dass sich schriftliches computergestütztes Feedback stärker fördernd auf die Lernleistung auswirkt als mündliches Feedback. Die Autoren erklären diesen Befund damit, dass schriftliches Feedback die Aufmerksamkeit eher auf die Aufgabe lenkt, mündliches Feedback hingegen – durch die Präsenz des Feedbackgebers – eher Metaprozesse anregt und damit die Aufmerksamkeit von der Aufgabe ablenkt (KRAUSE, STARK & MANDL, 2003). Künftig sollte das optimierte Rückmeldesystem der Praktischen Fahrerlaubnisprüfung daher aus einem sofortigen mündlichen Feedback zu den erbrachten Prüfungsleistungen im Anschluss an die Prüfungsfahrt und einer ausführlichen schriftlichen Rückmeldung bestehen, um die Funktion von Feedback in Lehr-Lernprozessen auch beim Erwerb der Fahrkompetenz optimal auszunutzen.

### **Das Rückmeldesystem in der optimierten Praktischen Fahrerlaubnisprüfung**

Grundsätzlich sei angemerkt, dass die elektronische Dokumentation (s. Kapitel 2.2) sämtlicher wesentlicher – positiver und negativer – Prüfungsleistungen erstmalig eine differenzierte Leistungsrückmeldung an den Bewerber und somit eine bessere Qualität und Umsetzung der prüfungsdidaktischen Funktion der Praktischen Fahrerlaubnisprüfung ermöglicht (STURZBECHER et al., 2010; STURZBECHER et al., 2014). Die prüfungsdidaktische Funktion der Prüfung besteht darin, dem Fahrerlaubnisbewerber eine substanzielle und nachlesbare Rückmeldung zu den Stärken und Schwächen seiner Fahrkompetenz zu geben, um ihm ein zielgerichtetes Weiterlernen zu ermöglichen.

Die bereits beschriebene Optimierung der Prüfungsdokumentation (s. Kapitel 2.3) eröffnet auch eine Reihe von Chancen für eine verbesserte Leistungsrückmeldung an die Fahrerlaubnisbewerber. So erlaubt die Darstellung der Prüfungsleistungen in einer Matrix die Zuordnung der ereignisbezogenen Bewertungen (Fehler, überdurchschnittliche Leistungen) zu den konkreten Fahraufgaben und Fahrkompetenzbereichen. Durch die Abbildung von drei aufeinander aufbauenden Bewertungsebenen (ereignisbezogene Bewertung, kompetenzbezogene Bewertung, Prüfungsergebnis) besteht die Mög-

lichkeit, die Prüfungsleistungen der Bewerber nachvollziehbarer und differenzierter als bisher zu dokumentieren, auszuwerten und an den Bewerber zurückzumelden. Die neue kompetenzbezogene Bewertung erlaubt es nun, ein Nichtbestehen der Prüfung nicht allein mit einzelnen Fahrfehlern zu begründen, sondern mit einer gezeigten unzureichenden Fahrkompetenz in bestimmten Kompetenzbereichen. Diese Fahrkompetenzdefizite können vom Fahrerlaubnisprüfer im anschließenden Auswertungsgespräch mit Beispielen belegt und verallgemeinernd angesprochen werden, um Wege für einen effektiven und verkehrssicherheitswirksamen Fahrkompetenzaufbau aufzuzeigen.

Wie soll das Rückmeldesystem der optimierten Praktischen Fahrerlaubnisprüfung aussehen? Künftig sollen alle Fahrerlaubnisbewerber – sowohl bei bestandener als auch bei nicht bestandener Praktischer Fahrerlaubnisprüfung – eine qualifizierte Rückmeldung zu ihrem jeweiligen Leistungsstand im Sinne von Stärken und Schwächen erhalten. Eine solche differenzierte und förderorientierte Rückmeldung zu den Prüfungsleistungen soll den Bewerber beim weiteren Erwerb von Fahrkompetenz unterstützen. Diese Rückmeldung kann der Bewerber sicherheitsfördernd für das anschließende Weiterlernen beim selbstständigen Fahren bzw. lerngewinnbringend für eine ggf. notwendige Prüfungswiederholung verwenden. Ausgehend von den Rückmeldungen an den Fahrerlaubnisbewerber zu seinen Prüfungsleistungen können qualifizierte Leistungsrückmeldungen – je nach Differenzierungsgrad – auch den Fahrlehrern bzw. den Fahrschulen dienen, indem sie Rückschlüsse auf die Fahrausbildungsqualität zulassen und Optimierungspotenziale in diesem Bereich aufzeigen. Somit tragen eine verbesserte Dokumentationspraxis und ein optimiertes Prüfprotokoll für eine differenzierte Rückmeldung auch zur besseren Verzahnung von Ausbildungs- und Prüfungssystem bei.

Die Ergebnisse der bereits vorgestellten Machbarkeitsstudie (FRIEDEL et al., 2012) zeigen, dass die Mehrheit der Bewerber ein hohes Interesse an einer aussagekräftigen Leistungsbeurteilung und Lernhinweisen besitzt. Dies erscheint nachvollziehbar, weil die Bewerber zum Prüfungszeitpunkt noch als unerfahrene Fahrer anzusehen sind und sich in der besonders risikoreichen Übergangsphase zum selbstständigen Fahren befinden. Daher liegen professionelle (mündliche und schriftliche) ereignis- und kompetenzbezogene Leistungsrückmeldun-

gen, die eine realistische Selbsteinschätzung und eine Verbesserung der Fahrkompetenz fördern, sowohl im Bewerberinteresse als auch im Interesse der Öffentlichkeit. Die Leistungsrückmeldungen sollen förderorientiert gestaltet sein sowie

- 1) ein mündliches Auswertungsgespräch und
- 2) eine ergänzende schriftliche Leistungsrückmeldung mit Lernhinweisen umfassen.

Zu 1): In dem direkt an die Prüfungsfahrt anschließenden mündlichen Auswertungsgespräch wird dem Fahrerlaubnisbewerber vom Fahrerlaubnisprüfer zunächst sein Prüfungsergebnis mitgeteilt. Danach sollen die während der Prüfungsfahrt festgestellten Kompetenzdefizite, aber auch gute Leistungen angesprochen werden, um die Ausgewogenheit der Prüfungsentscheidung zu illustrieren und dem Bewerber Anreize für das Weiterlernen zu geben. Das Gespräch ist mit einem zusammenfassenden Gesamteindruck von der Fahrkompetenz des Bewerbers abzuschließen. Inwieweit die erwünschten Lern- bzw. Fördereffekte tatsächlich erzielt werden, hängt von den bereits dargestellten Einflussfaktoren auf die Feedbackrezeption und nicht zuletzt vom Auftreten des Feedbackgebers bzw. im vorliegenden Fall des Fahrerlaubnisprüfers ab. Bei der Praktischen Fahrerlaubnisprüfung dürften insbesondere die vom Fahrerlaubnisbewerber empfundene Fairness des Fahrerlaubnisprüfers bei der Vermittlung der Prüfungsanforderungen, seine (ihm vom Bewerber unterstellte) Fachkompetenz und seine (vom Bewerber erfahrene) prüfungsdidaktische Kompetenz ausschlag-

gebend dafür sein, ob der Fahrerlaubnisprüfer eine gewisse Autorität erlangt oder nicht. Ist diese Autorität bzw. Prüfkompetenz aus Sicht des Feedbackempfängers bzw. Bewerbers nicht gegeben, wird die Rückmeldung gegebenenfalls als unwichtig angesehen oder gar ignoriert. Einen umfassenden Überblick über einschlägige Forschungsergebnisse liefern DITTON und MÜLLER (2014).

Zu 2): Die ergänzende schriftliche Leistungsrückmeldung enthält einen Überblick über die erbrachten Prüfungsleistungen und Hinweise zum zielgerichteten Weiterlernen. Der jeweilige Umfang und Aufbau der Rückmeldung ist von den individuellen Leistungen des Fahrerlaubnisbewerbers abhängig: Mit der Anzahl und Verschiedenartigkeit der dokumentierten Fehler steigt auch die Anzahl der notwendigen Erläuterungen und der Lernhinweise zur korrekten Bewältigung der einzelnen Fahraufgaben.

Auf der ersten Seite der schriftlichen Leistungsrückmeldung (s. Bild 2) finden sich die personenbezogenen Daten des Bewerbers (Name, Geburtsdatum) und allgemeine Angaben zur Prüfung (Datum, Prüfungsbeginn und Prüfungsende, Fahrerlaubnisklasse der Prüfung, Name des Prüfers und Angaben zur Prüfungsstrecke). Neben dem Prüfungsergebnis werden die erbrachten Leistungen in den einzelnen Prüfungsteilen bzw. Prüfungselementen mit den entsprechenden Bewertungen dargestellt: Dazu gehören je nach Fahrerlaubnisklasse „Verbinden und Trennen“, „Abfahrtskontrolle/Handfertigkeit“

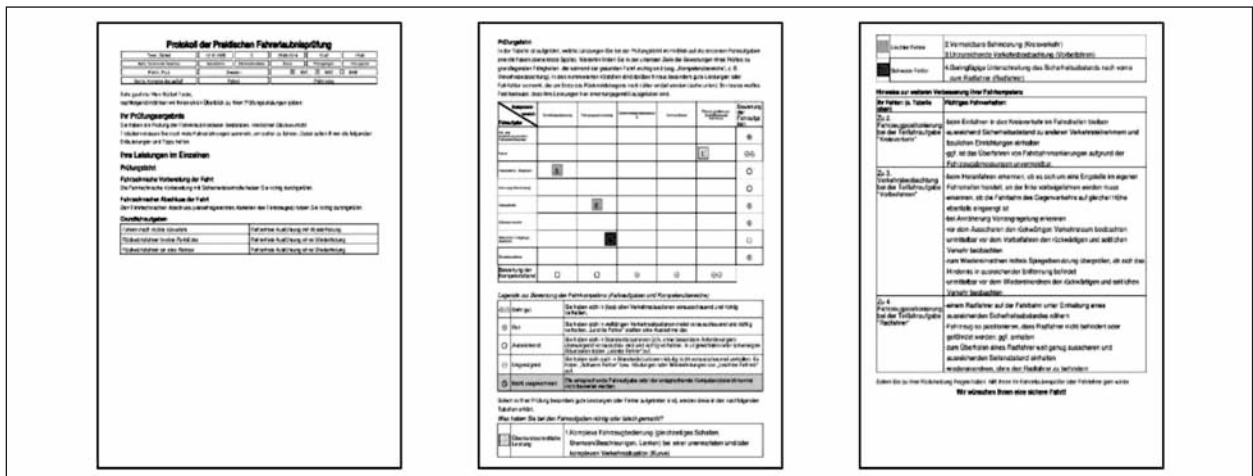


Bild 2: Beispiel einer schriftlichen Rückmeldung



ten“, „Fahrtechnische Vorbereitung, „Fahrtechnischer Abschluss“ und „Grundfahraufgaben“. Sofern die Prüfungsfahrt vorzeitig beendet oder abgebrochen wurde, ist dies unter Angabe einer Begründung als Prüfungsergebnis vermerkt.

Auf der zweiten Seite finden sich die Details zur Bewertung der Prüfungsfahrt. In der oberen Hälfte ist die Matrix des e-Prüfprotokolls abgebildet. Daraus lassen sich die Ergebnisse der ereignisbezogenen Bewertung (d. h. die Fehler und überdurchschnittlichen Leistungen bei den Fahraufgaben mit Bezug zu den einzelnen Fahrkompetenzbereichen) und die kompetenzbezogenen Gesamtbewertungen zu den Fahraufgaben und Kompetenzbereichen entnehmen. Die Matrix bietet somit einen Gesamtüberblick über die bei der Prüfungsfahrt erbrachten Fahrleistungen.

Die detaillierten Erläuterungen zu den Fehlern und besonderen Leistungen werden im unteren Teil der zweiten Seite überblicksartig in einer Tabelle dargestellt. Die entsprechenden Fehlerkategorien sind farblich bzw. mit Symbolen<sup>6</sup> gekennzeichnet. Die im Fahraufgabenkatalog festgelegten Definitionen der Fahraufgaben und Fahrkompetenzbereiche sollten zum besseren Verständnis für die Bewerber bereitgestellt werden (z. B. auf der Homepage der jeweiligen technischen Prüfstelle oder der TÜV | DEKRA | arge tp 21). Darüber hinaus sollten die Fahraufgaben und Kompetenzbereiche in der Fahrschulausbildung aufgegriffen und erläutert werden.

Die Hinweise zum weiteren Aufbau von Fahrkompetenz auf der dritten Seite werden entsprechend der bei der Prüfungsfahrt gemachten Fehler und der im Fahraufgabenkatalog (s. Kapitel 2.1) definierten Anforderungs- und Bewertungsstandards zusammengestellt. Dabei wird zu jedem Fehler die korrekte Handlungsabfolge zur Bewältigung der Fahraufgabe beschrieben. Abschließend finden sich die Legende zu den Symbolen der Kompetenzeinschätzung und die Defini-

tion der Bewertungsstufen. Ein beispielhafter Ausdruck der schriftlichen Rückmeldung ist dem Bericht als Anlage 2 beigelegt.

Das weiterentwickelte Rückmeldesystem wurde im Revisionsprojekt erprobt (s. Kapitel 4.5). Ergänzend zur Erprobung wurde eine Bewerberbefragung durchgeführt. Dabei wurden alle Fahrerlaubnisbewerber, die in den Modellregionen ihre Prüfung abgelegt haben, zur Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung und insbesondere zum Rückmeldesystem befragt (s. Kapitel 4.7). Die Befragungsergebnisse finden sich im Kapitel 5.5.

## 2.5 Weiterbildungskonzept

Eine systematische (Fahr-)Verhaltensbeobachtung, wie sie bei der Praktischen Fahrerlaubnisprüfung durchgeführt wird, erfolgt innerhalb eines theoretischen Bezugsrahmens und dient der Beantwortung bestimmter Fragen (MÜLLER, DAVID & STRAATMANN, 2011). „Das Mittel hierzu ist das Beobachtungsschema oder -instrument. Es dient gleichermaßen zur Lenkung wie zur Protokollierung. Es enthält die Beobachtungssitems (...), die Kategorien der Beobachtung (...) sowie generelle Angaben, z. B. der Dauer, des Ortes, der Zahl der Personen u. a., sofern diese nicht bereits zu den Beobachtungssitems gehören“ (FRIEDRICHS, 1973, S. 275). Im Falle der Praktischen Fahrerlaubnisprüfung handelt es sich beim Beobachtungsinstrument um das e-Prüfprotokoll, und der theoretische Bezugsrahmen ist durch die vorgegebenen Fahraufgaben (Beobachtungssitems) und Fahrkompetenzbereiche (Beobachtungskategorien) gegeben.

Die Qualität einer Verhaltensbeobachtung und somit auch die Güte der Ergebnisse sind sowohl von der Eignung des Beobachtungsinstruments als auch von den Kompetenzen der Beobachter und deren Schulung abhängig (FIEGUTH, 1977; BUNGARD et al., 1996). Bei der Beobachterschulung lernen die Beobachter das Beobachtungsinstrument (im vorliegenden Fall das e-Prüfprotokoll) in seinem theoretischen Bezugsrahmen kennen und anwenden; damit stellt die Beobachterschulung ein zentrales Element der Qualitätssicherung einer Verhaltensbeobachtung dar (MÜLLER, DAVID & STRAATMANN, 2011).

Das Ziel von Beobachterschulungen besteht darin, dass die Beobachter mit dem technischen Ablauf des Beobachtungsverfahrens und mit ihren eige-

<sup>6</sup> Je nachdem, bei welcher technischen Prüfstelle die Prüfung abgelegt wird, gibt es unterschiedliche Bereitstellungswege für die schriftliche Rückmeldung. Sie wird entweder in Schwarzweiß ausgedruckt und übergeben oder steht in Farbe als Download zur Verfügung.



nen Aufgaben – dem Protokollierungs- und Bewertungsprozedere – vertraut gemacht werden. In diesem Zusammenhang sollen sich die Teilnehmer von Beobachterschulungen bewusst mit den Zielen der Beobachtung sowie mit potenziellen kognitiven Wahrnehmungs- und Verarbeitungsverzerrungen auseinandersetzen (MÜLLER, DAVID & STRAATMANN, 2011; STURZBECHER et al., 2014). Darüber hinaus sollen sie das Beobachten und die Anwendung der Bewertungskriterien sowie den Umgang mit den bereitgestellten Testmaterialien praktisch einüben, damit in der Realsituation alle Beobachtungs- und Beurteilungsprozesse möglichst routiniert und reibungslos verlaufen.

Aufgrund der genannten Aspekte nimmt die Beobachterschulung (bzw. im vorliegenden Fall die Schulung der Fahrerlaubnisprüfer) auch im Prozess zur Optimierung der Praktischen Fahrerlaubnisprüfung einen besonderen Stellenwert ein. Die Erarbeitung und Erprobung eines entsprechenden Weiterbildungskonzepts ist Bestandteil des Revisionsprojekts. In diesem Weiterbildungskonzept werden die notwendigen Rahmenbedingungen für die Durchführung der Schulung skizziert sowie die Lehr-Lernziele und Lehr-Lerninhalte beschrieben.

### **Ziele und Inhalte der Weiterbildung der Fahrerlaubnisprüfer im Revisionsprojekt**

Das Ziel der Beobachterschulung bzw. Weiterbildung im Revisionsprojekt besteht darin, die teilnehmenden Fahrerlaubnisprüfer in den Modellregionen vor Beginn der Erprobung der optimierten Praktischen Fahrerlaubnisprüfung in die theoretischen und methodischen Grundlagen sowie in die Handhabung des e-Prüfprotokolls einzuweisen. Die Weiterbildung unterteilt sich in:

#### 1) Theoretisch-konzeptionelle Unterweisung

Im Rahmen der theoretisch-konzeptionellen Unterweisung erhalten die Fahrerlaubnisprüfer zuerst einen Überblick über die bisherigen Arbeiten zur Weiterentwicklung der Praktischen Fahrerlaubnisprüfung. Nachfolgend werden dann vertiefend die konzeptionellen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung und des e-Prüfprotokolls vorgestellt. Dabei werden zunächst die fachlichen und methodischen Qualitätsanforderungen an die Praktische Fahrerlaubnisprüfung angesprochen: ihr testpsychologisches Wesen als systematische Fahrverhaltensbeobachtung und Arbeitsprobe; ihre Durch-

führung als adaptive Prüfstrategie mit einem zirkulären Prozess des Planens, Beobachtens, Bewertens, Dokumentierens und Entscheidens sowie nicht zuletzt der notwendige empirische Nachweis ihrer methodischen Güte. Danach werden den Fahrerlaubnisprüfern anhand des Fahraufgabenkatalogs die Anforderungsstandards (im Sinne von situationsspezifischen Fahraufgaben und situationsübergreifenden Kompetenzbereichen) sowie die Bewertungsstandards (in Form von ereignisbezogenen und kompetenzbezogenen Bewertungskriterien) der optimierten Praktischen Fahrerlaubnisprüfung vermittelt. Weiterhin sollen die Weiterbildungsteilnehmer typische Beobachtungs- und Beurteilungsfehler kennen und kontrollieren lernen sowie eine selbstkritische Einstellung zur eigenen Urteilsbildung gewinnen. Schließlich soll der mögliche Nutzen der optimierten Praktischen Fahrerlaubnisprüfung und des e-Prüfprotokolls für die Verkehrssicherheit und Fahranfängervorbereitung im Allgemeinen sowie für einzelne Zielgruppen (Fahrerlaubnisbewerber, Fahrlehrer, Fahrerlaubnisprüfer) im Besonderen herausgearbeitet werden.

#### 2) Praxisorientiertes Beobachtertraining

Zu Beginn des Beobachtertrainings findet eine Einführung in die Dokumentationsstandards der optimierten Praktischen Fahrerlaubnisprüfung sowie in den grundsätzlichen Aufbau und die Bedienung des e-Prüfprotokolls statt. Darauf aufbauend setzen sich die Fahrerlaubnisprüfer anhand von Praxisbeispielen mit den Inhalten des Fahraufgabenkatalogs auseinander. Dabei lernen sie zunächst, bestimmte Verkehrssituationen gemäß Fahraufgabenkatalog richtig einzuordnen, zu beobachten und zu bewerten. Im Ergebnis dieses Übungsteils sollen die Weiterbildungsteilnehmer die Bewertungskriterien sicher und routiniert auf unterschiedliche Fahraufgaben anwenden können. Schließlich sollen sie lernen und üben, wie man anhand des Gesamtüberblicks über die dokumentierten Prüfungsleistungen (Matrix im e-Prüfprotokoll, s. o.) eine korrekte Prüfungsentscheidung treffen und im Anschluss an die Prüfungsfahrt ein förderorientiertes Auswertungsgespräch führen kann. In diesem Zusammenhang wird auch auf die optimierte schriftliche Rückmeldung eingegangen, welche künftig an alle Bewerber ausgehändigt werden soll (s. Kapitel 2.4): Die Fahrerlaubnis-

prüfer sollen im Rahmen der Weiterbildung den Aufbau und die Inhalte der schriftlichen Rückmeldung kennen und erläutern lernen.

Ausgehend von den im Revisionsprojekt für die Prüferweiterbildung erarbeiteten Lehr-Lernmedien (z. B. Lehrpräsentationen zu den theoretischen und methodischen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung, Fahraufgabenkatalog mit Lesehilfe, Benutzerhandbuch zum e-Prüfprotokoll, Handreichung zu den Durchführungs- und Bewertungsstandards, Simulationen der Dokumentationsabläufe mit dem e-Prüfprotokoll, Realfilme von videografierten Prüfungsfahrten) und den gewonnenen Weiterbildungserfahrungen, empfiehlt sich – nach Abschluss des Revisionsprojekts – die Erarbeitung eines Lehrmoduls für die reguläre Weiterbildung der Fahrerlaubnisprüfer. Ein solches professionelles Lehrmodul erscheint im Falle der bundesweiten Implementierung der optimierten Praktischen Fahrerlaubnisprüfung dringend notwendig, um eine erwachsenenpädagogisch anspruchsvolle und bundesweit einheitliche Schulung der Fahrerlaubnisprüfer aller technischen Prüfstellen sicherzustellen; eine einheitliche anspruchsvolle Weiterbildung ist wiederum als Voraussetzung für eine qualifizierte Prüfungsdurchführung anzusehen. Das Lehrmodul sollte – ergänzend zu den Lehr-Lernmedien aus dem Revisionsprojekt – filmische Darstellungen der Prüfungsabläufe, Animationen zum e-Prüfprotokoll, Übungs- und Kontrollaufgaben für den Prüfkompetenzerwerb, eine Arbeitshilfe für Fahrerlaubnisprüfer sowie nicht zuletzt ein Lehrkonzept für die Lehrbeauftragten der technischen Prüfstellen beinhalten.

### 3 Untersuchungsgegenstände

#### 3.1 Überblick

Das vorliegende Revisionsprojekt dient dem Zweck, wissenschaftlich fundierte Aussagen über die Qualität und ggf. die Optimierungspotenziale des Betriebskonzepts zur Durchführung und Evaluation der optimierten Praktischen Fahrerlaubnisprüfung in Deutschland zu gewinnen (s. Kapitel 1). Im Rahmen der hierfür notwendigen konzeptionellen Vorarbeiten wurden bereits die Elemente der angestrebten Prüfungsevaluation – instrumentelle Evaluation, Auswertung von Prüfungsergebnissen und Kundenbefragungen – erläutert und begründet (s. Kapitel 2.3). Diese Elemente stehen in einem

direkten Zusammenhang zu den Untersuchungsgegenständen im Revisionsprojekt. Zu den Untersuchungsgegenständen zählen die Objektivität, Reliabilität und Validität sowie die populationsspezifische Äquivalenz des Beobachtungsverfahrens. Die Ermittlung dieser Verfahrensparameter ist nötig, um Aussagen über die Verfahrensgüte der Prüfung treffen zu können. Die diesbezüglichen Untersuchungen lassen sich unter dem Begriff „Instrumentelle Evaluation“ zusammenfassen. Weiterhin stellt die alltägliche Durchführungsqualität der optimierten Praktischen Fahrerlaubnisprüfung einen Untersuchungsgegenstand dar, der durch die systematische Analyse von Prüfverhaltens- und Bewertungsmustern auf der Basis von Prüfungsergebnissen erhellt werden soll. Schließlich ist die Kundenzufriedenheit als weiterer Untersuchungsgegenstand anzusehen. Nachfolgend werden die genannten Untersuchungsgegenstände mit Bezug zum Revisionsprojekt ausführlich beschrieben. Im Kapitel 4 finden sich dann die zugehörigen Darstellungen der jeweiligen Untersuchungsanlage; das Kapitel 5 enthält die entsprechenden Untersuchungsergebnisse.

#### 3.2 Objektivität

Bei beobachtungs-basierten Beurteilungsverfahren wie der Praktischen Fahrerlaubnisprüfung können zahlreiche Beobachtungs- und Beurteilungsfehler auftreten, insbesondere wenn das Verfahren zur Erhöhung der inhaltlichen Aussagekraft nur übergreifend standardisiert ist (STURZBECHER et al., 2014). Aus diesem Grund wird im Rahmen des vorliegenden Projekts untersucht, ob der Prozess der Dokumentation und Bewertung von Prüfungsleistungen mittels e-Prüfprotokoll wissenschaftlich begründeten Qualitätskriterien genügt.

Die Praktische Fahrerlaubnisprüfung ist ein psychometrisches Messverfahren zur Erfassung von Fahrkompetenz. Die drei Hauptgütekriterien psychometrischer Messverfahren sind Objektivität, Reliabilität und Validität (LIENERT & RAATZ, 1998). Der Erfüllungsgrad dieser Kriterien ist ein Qualitätsindikator für die optimierte Praktische Fahrerlaubnisprüfung und somit ein wichtiger Anhaltspunkt bei der Entscheidung über die bundesweite Implementierung der Prüfung (s. Kapitel 6). Die Bedeutungen und die verschiedenen Formen dieser Gütekriterien wurden bereits bei STURZBECHER et al. (2014) ausführlich diskutiert. Im Folgenden beschränken sich die Erläuterungen daher auf jene Aspekte, wel-

che für die im Revisionsprojekt durchgeführten Untersuchungen von Relevanz sind. Dabei sei zunächst das Gütekriterium der Objektivität angesprochen.

Eine Messung gilt als objektiv, wenn das Ergebnis der Messung unabhängig von der messenden Person (also dem Beobachter) ist, d. h. zwei Beobachter sollten bei der Wahrnehmung und Beurteilung desselben Objekts idealer Weise zu gleichen Ergebnissen gelangen. Bei systematischen Verhaltensbeobachtungen ist die Ermittlung der Beobachterübereinstimmung<sup>7</sup> bei der zeitgleichen Beobachtung und Bewertung eines Objekts (bzw. im Fall der Praktischen Fahrerlaubnisprüfung des Fahrverhaltens des Bewerbers) durch mehrere Beobachter die vorherrschende Methode zur Kontrolle der Objektivität und der Reliabilität des Messverfahrens (SPRUNG & SPRUNG, 1984). Im Revisionsprojekt werden daher verschiedene Beobachterübereinstimmungsuntersuchungen durchgeführt (s. Kapitel 4), bei denen zwei oder mehrere Fahrerlaubnisprüfer dieselben Praktischen Fahrerlaubnisprüfungen beurteilen. Die Übereinstimmung zwischen den Prüfern wird dabei anhand spezieller Koeffizienten (z. B. Cohen's Kappa) bestimmt.

Entscheidend für die objektive Handhabung eines Messverfahrens sind detaillierte und verständliche Anweisungen an die Beobachter (STURZBECHER et al., 2014). Im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung kommt daher der Verständlichkeit des Dokumentations- und Bewertungsverfahrens sowie – damit verbunden – der (inhaltlichen) Ausgestaltung der Fahrerlaubnisprüferschulungen bzw. Fortbildungen (s. Kapitel 2.5) eine entscheidende Rolle zu. Die Fahrerlaubnisprüfer müssen systematisch und ausführlich in die inhaltliche und methodische Konzeption des Beobachtungsverfahrens (Anforderungs- und Bewertungsstandards) sowie in die Handhabung des e-Prüfprotokolls (Durchführungs- und Dokumentationsstandards) eingewiesen werden, um die einheitliche Anwendung der optimierten Praktischen Fahrerlaubnisprüfung zu fördern. Die Vereinheitlichung der Prüfungsdurchführung, -dokumentation und

-bewertung wird im vorliegenden Fall durch den Umstand erschwert, dass die optimierte Praktische Fahrerlaubnisprüfung in einer lebensweltlichen Domäne, dem Straßenverkehr, durchgeführt wird und sich die Prüfungsanforderungen daher nur begrenzt steuern lassen. Dies sollte bei der Interpretation der Untersuchungsergebnisse zur Objektivität der Prüfung berücksichtigt werden.

### 3.3 Reliabilität

Die Reliabilität bzw. Zuverlässigkeit charakterisiert den Grad der Messgenauigkeit eines Instruments. Sie ist umso größer, je kleiner der Anteil von Messfehlern an der Messung ist (BORTZ & DÖRING, 2006). Perfekte Reliabilität liegt vor, wenn mehrfache Messungen desselben Gegenstandes zum exakt gleichen Ergebnis führen. Ein messfehlerverursachender Störfaktor könnten beispielsweise Besonderheiten im Bewertungsverhalten der jeweiligen Prüfer – z. B. strenges oder mildes Bewertungsverhalten – sein. Aus diesem Beispiel wird deutlich, dass die Reliabilität eines beobachterbasierten Messverfahrens kaum von seiner Objektivität zu trennen ist. Daher werden die im Revisionsprojekt ermittelten Beobachterübereinstimmungsergebnisse als Indikator für die Erfüllung sowohl des Objektivitätskriteriums als auch des Reliabilitätskriteriums interpretiert; die beiden Gütekriterien werden bei der Auswertung auch gemeinsam thematisiert.

Grundsätzlich existieren verschiedene Methoden zur Erfassung der Reliabilität eines Messinstruments.<sup>8</sup> Diese Methoden unterscheiden sich je nach Art der Schätzung des Messfehlers. Im vorliegenden Fall einer systematischen Fahrverhaltensbeobachtung stellt – neben der Ermittlung der Beobachterübereinstimmung – die Berechnung der Bewertungsstabilität eine praktikable Methode zur Erfassung des Gütekriteriums dar. Die Bewertungsstabilität wird auch als Retest-Reliabilität, Intra-Rater-Reliabilität oder Test-Retest-Reliabilität bezeichnet. Bei dieser Methode wird ein Messverfahren zweimal durch dieselben Beobachter auf dieselben Objekte angewendet (SCHNELL, HILL & ESSER, 2005). Die Korrelation zwischen den beiden Messwertreihen entspricht dabei der Bewertungsstabilität. Im vorliegenden Revisionsprojekt soll die Bewertungsstabilität von den Fahrerlaubnisprüfern anhand der Beobachtung und Beurteilung von videografierten Prüfungsfahrten bestimmt werden (s. Kapitel 4.4). Die Bewertungsstabilität

<sup>7</sup> Die Beobachterübereinstimmung wird auch als Inter-Rater-Reliabilität bezeichnet.

<sup>8</sup> Hierzu zählen die Retest-Reliabilität, die Paralleltest-Reliabilität, die Testhalbierungs-Reliabilität und die Interne Konsistenz. Eine ausführliche Darstellung dieser Methoden findet sich bei BORTZ und DÖRING (2006).

würde dabei überschätzt werden, wenn sich die Fahrerlaubnisprüfer bei den Wiederholungsmessungen an die vorherigen Messungen erinnern könnten. Daher sind ausreichende Abstände zwischen den Messzeitpunkten einzuhalten.

### 3.4 Validität

Ein Messverfahren gilt als valide, wenn es wirklich das misst, was es zu messen beansprucht (LIENERT & RAATZ, 1998). Im Fall der Praktischen Fahrerlaubnisprüfung stellt sich daher hinsichtlich der Validität die Frage, ob sie auch tatsächlich die Fahrkompetenz der Bewerber erfasst. Die Validität ist das wichtigste und zugleich das am schwierigsten zu bestimmende Gütekriterium einer Messung. Man unterscheidet drei Formen der Validität: Inhalts-, Kriteriums- und Konstruktvalidität<sup>9</sup> (SCHNELL, HILL & ESSER, 2005).

Inhaltsvalidität bedeutet, dass ein Messverfahren nach Meinung von Fachexperten das zu erfassende Merkmal augenscheinlich abbildet. Dies kann hinsichtlich der optimierten Praktischen Fahrerlaubnisprüfung durch die konsequente Einbindung von Experten bei der Ausarbeitung der Anforderungs- und Bewertungsstandards (s. Kapitel 2.1) als gegeben angesehen werden. Da Inhaltsvalidität aber auf der subjektiven Einschätzung von Individuen beruht, gilt sie als ein nur beschränkt belastbares Kriterium. Inhaltsvalidität stellt daher ein notwendiges, aber kein hinreichendes Kriterium für den Validitätsnachweis der optimierten Praktischen Fahrerlaubnisprüfung dar.

Kriteriumsvalidität liegt vor, wenn das Messergebnis mit einem inhaltlich korrespondierenden Außenkriterium korreliert. Im Falle der Praktischen Fahrerlaubnisprüfung ist die Fähigkeit zur sicheren Teilnahme am Straßenverkehr maßgeblich für die Erteilung der Fahrerlaubnis. Das Prüfungsergebnis sollte daher beispielsweise mit der späteren Verkehrsbewährung oder den Unfallzahlen der Bewerber zusammenhängen. Bei der Auswahl der Außenkriterien ist allerdings auch darauf zu achten, dass diese im Rahmen der zeitlichen und finanziellen Möglichkeiten des jeweiligen Projektes erfass-

bar sind. So werden Verkehrsunfälle – neben der Fahrkompetenz des Kraftfahrzeugführers – auch durch viele andere Faktoren beeinflusst; sie treten darüber hinaus eher selten auf. Deshalb finden sich mögliche statistische Zusammenhänge zwischen den Prüfungsleistungen der Bewerber und ihrer Unfallhäufigkeit nur bei langfristig angelegten, also mehrere Jahre dauernden Untersuchungen. Das Unfallrisiko ist demzufolge für das zeitlich eng begrenzte Revisionsprojekt kein brauchbares Außenkriterium. Ein praktikableres Außenkriterium zur Validierung der Fahrkompetenzmessung stellt die Einschätzung der Fahrkompetenz der Bewerber durch Experten dar, die vom Fahrerlaubnisprüfer unabhängig sind und sich auf andere Weise ein fachkundiges Kompetenzurteil erworben haben. Stimmen diese Einschätzungen, die mit überschaubarem Aufwand erhoben werden können, in hohem Maße mit den Prüfungsbewertungen der Fahrerlaubnisprüfer überein, wäre dies ein Indiz für die Prüfungsvalidität. Daher wurden im Revisionsprojekt für eine ausreichend große Stichprobe von Bewerbern Fahrkompetenzeinschätzungen von den ausbildenden Fahrlehrern erhoben und ausgewertet (s. Kapitel 5.2).

Zur Kontrolle der Objektivität, Reliabilität und Validität der optimierten Praktischen Fahrerlaubnisprüfung wurden verschiedene Untersuchungen im realen Straßenverkehr und anhand videografierter Prüfungsfahrten vorgenommen. Dieser multimethodale Ansatz zur instrumentellen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung wurde gewählt, um die unvermeidlichen methodischen Schwächen einzelner Untersuchungsansätze durch die Stärken anderer Ansätze ausgleichen zu können. So haben z. B. Untersuchungen im realen Straßenverkehr den Vorteil einer besonders hohen ökologischen Validität, d. h. der Beobachtungs- und Bewertungsgegenstand – das Fahrverhalten des Bewerbers – wird im „natürlichen“ Umfeld erfasst, und das Prüfungsverhalten der Prüfer findet unter den alltäglichen Arbeitsgegebenheiten der „echten“ Prüfungen statt. Somit lassen sich auch die Ergebnisse von Untersuchungen in natürlicher Umgebung meist gut auf die Realität übertragen. „Echte“ Prüfungen lassen sich aber nicht wiederholen; und im Prüfungsfahrzeug finden höchstens zwei Prüfer Platz, deren Beobachtungsbedingungen sich auch noch systematisch unterscheiden. Videografierte Prüfungen bieten hingegen den Vorteil, dass sie von beliebig vielen Beobachtern beliebig oft beobachtet und beurteilt werden können, wodurch

<sup>9</sup> Die Konstruktvalidierung verknüpft eine pragmatisch orientierte Kriteriumsvalidierung mit wissenschaftstheoretischen Erwägungen und schließt alle anderen Validitätsarten ein. Sie trägt zur theoretischen Begründung und genauen Bestimmung des Testkonstrukts bei.



sich bei Güteuntersuchungen größere Fallzahlen und Messwiederholungen realisieren lassen: Je größer die Fallzahlen ausfallen, desto geringer ist der unerwünschte Einfluss von Zufallsfaktoren auf die Untersuchungsergebnisse. Messwiederholungen ermöglichen wiederum die Berechnung von intrapersonellen Stabilitätswerten, die neben der (interpersonellen) Beobachterübereinstimmung ein wichtiges Reliabilitätskriterium darstellen (s. o.).

### 3.5 Populationsspezifische Äquivalenz

Mobilität ist in der heutigen Gesellschaft ein hochgeschätztes Gut, und der Zugang zum motorisierten Straßenverkehr ist für viele Menschen sowohl in privater als auch in beruflicher Hinsicht von großer Bedeutung. Daher muss die Möglichkeit der motorisierten Verkehrsteilnahme jedem offen stehen, der die nötige Fahr- und Verkehrskompetenz im Rahmen der staatlich geregelten Fahrerlaubnisprüfung unter Beweis gestellt hat. Anders ausgedrückt: Es muss ausgeschlossen werden, dass Bevölkerungsgruppen aufgrund von Merkmalen, welche die Prüfungsleistungen nicht beeinflussen sollten, systematisch benachteiligt werden. Genügt eine Prüfung diesem Anspruch, spricht man von „populationsspezifischer Äquivalenz“.

Darüber hinaus kann die Analyse der populationsspezifischen Äquivalenz auch Hinweise auf andere Gütemerkmale der optimierten Praktischen Fahrerlaubnisprüfung geben, denn offensichtlich schränkt ein Einfluss inhaltlich irrelevanter Bewerbermerkmale auf die Prüfungsbewertung die Validität der Fahrkompetenzmessung ein. So erscheinen

Zusammenhänge zwischen dem Geschlecht der Bewerber und ihren Prüfungsergebnissen gesellschaftlich nicht erwünscht, da sie die Chancengleichheit beeinträchtigen. Fänden sich in den Güteuntersuchungen derartige Befunde, spräche das gegen die populationsspezifische Äquivalenz der optimierten Praktischen Fahrerlaubnisprüfung und würde die Frage aufwerfen, wie man Chancengleichheit herstellen könnte.

Für die Beantwortung dieser Frage wäre zunächst durch vertiefende Studien zu klären, ob es tatsächlich Geschlechtsunterschiede bei der Fahrkompetenz gibt und wodurch diese ggf. begründet sind oder ob gefundene geschlechtsspezifische Unterschiede bei den Prüfungsleistungen ggf. aus Beobachtungs- und Bewertungsfehlern der Prüfer resultieren, die Prüfung also nicht valide ist. Auf diese Weise können Untersuchungen zur populationsspezifischen Äquivalenz zugleich Validitätsmängel aufdecken oder auch Validitätsnachweise erbringen. So ist beispielsweise zu erwarten, dass die Gruppe der Fahrerlaubnisbewerber mit Fahrerlaubnissbesitz systematisch bessere Prüfungsleistungen erbringt als die Gruppe der Bewerber ohne Fahrerfahrung. Ein solcher Befund – also ein nachgewiesener Zusammenhang zwischen fahrkompetenzbegünstigenden Merkmalen der Bewerber und guten Prüfungsleistungen – spräche ggf. nicht gegen die populationsspezifischen Äquivalenz der Prüfung, sondern vielmehr für ihre Validität (sog. „konvergente Validierung“). Zeigen sich solche theoretisch begründbaren und daher zu erwartenden Zusammenhänge nicht, wäre auch dies als Ausgangspunkt für weitere Forschungsarbeiten zur methodischen Verbesserung der optimierten Praktischen Fahrerlaubnisprüfung anzusehen.

Bewerbermerkmal	Untersuchungsgegenstand
Fahrerfahrung	• (Konvergente) Validierung
Konsum von Computerspielen mit Bezug zum Straßenverkehr	• (Konvergente) Validierung
Bildungsniveau	• Populationsspezifische Äquivalenz • (mögliche Einschränkung der) Validität
Prüfungsangst	• Populationsspezifische Äquivalenz • (mögliche Einschränkung der) Validität
Migrationshintergrund	• Populationsspezifische Äquivalenz • (mögliche Einschränkung der) Validität
Geschlecht	• Populationsspezifische Äquivalenz • (mögliche Einschränkung der) Validität
Alter	• Populationsspezifische Äquivalenz • (mögliche Einschränkung der) Validität

Tab. 2: Unabhängige Variablen der Untersuchungen zur populationsspezifischen Äquivalenz



Ausgehend von den Empfehlungen von STURZBECHER et al. (2014) zur Evaluation der optimierten Praktischen Fahrerlaubnisprüfung und von den Vorgehensweisen bei der Evaluation der theoretischen Fahrerlaubnisprüfung (STURZBECHER, KASPAR, BÖNNINGER & RÜDEL, 2008), wurde eine Reihe von Bewerbermerkmalen identifiziert, deren populationsspezifische Äquivalenz untersucht werden soll. In der Tabelle 2 werden diese Merkmale aufgelistet und den Untersuchungsgegenständen zugeordnet.

### Fahrerfahrung

Vor der Praktischen Fahrerlaubnisprüfung können Bewerber durch eine intensive fahrpraktische Fahrerschulung oder den Vorbesitz anderer Fahrerlaubnisklassen bereits Fahrerfahrungen sammeln. Fahranfängertypische Defizite beispielsweise bei der Verkehrsbeobachtung können so vorab abgebaut werden. Ein entscheidender Unterschied zwischen erfahrenen Fahrern und Novizen sind die Verfügbarkeit und der Anwendungsbezug von Handlungsskripten. Nach BÖNNINGER und STURZBECHER (2005, S. 36 f.) werden „in diesen Skripten Informationen über beteiligte Personen und Akteure (z. B. Verkehrsteilnehmer), über Aktionen und ihre Abfolge (z. B. Verhaltenssequenzen der verschiedenen Verkehrsteilnehmer), Eingangs- und Endzustände von Geschehnissen (z. B. Ausgangspunkt und Folgen von Risikosituationen oder Unfällen) sowie „Slots“ (Leerstellen für situative Besonderheiten) bezüglich Klassen stereotypisierter sozialer Ereignisse gespeichert und miteinander verknüpft“. Die darauf beruhende Automatisierung der Fahrzeugführung setzt kognitive Ressourcen frei, sodass der Bewerber in Risikosituationen weniger unter Stress gerät und umsichtiger reagieren kann. Im Sinne einer konvergenten Validierung ist demnach ein positiver Zusammenhang zwischen der Fahrerfahrung zum Zeitpunkt der Prüfung und den Prüfungsergebnissen zu erwarten.

### Konsum von Computerspielen mit Bezug zum Straßenverkehr

Wie bereits erwähnt, stellt die Verfügbarkeit kognitiver Handlungsskripte einen entscheidenden Fahrkompetenzunterschied zwischen Fahranfängern und erfahrenen Fahrern dar. Weiterhin erscheint es plausibel, dass sich Fahranfänger solche Skripte auch beim Spielen verkehrsbezogener Computerspiele aneignen. Nach FRITZ (2011) lassen sich

Handlungsskripte aus virtuellen Welten auf reale Welten übertragen, vorausgesetzt die Situationen und Gegebenheiten der realen und der virtuellen Welt ähneln sich. Trotz offensichtlicher Unterschiede zwischen simulierten und realen Verkehrssystemen ähneln sich sowohl die grundlegenden Gestaltungsmerkmale der Systeme als auch die basalen psychischen und – mehr oder weniger je nach Art der Fahrsimulation – motorischen Anforderungen der Bewältigung von Verkehrssituationen in diesen Systemen (z. B. hinsichtlich der Aufmerksamkeitssteuerung, der Situationsbeurteilung und dem Erfordernis des schnellen Reagierens). VORDERER und KLIMT (2006) konnten im Rahmen einer Fahrsimulatorstudie zeigen, dass durch das Spielen von Autorennspielen kurzzeitig Informationsverarbeitungsprozesse für das Fahren am Simulator trainiert werden können. Es besteht daher Anlass zu der Annahme, dass der regelmäßige Konsum verkehrsbezogener Computerspiele die Ausbildung von Handlungsskripten, die für die Teilnahme am realen Straßenverkehr hilfreich sind, zumindest begünstigt.

### Bildungsniveau

Lesefähigkeiten und das kognitiv-mnestische Vermögen, explizites Wissen<sup>10</sup> zu speichern, stellen eine wesentliche Voraussetzung für gute schulische Leistungen wie auch für das Bestehen der theoretischen Fahrerlaubnisprüfung dar. Für die Praktische Fahrerlaubnisprüfung spielen diese Fähigkeiten eine geringere Rolle; hier stehen implizites Wissen und seine Verknüpfung mit explizitem Wissen zu psychomotorischem Prozesswissen<sup>11</sup> im Vordergrund. Die schulischen Lernstrategien

<sup>10</sup> Unter „explizitem Wissen“ versteht man im Rahmen des Langzeitwissens das Sachwissen („deklaratives Wissen“). Es umfasst semantisches bzw. abstraktes Wissen zu Begriffen, Objekten, Fakten, Sachverhalten oder Regeln sowie episodisches bzw. situationsgebundenes Wissen in Form von Situationsprototypen und Handlungsskripten als zentrale Elemente der Top-down-Handlungsplanung. Der Terminus „explizit“ bedeutet, dass dieses Wissen in aller Regel berichtet werden kann und somit auch durch verbale Instruktion vermittelbar bzw. in verbalisierter Form prüfbar ist (GRATTENTHALER, KRÜGER & SCHOCH, 2009).

<sup>11</sup> Prozesswissen integriert explizites und implizites Wissen (d. h. prozedurale Komponenten des Langzeitwissens, die in Form von motorischen Schemata und durch Übung bzw. einen Erfahrungsaufbau unter wechselnden Handlungsbedingungen erworben werden); sein Erwerb führt im Ergebnis zu (automatisierten) psychomotorischen Fertigkeiten (GRATTENTHALER et al., 2009).

und die eher erfahrungsbasierten Erwerbsstrategien fahrpraktischer Kompetenzen ähneln sich also kaum. Daher sollte kein Zusammenhang zwischen dem schulischen Bildungsniveau des Bewerbers und den Leistungsbewertungen durch den Prüfer bestehen. Findet sich dennoch ein Zusammenhang zwischen dem Bildungsniveau und den Prüfungsbewertungen, sollte genauer untersucht werden, ob eine Diskriminierung bestimmter Bildungsschichten vorliegt oder ob sich der Zusammenhang inhaltlich aufklären lässt. Einen Erklärungsansatz dafür bietet die Sprachkompetenz, die eine entscheidende Voraussetzung für den Bildungserfolg von Migranten darstellt (ESSER, 2006): Da Sprachkompetenz das Verstehen der Prüferanweisungen erleichtert, könnte sie sich auch positiv auf die Prüfungsleistungen in der Praktischen Fahrerlaubnisprüfung auswirken.

### **Prüfungsangst**

Es existieren nur wenige Studien, in denen konkrete Werte zur Prävalenz (d. h. Auftretenshäufigkeit) von Prüfungsangst unter Schülern und Studierenden genannt werden, wobei die einzelnen Angaben weit voneinander abweichen (5 % bis 41 %). Das hängt nach FEHM und FYDRICH (2011) mit dem Fehlen einer einheitlichen Definition und eines allgemein anerkannten Messinstrumentes für Prüfungsangst zusammen. Konsens besteht jedoch darüber, dass Prüfungsangst ein relativ häufig auftretendes Phänomen ist. Sie erreicht gewöhnlich in den Minuten vor und nach Prüfungsbeginn ihr Maximum und klingt danach ab. Viele Prüfer verfolgen daher bereits heute die Strategie, Fehler der Bewerber innerhalb der ersten Minuten der Prüfungsfahrt weniger streng zu beurteilen. Dieses Vorgehen ist aber weder institutionalisiert noch standardisiert. Die Evaluation der optimierten Praktischen Fahrerlaubnisprüfung bietet nun eine Gelegenheit zu überprüfen, wie weit diese Angst tatsächlich verbreitet ist und wie groß ggf. ihr Einfluss auf die Prüfungsergebnisse ausfällt. Ziel sollte es sein, einen etwaigen Einfluss der Prüfungsangst auf die Prüfungsergebnisse zu minimieren: Bei einer validen Prüfung sollte nicht im Vordergrund stehen, ob ein Bewerber unter Prüfungsbedingungen sicher die Anforderungen des motorisierten Straßenverkehrs bewältigen kann, sondern ob er die Verkehrsanforderungen unter Alltagsbedingungen meistert. Daher müssen Leistungsverzerrungen durch Prüfungsangst möglichst vermieden werden.

### **Migrationshintergrund**

Migranten werden gemäß empirisch gestützter Aussagen im 9. Bericht über die „Lage der Ausländerinnen und Ausländer in Deutschland“ (Bundesregierung, 2012) in zentralen Lebensbereichen wie Arbeit und Bildung diskriminiert. Im Rahmen des Revisionsprojekts soll daher überprüft werden, ob diese Diskriminierung auch im Zusammenhang mit der Praktischen Fahrerlaubnisprüfung auftritt. Findet sich ein Zusammenhang zwischen dem Vorliegen eines Migrationshintergrunds bei den Bewerbern und ihren Prüfungsbewertungen durch den Prüfer, wäre zu kontrollieren, ob dafür tatsächlich Diskriminierungsprozesse oder eher inhaltliche Hintergründe ursächlich sind. Ein möglicher inhaltlicher Erklärungsansatz wäre, dass Migranten durch mangelnde Sprachkompetenz häufiger Probleme haben, die Anweisungen der Fahrerlaubnisprüfer zu verstehen.

Es existieren verschiedene Definitionen darüber, wer als Migrant einzustufen ist. Für die Untersuchung von Diskriminierungsprozessen ist aber letztendlich nicht relevant, ob eine Person definitiv als Migrant zu betrachten ist, sondern ob sie als solcher wahrgenommen wird: Die Zuschreibung eines Migrationshintergrundes geschieht in der Regel stereotyp anhand von soziokognitiven Konstruktionsprozessen. Dementsprechend wird der Fahrerlaubnisprüfer um seine Einschätzung gebeten, ob er den jeweiligen Bewerber als Migranten einstuft (z. B. aufgrund seines Erscheinungsbildes und seiner Sprachfertigkeiten).

### **Geschlecht**

Im Straßenverkehr sind Stereotype im Hinblick auf geschlechtsspezifische Verhaltensweisen weit verbreitet. Deuten die vorliegenden Äquivalenzuntersuchungen im Rahmen der Erprobung der optimierten Praktischen Fahrerlaubnisprüfung entsprechend auf geschlechtsspezifische Prüfungsergebnisse hin, muss untersucht werden, ob die systematischen Benachteiligungen auf Geschlechtsrollenstereotypen basieren oder ob sie durch tatsächliche geschlechtsspezifische Eigenschaften vermittelt werden. Hier wäre beispielsweise vorstellbar, dass Männer sich systematisch früher und intensiver mit automobil- bzw. verkehrsbezogenen Themen beschäftigen, was eine schnellere Aneignung der Fahrkompetenz zur Folge haben könnte. So ist einerseits bekannt, dass junge Männer vor dem Fahrerlaubniswerb häufiger als Frauen illegal

Fahrerfahrungen sammeln (STURZBECHER, KLEEBERG-NIEPAGE & HOFFMANN, 2011). Andererseits bereiten sich Frauen in der Regel intensiver auf Prüfungen vor als Männer (CHETCUTI, 2009), was wiederum bessere Prüfungsleistungen des weiblichen Geschlechts erklären könnte.

### Alter

Zeigen sich im Rahmen der Datenauswertung starke signifikante Zusammenhänge zwischen dem Alter der Bewerber und ihren Prüfungsergebnissen, muss genauer untersucht werden, ob bestimmte Altersgruppen bei der Prüfungsdurchführung bzw. bei der Bewertung der Prüfungsleistungen benachteiligt werden oder ob sich die Zusammenhänge durch altersspezifische Eigenschaften begründen lassen. Beispielsweise erscheint es vorstellbar, dass eine größere Lebenserfahrung und Gelassenheit im Alter die Aneignung und Demonstration von Fahrkompetenz erleichtert, während der Rückgang körperlicher Fähigkeiten einen gegenteiligen Zusammenhang erklären könnte. Das Alter ist zudem eine wichtige Kontrollvariable für andere Zusammenhangsanalysen. So ist bekannt, dass ältere Fahrschüler im Durchschnitt mehr Übungsstunden bis zur Erlangung der Fahrerlaubnis benötigen als jüngere Fahrschüler. Ein Zusammenhang von Fahrerfahrung und Prüfungsergebnissen könnte daher über das Alter vermittelt sein.

### 3.6 Alltägliche Durchführungsqualität

Die in den vorangegangenen Kapiteln beschriebenen Untersuchungsgegenstände Objektivität, Reliabilität, Validität und populationsspezifische Äquivalenz sind Teil der instrumentellen Evaluation und beziehen sich auf die methodische Güte der optimierten Praktischen Fahrerlaubnisprüfung als Beobachtungsverfahren. In den nachfolgenden Kapiteln steht hingegen die prozessuale Evaluation im Mittelpunkt. Deren Ziel ist der Nachweis einer bundesweit hohen Qualität bei der alltäglichen Durchführung der optimierten Praktischen Fahrerlaubnisprüfung.

Eine mögliche Methode zur Analyse der alltäglichen Durchführungsqualität der optimierten Praktischen Fahrerlaubnisprüfung ist die Auswertung von Prüfungsergebnissen. Durch die systematische elektronische Dokumentation von Prüfungsleistungen mittels e-Prüfprotokoll wird gegenüber der bisherigen Papier-Bleistift-Variante eine wesentlich umfas-

sendere Datenbasis geschaffen, die im Rahmen einer kontinuierlichen Evaluation eine detaillierte Analyse der Prüfungsdurchführung und -bewertung ermöglicht. Die technischen Prüfstellen tragen damit einerseits den Forderungen der Richtlinie 2006/126/EG Rechnung, nach denen die Ergebnisse der abgenommenen Prüfungen regelmäßig überprüft werden sollen (Anhang 4, Nr. 4.1.2). Andererseits genügen sie damit den „Anforderungen an Träger von technischen Prüfstellen“ (BASt, 2009, Nr. 6.9).

Bei der Auswertung der Prüfungsergebnisse können in Anlehnung an STURZBECHER et al. (2014) drei Ebenen unterschieden werden:

- 1) Auf der ersten Ebene erfolgt eine deskriptive Auswertung der Prüfungsanforderungen und -bewertungen. Hierbei werden erstens Häufigkeiten bezüglich der Durchführung von Fahraufgaben analysiert. Dies ermöglicht Rückschlüsse darauf, inwiefern ein bundesweit einheitliches Anforderungsniveau realisiert wird.<sup>12</sup> Zweitens werden die Bestehensquoten im Hinblick auf ihre regionale Verteilung analysiert. Hierbei erscheint vor allem die Frage von Interesse, ob es systematische regionale Unterschiede gibt. Drittens werden die Häufigkeiten der kompetenzbezogenen Gesamtbewertungen der Fahraufgaben und Beobachtungskategorien ausgewertet, was Rückschlüsse auf besonders anspruchsvolle bzw. weniger anspruchsvolle Fahraufgaben und Kompetenzbereiche zulässt. Viertens schließlich werden die Häufigkeiten ereignisbezogener Bewertungen, also des Auftretens von Fehlern oder überdurchschnittlichen Leistungen, analysiert. Das Wissen darüber, welche Fahraufgaben bzw. Kompetenzbereiche den Fahrerlaubnisbewerbern besondere Schwierigkeiten bereiten und welche Fehler dabei besonders häufig begangen werden, kann als wichtige Grundlage für die Steuerung der Fahranfängervorbereitung zum gezielten Abbau sicherheitsrelevanter Fahrkompetenzdefizite dienen.

<sup>12</sup> Beispielsweise kann verglichen werden, ob bestimmte Fahraufgaben in Abhängigkeit vom Prüfort besonders häufig nicht vorkommen. Für detailliertere Prüfortanalysen (z. B. hinsichtlich der Häufigkeit von Fahraufgaben) müsste die Prüfungsdokumentation erweitert werden, da in der regulären Prüfung eine explizite Dokumentation „normaler“ (d. h. fahranfängertypischer) Leistungen bislang nicht vorgesehen ist.

- 2) Auf der zweiten Ebene werden die Zusammenhänge zwischen den Prüfungsanforderungen, den Prüfungsbewertungen, den Prüfungsrahmenbedingungen und den Prüfungsentscheidungen analysiert. Hierbei sind die Identifikation prognostischer Zusammenhänge und die Überprüfung von Plausibilitäten von besonderem Interesse. So werden die Zusammenhänge zwischen den kompetenzbezogenen Gesamtbewertungen der Fahraufgaben und der Beobachtungskategorien auf der einen sowie den Prüfungsentscheidungen auf der anderen Seite auf statistische Signifikanz hin überprüft und die Effektstärken in ihrer Bedeutsamkeit verglichen. Damit lässt sich beispielsweise ermitteln, bei welchen Fahraufgaben oder Beobachtungskategorien sich eine als „Ungenügend“ bewertete Leistung besonders negativ auf die Prüfungsentscheidung auswirkt bzw. eher noch kompensiert werden kann. Gleichzeitig zeigen unplausible Ergebnisse wie ein negativer Zusammenhang zwischen den Gesamtbewertungen eines Kompetenzbereiches und den Prüfungsentscheidungen Hinweise auf Fehler in der Prüfungsdurchführung bzw. validitätsmindernde Aspekte der optimierten Praktischen Fahrerlaubnisprüfung auf, die gegebenenfalls in der Folge genauer untersucht werden müssten. Auf der Ebene der ereignisbezogenen Bewertungskriterien wird beispielhaft untersucht, welche Fehler bzw. überdurchschnittlichen Leistungen sich besonders stark auf die zugehörigen Gesamtbewertungen der Fahraufgaben bzw. Beobachtungskategorien und auf die Prüfungsentscheidungen auswirken. Daraus ergeben sich detaillierte Hinweise auf typische Ereignisse, die für die Aneignung von Fahrkompetenz als kritisch anzusehen sind.
- (d. h. die Wachheit bzw. andauernde Aufmerksamkeit) im Tagesverlauf bei den meisten Menschen am Vormittag am höchsten ist und am Nachmittag vorübergehend stark nachlässt (HORNE & REYNER, 1995).
- 3) Auf der dritten Ebene werden Merkmalsmuster im Hinblick auf die dokumentierten Prüfungsleistungen und -bewertungen untersucht; basierend darauf wird in einem weiteren Schritt versucht, Prüfertypen zu identifizieren. Bei der Analyse der Merkmalsmuster und Prüfertypen liegt der Fokus auf der Anzahl dokumentierter Fehler und überdurchschnittlicher Leistungen, der Auswahl der Fahraufgaben (Prüfungsanforderungen) sowie auf den acht kompetenzbezogenen Gesamtbewertungen der Fahraufgaben, den fünf kompetenzbezogenen Gesamtbewertungen der Kompetenzbereiche und den Prüfungsentscheidungen. Mit Bezug zu den Kompetenzbereichen soll beispielsweise untersucht werden, wie stark die Bewertungen der Kompetenzbereiche korrelieren und ob sich dies in Bewertungsmustern widerspiegelt. So wäre beispielsweise anzunehmen, dass die Bewerberleistungen in den Fahrkompetenzbereichen „Fahrzeugpositionierung“ und „Geschwindigkeitsanpassung“ stärker miteinander korrelieren als die Leistungen in den Bereichen „Fahrzeugpositionierung“ und „Kommunikation“. Diese Vermutung ist darin begründet, dass sowohl die „Fahrzeugpositionierung“ als auch die „Geschwindigkeitsanpassung“ in erster Linie auf motorischen und koordinativen Fähigkeiten beruhen; die Bewerberleistungen im Bereich „Kommunikation“ hängen hingegen vor allem von der Qualität der Verkehrsbeobachtung und der soziokognitiven Perspektivenübernahmefähigkeiten ab.

Abschließend wird der Einfluss bestimmter Prüfungsrahmenbedingungen auf die Prüfungsbewertungen untersucht. So ist beispielsweise zu vermuten, dass Prüfungen bei Dunkelheit bzw. reduzierter Helligkeit aufgrund der eingeschränkten Sicht erhöhte Anforderungen an die Bewerber stellen, was mit durchschnittlich schlechteren kompetenzbezogenen Gesamtbewertungen der Fahraufgaben und Beobachtungskategorien sowie mit niedrigeren Bestehensquoten einhergehen könnte. Weiterhin könnten Prüfungen zwischen 9 und 12 Uhr mit besseren Ergebnissen beendet werden als Prüfungen zwischen 14 und 16 Uhr, da die Vigilanz

Die Analyse der Bewertungsmuster bei der kompetenzbezogenen Gesamtbewertung der Fahraufgaben erfolgt methodisch gesehen analog zu der Analyse der Muster bei der kompetenzbezogenen Gesamtbewertung der Kompetenzbereiche, da die zugrunde liegenden Variablen auf derselben vierstufigen Skala basieren. Zusammenhänge zwischen den Bewertungen sind beispielsweise bezüglich der Fahraufgaben „Kreuzungen, Einmündungen“ und „Kreisverkehr“ vorstellbar, da diese beiden Klassen prototypischer Verkehrssituationen ähnliche Handlungsanforderungen an den Bewerber stellen und somit auch ähnliche Fahrkompetenzen verlangen.



Ein geeignetes Verfahren, um das Vorhandensein von Prüfertypen zu untersuchen, stellt die Clusteranalyse dar. Mittels dieser Methode werden die Fahrerlaubnisprüfer in Abhängigkeit von ihren Merkmalsausprägungen bei ausgewählten Variablen in Gruppen unterteilt, die gruppenintern möglichst homogen und im Gruppenvergleich möglichst heterogen ausfallen (BACKHAUS, ERICHSON, PLINKE & WEIBER, 2006). So sollen mithilfe der Clusteranalyse – unter Einbeziehung der kompetenzorientierten Gesamtbewertungen und der Prüfungsentscheidungen – explorativ Prüfertypen identifiziert und beschrieben werden.

Grundsätzlich verlangt der Gesetzgeber, dass alle Fahrerlaubnisbewerber nach einheitlichen Anforderungs- und Bewertungsstandards geprüft werden. Falls die oben beschriebenen Analysen Bewertungsmuster bzw. Prüfertypen aufdecken, die von dieser Vorgabe abweichen, kann dies eine wichtige Grundlage für das interne Qualitätsmanagement der technischen Prüfstellen sowie für die Aus- und Fortbildungen der Fahrerlaubnisprüfer darstellen. So wäre vorstellbar, dass bestimmte Fahraufgaben in vielen Prüfungen nicht durchgeführt werden und ihre angedachte Bedeutung für die Erteilung der Fahrerlaubnis empirisch keine ausreichende Umsetzung findet. Andererseits könnte sich zeigen, dass bestimmte Prüfertypen – außerhalb sinnvoller Toleranzbereiche – zu mild oder zu streng in der Prüfungsbewertung und daran anknüpfend dem Erteilen der Fahrerlaubnis sind. Dies hätte unter Umständen eine Benachteiligung der betroffenen Bewerber zur Folge. Solche Erkenntnisse könnten den technischen Prüfstellen weitere Anknüpfungspunkte für eine optimale Erfüllung des gesetzlichen Auftrags im Rahmen der Praktischen Fahrerlaubnisprüfung bieten.

### 3.7 Kundenzufriedenheit

Bei der Praktischen Fahrerlaubnisprüfung handelt es sich um eine von den technischen Prüfstellen erbrachte Dienstleistung (STURZBECHER & MÖRL, 2008). Daher gelten für die optimierte Praktische Fahrerlaubnisprüfung auch die im Dienstleistungsbereich üblichen Verfahren zur Qualitätssicherung. Dazu gehören Kundenbefragungen als wichtige Verfahrensgruppe zur Erfassung der Dienstleistungsqualität (MEFFERT & BRUHN, 2000).

Kundenbefragungen dienen der gezielten und strukturierten Erhebung von Informationen über die

Nutzererwartungen an die Gestaltung und Qualität einer Dienstleistung. Darüber hinaus werden mit Kundenbefragungen die mit einer bestimmten Dienstleistung verbundenen Nutzungserfahrungen und die Nutzerzufriedenheit der Kunden erfasst (SCHNEIDER & KORNMEIER, 2006). Die Erfassung der Zufriedenheit der Nutzer einer Dienstleistung stellt in der Regel das Kernelement von Kundenbefragungen dar.

Unter „Zufriedenheit“ wird nach v. HOLTZ (1998) „ein individuelles psychologisches Phänomen, das sich aus emotionalen, kognitiven und intentionalen Bestandteilen zusammensetzt“, verstanden (S. 23). Zufriedenheit entsteht als Ergebnis des Vergleichs einer Soll-Komponente mit einer Ist-Komponente. Die Soll-Komponente umfasst dabei Bedürfnisse und Erwartungen; sie stellt ein individuelles Bezugssystem im Sinne eines Vergleichsmaßstabs dar, das zur Beurteilung der Ist-Komponente dient. Als Ist-Komponente fungiert in diesem Vergleich die vom Individuum subjektiv wahrgenommene Realität. Ziel solcher Zufriedenheitsmessungen ist es, aus der Diskrepanz zwischen Soll-Werten und Ist-Werten Schlussfolgerungen zu Optimierungspotenzialen und Optimierungsmaßnahmen abzuleiten. In der Kundenzufriedenheitsforschung kann mit Blick auf die theoretischen Grundlagen das „Konfirmation-Diskonfirmation-Paradigma“ (HOMBURG, 2008) als eine Basistheorie angesehen werden: Bewertet der Kunde den Ist-Zustand als dem Soll-Zustand entsprechend, so entsteht Zufriedenheit („Konfirmation“). Übertrifft der Ist-Zustand aus Sicht des Kunden den Soll-Zustand, so wird dies als „positive Diskonfirmation“ bezeichnet. „Negative Diskonfirmation“ entsteht, wenn der Ist-Zustand den Soll-Zustand nicht erreicht, der Kunde also unzufrieden ist.

Für die Durchführung aussagekräftiger Kundenbefragungen sind besonders hohe Anforderungen an die Validität und Ökonomie des Befragungsverfahrens zu stellen. Substanzielle belastbare Ergebnisse sind insbesondere von multiperspektivischen Kundenbefragungssystemen zu erwarten, die heterogene Kundenanforderungen ggf. aus der Perspektive unterschiedlicher Kundengruppen abbilden. Eine methodische Herausforderung besteht vor allem darin, den Zeitraum zwischen Kundenkontakt und Kundenbefragung weder zu kurz noch zu lang zu wählen, um aus forschungsmethodischer Sicht eine relativ reflektierte Einschätzung der Befragten zu erhalten. Wird die Befragung direkt im Anschluss an die Leistungserbringung durchgeführt, sind zwar



die Nutzungserfahrungen des Kunden noch sehr präsent; allerdings können die mittel- und langfristigen Folgen der Dienstleistung noch nicht hinreichend abgeschätzt werden. Ist der zeitliche Abstand der Befragung von der Leistungserbringung hingegen zu groß, kann es verstärkt zu Erinnerungslücken und (fehlerhaften) mentalen Rekonstruktionen kommen. Als Erhebungsinstrumente kommen bei Kundenbefragungen telefonische und direkte persönliche Interviews vor Ort sowie schriftliche postalische Befragungen und Onlinebefragungen in Betracht, deren methodische Vor- und Nachteile im Einzelfall abzuwägen sind (SCHARNBACHER & KIEFER, 1998).

Die Praktische Fahrerlaubnisprüfung stellt aufgrund ihrer Selektionsfunktion einen Teil des Sicherungssystems dar, das im öffentlichen Interesse die Verkehrssicherheit im Straßenverkehr garantiert. Von staatlicher Seite wird gefordert, dass die Qualitätspolitik der Träger von technischen Prüfstellen den Erwartungen und Erfordernissen der Kunden Rechnung tragen muss (BASt, 2009, Nr. 3.1). Es gilt daher, alle Hinweise der am Fahrerlaubnisprüfungsprozess Beteiligten – und nicht zuletzt der „Kunden“ der Fahrerlaubnisprüfung – zu erfassen und ernst zu nehmen, um das Sicherungssystem im Allgemeinen und die Praktische Fahrerlaubnisprüfung im Besonderen unter inhaltlichen und methodischen Aspekten optimieren und weiterentwickeln zu können (STURZBECHER, BIEDINGER et al., 2010). Hierfür eignen sich unter fachwissenschaftlichen und ökonomischen Aspekten am besten Kundenbefragungen. Im Gegensatz zu offenen Expertenbeobachtungen bzw. Qualitätsaudits, die im Regelfall die maximale Leistungsfähigkeit des Dienstleistungserbringers abbilden, weil dieser – um die Beobachtung und Bewertung wissend – „alles gibt“, spiegelt sich in Kundenbefragungen das typische Leistungsverhalten in Alltagssituationen ohne Anreize für eine maximale Leistungserbringung wider (SACKETT, ZEDECK & FOGLI, 1988; SCHULER, 2001). Schließlich ist die Erfassung der Kundenperspektive insbesondere bei einer beabsichtigten Neueinführung von (optimierten) Dienstleistungen als eine wichtige Informationsquelle für die professionelle Gestaltung der erforderlichen Innovationsprozesse anzusehen (PILLER, 2006). Aus den genannten Gründen haben sich Kundenbefragungen im Bereich der Praktischen Fahrerlaubnisprüfung als ein ergänzendes Instrument zu den internen und externen Audits in Form von Experteneinschätzungen längst

etabliert, um zu aussagekräftigen Qualitätseinschätzungen zu gelangen (BÖNNINGER et al., 2010).

Kundenbefragungen sind im Evaluationssystem zur optimierten Praktischen Fahrerlaubnisprüfung (s. Kapitel 2.3) als ein eigenständiges Element anzusehen. Wie bereits erwähnt, dienen sie dazu, das Niveau und die Einheitlichkeit der Qualität der (alltäglichen) Prüfungsdurchführung (einschließlich der Leistungsbewertung) repräsentativ für Deutschland methodisch belastbar zu erhellen. Genauso wie die Auswertung von Prüfungsergebnissen stellen Kundenbefragungen demnach ein Instrument der prozessualen Evaluation dar.

Da sowohl die Fahrerlaubnisbewerber als auch im weiteren Sinne – nämlich über die Forderung nach einer kundenfreundlichen Gestaltung der Prüfungsabläufe, an der die Fahrschulen beteiligt sind – die Fahrlehrer als Kunden der technischen Prüfstellen anzusehen sind, wurden im vorliegenden Revisionsprojekt Befragungsinstrumente für Bewerber und Fahrlehrer entwickelt. Als methodische Grundlage diente das bereits existierende multiperspektivische Methodensystem der TÜV | DEKRA arge tp 21 zur Erfassung der Zufriedenheit mit der Fahrerlaubnisprüfung (STURZBECHER & MÖRL, 2008). Dieses Methodensystem enthält u. a. standardisierte Fragebögen für Fahrerlaubnisbewerber und für Fahrlehrer.

Das im Revisionsprojekt entwickelte Befragungsinventar knüpft an den vorhandenen wissenschaftlich begründeten und erprobten Fragebögen an und greift einzelne Teile davon auf. Eine einfache Nachnutzung verbietet sich aber, weil mit den jeweiligen Befragungen unterschiedliche Zielstellungen verfolgt werden. Während die Befragungsinstrumente aus dem Methodensystem der TÜV | DEKRA arge tp 21 (STURZBECHER & MÖRL, 2008) dem unternehmensinternen Qualitätsmanagement der technischen Prüfstellen dienen und daher beispielsweise alle Rahmenbedingungen der Prüfungsdurchführung berücksichtigen, fokussieren die im Rahmen des Revisionsprojekts entwickelten und erprobten Fragebögen auf die für den Gesetzgeber relevanten Qualitätsmerkmale der Maßnahme „Optimierte Praktische Fahrerlaubnisprüfung“. Die Fragebögen wurden demnach im Revisionsprojekt entsprechend der Intentionen des Auftraggebers BASt gekürzt und weiterentwickelt.

Im Entwurf zum „Handbuch zum Fahrerlaubnisprüfungssystem (Praxis)“ von 2011 und im Bericht zum

BASt-Forschungsprojekt „Optimierung der Praktischen Fahrerlaubnisprüfung“ (STURZBECHER et al., 2014) wurde vorgesehen bzw. beschrieben, dass künftig mit einem empfohlenen Fünf-Jahres-Turnus bundesweite Bewerberbefragungen durchgeführt werden. Die Möglichkeiten zur methodischen Umsetzung solcher Bewerberbefragungen sollten ursprünglich im Rahmen des vorliegenden Revisionsprojekts erarbeitet und exemplarisch erprobt werden. Entgegen dieser Planung wurde die Bewerberbefragung im Revisionsprojekt jedoch ausschließlich auf die Erfassung der Zufriedenheit (Akzeptanz und Lernwirksamkeit) der Bewerber in den Modellregionen mit dem neuen Rückmeldesystem für die Prüfungsleistungen fokussiert, um gezielt Hinweise zur Optimierung des Rückmeldesystems zu gewinnen. Daneben wurden zusätzlich Bewerbermerkmale erfasst, welche die Zufriedenheit mit dem Rückmeldesystem beeinflussen könnten. Die Details zur Untersuchungsanlage werden im Kapitel 4.8 erläutert; die Ergebnisse der Bewerberbefragung finden sich im Kapitel 5.6.

Ebenfalls im Evaluationssystem zur optimierten Praktischen Fahrerlaubnisprüfung vorgesehen ist eine im Fünf-Jahres-Turnus durchzuführende bundesweite Fahrlehrerbefragung. Das Ziel im Revisionsprojekt bestand darin, die für eine solche Fahrlehrerbefragung notwendigen methodischen Materialien und Strategien zu entwickeln sowie das erarbeitete methodische Instrumentarium und mögliche Untersuchungsabläufe zu erproben. Die diesbezügliche Untersuchungsanlage wird im Kapitel 4.9 beschrieben; die Untersuchungsergebnisse werden im Kapitel 5.6 des vorliegenden Berichts dargestellt.

## 4 Anlage und Durchführung der Erprobungsuntersuchungen

### 4.1 Überblick

In den vorangegangenen Kapiteln wurden die Ausgangslage und die Zielstellung, die konzeptionellen Vorarbeiten und – darauf basierend – die Untersuchungsgegenstände im Revisionsprojekt zur optimierten Praktischen Fahrerlaubnisprüfung beschrieben. Die Untersuchungsgegenstände und die damit in Verbindung stehenden Forschungsfragen wurden in Anlehnung an die von STURZBECHER et al. (2014) begründeten Säulen zur Evaluation der optimierten Prüfung – instrumentelle Evaluation,

Auswertung von Prüfungsergebnissen und Kundebefragungen – erarbeitet (s. Kapitel 2). Zu den Untersuchungsgegenständen zählen die Objektivität, die Reliabilität, die Validität, die populationsspezifische Äquivalenz und die alltägliche Durchführungsqualität der Prüfung sowie die Kundenzufriedenheit mit der Prüfung. Zur Bearbeitung dieser Untersuchungsgegenstände wurde im Rahmen des vorliegenden Projekts eine Reihe von Untersuchungen konzeptioniert und durchgeführt. In diesen Untersuchungen wurden teilweise mehrere Untersuchungsgegenstände aufgegriffen.

Eine erste Gruppe von Untersuchungen diente der Kontrolle der Objektivität, Reliabilität und Validität und somit der instrumentellen Evaluation der Prüfung. Diese Untersuchungen wurden mit simulierten und realen Prüfungen im Straßenverkehr sowie anhand videografierter Prüfungsfahrten durchgeführt. Der damit gewählte multimethodale Ansatz wurde gewählt, um die (unvermeidlichen) methodischen Einschränkungen einzelner Ansätze durch die Stärken anderer Ansätze in Bezug auf die Bestimmung der Verfahrensgüte ausgleichen zu können.

Eine zweite Gruppe von Untersuchungen stellten die Felderprobung, die Bewerberbefragung und die Fahrlehrerbefragung dar. Im Rahmen der Felderprobung wurden in viereinhalb Monaten ca. 9.000 Prüfungen mittels e-Prüfprotokoll durchgeführt. Die an der Felderprobung teilnehmenden Fahrerlaubnisprüfer wurden im Rahmen von drei Veranstaltungen zum Erfahrungsaustausch und zusätzlichen Telefoninterviews zwei Mal zu ihren Erfahrungen im Umgang mit der optimierten Prüfung befragt. Die Bewerberbefragung zur populationsspezifischen Äquivalenz erfolgte im Rahmen der Felderprobung: Alle an der Felderprobung teilnehmenden Fahrerlaubnisbewerber wurden im direkten Anschluss an die Prüfung von ihren Fahrerlaubnisprüfern persönlich zu personenbezogenen Merkmalen befragt. Die Analyse der populationsspezifischen Äquivalenz erlaubt auch Rückschlüsse auf die Validität der Prüfung. Die zweite Bewerberbefragung (v. a. zur Qualität der Leistungsrückmeldungen) und die Befragung der Fahrlehrer zur Kundenzufriedenheit erfolgten (bis auf Ausnahmen) online.

Das vorliegende Kapitel dient der detaillierten Erläuterung des methodischen Vorgehens bei der Durchführung und Auswertung der skizzierten Untersuchungen. Die diesbezüglichen Untersuchungsergebnisse finden sich dann im Kapitel 5.

## 4.2 Untersuchungen zur instrumentellen Evaluation im realen Straßenverkehr

Die Aussagekraft von Validitätsuntersuchungen wächst mit der Natürlichkeit der Untersuchungsbedingungen (BORTZ & DÖRING, 2006). Aus diesem Grund sollten Untersuchungen zur Verfahrensgüte der optimierten Praktischen Fahrerlaubnisprüfung möglichst im realen Straßenverkehr, mit prüfungsreifen Fahranfängern, deren Fahrlehrern und Ausbildungs- bzw. Prüfungsfahrzeugen sowie am Prüfungsort tätigen Fahrerlaubnisprüfern stattfinden. Ferner ist es für die Repräsentativität der Ergebnisse ratsam, Untersuchungen in verschiedenen Regionen Deutschlands und bei unterschiedlichen technischen Prüfstellen durchzuführen, um den Einfluss regionaler und unternehmensspezifischer Besonderheiten besser abschätzen zu können. Diese Maßgaben der natürlichen Untersuchungsbedingungen sowie der regionalen und prüfstellenbezogenen Differenzierung wurden berücksichtigt und umgesetzt: Im Rahmen des vorliegenden Revisionsprojekts haben Untersuchungen im realen Straßenverkehr in Form von 1) simulierten Prüfungen in Essen, 2) realen Prüfungen der Fahrerlaubnisklasse B in Berlin und 3) realen Prüfungen verschiedener Fahrerlaubnisklassen in Oranienburg stattgefunden. Die Ergebnisse dieser Untersuchungen sollen Hinweise liefern, ob bei fachgerechter Anwendung der optimierten Praktischen Fahrerlaubnisprüfung objektive, reliable und valide Prüfungsergebnisse erzielt werden können.

Zu 1): Vom 02. bis 04. Juni 2014 wurden simulierte Prüfungen der Fahrerlaubnisklasse B in Essen durchgeführt und – zur Erfassung der Beobachterübereinstimmung – durch zwei Fahrerlaubnisprüfer mittels des e-Prüfprotokolls gleichzeitig und unabhängig voneinander dokumentiert und bewertet. Als Bewerber und Fahrlehrer fungierten ortsansässige Fahrschüler und deren Fahrlehrer. Ursprünglich waren 20 simulierte Prüfungen geplant, wegen Ausfällen – z. B. Nichterscheinen der Fahrschüler – sind schlussendlich 15 Prüfungen durchgeführt worden.

Zur Validierung der Fahrkompetenzmessung wurde ein Fragebogen zur Einschätzung der Fahrkompetenz des Bewerbers entwickelt, der als Leitfaden für Fahrlehrer-Interviews diente und die vielschichtigen Anforderungen der optimierten Praktischen Fahrerlaubnis-

prüfung an die Bewerber gemäß STURZBECHER et al. (2014) vollständig abbildet. Zu diesen Anforderungen gehören u. a. situationspezifische Fahraufgaben wie „Vorbeifahren und Überholen“ oder „Kreisverkehr“ sowie situationsübergreifende Fahrkompetenzbereiche wie „Fahrzeugpositionierung“ oder „Geschwindigkeitsanpassung“. Die Fahrlehrer wurden gebeten, die Bewältigung der Fahraufgaben und die fünf Fahrkompetenzbereiche<sup>13</sup> für jeden Bewerber zu bewerten. Zusätzlich sollten die Fahrlehrer angeben, welches Prüfungsergebnis sie erwarten. Die Befragung der Fahrlehrer erfolgte in persönlichen Interviews („face-to-face“), die von einem wissenschaftlichen Mitarbeiter des Forschungsnehmers durchgeführt wurden. Die Fahrlehrer wurden sowohl im Vorfeld der Prüfungen als auch im Anschluss an die Prüfungen um die Einschätzung der Bewerber gebeten. So entstanden jeweils eine auf den Ausbildungserfahrungen basierende und eine auf den Prüfungsleistungen basierende Fahrlehrereinschätzung des Bewerbers.

Für die Überprüfung der Validität der optimierten Praktischen Fahrerlaubnisprüfung wurden weiterhin Untersuchungen mittels der sog. „Technik der bekannten Gruppen“ (SCHNELL, HILL & ESSER, 2005) durchgeführt. Hierbei werden die Probanden in Gruppen unterteilt, die sich aufgrund von Ausbildungsmerkmalen im Fahrkompetenzniveau unterscheiden müssten. Würden die auf diese Weise vorherbestimmten „Sehr fahrkompetenten Schüler“ bessere Prüfungsergebnisse als die „Wenig fahrkompetenten Schüler“ erzielen, wären die Fahrerlaubnisprüfer mittels der optimierten Prüfung offensichtlich in der Lage, Fahrerlaubnisbewerber im Hinblick auf ihr Fahrkompetenzniveau fachgerecht zu unterscheiden.

Zu 2): Vom 23. bis 25. Juni 2014 wurden in Berlin reale Praktische Fahrerlaubnisprüfungen der Klasse B durchgeführt. Die insgesamt 20 Prüfungen wurden durch zwei Fahrerlaubnisprüfer mittels des e-Prüfprotokolls gleichzei-

<sup>13</sup> Zu den fünf Kompetenzbereichen zählen „Verkehrsbeobachtung“, „Fahrzeugpositionierung“, „Geschwindigkeitsanpassung“, „Kommunikation“ und „Fahrzeugbedienung/ Umweltbewusste Fahrweise“.

tig und unabhängig voneinander dokumentiert und bewertet. Dies ermöglicht in einem weiteren Schritt die Analyse der Beobachterübereinstimmung. Die Untersuchung der Beobachterübereinstimmung anhand realer Prüfungen bietet gegenüber simulierten Prüfungen den Vorteil einer erhöhten ökologischen Validität, d. h. die Untersuchungsbedingungen entsprechen in verstärktem Maße den natürlichen Bedingungen einer Praktischen Fahrerlaubnisprüfung, wodurch sich die Untersuchungsergebnisse besser verallgemeinern lassen.

Zu 3): Die Untersuchung zur Beobachterübereinstimmung bei realen Prüfungen verschiedener Fahrerlaubnisklassen wurde vom 21. bis 29. Juni 2014 in Oranienburg durchgeführt. Auch diese 20 Prüfungen<sup>14</sup> wurden durch zwei Fahrerlaubnisprüfer mittels des e-Prüfprotokolls gleichzeitig und unabhängig voneinander dokumentiert und bewertet. Ergänzend zu den beiden zuvor beschriebenen Studien verspricht diese Untersuchungsanlage Hinweise darauf, ob eine ausreichende Beobachterübereinstimmung der optimierten Prüfung auch – über die Fahrerlaubnisklasse B hinaus – für andere Fahrerlaubnisklassen angenommen werden kann.

In allen drei Untersuchungen wurden die beiden jeweils teilnehmenden Fahrerlaubnisprüfer vorab in der Handhabung des e-Prüfprotokolls geschult, denn „die Übereinstimmung oder Reliabilität sind keine Eigenschaften von Ratingskalen selbst, sondern Eigenschaften eines bestimmten Gebrauchs von Ratingskalen“ (WIRTZ & CASPAR, 2002, S. 24 f.). Die Handhabung eines Messinstruments ist somit von der Qualität der Versuchsleiterschulung sowie von der Kompetenz, dem Erfahrungshintergrund und der Einstellung der Versuchsleiter abhängig (ebenda). Die Übertragbarkeit der Untersuchungsergebnisse der instrumentellen Evaluation auf andere Fahrerlaubnisprüfer ist also umso höher, je ähnlicher diese den teilnehmenden Fahr-

erlaubnisprüfern in Bezug auf die genannten Einflussfaktoren sind.

Die Prüfungsanforderungen wurden in allen drei Untersuchungen grundsätzlich entsprechend den Vorgaben der Prüfungsrichtlinie ausgestaltet. Für die Sitzpositionen sowie für das Kommunikationsverhalten der Fahrerlaubnisprüfer untereinander und mit dem Bewerber wurden vorab Regeln festgelegt, da beide Aspekte die Durchführung und Dokumentation und somit auch die Bewertung der Prüfung beeinflussen können: Dem Prüfungsalltag entsprechend, führte der hinten rechts im Fahrzeug sitzende Fahrerlaubnisprüfer die Prüfung mittels des e-Prüfprotokolls durch und gab dem Fahrschüler die üblichen Prüfungsanweisungen. Hinten links saß der zweite, „stille“ Fahrerlaubnisprüfer, der die Prüfungsleistungen ebenfalls mittels des e-Prüfprotokolls dokumentierte und bewertete sowie die Prüfungsentscheidung traf. Die beiden Fahrerlaubnisprüfer waren angehalten, ihre Prüfungsbeobachtungen, ihre Prüfungsbewertungen und ihre Prüfungsentscheidungen in keiner Form zu erörtern oder gar abzustimmen. Durch den Vergleich der Prüfungsergebnisse der beiden Fahrerlaubnisprüfer lässt sich die Beobachterübereinstimmung berechnen, die Rückschlüsse auf die Objektivität und Reliabilität der Fahrkompetenzmessung mittels der optimierten Praktischen Fahrerlaubnisprüfung zulässt (s. Kapitel 3.1 und 3.2).

### 4.3 Untersuchungen zur instrumentellen Evaluation mittels videografierten Prüfungen

Beim Betrachten einer videografierten Prüfungssituation ist im Vergleich zu einer realen Prüfungssituation die akustische und visuelle Wahrnehmung durch den Beobachter begrenzt. Die Beurteilungen videografierten und realen Fahrverhaltens können sich daher unterscheiden; die ökologische Validität (s. o.) einer videobasierten Untersuchung ist somit potenziell eingeschränkt. Um derartigen Einschränkungen vorzubeugen, müssen die beurteilungsrelevanten Informationen im Videografierungsprozess möglichst vollständig erfasst und bei der Videovorführung möglichst realitätsnah wiedergegeben werden (z. B. durch verschiedene Kameraperspektiven).

Trotz der potenziell eingeschränkten ökologischen Validität sind videografierte Prüfungen für die instrumentelle Evaluation der optimierten Prak-

<sup>14</sup> Die 20 Prüfungen verteilen sich wie folgt auf die verschiedenen Fahrerlaubnisklassen: A (n = 1), A1 (n = 3), A2 (n = 2), B (n = 8), BE (n = 2), C (n = 2), CE (n = 2). Die Prüfungen wurden zusammengefasst analysiert, da die Fälle zu den einzelnen Fahrerlaubnisklassen nicht ausreichen, um bei einer Analyse pro Klasse zu aussagekräftigen Ergebnissen zu gelangen.



tischen Fahrerlaubnisprüfung unverzichtbar, denn sie bieten gegenüber anderen Untersuchungsdesigns verschiedene Vorteile: Erstens erlauben sie die gleichzeitige Ermittlung der Beobachterübereinstimmung bei einer größeren Zahl von Beobachtern, was repräsentativere Ergebnisse erzeugt als die Analyse der Übereinstimmung von nur zwei Beobachtern wie bei den Untersuchungen im realen Straßenverkehr (s. Kapitel 4.1). Zweitens erfolgen diese Beobachtungen auf Basis der exakt gleichen Informationen; eine Erarbeitung und Beeinträchtigung der Übereinstimmungswerte durch abweichende Sitzpositionen der Fahrerlaubnisprüfer ist beispielsweise ausgeschlossen. Drittens kann eine videografierte Prüfung mehrmals durch denselben Fahrerlaubnisprüfer bewertet und so die Bewertungsstabilität berechnet werden. Viertens kann untersucht werden, ob die Beobachterübereinstimmung zwischen denselben Beobachtern bei mehrfacher Beurteilung derselben Prüfungen zunimmt, was ein Validitätsindiz für die Praktische Fahrerlaubnisprüfung darstellen würde (s. u.).

### **Erprobung verschiedener Videografierungskonzepte**

Die Erarbeitung eines Videografierungskonzepts sollte sich – innerhalb der gegebenen technischen und finanziellen Möglichkeiten – vor allem nach dem Nutzungszweck der Filme richten. Im vorliegenden Fall handelt es sich dabei um die Untersuchung der Beobachterübereinstimmung und Bewertungsstabilität bei der optimierten Praktischen Fahrerlaubnisprüfung.

Naturngemäß können Filme die audiovisuell wahrnehmbaren Merkmale von realen Situationen nur begrenzt wiedergeben und haptisch spürbare Merkmale gar nicht vermitteln. Innerhalb dieser Grenzen ist jedoch ein erheblicher Ausgestaltungsspielraum vorhanden, der zur Erzeugung möglichst realitätsnaher Sinneseindrücke bei den Betrachtern der Filme genutzt werden kann. Die videografische Erfassung der leistungsbewertungsrelevanten Prüfungsereignisse sollte daher in visueller und auditiver Hinsicht der realen Prüfungssituation möglichst nahekommen. Hinsichtlich auditiver Informationen ist dies durch Installation eines Mikrophons im hinteren Teil des Prüfungsfahrzeugs und anschließende Abspiegelung der Aufnahme – z. B. über Lautsprecher – relativ einfach zu bewerkstelligen. Bezüglich der Präsentation optischer Informationen muss jedoch wohl überlegt werden, welche Sichtperspek-

tiven des Fahrerlaubnisprüfers mit welchen technischen Hilfsmitteln erfasst bzw. später präsentiert werden sollten, um dem Fahrerlaubnisprüfer ein möglichst natürliches Bewertungsverhalten zu ermöglichen. Vor diesem Hintergrund wurden drei Konzepte zur Videografierung Praktischer Fahrerlaubnisprüfungen entwickelt<sup>15</sup> und mithilfe einer Medienproduktionsfirma im Rahmen von Testaufnahmen umgesetzt:<sup>16</sup>

#### **1. Fahrerlaubnisklasse B – 6 Perspektiven**

Ein Großteil (ca. 80 %) der in Deutschland abgenommenen Fahrerlaubnisprüfungen betrifft die Fahrerlaubnisklasse B (Personenkraftwagen). Untersuchungen zur Güte der optimierten Praktischen Fahrerlaubnisprüfung sollten daher Klasse B-Prüfungen einbeziehen bzw. darauf fokussieren. Während einer Klasse B-Prüfung sitzt der Fahrerlaubnisprüfer hinten rechts im Fahrzeug mit Blick nach vorn. Dabei kann er den vor ihm liegenden Straßenverkehr (1. Perspektive) und das Verhalten des Bewerbers (2. Perspektive) gut beobachten. Bei Bedarf kann er zusätzlich links (3. Perspektive), rechts (4. Perspektive) und rückwärtig (5. Perspektive) aus dem Fenster schauen, um beispielsweise Informationen über den seitlichen oder rückwärtigen Verkehr zu gewinnen. Ferner ist der Blick auf die Armaturen (6. Perspektive) für die Beurteilung der Prüfungsleistungen des Bewerbers notwendig.

<sup>15</sup> Die Videografierung der Praktischen Prüfungen betrifft lediglich die Prüfungsfahrt. Die Grundfahraufgaben sowie die Fahrtechnische Vorbereitung und der Fahrtechnische Abschluss wurden im Rahmen des Optimierungsprozesses zur Praktischen Fahrerlaubnisprüfung in ihrem Ablauf bisher nicht verändert und sind daher kein Gegenstand der Güteuntersuchungen. Abgesehen davon bieten die Kriterien zur Bewertung dieser Prüfungselemente wenig Interpretationsspielraum, weshalb von einer hohen Einheitlichkeit der Bewertungen und somit von einer hohen Verfahrensobjektivität und -zuverlässigkeit ausgegangen werden kann.

<sup>16</sup> Die Videoaufnahmen wurden vom Projektnehmer (IFK) ausschließlich zu internen wissenschaftlichen Zwecken im Rahmen der Messungen zur instrumentellen Evaluation eingesetzt. Erstellt wurden die Aufnahmen von einer Medienproduktionsfirma, welche die Einhaltung aller gesetzlichen datenschutzrechtlichen Vorgaben garantiert hat.

Die Videoaufnahmen wurden für Dritte unzugänglich aufbewahrt und nicht weitergegeben. Darüber hinaus wurden sie nicht in Ergebnispräsentationen oder Berichten veröffentlicht. Zusammenfassend bleibt festzuhalten, dass alle Maßnahmen getroffen wurden, um die Anforderungen des Bundesdatenschutzgesetzes (BDSG), insbesondere des § 9 BDSG sowie der Anlage zu § 9 Satz 1 des BDSG, zu erfüllen.



**Bild 3:** Videografierung von Klasse B-Prüfungen – Konzept 1

Das Bild 3 beinhaltet ein Standbild der Probeaufnahmen mit den sechs genannten Perspektiven.<sup>17</sup> Die dargestellten Perspektiven wurden für die Erstellung des Prüfungsfilms in einem Bild zusammengeschnitten<sup>18</sup> und zeitlich synchronisiert. Die Sicht zentral aus der Frontscheibe (1. Perspektive) stellt die wesentliche Informationsquelle für die Beurteilung der Fahrleistungen des Bewerbers dar und wird daher zentral und groß dargestellt. Die anderen Perspektiven wurden so um die zentrale Perspektive herum angeordnet, dass sie erstens die zentrale Perspektive möglichst wenig stören und zweitens ihre Anordnung der Realität möglichst nahekommt (z. B. befindet sich die Sicht links hinten aus dem Fahrzeug unten links im Zusammchnitt).

<sup>17</sup> Die Heckkamera ist während der Probeaufnahmen ausgefallen, sodass ein Standbild eingefügt werden musste, das nicht dem örtlichen Kontext der anderen fünf Aufnahmen entspricht.

<sup>18</sup> Zur weiteren Erhöhung der ökologischen Validität wurde zunächst erwogen, die verschiedenen Perspektiven bei der Präsentation der Videos auf verschiedenen Bildschirmen darzustellen. Die Bildschirme sollten dabei – im Sinne einer möglichst realitätsgetreuen Wiedergabe der Perspektiven – um die Zuschauer (d. h. Prüfer) herum angeordnet werden. Nach reiflicher Prüfung der technischen, organisatorischen und finanziellen Rahmenbedingungen wurde dieses Vorgehen aber zugunsten der Darstellung aller Perspektiven in einem (großen) Bild aufgegeben.

## 2. Fahrerlaubnisklasse B – 2 Perspektiven

Die im oben beschriebenen Konzept 1 einbezogenen Perspektiven beinhalten nahezu alle für die Bewertung einer Fahrerlaubnisprüfung relevanten visuellen Informationen. Allerdings birgt die gleichzeitige Darstellung von sechs Perspektiven das Risiko, die Fahrerlaubnisprüfer beim Betrachten der Filme zu überfordern. Außerdem beinhaltet keine der Perspektiven die „eigentliche“ Sicht des Fahrerlaubnisprüfers von hinten rechts im Fahrzeug nach vorne auf die Straße. Daher wurde ein zweites, stark vereinfachtes Konzept mit nur zwei Perspektiven getestet. Bei diesem Konzept wird die Hauptperspektive von einer Kamera zwischen den Kopfstützen des Vordersitzes erfasst; sie zeigt den vorderen Teil des Wagens – inklusive Fahrerlaubnisbewerber, Spiegel und Schaltknäuf – sowie die Sicht auf die Straße. Die zweite Perspektive beinhaltet eine Nahaufnahme der Armaturen, da diese aus der Hauptperspektive nicht ausreichend gut zu erkennen sind. Das Bild 4 zeigt ein Standbild des zweiten Konzepts. Die dargestellte Hauptperspektive entspricht im Wesentlichen der „eigentlichen“ Sicht des Fahrerlaubnisprüfers.<sup>19</sup> Hierbei besteht allerdings die Gefahr, dass – insbesondere bei wechselnden Lichtverhältnissen wie zum Beispiel im Wald oder im Tunnel – der Lichtkontrast zwischen Fahrzeuginnenraum und Außenwelt zu einer schlechten Aufnahmequalität der Außenwelt durch



**Bild 4:** Videografierung von Klasse B-Prüfungen – Konzept 2

Überblendungseffekte führen könnte. Außerdem ist der Blick nach draußen durch das Fahrzeuggehäuse beschränkt, sodass möglicherweise relevante Informationen links, rechts oder hinter dem Fahrzeug sowie auf der Straße oder über dem Fahrzeug (z. B. Ampeln, Verkehrsschilder) verdeckt sein könnten

### 3. Fahrerlaubnisklasse A – 2 Perspektiven

Bei Prüfungen der Fahrerlaubnisklasse A (Krafträder) begleiten der Fahrlehrer (am Steuer) und der Fahrerlaubnisprüfer (auf dem Beifahrersitz) den Bewerber auf dem Kraftrad (Führungsfahrzeug) in einem Personenkraftwagen. Der Fahrerlaubnisprüfer gibt dem Bewerber die Prüfungsanweisungen über Funk. Der Fahrerlaubnisprüfer blickt zur Beurteilung der Prüfungsleistungen meist nach vorn (1. Perspektive), was sich vergleichsweise einfach per Kamera erfassen lässt. Zusätzlich müssen – zur Einschätzung der Geschwindigkeit des Kraft-



**Bild 5:** Videografierung von Klasse A-Prüfungen

rades – lediglich die Armaturen des Personenkraftwagens erfasst werden (2. Perspektive). Das Bild 5 zeigt ein Standbild einer videografierten Kraftradprüfung und die weitreichende Prüferperspektive.

### Auswahl eines Videografierungskonzepts

Bei der Abwägung, welches Videografierungskonzept schließlich verwendet werden soll, wurden nicht nur theoretische und methodische Argumente berücksichtigt, sondern auch die Einschätzungen von vier erfahrenen Fahrerlaubnisprüfern. Diesen Fachexperten wurden zur Beurteilung der Verwendbarkeit der Konzepte folgende vier Kriterien vorgegeben:

- Vollständigkeit prüfungsrelevanter Informationen,
- Erkennbarkeit prüfungsrelevanter Informationen,
- Beobachterfreundlichkeit (im Sinne einer intuitiv verständlichen und natürlichen Form der Darbietung) und
- erwartete Akzeptanz unter Fahrerlaubnisprüfern.

Im Ergebnis der Diskussion wurde festgestellt, dass die Videos der Klasse A-Prüfungen (3. Konzept) nicht verwendet werden sollen, weil – entgegen der ursprünglichen Annahmen – die Sicht nach vorn aus dem Fahrzeug nicht alle wichtigen Informationen aus dem Straßengeschehen beinhaltet (z. B. müssen die Prüfer häufig auch nach links und rechts blicken, um zu beurteilen, ob der Bewerber ggf. kreuzende Fahrzeuge ausreichend berücksichtigt).

Darüber hinaus sind die Fahrzeugbedienung des Bewerbers und seine Verkehrsbeobachtung beim Beobachten des Videos noch schwieriger zu beur-

<sup>19</sup> Ferner wurde überlegt, die „eigentliche“ Sicht des Fahrerlaubnisprüfers über eine Helmkamera zu erfassen, um die bewertungsrelevanten Informationen noch realitätsgetreuer darstellen zu können. Dies wurde aus folgenden Gründen jedoch wieder verworfen: Erstens entsprechen die Kopfbewegungen der Fahrerlaubnisprüfer nur teilweise ihrem Blickverhalten. So wäre beispielsweise vorstellbar, dass der Kopf des Fahrerlaubnisprüfers zur Dokumentation einer Prüfungsleistung nach unten gerichtet ist, während er im Wahrnehmungsbereich seiner Augen trotzdem das Geschehen auf der Straße zumindest punktuell erfasst. Diese Informationen blieben den Betrachtern der Filme verwehrt. Zweitens führt eine Helmkamera wegen der fehlenden Fixierung gewöhnlich zur sehr unruhigen Aufnahmen. Drittens würde der Aufmerksamkeitsfokus der Filmbetrachter durch die Kopfbewegungen des Fahrerlaubnisprüfers auf bestimmte visuelle Informationen gelenkt. Es läge also – anders als bei einer echten Prüfung – nicht mehr im Ermessen des Fahrerlaubnisprüfers, wohin er zur Beurteilung der Prüfungsleistung blickt und was er somit als bewertungsrelevant ansieht. Die ökologische Validität der Aufnahmen wäre folglich eingeschränkt.





**Bild 6:** Videografierung von Klasse B-Prüfungen – Abschließendes Konzept

teilen, als dies in realen Klasse A-Prüfungen durch die räumliche Distanz zwischen Kraftrad und Prüfer-Pkw ohnehin der Fall ist.

Schließlich wurde entschieden, für die Videografierung von Klasse B-Prüfungen weder das erste noch das zweite oben genannte Konzept zu verwenden, sondern eine neue Konzeptversion, welche die Vorzüge der ersten beiden Konzepte verbindet (Bild 6).

Zentral und groß dargestellt (Perspektive 1) wird in dieser Version die Hauptperspektive aus dem zweiten Videografierungskonzept (s. o.), die aus Sicht der Fachexperten sehr realitätsnah das Hauptblickfeld des Prüfers während einer Prüfung abbildet. Allerdings ist gegenüber der Hauptperspektive aus dem ersten Videografierungskonzept (s. o.) die Sicht auf Umweltbereiche bzw. Verkehrsgegebenheiten begrenzt, die sich oberhalb des Prüfungsfahrzeugs befinden bzw. auf der Straße ereignen. Dies wurde jedoch nicht als Einschränkung empfunden, sondern als Verbesserung der ökologischen Validität gewertet: Zum einen entspricht diese Anordnung den realen Sichtmöglichkeiten eines Fahrerlaubnisprüfers während einer Klasse B-Prüfung. Zum anderen bilden – wie bei einer realen Prüfung – die Sicht des Fahrerlaubnisprüfers auf den Bewerber und die Sicht vorn aus der Frontscheibe eine Einheit und werden nicht – wie im ersten Videografierungskonzept – künstlich in zwei Perspektiven getrennt.

Mit der Hauptperspektive wird aus Sicht der Experten das Geschehen links und rechts des Prüfungsfahrzeugs noch nicht genügend erfasst, um beispielsweise kreuzende Fahrzeuge ausreichend früh zu sehen und die Korrektheit der Bewerberreaktionen adäquat beurteilen zu können. Daher wird die Sicht aus dem linken Rücksitzfenster (Perspektive 2) und dem rechten Rücksitzfenster (Perspektive 3) mit zwei zusätzlichen Kameras erfasst. Die daraus resultierenden Perspektiven sind – ähnlich wie bei einer realen Prüfungsfahrt – links unterhalb bzw. rechts unterhalb der Hauptperspektive angeordnet, um das intuitive Verständnis der dargebotenen Informationen zu erhöhen. Ferner wurden die direkte Sicht auf die Armaturen (Perspektive 4) und aus dem Rückfenster (Perspektive 5) als wichtige ergänzende Informationen gewünscht.

Im Ergebnis entstand ein neues Videografierungskonzept mit fünf (Kamera-)Perspektiven.<sup>20</sup> Nach

<sup>20</sup> An dieser Stelle könnte kritisiert werden, dass die Beobachtung von fünf simultan dargestellten Perspektiven die Fahrerlaubnisprüfer überfordern könnte. Dem ist entgegenzuhalten, dass sich die Prüfer bei der Betrachtung der Filme – wie bei einer realen Prüfung – auf die Hauptperspektive konzentrieren, während die anderen Perspektiven – z. B. das Geschehen links, rechts oder hinter dem Fahrzeug – nur bei Bedarf ergänzend berücksichtigt werden. Die Anforderungen an die Beobachtung einer videografierten Prüfungsfahrt sind so gesehen nicht größer als die Anforderungen an die Beobachtung einer realen Prüfungsfahrt.



Ansicht der Experten werden mit diesem Konzept alle für die ökologisch valide Bewertung von Praktischen Fahrerlaubnisprüfungen nötigen Informationen ausreichend vollständig, intuitiv verständlich und realitätsnah erfasst.<sup>21</sup> Dem Konzept wurde daher eine hohe Akzeptanz bei den Prüfern vorausgesagt, welche die Prüfungsvideos beurteilen sollen.

### **Erstellung der Prüfungsvideos**

Gemäß der Vorgaben von STURZBECHER et al. (2014) sind für die videogestützten Untersuchungen zur Beobachterübereinstimmung und Bewertungsstabilität der optimierten Praktischen Fahrerlaubnisprüfung mindestens vier (simulierte) Prüfungen zu videographieren. Die entsprechenden Videoaufnahmen wurden am 30. Juni 2014 in Berlin mit einem erfahrenen Fahrerlaubnisprüfer, einem versierten Fahrlehrer sowie mehreren prüfungsreifen Fahrschülern begonnen (Fahrerlaubnisklasse B). Die Bedingungen der simulierten Prüfungen (z. B. Prüfstrecke, Prüfdauer und Prüfungsinhalte) entsprachen genau der Prüfungsrichtlinie. Alle simulierten Prüfungen fanden mit demselben Prüfungsfahrzeug statt. Die Kameraeinstellungen wurden von der beauftragten Medienproduktionsfirma vorgenommen und vom teilnehmenden Fahrerlaubnisprüfer im Hinblick auf eine natürliche Sichtweise überprüft. Die Arbeiten am ersten Drehtag wurden u. a. durch technische Probleme mit den Kameras (unzureichende Batterielaufzeiten) erschwert, sodass im Ergebnis nur drei vollständige Prüfungsvideos vorlagen. Infolgedessen wurden am 31. Juli 2014 zwei weitere simulierte Prüfungen der Fahrerlaubnisklasse B in Oranienburg videografiert. Am zweiten Drehtag traten an einigen Stellen der Prüfungsfahrten durch starke Sonneneinstrahlung Überblendungseffekte auf. Diese haben – je nach Streckenprofil (z. B. Fahrtrichtung, Bebauung, Bepflanzung) und Bewölkung – die Aufnahmequalität aus der Hauptperspektive punktuell beeinträchtigt (s. o. und Fußnote).

Im Anschluss an die beiden Drehtage wurden die Aufnahmen bildlich und klanglich nachbearbeitet, um eine optimale Qualität zu erreichen. So wurden

die Anweisungen der Fahrerlaubnisprüfer beispielsweise in einigen Fällen nachgesprochen, um Verständnisschwierigkeiten bei der Videobeobachtung vorzubeugen. Weiterhin wurden auf Anraten von Fachexperten die für die Bewertung der Prüfungsleistungen relevanten Verkehrsschilder zum Zeitpunkt ihres Erscheinens oben rechts im Bild als Piktogramme eingeblendet. Dies war notwendig, weil die Kameras bei den Videoaufnahmen fixiert waren und die Fahrerlaubnisprüfer daher beim Beobachten der Videos – im Gegensatz zur realen Prüfung – nicht die Möglichkeit hatten, ihr Blickfeld zu verändern, falls Fahrzeugteile oder der Fahrlehrer ein Verkehrsschild überdecken. Infolgedessen wären die Betrachter der Videos über geltende Verkehrsregeln möglicherweise im Unklaren geblieben und hätten die Bewerberleistungen nicht adäquat beurteilen können. Insgesamt wurden vier Prüfungsvideos fertiggestellt.

### **Vorführung und Bewertung der Prüfungsvideos**

Die vier erstellten Prüfungsvideos wurden den an der Felderprobung teilnehmenden Fahrerlaubnisprüfern an drei Messzeitpunkten mit der Bitte vorgeführt, die beobachteten Bewerberleistungen per e-Prüfprotokoll zu dokumentieren und zu bewerten. Das Ziel war die Ermittlung der Beobachterübereinstimmung zwischen den Prüfern und der Bewertungsstabilität bei den einzelnen Prüfern (s. Kapitel 3.1 und 3.2). Im Rahmen aller Vorführungen diente die Beobachtung und Bewertung des ersten Videos als Übungsaufgabe (sog. „Aufwärmphase“). Aufgrund krankheits- bzw. organisatorisch bedingter Ausfälle konnten nicht alle an der Felderprobung teilnehmenden Fahrerlaubnisprüfer bei allen Messungen zugegen sein. Die Stichproben erwiesen sich aber in allen Fällen als ausreichend für die Ermittlung aussagekräftiger Befunde. Die Vorführungen erfolgten über Beamer auf großen Leinwänden. Die Fahrerlaubnisprüfer waren angehalten, ihre Leistungsdokumentationen und Leistungsbewertungen in keiner Form untereinander zu diskutieren oder gar abzustimmen.

Die erste Messung fand kurz vor der Felderprobung im Rahmen der Versuchsleiterschulungen statt (s. Kapitel 4.5). Die zweite und die dritte Messung erfolgten im Zuge von Treffen zum Erfahrungsaustausch bezüglich des Umgangs mit dem e-Prüfprotokoll (s. Kapitel 4.6) zur Mitte bzw. nach Ende der Felderprobung. Da der Mensch in seinem Streben nach Verhaltenskonsistenz seine vormaligen Ein-

<sup>21</sup> Bei den Testaufnahmen ergaben sich zuweilen Überblendungseffekte durch Lichtkontraste zwischen der Außenwelt und dem Innenraum. Die daraus resultierenden punktuellen Einschränkungen der Aufnahmequalität wurden von den Fachexperten als tolerierbar beurteilt.

schätzungen eher bestätigt als korrigiert, was ggf. zu einer Überschätzung der Bewertungsstabilität führen würde, waren die Abstände zwischen den Vorführungen so zu wählen, dass Erinnerungseffekte der Fahrerlaubnisprüfer minimiert werden. Die drei Messungen im vorliegenden Revisionsprojekt erfolgten daher zu Beginn, zur Mitte und zum Ende der Felderprobung, um die zeitlichen Abstände zu maximieren. Erkenntnisse aus anderen Studien der Verkehrsforschung lassen diese – ca. achtwöchigen – Abstände als ausreichend erscheinen: BÉDARD, PARKKARI, WEAVER, RIENDEAU & DAHLQUIST (2010) wählten für ihre Fahrsimulatorstudie ein Retest-Intervall von vier Wochen, STURZBECHER (2004) verwendete bei Reliabilitätsuntersuchungen im Rahmen der Pädagogisch qualifizierten Fahrschulüberwachung (PQFÜ) einen Abstand von acht Wochen. Zur Reduzierung möglicher Erinnerungseffekte wurde ferner die Reihenfolge der Videopräsentationen bei den Messzeitpunkten variiert.

Die Ermittlung der Beobachterübereinstimmung zu verschiedenen Messzeitpunkten ermöglicht – zusätzlich zur Kontrolle der Objektivität und Reliabilität – Hinweise auf die Validität der Prüfung: Nimmt die Beobachterübereinstimmung über die Messzeitpunkte zu, ergibt sich ein Validitätsindiz, da mit zunehmender Übung der Fahrerlaubnisprüfer im Umgang mit der optimierten Prüfung der Einfluss individueller Fehler auf die Fahrkompetenzmessung abnehmen sollte.

Nach der Dokumentation und Bewertung der videografierten Prüfungsleistungen durch die teilnehmenden Fahrerlaubnisprüfer wurden die so generierten Daten von der TÜV | DEKRA arge tp 21 aus dem e-Prüfprotokoll ausgelesen und für die Datenauswertung bereitgestellt.

#### 4.4 Felderprobung

Neben dem Nachweis der instrumentellen Güte stellt die Analyse der alltäglichen Durchführungsqualität anhand realer Prüfungsergebnisse einen zentralen Bestandteil der Evaluation der optimierten Praktischen Fahrerlaubnisprüfung dar (STURZBECHER et al., 2014). Diese Analyse erfolgte auf Basis von 8.788 Prüfungen (bereinigter Datensatz, alle Fahrerlaubnisklassen), die vom 15. September 2014 bis zum 06. Februar 2015 in verschiedenen Modellregionen der technischen Prüfstellen unter Realbedingungen durchgeführt wurden. In diesen Prüfungen wurden die Prüfungsleistungen mittels

e-Prüfprotokoll dokumentiert und bewertet; die so erfassten Daten lieferten die Grundlage der wissenschaftlichen Auswertungen. Die Erprobungsuntersuchung fand mit insgesamt 43 Fahrerlaubnisprüfern in der DEKRA-Niederlassung Oranienburg, in der TÜV Rheinland-Niederlassung Berlin, in der TÜV NORD-Region Hannover (Stadt/Land) sowie an verschiedenen TÜV SÜD-Standorten in Bayern, Baden-Württemberg und Hamburg statt.

Die technischen Prüfstellen disponierten die Prüfungen für den Zeitraum der Erprobung in den Modellregionen mit einem zusätzlichen Zeitanteil, sodass die teilnehmenden Fahrerlaubnisprüfer zeitliche Reserven für die Umsetzung der geforderten optimierten Inhalte und Prozesse zur Verfügung hatten. So wurde beispielsweise zugunsten einer detaillierteren Dokumentation kompetenzbezogener Bewertungen und einer ausführlichen Leistungsrückmeldung an die Fahrerlaubnisbewerber sowie für die vom Prüfer durchzuführenden Bewerberbefragungen zusätzliche Zeit im Anschluss an die Prüfungsfahrt eingeplant (s. Kapitel 6).

Bevor die teilnehmenden Fahrerlaubnisprüfer mit der Erprobung des e-Prüfprotokolls im Feld begannen, wurden sie hinsichtlich der konzeptionellen und inhaltlichen Hintergründe sowie der praktischen Durchführung der optimierten Praktischen Fahrerlaubnisprüfung intensiv geschult. Dabei standen die Inhalte und die Bedienung des e-Prüfprotokolls im Mittelpunkt der Schulung. Als Ausgangspunkt für die Fortbildung diente das im Projekt erarbeitete Konzept für eine bundesweit einheitliche Fortbildung der Fahrerlaubnisprüfer zur Einführung des e-Prüfprotokolls (s. Kapitel 2.5). Dieses Fortbildungskonzept wurde erstmalig zur Vorbereitung der an der Erprobung teilnehmenden Fahrerlaubnisprüfer eingesetzt; dafür wurden einige wenige inhaltliche Anpassungen an das laufende Projekt vorgenommen. Aufgrund der begrenzten Zeit konnten beispielsweise nicht alle gemäß Konzept vorgesehenen Inhalte im Theorieteil geschult werden. Im Revisionsprojekt umfassten die Schulungsveranstaltungen zwei Tage und wurden in jeder technischen Prüfstelle separat durchgeführt.

Am ersten Schulungstag wurden die theoretischen, methodischen und inhaltlichen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung vermittelt; Schwerpunkt war der Fahraufgabenkatalog. Daran anschließend wurden die Dokumentationsprozesse mittels e-Prüfprotokoll erläutert. In diesem

Zusammenhang wurden auch die Besonderheiten der Bewertungskriterien in Bezug auf die ereignisbezogene Bewertung einerseits und die kompetenzbezogene Bewertung andererseits erklärt. Im Zuge der Einführung in die Bedienung des e-Prüfprotokolls nutzten die Fahrerlaubnisprüfer ihre jeweiligen Eingabegeräte (Tablet-Computer), um sich mit dem Protokollaufbau vertraut zu machen und Übungen zur gezielten Dokumentation von Prüfungsereignissen (dokumentationswürdiges Bewerberverhalten) durchzuführen. Am zweiten Schultag standen weitere Übungen zum Umgang mit dem e-Prüfprotokoll im Fokus der Fortbildungsveranstaltung. Darüber hinaus fand im Rahmen von vertiefenden Praxisübungen anhand videografiertes Prüfungsfahrten auch die erste Messung der Beobachterübereinstimmung statt (s. Kapitel 3). Dabei hatten die Fahrerlaubnisprüfer die Möglichkeit, anhand der gezeigten Videos das Bewerberverhalten in Echtzeit zu dokumentieren.

Nach den beiden Schultagen waren die Fahrerlaubnisprüfer angehalten, die Nutzung des e-Prüfprotokolls bis zum Beginn der Felderprobung individuell weiter zu üben und die Inhalte zu verinnerlichen. Hierfür wurde den Prüfern Übungs- und Vorbereitungsmaterial zur Unterstützung des Lernprozesses bereitgestellt. Für die Zukunft wird empfohlen, die Schulungsveranstaltung um einen Tag zu verlängern, um für die Prüfer ausreichend intensive Übungsstunden im Beisein der Schulungsleiter gewährleisten zu können. Darüber hinaus sollte vor dem Praxiseinsatz genügend Zeit für die selbstständige Aneignung der Inhalte der optimierten Praktischen Fahrerlaubnisprüfung (v. a. Fahraufgabenkatalog) und für das Erlernen der Bedienung des e-Prüfprotokolls eingeplant werden.

Während der Felderprobung wurden sowohl die Bewerberbefragung zur populationsspezifischen Äquivalenz (s. Kapitel 3.4) als auch die Befragung der Bewerber zu ihrer Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung (s. Kapitel 3.6) durchgeführt. Daher wurde die Schulungsveranstaltung auch dafür genutzt, die Prüfer über die Inhalte und die Durchführung dieser Befragungen zu informieren. Einerseits wurden die Prüfer instruiert, wie sie die Befragung zur Äquivalenz mit den Bewerbern im Anschluss an ihre Prüfungsfahrt durchführen sollen. Dazu wurden die Befragungsziele und die einzelnen Fragen erläutert sowie Hilfestellungen zum Führen des Einführungsgesprächs mit den Bewerbern gegeben. Andererseits

wurden die Prüfer gebeten, nach der Praktischen Fahrerlaubnisprüfung die Bewerber auf die Onlinebefragung zur Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung hinzuweisen und den Bewerbern ihre Zugangscodes auszuhändigen. Dafür erhielten die Prüfer die Abreißblöcke mit den Zugangscodes sowie die Befragung betreffende Informationen (z. B. zu Datenschutzbestimmungen).

Mit der Felderprobung wurden alle zur optimierten Praktischen Fahrerlaubnisprüfung zugehörigen Verfahren und Abläufe erprobt. Während der Erprobung wurden die Prüfungsfahrten nach geltendem Recht unter Einsatz des e-Prüfprotokolls durchgeführt. Unterschiede zwischen der gegenwärtig praktizierten und der optimierten Praktischen Fahrerlaubnisprüfung finden sich insbesondere hinsichtlich der Dokumentation und der kompetenzbezogenen Bewertung der Prüfungsleistungen im e-Prüfprotokoll. Neben der elektronischen Dokumentation stellt die explizit geforderte zusammenfassende kompetenzbezogene Bewertung der acht Fahraufgaben und fünf Kompetenzbereiche nach der Prüfungsfahrt die größte Umstellung für die Fahrerlaubnisprüfer dar. Darüber hinaus wurde die Rückmeldung an den Bewerber umfassend überarbeitet und erweitert. Dies betrifft einerseits das Auswertungsgespräch, in dem eine ausführlichere und differenzierte kompetenzbezogene Rückmeldung an den Bewerber gegeben werden soll. Andererseits stellt auch die schriftliche Rückmeldung, die nun jeder Bewerber – unabhängig davon, ob er die Prüfung bestanden hat oder nicht – nach der Prüfung erhält, eine erhebliche Neuerung dar.

Der Aufbau und die Inhalte der schriftlichen Rückmeldung wurden bereits im Kapitel 2.4 beschrieben. Bei DEKRA wurde diese Rückmeldung unmittelbar nach der Prüfung im Prüfungsfahrzeug ausgedruckt und dem Fahrerlaubnisbewerber ausgehändigt. Bei den anderen technischen Prüfstellen wurde die schriftliche Rückmeldung in Form eines pdf-Dokuments auf einem von der TÜV|DEKRA angebotenen Web-Portal bereitgestellt. Die Fahrerlaubnisbewerber erhielten nach Abschluss ihrer Prüfung vom Fahrerlaubnisprüfer – neben Erläuterungen zum Aufbau und zu den Inhalten der schriftlichen Rückmeldung – einen individuellen Zugangscode, mit welchem sie ihre schriftliche Rückmeldung herunterladen konnten.

Die gesamte Felderprobung wurde sowohl wissenschaftlich als auch technisch begleitet. Die

Fahrerlaubnisprüfer wurden im Verlauf der Erprobung mehrfach zu ihren Erfahrungen mit der optimierten Praktischen Fahrerlaubnisprüfung im Rahmen von Telefoninterviews befragt (s. u.). Darüber hinaus wurde am Ende des Erprobungszeitraums in jeder technischen Prüfstelle ein Erfahrungsaustausch bzw. eine Gruppendiskussion durchgeführt (s. Kapitel 4.4). Der technische Support für die an der Erprobung teilnehmenden Fahrerlaubnisprüfer wurde sowohl durch die technischen Prüfstellen (unternehmensinterne IT- und Organisationsprozesse) als auch durch die TÜV|DEKRA arge tp 21 (z. B. e-Prüfprotokoll) sichergestellt.

Im Ergebnis der Gruppendiskussionen in den technischen Prüfstellen bleibt festzuhalten, dass die Erprobung der optimierten Praktischen Fahrerlaubnisprüfung sowie insbesondere die Integration des e-Prüfprotokolls und die Umsetzung der optimierten Prozessabläufe (z. B. Rückmeldesystem) im Prüfungsalltag aus Sicht aller Erprobungsbeteiligten als Erfolg gewertet wurde. Es konnten alle Prüfungen mittels der neuen Systeme durchgeführt werden; technisch bedingte „Ausfälle“ traten nicht auf. Die erarbeiteten Inhalte und nicht zuletzt das Konzept der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien erwiesen sich als tragfähig und praktikabel. Zudem ist es auch als Gewinn anzusehen, dass die Felderprobung wichtige Hinweise zur Weiterentwicklung der optimierten Praktischen Fahrerlaubnisprüfung erbrachte. Die Rückmeldungen der teilnehmenden Fahrerlaubnisprüfer werden im Detail im nächsten Kapitel beschrieben.

#### **4.5 Rückmeldungen der Fahrerlaubnisprüfer zur Praxiserprobung des e-Prüfprotokolls**

Die erste Erhebung der individuellen Praxiserfahrungen der Fahrerlaubnisprüfer wurde etwa sechs Wochen nach Erprobungsbeginn durchgeführt. Hierzu wurde mit jedem Prüfer ein teilstandardisiertes Telefoninterview durchgeführt. Dabei wurden Fragen zu den Inhalten der optimierten Praktischen Fahrerlaubnisprüfung (v. a. zur Vollständigkeit der Bewertungskriterien) sowie zur Gebrauchstauglichkeit (z. B. zum Layout) des e-Prüfprotokolls und der eingesetzten Hardware gestellt. Im Zuge dessen hatten die Fahrerlaubnisprüfer die Gelegenheit, ihre Ergänzungs- bzw. Änderungswünsche vorzu-

tragen. Diese werden nach Abschluss des Revisionsprojekts bei der Vorbereitung des Produktsystems für den flächendeckenden Einsatz der optimierten Praktischen Fahrerlaubnisprüfung umgesetzt.

Nach Abschluss der Erprobung fand bei jeder technischen Prüfstelle ein Erfahrungsaustausch mit den an der Erprobung teilnehmenden Fahrerlaubnisprüfern statt. Dabei kam – in Anlehnung an die Telefoninterviews – ein standardisierter Fragebogen mit vier Inhaltsbereichen zum Einsatz:

- Zufriedenheit mit der Bedienbarkeit und dem Layout des e-Prüfprotokolls (z. B. mit der Erkennbarkeit und Bedienbarkeit von Text- und Schaltflächen),
- Zufriedenheit mit den Inhalten des e-Prüfprotokolls (z. B. mit den Dokumentationsmöglichkeiten zu den Prüfungsleistungen in den einzelnen, Prüfungsteilen),
- Gesamtzufriedenheit mit der Software,
- Zufriedenheit mit der Hardware (z. B. mit den Reaktionszeiten, den Akkulaufzeiten, und der Displayspiegelung).

Die Zufriedenheitsfragen wurden auf einer vierstufigen Skala beantwortet; dabei konnte der Fahrerlaubnisprüfer zwischen „Sehr zufrieden“, „Eher zufrieden“, „Eher unzufrieden“ und „Sehr unzufrieden“ wählen.

Die Befragungsergebnisse zeigen, dass die Erfahrungen der Fahrerlaubnisprüfer mit der optimierten Praktischen Fahrerlaubnisprüfung in den einzelnen technischen Prüfstellen ähnlich sind. Neben den Vorteilen – beispielsweise hinsichtlich der Unterstützung der Prüfungsdurchführung durch das e-Prüfprotokoll – wurden auch Empfehlungen zur Überarbeitung der Prüfungsinhalte und des Rückmeldesystems sowie hinsichtlich des e-Prüfprotokolls (Programmfunktionalitäten) gegeben.

In Bezug auf die Prüfungsinhalte merkten die Fahrerlaubnisprüfer an, dass vereinzelt konkrete Verkehrssituationen nicht im Fahraufgabenkatalog und damit auch nicht im e-Prüfprotokoll hinterlegt waren. Dies betrifft letztlich auch die im Fahraufgabenkatalog enthaltenen Bewertungskriterien (Beschreibungen der leichten und schweren Fehler sowie der überdurchschnittlichen Verhaltensweisen), welche nach Abschluss des Revisionsprojektes geprüft und gegebenenfalls ergänzt



werden sollten.<sup>22</sup> Betont soll an dieser Stelle werden, dass es sich mit Blick auf den Umfang des Fahraufgabenkataloges um eine äußerst geringe Anzahl an bisher nicht berücksichtigten Bewertungskriterien handelt, welche gleichwohl im Einzelfall die Zufriedenheit des Fahrerlaubnisprüfers mit dem e-Prüfprotokoll schmälern. Obwohl im e-Prüfprotokoll die Möglichkeit besteht, individuelle Bemerkungen zu dokumentieren, ist es Anspruch und Ziel sowohl der Entwickler als auch der Anwender, dass sich jedes beobachtete Bewerberverhalten in den im Fahraufgabenkatalog definierten und damit im e-Prüfprotokoll hinterlegten Bewertungskriterien wiederfindet. Diesem Ziel kann man sich anhand der Anregungen der an der Erprobung teilgenommenen Fahrerlaubnisprüfer weitgehend nähern; zu erreichen ist es erst im Rahmen einer bundesweiten kontinuierlichen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung.<sup>23</sup>

Die Fahrerlaubnisprüfer gaben darüber hinaus wichtige Hinweise zur Optimierung der Software des e-Prüfprotokolls. Beispielsweise wurden eine bessere Erkennbarkeit bestimmter Texte (Schriftgröße) und eine Vereinfachung einiger Eingabefunktionen zur schnelleren Dokumentation bestimmter Prüfungsleistungen (z. B. Grundfahraufgaben der Zweiradklassen) gewünscht. Zudem wurden Empfehlungen hinsichtlich der Einbindung des e-Prüfprotokolls in die bestehenden Verwaltungssysteme der technischen Prüfstellen gegeben. Schließlich unterbreiteten die Fahrerlaubnisprüfer auch Vorschläge zur Überarbeitung der schriftlichen Rückmeldung, insbesondere in Bezug auf eine Vereinfachung der Sprache und die Ausgestaltung der Empfehlungen zum weiteren Fahrkompetenzerwerb.

Sowohl die inhaltlichen als auch die auf die Software abzielenden Optimierungshinweise der Fahrerlaubnisprüfer werden nach Abschluss des Revisionsprojekts zur empirisch gestützten Überarbeitung des Fahraufgabenkataloges, des e-Prüfprotokolls und des Rückmeldesystems genutzt. Damit sind die Erfahrungen aus der Erprobung ein wichti-

ger Ausgangspunkt für die Vorbereitung einer flächendeckenden Implementierung der optimierten Praktischen Fahrerlaubnisprüfung (s. Kapitel 6), sofern sie vom Gesetzgeber beschlossen wird.

#### 4.6 Bewerberbefragung zur populationspezifischen Äquivalenz

Wie bereits dargelegt, soll während der Felderprobung die populationspezifische Äquivalenz der optimierten Praktischen Fahrerlaubnisprüfung analysiert werden. Grundsätzlich gesehen, geht es dabei darum, Zusammenhänge zwischen bestimmten Merkmalen und den Prüfungsergebnissen der Bewerber aufzudecken (s. Kapitel 3.4). Daher müssen für die Analyse der populationspezifischen Äquivalenz die Merkmale und Prüfungsergebnisse der Bewerber in einem Datensatz verknüpft vorliegen. Bei der Felderprobung könnte dies am effizientesten erreicht werden, wenn die Bewerbermerkmale jeweils im Anschluss an die einzelnen Prüfungen vom Fahrerlaubnisprüfer im Rahmen eines kurzen Interviews erhoben werden, dessen Ergebnisse auf dem Prüfcomputer dokumentiert werden – selbstverständlich unter Einhaltung der geltenden Datenschutzbestimmungen.

Welche Vorteile würden der Befragungszeitpunkt und die Befragungsart bieten? Würden die Daten bereits im Vorfeld der Prüfung erhoben, müsste davon ausgegangen werden, dass die Bewerber sozial erwünschte Antworten geben. So könnten die Bewerber annehmen, dass es Antworten gibt (z. B. bezüglich der Prüfungsangst), die ihre Prüfungschancen begünstigen oder schmälern. Die Folge wären verzerrte Daten. Daher sollten die Fragen erst nach Bekanntgabe der Prüfungsentcheidung an den Bewerber gestellt werden. Dies könnte im direkten Anschluss an die Rückmeldung geschehen oder im Nachgang zur Prüfung (z. B. durch eine postalische, eine telefonische oder eine online-basierte Befragung). Im letztgenannten Fall stünde allerdings zu befürchten, dass die Bewerber bereits das Interesse an der Prüfung verloren haben und die Teilnahme an der Befragung verstärkt ablehnen. Die Folge wären niedrige Ausschöpfungsquoten und eine weniger repräsentative Stichprobe durch Selbstselektion. Außerdem wäre eine Verknüpfung mit den Prüfungsergebnissen technisch erschwert sowie mit nicht unerheblichem Mehraufwand und Fehlerpotenzial verbunden. Würde die Befragung hingegen im direkten

<sup>22</sup> Der Fahraufgabenkatalog wird durch die AG Fahraufgabenbeschreibung überarbeitet, welche schon für die bisherige Erarbeitung verantwortlich ist.

<sup>23</sup> Vergleichend sei hier auf die kontinuierliche Evaluation und Optimierung der Inhalte der theoretischen Fahrerlaubnisprüfung verwiesen, welche nach Einführung des optimierten Prüfungssystems (PC) begann.

Anschluss an die Rückmeldung an den Bewerber anhand des Prüfungscomputers des Fahrerlaubnisprüfers erfolgen, könnte die Verknüpfung der Daten automatisch und fehlerfrei vorgenommen werden.

Eine Befragung durch den Fahrerlaubnisprüfer würde noch weitere Vorteile mit sich bringen: Erstens ist der Fahrerlaubnisprüfer bereits mit den Fragehintergründen vertraut, sodass Verständnis- bzw. Interpretationsschwierigkeiten minimiert werden könnten; dies würde zur Fehlerfreiheit und Vergleichbarkeit der Daten beitragen. Zweitens würde das persönliche Interview den Bewerber – anders als bei einer schriftlichen Befragung – nur zwei bis drei Minuten seiner Zeit kosten, da keine Mühe auf das Öffnen, Lesen und Versenden des Schreibens verwendet werden muss; dies wiederum würde die Teilnahmebereitschaft erhöhen. Drittens schließlich ist der Bewerber durch den persönlichen Kontakt zum Fahrerlaubnisprüfer vermutlich eher teilnahmebereit, denn eine persönliche Bitte wird seltener ignoriert oder abgelehnt als ein Schreiben oder ein Anruf. Daher haben persönliche Interviews generell eine höhere Ausschöpfungsquote als schriftliche Befragungen (BORTZ & DÖRING, 2006). Eine höhere Ausschöpfungsquote bedeutet gewöhnlich auch eine geringere Selbstselektion der Befragten und in der Folge repräsentativere Daten.

Aus den genannten Gründen wurden die Bewerbermerkmale im vorliegenden Projekt mittels eines Interviews erhoben, das durch den Fahrerlaubnisprüfer im direkten Anschluss an die Bewerberrückmeldung mithilfe seines Prüfcomputers durchgeführt wird. Hierfür erhielten die Fahrerlaubnisprüfer

vorab eine Einweisung zu den Inhalten und zur korrekten Durchführung der Befragung. Weiterhin wurde – zur Erhöhung der Teilnahmebereitschaft – den Bewerbern der wissenschaftliche Hintergrund der Befragung erläutert; außerdem wurden die Einhaltung der geltenden Datenschutzbestimmungen sowie die Anonymität und Exklusivität (keine Weitergabe an Dritte) bei der Datenauswertung zugesagt. Schließlich wurde auf die Freiwilligkeit der Teilnahme hingewiesen und zugesichert, dass die Befragung in keinerlei Zusammenhang mit dem Prüfungsergebnis steht.

Bezüglich des Befragungsablaufs muss zwischen jenen Fragen unterschieden werden, die nur vom Bewerber beantwortet werden können (Selbsteinschätzungen), und solchen, die der Fahrerlaubnisprüfer selbst beantworten kann (Fremdeinschätzungen). Letztere wurden zum Abschluss vom Fahrerlaubnisprüfer allein bearbeitet, um die Verfahrensökonomie zu erhöhen. Bei den Fragen an den Bewerber wurden weniger sensible Merkmale wie die Fahrerfahrung zuerst erfragt; sensiblere Merkmale wie der Bildungshintergrund folgten danach. Durch das Stellen weniger sensibler Fragen zu Beginn sollte der Bewerber Vertrauen für die Beantwortung der sensibleren Fragen fassen. Die Tabelle 3 zeigt eine Übersicht über alle erhobenen Bewerbermerkmale und die Reihenfolge ihrer Erfassung.<sup>24</sup> Die inhaltliche Begründung für die Erhebung dieser Merkmale findet sich im Kapitel 3.4.

<sup>24</sup> Der gesamte Fragebogen mit allen Frage- und Antwortformulierungen, den Anweisungen des Interviewers sowie der Filterfragenlogik findet sich als Anlage 3.

Fragebereich	Personenbezogenes Merkmal des Bewerbers
Fragen an den Bewerber (Selbsteinschätzungen)	Vorbesitz Fahrerlaubnis
	Kilometer Fahrpraxis (bei Vorbesitz einer Fahrerlaubnis)
	Stunden praktischer Fahrausbildung (wenn kein Vorbesitz einer Fahrerlaubnis)
	Regelmäßiger Konsum straßenverkehrsbezogener Computerspiele
	Durchschnittlicher Konsum straßenverkehrsbezogener Computerspiele pro Woche (bei regelmäßigem Konsum straßenverkehrsbezogener Computerspiele)
	Angestrebter bzw. erreichter höchster Schulabschluss
	Ausmaß der Prüfungsangst
Fragen an den Fahrerlaubnisprüfer (Fremdeinschätzungen)	Migrationshintergrund
	Geschlecht
	Alter

Tab. 3: Übersicht der erhobenen Bewerbermerkmale

#### 4.7 Bewerberbefragung zur Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung

Die Gründe dafür, warum Kundenbefragungen eine Säule des Evaluationssystems zur optimierten Praktischen Fahrerlaubnisprüfung darstellen, wurden in Kapitel 3.7 dargelegt. Nachfolgend soll nun beschrieben werden, welche Fragen an die Fahrerlaubnisbewerber gestellt wurden und wie die Bewerberbefragung im Revisionsprojekt durchgeführt wurde.

Im Projektantrag zum Revisionsprojekt war ursprünglich vorgesehen, unter methodischen Gesichtspunkten die Durchführung einer bundesweiten Bewerberbefragung zur Zufriedenheit mit der Praktischen Fahrerlaubnisprüfung zu erproben. In Abstimmung mit der BAST wurden allerdings im Revisionsprojekt tatsächlich nur Bewerber befragt, die während der Felderprobung in den Modellregionen eine optimierte Praktische Fahrerlaubnisprüfung absolviert und auf dieser Grundlage eine schriftliche Rückmeldung zu ihren Prüfungsleistungen erhalten hatten. Der Grund für diese Änderung resultierte aus der Notwendigkeit, die Qualität der verkehrspädagogischen Gestaltung des Rückmeldesystems und seinen Nutzen zu untersuchen. Mit der Veränderung des Untersuchungsgegenstands der Bewerberbefragung wurde dem Erfordernis Rechnung getragen, dem Verordnungsgeber wissenschaftlich fundierte Entscheidungsgrundlagen für die Frage bereitzustellen, ob bei einer flächendeckenden Einführung des Rückmeldesystems den erhöhten Kosten ein adäquater Nutzen für die Verkehrssicherheit gegenübersteht.

Entgegen der Empfehlungen von STURZBECHER et al. (2014), die Zufriedenheit der Bewerber im Rahmen einer telefonischen Befragung zu erfassen, wurde die Befragung aus Kostengründen als Online-Befragung mit einem standardisierten Fragebogen durchgeführt. Mit dieser Befragungsmethode stehen die Daten ebenfalls – wie bei einer computerunterstützten telefonischen Befragung – direkt nach der Befragung zur Auswertung zur Verfügung. Eine ausführliche Abwägung der Vor- und Nachteile der zur Verfügung stehenden Befragungsmethoden hat mit Bezug zur Fahrerlaubnisprüfung bereits bei STURZBECHER und MÖRL (2008) stattgefunden.

Die Grundgesamtheit der Befragten setzt sich aus allen Fahrerlaubnisbewerbern zusammen, die im

Erprobungszeitraum in den Modellregionen eine optimierte Praktische Fahrerlaubnisprüfung absolviert haben. Die Bewerber wurden am Prüfungstag nach Abschluss der Prüfung vom Prüfer über den Zweck, die Inhalte und den datenschutzrechtlichen Hintergrund der Befragung informiert und erhielten dazu ein Informationsblatt. Auf diesem Blatt fand sich auch ein Online-Zugangscode mit dazugehöriger Internetadresse. Die Bewerber konnten den Fragebogen dann zu gegebener Zeit im Internet ausfüllen. Mit dieser Vorgehensweise wurde sichergestellt, dass die ausgewählten Bewerber im Evaluationszeitraum auch tatsächlich an einer optimierten Praktischen Fahrerlaubnisprüfung teilgenommen hatten.

Der Fragebogen enthält drei Inhaltsbereiche:

- Praktische Fahrerlaubnisprüfung und Prüferverhalten,
- mündliche und schriftliche Rückmeldung und
- sonstige und persönliche Angaben.

Im Mittelpunkt der Bewerberbefragung standen Fragen zur Zufriedenheit mit dem mündlichen Auswertungsgespräch durch den Prüfer einerseits und mit der schriftlichen Rückmeldung andererseits. Neben den Fragen zum Rückmeldesystem wurden aber auch Merkmale erfasst, welche die Zufriedenheit mit dem Rückmeldesystem ggf. beeinflussen könnten.

Grundsätzlich ist davon auszugehen, dass sich die Gesamtzufriedenheit mit einem Produkt oder einer Dienstleistung aus verschiedenen Inhaltsfacetten zusammensetzt, die sich meist gegenseitig beeinflussen. Um zu erfassen, welche Zufriedenheitsfacetten den größten Einfluss haben und welche ggf. miteinander im Zusammenhang stehen, müssen möglichst alle relevanten Zufriedenheitsaspekte erfasst werden. So ist zum Beispiel zu untersuchen, welchen Einfluss das Prüferverhalten – der Fahrerlaubnisprüfer stellt quasi das „Messinstrument“ der optimierten Praktischen Fahrerlaubnisprüfung dar – auf die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung im Allgemeinen und auf die Zufriedenheit mit dem Rückmeldesystem im Besonderen hat. Wird der Prüfer beispielsweise vom Bewerber als unsympathisch bzw. unfreundlich empfunden, könnte das unter Umständen andere Bewertungsaspekte der Zufriedenheit in anderen Bereichen überstrahlen. In diesem Fall spricht man von einem hervorstechenden Merkmal,

das den Betrachter scheinbar so „blendet“, dass er nicht mehr differenziert bewerten kann („Halo-Effekt“). Darüber hinaus ist bekannt, dass der stärkste Faktor zur Erklärung der Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung durch Einschätzungen gebildet wird, die sich auf das Verhalten des Fahrerlaubnisprüfers beziehen (STURZBECHER & MÖRL, 2008). Für die Interpretation der Befragungsergebnisse und um sichere Aussagen zur Zufriedenheit mit dem Rückmeldesystem der optimierten Praktischen Fahrerlaubnisprüfung treffen zu können, war es daher dringend geboten, auch die Zufriedenheit mit dem Prüferverhalten zu erfassen.

Jeweils am Ende des entsprechenden Inhaltsbereiches wird gefragt, wie zufrieden der Bewerber insgesamt mit dem Prüfer, mit dem mündlichen Auswertungsgespräch und mit der schriftlichen Rückmeldung war. Auf diese Weise, also nach der Einschätzung der einzelnen Zufriedenheitsfacetten, wird ein reflektiertes, abgewogenes Urteil und kein „Bauchgefühl“ erfasst (STURZBECHER & MÖRL, 2008). Die Erhebung der Gesamtzufriedenheit der Bewerber mit einzelnen Bereichen ermöglicht es zudem, bei der Auswertung der Befragungsergebnisse die einzelnen Bereiche entsprechend ihrer relativen Bedeutsamkeit für die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung zu gewichten.

Für die Beantwortung der Fragen wurde ein vierstufiges Antwortmuster mit den Antwortvorgaben „Sehr zufrieden“, „Eher zufrieden“, „Eher unzufrieden“ oder „Sehr unzufrieden“ gewählt. Somit existiert keine „neutrale“ Mittelkategorie. Dies bietet den Vorteil, dass sich der Befragte bei seiner Zufriedenheitseinschätzung für eine Richtung entscheiden muss und die Antworten eine höhere Aussagekraft erhalten.

Sowohl zum schriftlichen Prüfprotokoll als auch zur Praktischen Fahrerlaubnisprüfung insgesamt wurde darüber hinaus jeweils eine offene Frage zu Verbesserungsvorschlägen gestellt. Die Verwendung derartiger Fragen bringt erstens Respekt vor dem Befragten zum Ausdruck und erhöht somit seine Akzeptanz für die Befragung. Zweitens erlauben es offene Fragen dem Befragten, bereits gegebene Rückmeldungen mit zusätzlichen Hinweisen zu detaillieren. Aus diesem Grund besitzen diese Fragen – trotz des damit verbundenen hohen Erhebungs- und Auswertungsaufwands – eine hohe Bedeutung für die Weiterentwicklung

der Praktischen Fahrerlaubnisprüfung. Die Items zur Zufriedenheitsmessung wurden schließlich mit Fragen zu Alter, Geschlecht und Schulabschluss der Bewerber ergänzt. Mithilfe dieser Daten kann eine nach sozialen Teilgruppen differenzierte Auswertung der Bewerberbefragung und eine Abschätzung von (unerwünschten) Einflussfaktoren auf die Durchführungsqualität der Praktischen Fahrerlaubnisprüfung vorgenommen werden. Der vollständige Fragebogen der Bewerberbefragung findet sich als Anlage 4 zum vorliegenden Bericht.

#### 4.8 Fahrlehrerbefragung

Neben den Fahrerlaubnisbewerbern sollen künftig auch die Fahrlehrer als eine weitere Kundengruppe der technischen Prüfstellen (s. o.) in einem Fünfjahresturnus bundesweit zu ihrer Zufriedenheit mit der Praktischen Fahrerlaubnisprüfung befragt werden (STURZBECHER et al., 2014). Die exemplarische Erprobung der Inhalte und methodischen Abläufe einer solchen Fahrlehrerbefragung war Bestandteil des vorliegenden. Nachfolgend werden die Untersuchungsanlage und die Befragungsinhalte erläutert.

Im Vorfeld der Befragung wurde von Vertretern der Fahrlehrerschaft darauf hingewiesen, dass in einigen Regionen eine substanzielle Zahl von Fahrschulen nicht über einen Internetzugang verfügt. Um diesem Umstand Rechnung zu tragen, wurden für die Rekrutierung unterschiedliche Anspracheformen gewählt bzw. zur Datenerhebung verschiedene schriftliche Befragungsverfahren (Online-Befragung, postalische Befragung) eingesetzt.

In Abhängigkeit von der Modellregion und der zuständigen technischen Prüfstelle wurden die Fahrschulen per Post oder per Email angeschrieben. Dieses Anschreiben enthielt die notwendigen Informationen zum Projekt und zur Befragung sowie einen Zugangscode für die Onlinebefragung mit dem entsprechenden Internetlink (TÜV NORD, TÜV Rheinland, TÜV SÜD) bzw. einen schriftlichen Fragebogen nebst frankierten Rückumschlag (DEKRA). Darüber hinaus wurden die Fahrschulinhaber über die Maßnahmen zur Einhaltung der geltenden Datenschutzbestimmungen informiert. Bei der Auswertung der Befragung war eine Rückverfolgung bzw. Verknüpfung der bewertenden Fahrschulen bzw. Fahrlehrer und der bewerteten Prüfer grundsätzlich ausgeschlossen.



Bei der inhaltlichen und methodischen Gestaltung der Indikatoren wurde weitgehend auf das im Jahr 2008 von den technischen Prüfstellen und Vertretern der Fahrlehrerschaft gemeinsam entwickelte System zur Kundenbefragung zurückgegriffen. Dieses Kundenbefragungssystem zur Zufriedenheit mit den technischen Prüfstellen im Bereich der Praktischen Fahrerlaubnisprüfung enthält u. a. einen Fragebogen für Fahrlehrer (STURZBECHER & MÖRL, 2008). Dieser Fragebogen diente als Ausgangspunkt für die Erarbeitung des Fragenkatalogs für die vorliegende Untersuchung; Aufbau und Layout wurden übernommen. Die Fragen wurden aufgrund unterschiedlicher Auftraggeber und Zielsetzungen für das Revisionsprojekt entsprechend angepasst. Notwendige Anpassungen wurden bereits im Bericht „Optimierung der Praktischen Fahrerlaubnisprüfung“ (STURZBECHER et al., 2014) beschrieben.

Weitere Änderungen wurden entsprechend des Untersuchungsgegenstands im Rahmen des Revisionsprojekts vorgenommen. Der verwendete Fragebogen bestand schließlich aus 15 Fragen und untergliederte sich in vier Inhaltsbereiche:<sup>25</sup>

- Fragen zu organisatorischen Rahmenbedingungen („Prüfauftragsverwaltung“, „Terminmanagement“; bei der Einschätzung dieser Inhalte sollten sich die Befragten auf die Zusammenarbeit mit ihrer technischen Prüfstelle in den letzten zwölf Monaten beziehen),
- Fragen zur Durchführung der Praktischen Fahrerlaubnisprüfung und zum Prüferverhalten (bei diesen Einschätzungen wurden die Befragten gebeten, an ihre letzte Praktische Fahrerlaubnisprüfung zu denken),
- Erwartungen im Hinblick auf eine optimierte Praktische Fahrerlaubnisprüfung (dabei standen die Wünsche der Befragten im Falle einer Einführung der optimierten Prüfung im Vordergrund) und
- sonstige und persönliche Angaben (hier wurden Fragen zum Berufsstatus – Inhaber vs. an-

gestellter Leiter – und zur Fahrschulgröße sowie zu Alter und Geschlecht der Befragten gestellt).

Inhaltlich und methodisch ähnelt der Fragebogen für die Fahrschulen weitgehend dem Bewerberfragebogen. Die Zufriedenheitsaspekte wurden ebenfalls mit einem vierstufigen Antwortmuster („Sehr zufrieden“, „Eher zufrieden“, „Eher unzufrieden“, „Sehr unzufrieden“) erhoben. Am Ende der einschlägigen Inhaltsbereiche wurde jeweils eine Frage nach der Gesamtzufriedenheit platziert, um ein reflektiertes Zufriedenheitsurteil des Befragten zu erfassen. Darüber hinaus wurden auch offene Fragen zu Verbesserungsvorschlägen bzw. Wünschen der Befragten gestellt. Der vollständige Fragebogen findet sich in der Anlage 5 des vorliegenden Berichts.

## 5 Ergebnisse

### 5.1 Objektivität und Reliabilität

#### 5.1.1 Überblick

Bei der Bestimmung der methodischen Güte eines Beobachtungsverfahrens lässt sich kaum zwischen Objektivitäts- und Reliabilitätsaspekten unterscheiden (HASEMANN, 1983; INGENKAMP & LISSMANN, 2008; MEES, 1977). Eine getrennte Darstellung der Untersuchungsergebnisse zur Objektivität und zur Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung erscheint daher nicht sinnvoll.

Bei Beobachtungsverfahren stellt die Beobachterübereinstimmung sowohl für die Objektivität als auch für die Reliabilität die bedeutendste Gütekenngroße dar. Die Beobachterübereinstimmung ist daher als wichtigster Untersuchungsgegenstand im Rahmen der instrumentellen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung anzusehen. Zur Ermittlung der Beobachterübereinstimmung wurden Untersuchungen im Realverkehr (s. Kapitel 4.1.1) und anhand videografiertes Prüfungen (s. Kapitel 4.1.2) durchgeführt, deren Ergebnisse im Folgenden dargestellt werden (s. Kapitel 5.1.2). Anhand der videografierten Prüfungen wurde darüber hinaus die Bewertungsstabilität bzw. Retest-Reliabilität berechnet, die ein weiterer Indikator für die Reliabilität der Prüfung ist. Die Ergebnisse dieser Analysen werden in Kapitel 5.1.3 vorgestellt.

<sup>25</sup> Beim TÜV NORD waren es nur 12 Fragen in drei Inhaltsbereichen. Wie bereits dargelegt, wurde der beim TÜV NORD verwendete Fragenkatalog später um drei Fragen bzw. einen Inhaltsbereich zur optimierten Praktischen Fahrerlaubnisprüfung ergänzt.

### 5.1.2 Überprüfung der Objektivität und Reliabilität anhand der Beobachterübereinstimmung

STURZBECHER et al. (2014) haben hergeleitet und begründet, warum 1. für die Prüfungsentscheidungen, 2. für die kompetenzbezogenen Gesamtbewertungen der Fahraufgaben und Beobachtungskategorien und 3. für die ereignisbezogenen Bewertungen (Fehler und überdurchschnittliche Leistungen bei der Durchführung von Fahraufgaben) verschiedene Mindestwerte der Übereinstimmung gefordert werden sollten. Im Folgenden werden daher die Ergebnisse zur Beobachterübereinstimmung aus den verschiedenen Untersuchungen zunächst für die Prüfungsentscheidungen, darauf folgend für die Gesamtbewertungen der Fahraufgaben und Beobachtungskategorien sowie abschließend für die ereignisbezogenen Bewertungen der Fahraufgaben vorgestellt und diskutiert.

#### Prüfungsentscheidung

Die Prüfungsentscheidung ist maßgeblich für die Zulassung eines Fahrerlaubnisbewerbers zur zukünftigen Teilnahme am motorisierten Straßenverkehr. Sie beinhaltet ein Gesamturteil über alle Facetten der Fahrkompetenz und stellt für den Bewerber die wichtigste Ebene der Prüfungsbewertung dar. Die Prüfungsentscheidung ist eine dichotome Variable mit den Ausprägungsmöglichkeiten „Bestanden“ und „Nicht bestanden“. Für die Quantifizierung der Übereinstimmung von zwei Beobachtern auf einer dichotomen Variable ist Cohen's Kappa das meistgebrauchte Maß (WIRTZ & CASPAR, 2002). Im Gegensatz zur intuitiven Berechnung einer prozentualen Übereinstimmung korrigiert Kappa die prozentuale Übereinstimmung um die zufällig erwartete Übereinstimmung zwischen den Beobachtern. Die zufällig erwartete Übereinstimmung wird anhand der Kontingenztafel der Beurteilungen berechnet, wobei die Randsum-

men der Beurteilungskategorien multipliziert werden (BORTZ, 1999). Cohen's Kappa kann Ausprägungen von „-1“ bis „+1“ annehmen. „-1“ bedeutet, dass absolut keine Übereinstimmung zwischen den Beobachtern vorliegt; „+1“ bedeutet dagegen eine Übereinstimmung der beiden Beobachter in allen Fällen („perfekte“ Übereinstimmung).

Wie der Tabelle 4 entnommen werden kann, erzielten die Beobachter (Fahrerlaubnisprüfer) beim Treffen der Prüfungsentscheidungen bei allen drei Untersuchungen im Realverkehr perfekte Übereinstimmungen. Damit haben sie den von STURZBECHER et al. (2014) geforderten Mindestwert von 0,90 deutlich übertroffen.

Bei der Berechnung von Cohen's Kappa erfolgt ein Vergleich der Datenreihen von zwei Beobachtern. Für die Analysen zur Beobachterübereinstimmung anhand videografiert Prüfungen ist dieser Koeffizient ungeeignet, da hierbei die Datenreihen von bis zu 38 Fahrerlaubnisprüfern miteinander verglichen werden müssen. FLEISS (1971) hat für derartige Fälle (also für die Berechnung der Übereinstimmung der Bewertungen von mehr als zwei Beobachtern) die Berechnungsverfahren von Cohen's Kappa erweitert und dadurch die Berechnung von Übereinstimmungswerten für eine beliebig große Anzahl von Beobachtern ermöglicht. Die Berechnung von Fleiss' Kappa kann auch bei nominalskalierten Daten wie den Daten zu den Prüfungsentscheidungen vorgenommen werden.

Bei der Beurteilung der Prüfungsvideos waren die Fahrerlaubnisprüfer unter anderem auch aufgefordert, Gesamtbewertungen zu den Fahraufgaben und Kompetenzbereichen (Beobachtungskategorien) auf einer vierstufigen ordinalskalierten Skala<sup>26</sup>

<sup>26</sup> Die Skalenstufen lauteten „Sehr gut“, „Gut“, „Ausreichend“ und „Ungenügend“.

	Cohen's Kappa	n	p	Abweichungen
Simulierte Prüfungen	1	15	0,00***	0
Reale Prüfungen der Fahrerlaubnisklasse B	1	20	0,00***	0
Reale Prüfungen verschiedener Fahrerlaubnisklassen	1	20	0,00***	0

\*\*\* signifikant auf dem 0,1-Prozent-Niveau;  
 \*\* signifikant auf dem 1-Prozent-Niveau;  
 \* signifikant auf dem 5-Prozent-Niveau; keine Sternchen: Koeffizient ist nicht signifikant

Tab. 4: Übereinstimmung zwischen den Fahrerlaubnisprüfern beim Treffen der Prüfungsentscheidungen – Prüfungen im Realverkehr

vorzunehmen. Würde man die Übereinstimmungen zwischen den Prüfern nun anhand von Fleiss' Kappa berechnen, würden die in den ordinalskalierten Werten verborgenen Ranginformationen nicht berücksichtigt werden, d. h. bei der Berechnung des Koeffizienten würde nicht unterschieden werden, ob die Bewertungen der Fahrerlaubnisprüfer leicht oder stark voneinander abweichen. Eine solche Unterscheidung wäre aber wünschenswert, da eine Abweichung zwischen „Sehr gut“ und „Gut“ die Objektivität und Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung ggf. weniger stark infrage stellen würde als eine Abweichung zwischen „Sehr gut“ und „Ungenügend“.

Ein weiteres Problem bei der Anwendung von Fleiss' Kappa stellt der Umgang mit fehlenden Werten dar. Sobald einer der Beobachter eine der Fahraufgaben oder Beobachtungskategorien nicht bewertet hat – z. B. weil er der Ansicht war, diese sei in der Prüfung nicht vorgekommen – kann der Koeffizient für diese Fahraufgabe oder Beobachtungskategorie nicht mehr berechnet werden („listenweiser Ausschluss“). In der vorliegenden Stichprobe kam dies bei den 38 Beobachtern relativ häufig vor, sodass vielfach mittels Fleiss' Kappa keine Übereinstimmungswerte berechnet werden konnten.

KRIPPENDORFF (2011) hat dagegen einen Übereinstimmungskoeffizienten entwickelt, der es ermöglicht, eine beliebig große Anzahl von Beobachtern einzubeziehen und bei der Berechnung auch Ranginformationen zu berücksichtigen. Darüber hinaus werden ggf. fehlende Werte nicht listenweise ausgeschlossen, sondern durch existierende Werte ersetzt. Dadurch lassen sich mit Krippendorff's Alpha für alle Fahraufgaben und Beobachtungskategorien bzw. Kompetenzbereiche Übereinstimmungswerte berechnen.

Als Nachteil von Krippendorff's Alpha ist der Umstand anzusehen, dass eine Berechnung der Konfidenzintervalle und damit der statistischen Signifikanz nicht möglich ist. Generell deuten aber Probeberechnungen mit Fleiss' Kappa darauf hin, dass die Zahl der Objekte und Beobachter ausreichend groß für die Bestimmung signifikanter Übereinstimmungskoeffizienten ist, wenn die Berechnungen für alle drei Filme gleichzeitig und nicht pro einzeltem Film durchgeführt werden. Krippendorff's Alpha ist daher im vorliegenden Fall insgesamt gesehen das am besten geeignete Maß für die Analyse der Beobachterübereinstimmung bei der Bewertung

videografierter Prüfungen durch eine große Zahl von Fahrerlaubnisprüfern.

Genauso wie das oben beschriebene Cohen's Kappa kann Krippendorff's Alpha Ausprägungen von „-1“ bis „+1“ annehmen. „-1“ bedeutet wiederum, dass absolut keine Übereinstimmung zwischen den Beobachtern vorliegt; „+1“ bedeutet eine perfekte Übereinstimmung. Ein Nachteil der Übereinstimmungskoeffizienten für mehr als zwei Beobachter besteht darin, dass die absolute Höhe der Ergebnisse von der Zahl der Beobachter und der Zahl der bewerteten Objekte abhängig ist. Eine Bewertung der berechneten Koeffizienten anhand der von STURZBECHER et al. (2014) begründeten Grenzwerte (s. o.) ist daher im Falle der videografierten Prüfungen nicht zielführend. Ein relativer Vergleich der Ergebnisse zwischen den Messzeitpunkten wäre hingegen sinnvoll, wenn die Rahmenbedingungen zwischen den drei Messzeitpunkten ähnlich waren. Durch den Vergleich der Messreihen lassen sich Lernverläufe hinsichtlich der objektiven und reliablen Durchführung der optimierten Praktischen Fahrerlaubnisprüfung durch die Fahrerlaubnisprüfer untersuchen.

Die in Tabelle 5 aufgeführten Ergebnisse zeigen, dass bei der Bewertung videografierteter Prüfungen im Gegensatz zur Bewertung der Prüfungen im Realverkehr zu keinem der Messzeitpunkte eine perfekte Übereinstimmung erzielt wurde. Allerdings ist in der zweiten Spalte der Tabelle zu erkennen, dass die Übereinstimmungswerte zwischen dem ersten und dem letzten Messzeitpunkt deutlich gestiegen sind. Die dritte Spalte (n) der Tabelle gibt die Zahlen der Fahrerlaubnisprüfer<sup>27</sup> an, die zu den jeweiligen Messzeitpunkten die jeweils drei Filme bewertet haben.

Zum Anstieg der Übereinstimmungswerte über den Erprobungszeitraum dürften verschiedene Gründe beigetragen haben:

- Erstens waren die an der Untersuchung teilnehmenden Fahrerlaubnisprüfer zum ersten Messzeitpunkt im Umgang mit der optimierten Prak-

<sup>27</sup> Aufgrund krankheitsbedingter Ausfälle und organisatorischer Gründe auf Seiten der technischen Prüfstellen konnten nicht alle Fahrerlaubnisprüfer zu allen Messzeitpunkten an der Untersuchung teilnehmen. Die Fallzahlen – 3 Videos bewertet durch 28 bzw. 31 bzw. 38 Prüfer – liegen aber nahe genug beisammen, um die Ergebnisse der verschiedenen Messzeitpunkte sinnvoll vergleichen zu können.

	Krippendorff's Alpha	n
Videografierte Prüfungen – 1. Messzeitpunkt	0,45	38
Videografierte Prüfungen – 2. Messzeitpunkt	0,48	28
Videografierte Prüfungen – 3. Messzeitpunkt	0,74	31

Tab. 5: Übereinstimmung zwischen den Fahrerlaubnisprüfern beim Treffen der Prüfungsentscheidung – Videografierte Prüfungen

tischen Fahrerlaubnisprüfung und dem e-Prüfprotokoll wahrscheinlich noch nicht ausreichend geübt.<sup>28</sup> Zwar hatten alle Prüfer am Tag vor der ersten Messung dazu eine ganztägige Schulung erhalten, aber die nötige Zeit zur Aneignung einer automatisierten Anwendung der Dokumentierungs-, Bewertungs- und Entscheidungsregeln mittels e-Prüfprotokoll war dabei nicht gegeben. Vereinfacht ausgedrückt, waren die Prüfer noch zu sehr mit der Handhabung der veränderten Standards und der technischen Neuerungen beschäftigt, um sich voll auf das videografierte Verkehrsgeschehen konzentrieren zu können.

- Zweitens muss berücksichtigt werden, dass sich die Dokumentierung und Bewertung videografierte Prüfungen – trotz aller Bestrebungen zur Maximierung der ökologischen Validität der Prüfungsvideos (s. Kapitel 4.3) – von der Dokumentierung und Bewertung realer Prüfungen unterscheidet. Auch hier mussten die Prüfer einen

Anpassungsprozess vollziehen, der zum ersten Messzeitpunkt – trotz einer Übungsphase mit Probefilmen – am wenigsten fortgeschritten war.

- Drittens schließlich waren die Fahrerlaubnisprüfer bei der Beobachtung der Prüfungsvideos – im Gegensatz zu realen Prüfungen in ihrem jeweiligen Prüfort – nicht mit dem Prüfort und den (möglichen) Streckenführungen vertraut; sie konnten die Prüfungsanforderungen bzw. die zu bewältigenden Fahraufgaben also vor allem bei der ersten Betrachtung der Filme nicht voraussehen.<sup>29</sup> Für die Fahrerlaubnisbewerber typischerweise schwierig zu bewältigende Verkehrssituationen bzw. für die Prüfungsbewertung besonders relevante Situationen wurden damit wahrscheinlich leichter übersehen als in realen Prüfungen. Dies gilt insbesondere vor dem Hintergrund der beiden zuvor beschriebenen Punkte bzw. „Störgrößen“ bei der Prüfungsdurchführung: Ablenkung vom Verkehrsgeschehen durch unzureichende Routine bei der Handhabung des e-Prüfprotokolls und mangelnde Erfahrung bei der Betrachtung von Prüfungsvideos.

<sup>28</sup> Die an der Untersuchung teilnehmenden Fahrerlaubnisprüfer haben gleichzeitig an der Felderprobung (s. Kapitel 4.4) der optimierten Praktischen Fahrerlaubnisprüfung teilgenommen und dabei die praktische Durchführung der Prüfung entsprechend der neuen Standards trainiert. Die erste Beurteilung der Prüfungsvideos fand kurz vor Beginn der Felderprobung statt; die zweite Beurteilung erfolgte zur Mitte und die dritte Beurteilung zum Ende der viereinhalbmönatigen Felderprobung.

<sup>29</sup> So weist bereits HAMPEL (1977) darauf hin, dass jeder Fahrerlaubnisprüfer bewusst oder unbewusst über einen Vorrat an „persönlichen“ Prüfstrecken verfügt, die er nach seiner (unsystematischen) Erfahrungsbildung für die Prüfungsdurchführung geeignet hält. STURZBECHER, BÖNINGER und RÜDEL (2008 bzw. 2010, S. 94) greifen diesen Gedanken unter dem Stichwort „Streckenprojektierung“ auf und weisen auf den entlastenden Charakter dieser vorausschauenden Streckenplanung hin: „Der Prüfer ist gut beraten, in seinem regionalen Prüfungsbereich Teilstrecken zu projektieren und sich zu merken, auf denen er Möglichkeiten zum Prüfen bestimmter Fahraufgaben mit hoher Wahrscheinlichkeit vorfindet und die er kombinieren kann; diesbezügliches Streckenwissen vermittelt ihm Sicherheit und entlastet ihn bei der Prüfungsfahrt von aufwändigen Planungsleistungen. Daher gehen die Prüfer – damals wie heute – auch bei der Gestaltung der Prüfungsfahrt auf diese Weise vor.“

Die fehlende Übung im Umgang mit der optimierten Praktischen Fahrerlaubnisprüfung und dem gewählten Videografierungsformat sowie unzureichende Kenntnisse über den Prüfort und die Streckenführung dürften die Präzision der Prüfer beim Erfassen und Bewerten entscheidungsrelevanter Prüfungsleistungen beeinträchtigt und die Beobachterübereinstimmung zwischen den Fahrerlaubnisprüfern verringert haben. Mit zunehmender Gewöhnung an das veränderte Regelwerk, das technische Instrumentarium, das gewählte Videografierungsformat und die aufgezeichnete Streckenführung wurden diese Störquellen aber reduziert, woraus sich die über die Messzeitpunkte steigenden Übereinstimmungswerte wahrscheinlich ergeben haben. Dies ist ein gewünschter Effekt, der auf die Güte der optimierten Praktischen Fahrerlaubnisprüfung hindeutet. Der Umstand, dass sich die Steigerung der Beobachterübereinstimmung aber erst zwischen dem zweiten und dritten



Messzeitpunkt deutlich zeigt, weist darauf hin, dass umfassende zeitintensive Beobachterschulungen und Lernprozesse notwendig sind, um eine objektive und reliable Anwendung der optimierten Praktischen Fahrerlaubnisprüfung zu gewährleisten. Dies erscheint vor dem Hintergrund der nicht unerheblichen Änderungen bei der Restrukturierung der Anforderungs- und Bewertungsstandards (s. Kapitel 2.1) sowie des Dokumentationsinstruments (s. Kapitel 2.2) plausibel.

### Gesamtbewertung der Fahraufgaben

Die situationsübergreifenden Gesamtbewertungen der Fahraufgaben für den gesamten Prüfungsverlauf wurden anhand einer vierstufigen Ordinalskala mit den Ausprägungen „Sehr gut“, „Gut“, „Ausreichend“ und „Ungenügend“ vorgenommen. Zur Berechnung des Zusammenhangs zweier Ordinalskalen kann zwischen justierten und unjustierten Maßen unterschieden werden (WIRTZ & CASPAR, 2002): Während unjustierte Maße die Übereinstimmung absoluter Werte widerspiegeln, beziehen sich justierte Maße<sup>30</sup> auf ihre relative Ordnung. Bei unjustierten Maßen steht also die tatsächliche Übereinstimmung der beiden Beobachter im Vordergrund; mittels justierter Maße wird lediglich die Distanz zum jeweiligen Mittelwert des Beobachters dargestellt. Bewerten beide Beobachter den Beobachtungsgegenstand höher als in ihrem jeweiligen Mittel, wirkt sich das bereits positiv auf einen justierten Koeffizienten aus, ohne dass die beiden Beobachter tatsächlich übereinstimmen müssen. Dies kann im Rahmen der Bewertungen von Fahrleistungen für den Nachweis von Fahrkompetenz als Voraussetzung für die Erteilung der Fahrerlaubnis nicht als ausreichend betrachtet werden. Hierbei sind tatsächliche Übereinstimmungen zu fordern, die durch unjustierte Maße wie Cohen's Kappa quantifiziert werden. Daher wird auch für die situationsübergreifenden Bewertungen der Fahraufgaben der Kappa-Koeffizient verwendet.

Beim Berechnen des „gewöhnlichen“ Kappa-Koeffizienten würde nicht zwischen geringen Abweichungen (z. B. wenn ein Fahrerlaubnisprüfer eine beobachtete Fahrleistung mit „Sehr gut“ und ein anderer mit „Gut“ bewertet) und größeren Abweichungen (z. B. wenn ein Fahrerlaubnisprüfer eine beobachtete Fahrleistung mit „Sehr gut“ und ein anderer mit „Ausreichend“ bewertet) unterschieden werden. Möchte man den Grad der Abweichung bei der Berechnung der Koeffizienten berücksichtigen, ist daher ein „gewichtetes“ Kappa zu verwenden. Hierbei kann zwischen einer linearen und einer quadratischen Gewichtung unterschieden werden. Bei der linearen Variante wird eine Verdopplung der Abweichung doppelt und bei der quadratischen Variante vierfach gewichtet. Bei der quadratischen Variante wirken sich starke Abweichungen zwischen den Beobachtern also besonders negativ auf den Übereinstimmungskoeffizienten aus. Dies scheint im Falle der Fahrkompetenzbewertung angebracht, da große Abweichungen zwischen den Fahrerlaubnisprüfern die Objektivität und Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung in besonderem Ausmaß infrage stellen würden.

Bei einer genaueren Betrachtung der Daten, die für die dritte Untersuchung der Beobachterübereinstimmung im Realverkehr (reale Prüfungen verschiedener Fahrerlaubnisklassen) vorliegen, fällt ein weiteres Phänomen auf, das es bei der Datenauswertung zu berücksichtigen gilt. Bei der Bewertung der Fahraufgaben und Beobachtungskategorien hat in 30 von insgesamt 260 Bewertungsfällen – dies entspricht einem Anteil von 11,5 Prozent – einer der beiden Fahrerlaubnisprüfer die Aufgabe bzw. Bewertungskategorie als „Nicht vorgekommen“ eingestuft, während der andere Fahrerlaubnisprüfer eine Bewertung vorgenommen hat, die Fahraufgabe bzw. Beobachtungskategorie also implizit als „Vorgekommen“<sup>31</sup> eingestuft hat. In den ersten beiden Untersuchungen zur Beobachterübereinstimmung im Realverkehr (simulierte Prüfungen bzw. reale Prüfungen der Fahrerlaubnisklasse B) ist dies – mit sehr geringem und daher vernachlässigbarem Einfluss auf die Ergebnisse – lediglich in 2 von 205 bzw. 5 von 260 Fällen vorgekommen. Bei einem Anteil von 11,5 Prozent der Fälle in der dritten Untersuchung können die genannten Abweichungen jedoch nicht mehr vernachlässigt werden. Dies wirft die Frage auf, wie diesem Sachverhalt bei der Analyse der Beobachterübereinstimmung am besten Rechnung getragen wer-

<sup>30</sup> Die am häufigsten verwendeten justierten Maße für ordinalskalierte Variablen sind die rangkorrelationsbasierten Koeffizienten Spearmans Rho und Kendall's Tau (WIRTZ & CASPAR, 2002).

<sup>31</sup> Die Bewertungskategorie „Vorgekommen“ existiert nicht im e-Prüfprotokoll. Sie wird jedoch an dieser Stelle und im Folgenden als Kategorie dargestellt, da dies die Möglichkeit eröffnet, die Erklärung des Sachverhalts sprachlich auf das Kernproblem zu reduzieren und damit zu erleichtern.

den könnte bzw. wie stark die Abweichung zwischen „Vorgekommen“ und „Nicht vorgekommen“ gewichtet werden sollte. Wählt man beispielsweise die kleinstmögliche Abweichung (d. h. um eine Kategorie), wäre zu fragen, ob dies als ausreichend betrachtet werden kann: Sollte der Unterschied zwischen „Sehr gut“ und „Gut“ tatsächlich genauso gewichtet werden wie beispielsweise zwischen „Nicht vorgekommen“ und „Gut“? Dies wäre wohl eher zu verneinen. Wählt man umgekehrt die größtmögliche Abweichung (d. h. um drei Kategorien), würde z. B. der Unterschied zwischen „Nicht vorgekommen“ und „Ausreichend“ genauso gewichtet werden wie der Unterschied zwischen „Ungenügend“ und „Sehr gut“. Auch dies wäre willkürlich und schwer nachvollziehbar.

Würde man sich – trotz dieser Fragen bzw. „Interpretationsschwierigkeiten“ – auf eine bestimmte Vorgehensweise zur Gewichtung festlegen, ergäbe sich ein weiteres methodisches Problem: Wie verfährt man mit Fällen, in denen die beiden Fahrerlaubnisprüfer eine Fahraufgabe bzw. Beobachtungskategorie übereinstimmend als „Nicht vorgekommen“ eingestuft haben? Wertet man dies als Übereinstimmung, wächst Cohen's Kappa (je nach Häufigkeit dieser Konstellation), obwohl die übereinstimmende Beobachtung, dass eine Fahraufgabe – z. B. Kreisverkehr – nicht vorgekommen ist, doch als selbstverständlich erachtet werden sollte. Außerdem würde sich solch eine übereinstimmende Beurteilung als „Nicht vorgekommen“ dann in gleicher Weise – d. h. mit der gleichen Gewichtung – positiv auf Cohen's Kappa auswirken wie eine Übereinstimmung bei der konkreten Bewertung einer Fahraufgabe (z. B. als Ungenügend). Beides erscheint wiederum als unangemessen. Umgekehrt wäre es jedoch auch methodisch fraglich, nur Abweichungen hinsichtlich der Übereinstimmung von Skalenstufen, nicht aber Übereinstimmungen hinsichtlich des Auftretens von Fahraufgaben bzw. Beobachtungskategorien zu berücksichtigen.

Wie die dargestellten Überlegungen zeigen, ist die statistische Berücksichtigung der Fälle, in denen die Fahrerlaubnisprüfer hinsichtlich des Auftretens einer Fahraufgabe bzw. Beobachtungskategorie nicht übereinstimmen, mit erheblichen methodischen Problemen behaftet. Im Folgenden werden die Berechnungen zur Beobachterübereinstimmung bei der Gesamtbewertung der Fahraufgaben und Beobachtungskategorien bei realen Prüfungen verschiedener Fahrerlaubnisklassen daher unter Ausschluss dieser Fälle vorgenommen. Dies führt

dazu, dass die gefundenen Übereinstimmungswerte für diese Untersuchungen vermutlich überschätzt werden. Die Tabelle 6 zeigt zunächst die Ergebnisse der Übereinstimmung zwischen den Fahrerlaubnisprüfern bei den Gesamtbewertungen der Fahraufgaben. Wie der zweiten Spalte (Kappa) der Tabelle 6 zu entnehmen ist, liegt in der ersten Untersuchung im Realverkehr (simulierte Prüfungen) die Übereinstimmung zwischen den Fahrerlaubnisprüfern bei der Gesamtbewertung aller Fahraufgaben über dem von STURZBECHER et al. (2014) geforderten Wert von 0,80.

Für alle Fahraufgaben zusammen beträgt der Kappa-Koeffizient sogar 0,94, was als sehr hoch interpretiert werden kann.<sup>32</sup> Auch für die zweite Untersuchung im Realverkehr (reale Prüfungen der Fahrerlaubnisklasse B) liegt die Übereinstimmung zwischen den Fahrerlaubnisprüfern bei der Gesamtbewertung der Fahraufgaben mehrheitlich über dem geforderten Wert von 0,80. Für alle Fahraufgaben zusammen beträgt der Kappa-Koeffizient 0,85. Bei der dritten Untersuchung im Realverkehr (reale Prüfungen verschiedener Fahrerlaubnisklassen) hingegen liegt die Übereinstimmung zwischen den Fahrerlaubnisprüfern bei der Gesamtbewertung der Fahraufgaben nur in drei von acht Fällen über oder knapp unter dem geforderten Wert von 0,80; für die restlichen fünf Fahraufgaben liegt er deutlich darunter. Berechnet man Kappa für alle Fahraufgaben zusammen, liegt der Koeffizient nur bei 0,48 (akzeptable Übereinstimmung).

In der dritten Spalte (n) der Tabelle ist die Anzahl der Bewertungspaare vermerkt, die den jeweiligen Berechnungen zugrunde liegen, und in der fünften Spalte die Zahl der Abweichungen zwischen den Fahrerlaubnisprüfern. Die erste Zahl gibt dabei den Grad der Abweichung an (Um wie viele Kategorien wichen die Bewertungen der Fahrerlaubnisprüfer voneinander ab?); in den Klammern dahinter ist vermerkt, wie häufig dieser Grad der Abweichung auftrat. Der p-Wert in der vierten Spalte gibt an, wie groß die Wahrscheinlichkeit ist, dass die jeweiligen Übereinstimmungswerte in der Stichprobe nur zufällig zustande gekommen sind, also fälschlicher-

<sup>32</sup> WIRTZ und CASPAR (2002) bezeichnen ein Kappa von 0,40 bis 0,60 als Indikator für eine akzeptable Übereinstimmung. Zwischen 0,60 und 0,75 sprechen die Autoren von einer guten Übereinstimmung und über 0,75 von einer sehr guten Übereinstimmung. Kappa-Werte unter 0,40 wären demnach als unzureichend zu interpretieren.

<b>Simulierte Prüfungen</b>	<b>Kappa</b>	<b>n</b>	<b>p</b>	<b>Abweichungen</b>
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	1,00	15	0,00***	0
Kurven	1,00	15	0,00***	0
Vorbeifahren, Überholen	1,00	15	0,00***	0
Kreuzungen, Einmündungen	0,89	15	0,00***	1(1)
Kreisverkehr	-	-	-	-
Schienenverkehr	-	-	-	-
Haltestellen, Fußgänger, Radfahrer	1,00	15	0,00***	0
Geradeausfahren	0,88	15	0,00***	1(1)
<b>Gesamt</b>	<b>0,94</b>	<b>94</b>	<b>0,00***</b>	<b>1(2)</b>
<b>Reale Prüfungen der Fahrerlaubnisklasse B</b>				
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	0,85	20	0,00***	1(2)
Kurven	1,00	20	0,00***	0
Vorbeifahren, Überholen	0,72	20	0,00***	1(2)
Kreuzungen, Einmündungen	0,82	20	0,00***	1(5)
Kreisverkehr	-	-	-	-
Schienenverkehr	1,00	15	0,00***	0
Haltestellen, Fußgänger, Radfahrer	0,64	20	0,00**	1(1)
Geradeausfahren	0,76	20	0,00***	1(2)
<b>Gesamt</b>	<b>0,85</b>	<b>137</b>	<b>0,00***</b>	<b>1(12)</b>
<b>Reale Prüfungen verschiedener Fahrerlaubnisklassen</b>				
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	0,81	13	0,00**	1(3)
Kurven	0,25	18	0,21	1(6)
Vorbeifahren, Überholen	-0,01	19	0,98	1(8); 2(2)
Kreuzungen, Einmündungen	0,34	19	0,14	1(10)
Kreisverkehr	0,50	19	0,02*	1(6)
Schienenverkehr	0,77	6	0,04*	1(1)
Haltestellen, Fußgänger, Radfahrer	0,77	17	0,00**	1(1)
Geradeausfahren	0,51	18	0,02*	1(9)
<b>Gesamt</b>	<b>0,48</b>	<b>129</b>	<b>0,00***</b>	<b>1(44); 2(2)</b>

**Tab. 6:** Übereinstimmung zwischen den Fahrerlaubnisprüfern bei den Gesamtbewertungen der Fahraufgaben – Prüfungen im Realverkehr

weise als systematisch interpretiert werden (Irrtumswahrscheinlichkeit). Für den Großteil der untersuchten Fahraufgaben liegt diese Wahrscheinlichkeit unter fünf Prozent. Gemäß sozialwissenschaftlicher Forschungskonvention sind diese Wahrscheinlichkeiten als niedrig und die Ergebnisse als statistisch signifikant zu interpretieren (BORTZ & DÖRING, 2006). Fahraufgaben, zu denen in Tabelle 6 keine statistischen Angaben aufgeführt sind, wurden in den Untersuchungen nicht ausreichend häufig geprüft, um die gewünschten Koeffizienten berechnen zu können. Bei den Berechnungen der Gesamtkoeffizienten für die Menge

aller Fahraufgaben wurden diese Fälle aber berücksichtigt.

Bei genauerer Betrachtung einiger Fahraufgaben mit relativ niedrigen Kappa-Werten fällt auf, dass diese nicht in allen Fällen mehr Abweichungen in den zugrunde liegenden Bewertungen der Fahrerlaubnisprüfer aufweisen als Fahraufgaben mit hohen Kappa-Werten. Die Ursache hierfür liegt in einer „Konstruktionsschwäche“ des Kappa-Koeffizienten begründet, die mit unausgeglichenen Randverteilungen der Beurteilungen in Zusammenhang steht.

Ein Vergleich der beiden Kontingenztafeln der Fahraufgaben „Kreuzungen, Einmündungen“ und „Haltestellen, Fußgänger, Radfahrer“ in der zweiten Untersuchung im Realverkehr (reale Prüfungen der Fahrerlaubnisklasse B) verdeutlicht dieses Problem (s. Tabelle 7 und 8). Wie der Tabelle 7 entnommen werden kann, stimmen die Einschätzungen der beiden Fahrerlaubnisprüfer bei der Bewertung der Fahraufgabe „Kreuzungen, Einmündungen“ in 15 von 20 Fällen überein (fett gedruckt). Der Kappa-Koeffizient liegt mit 0,82 über dem geforderten Wert von 0,80. Bei der Bewertung der Fahraufgabe „Haltestellen, Fußgänger, Radfahrer“ (s. Tabelle 8) hingegen stimmen die beiden Fahrerlaubnisprüfer sogar in 18 von 20 Fällen überein, und trotzdem liegt der Kappa-Wert mit 0,64 deutlich unter 0,80.

Warum ist der Kappa-Koeffizient im ersten Beispiel – trotz einer geringeren Zahl von Übereinstimmungen – deutlich größer? Ein genauerer Blick auf die Verteilung der Daten zeigt, dass sich die Übereinstimmungen im Fall der Fahraufgabe „Kreuzungen, Einmündungen“ relativ gleichmäßig auf die verschiedenen Bewertungskategorien verteilen. Bei der Fahraufgabe „Haltestellen, Fußgänger, Radfahrer“ hingegen entfallen 18 der 19 Übereinstimmungen auf die Bewertungskategorie „Gut“. Wie bereits oben erwähnt, wird bei der Berechnung von Kappa die tatsächliche Übereinstimmung um die zufällig erwartete Übereinstimmung bereinigt. Die zufällig erwartete Übereinstimmung berechnet sich auf Basis der multiplizierten Randsummen. Sie wird daher umso grö-

ßer, je stärker sich die Bewertungsübereinstimmungen auf eine oder wenige Kategorien konzentrieren. Infolgedessen fällt Kappa bei der Fahraufgabe „Haltestellen, Fußgänger, Radfahrer“ trotz einer hohen tatsächlichen Übereinstimmung vergleichsweise gering aus, da die zufällig erwartete Übereinstimmung ebenfalls hoch ist. Die Fahrerlaubnisprüferbewertungen im Bereich „Haltestellen, Fußgänger, Radfahrer“ waren daher nicht weniger objektiv und reliabel, sondern nur einseitiger verteilt als die Bewertungen der Fahraufgabe „Kreuzungen, Einmündungen“. In gleicher Weise haben die beiden Fahrerlaubnisprüfer in der zweiten Untersuchung die Fahraufgaben „Vorbeifahren, Überholen“ und „Geradeausfahren“ sehr homogen bewertet, was wiederum relativ geringe Kappa-Werte nach sich zieht. Ursächlich für das einseitige Bewertungsverhalten könnte ein Mangel an besonderen Vorkommnissen bei der Durchführung dieser Fahraufgaben sein, sodass die Bewerberleistungen in diesen Fällen erwartungsgemäß waren und daher größtenteils der Definition der gleichen Bewertungskategorie („Gut“) entsprachen. Im Ergebnis führt also eine ausgeprägte Gleichverteilung der zugrunde liegenden Daten in der zweiten Untersuchung zu einer Unterschätzung der Beobachterübereinstimmung durch den Kappa-Koeffizienten für die Fahraufgaben „Haltestellen, Fußgänger, Radfahrer“, „Vorbeifahren, Überholen“ und „Geradeausfahren“. Gleiches trifft auch auf die Fahraufgabe „Haltestellen, Fußgänger, Radfahrer“ in der dritten Untersuchung (reale Prüfungen verschiedener Fahrerlaubnisklassen) zu.

Prüfer 1	Prüfer 2			
	Ungenügend	Ausreichend	Gut	Gesamt
Ungenügend	<b>4</b>	2	0	6
Ausreichend	0	<b>3</b>	2	5
Gut	0	1	<b>8</b>	9
Gesamt	4	6	10	20

Tab. 7: Kontingenztafel für die Bewertungen der Fahraufgabe „Kreuzungen, Einmündungen“ der zweiten Untersuchung (reale Prüfungen der Fahrerlaubnisklasse B)

Prüfer 1	Prüfer 2			
	Ungenügend	Ausreichend	Gut	Gesamt
Ungenügend	<b>0</b>	0	0	0
Ausreichend	0	<b>1</b>	1	2
Gut	0	0	<b>18</b>	18
Gesamt	0	1	19	20

Tab. 8: Kontingenztafel für die Bewertungen der Fahraufgabe „Haltestellen, Fußgänger, Radfahrer“ der zweiten Untersuchung (reale Prüfungen der Fahrerlaubnisklasse B)



In der Tabelle 9 sind die Übereinstimmungswerte bei der Gesamtbewertung der Fahraufgaben im Rahmen der Untersuchung anhand videografiertener Prüfungen dargestellt. Zur Quantifizierung der Übereinstimmung wurde Krippendorff's Alpha gewählt, das für eine beliebig große Zahl von Beobachtern und ordinalskalierte Daten geeignet ist (s. o.). Bei den Gesamtbewertungen der Fahraufgaben ist keine systematische Zunahme oder Abnahme der Werte über die Messzeitpunkte erkennbar. Manche Fahraufgaben werden zunehmend einheitlich bewertet (z. B. Kreuzungen, Einmündungen), andere Fahraufgaben werden zunehmend uneinheitlich beurteilt (z. B. Kreisverkehr). Die Beobachterübereinstimmung über alle Fahraufgaben zusammen ist für den ersten und den letzten Messzeitpunkt identisch.

Die sehr niedrigen Übereinstimmungswerte einiger Fahraufgaben hängen erneut mit einer ausgeprägten Gleichverteilung der zugrunde liegenden Daten und der oben bereits erläuterten Konstruktionschwäche des Kappa-Koeffizienten zusammen, die in gleicher Weise für Krippendorff's Alpha zutrifft, da auch hier die tatsächliche Übereinstimmung um die statistisch erwartete korrigiert wird. Inhaltlich bedeutet dies, dass die betroffenen Fahraufgaben in den Filmen gar nicht bzw. selten vorkamen (z. B. Schienenverkehr) oder sehr homogen bewertet wurden aufgrund eines Mangels an Fehlern oder überdurchschnittlichen Leistungen (z. B. Kurven). Bei der Berechnung der Übereinstimmungswerte für alle Fahraufgaben zusammen fällt diese Konstruktionschwäche aber nicht ins Gewicht.

Zusammenfassend kann für die simulierten Prüfungen und für die realen Prüfungen der Fahrerlaub-

nisklasse B im Realverkehr festgestellt werden, dass auf Basis der vorliegenden Stichproben von einer guten bis sehr guten Objektivität und Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung mit Blick auf die Bewertung der Fahraufgaben ausgegangen werden kann, obwohl einige wenige Übereinstimmungswerte unterhalb des von STURZBECHER et al. (2014) geforderten Wertes von 0,80 liegen. Dieser Befund wird allerdings durch die Ergebnisse der Untersuchung von realen Prüfungen in verschiedenen Fahrerlaubnisklassen im Realverkehr geschwächt, im Rahmen derer die Beobachterübereinstimmungswerte nur für drei von acht Fahraufgaben über bzw. knapp unter dem geforderten Wert liegen.

Die Auswertung der videografierten Prüfungen indiziert im Fall der Fahraufgaben keine Zunahme der Beobachterübereinstimmung bei zunehmender Übung im Umgang mit der optimierten Praktischen Fahrerlaubnisprüfung. Dies ist aus lerntheoretischer Sicht unplausibel und dürfte den bereits oben erläuterten Problemen der Prüfer bei der Bewertung von Prüfungsvideos – fehlende Kenntnis des Prüfortes und der Prüfstrecke,<sup>33</sup> ungewohnte Auf-

<sup>33</sup> Während dem Prüfer in realen Prüfungen die mit bestimmten Streckenpunkten verbundenen Prüfungsanforderungen im Detail vertraut sind und er die Prüfungsstrecke darauf basierend selbst planen bzw. vorgeben kann, beinhalten die Prüfungsvideos für ihn nicht vorhersehbare bzw. überraschende Anforderungen. Dies erhöht die Wahrscheinlichkeit, dass der Prüfer in einem wichtigen Moment den Blick von der Strecke abwendet – z. B. um ein vorheriges Ereignis zu dokumentieren – und ein bewertungsrelevantes Ereignis verpasst.

	Krippendorff's Alpha		
	1. Messzeitpunkt (n = 38)	2. Messzeitpunkt (n = 28)	3. Messzeitpunkt (n = 31)
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	0,36	0,21	0,21
Kurven	0,00	0,03	0,04
Vorbeifahren, Überholen	0,03	0,18	0,16
Kreuzungen, Einmündungen	0,39	0,41	0,59
Kreisverkehr	0,36	0,18	0,13
Schienenverkehr	0,00	0,00	0,06
Haltestellen, Fußgänger, Radfahrer	0,09	0,00	0,20
Geradeausfahren	0,43	0,46	0,35
Gesamt	0,38	0,33	0,38

Tab. 9: Übereinstimmung zwischen den Fahrerlaubnisprüfern bei den Gesamtbewertungen der Fahraufgaben – Videografierte Prüfungen

bereitung bewertungsrelevanter Informationen in den Videos<sup>34</sup> – geschuldet sein.

Die Gesamtbewertungen der Kompetenzbereiche und vor allem die Prüfungsentscheidung erfolgen auf einem höheren Abstraktionsniveau als die Gesamtbewertungen der Fahraufgaben und erscheinen daher situationsunabhängiger bzw. weniger „störanfällig“. Sie zeichnen eher ein Gesamtbild, das auf einer wesentlich größeren Zahl einzelner

Ereignisse basiert. Das dem Untersuchungsdesign der videografierten Prüfungen geschuldete Übersehen einzelner Ereignisse sollte daher – aufgrund des höheren Abstraktionsniveaus – bei der Gesamtbewertung der Kompetenzbereiche bzw. Beobachtungskategorien und bei der Prüfungsentscheidung weit weniger ins Gewicht fallen als bei der Gesamtbewertung der Fahraufgaben. Außerdem sollte es sich durch Lerneffekte auch schneller kompensieren lassen.

### Gesamtbewertung der Beobachtungskategorien

Die situationsübergreifenden Bewertungen der Kompetenzbereiche bzw. Beobachtungskategorien wurden anhand der gleichen vierstufigen Skala vorgenommen wie die Gesamtbewertungen der Fahraufgaben. Zur Analyse der Beobachterübereinstimmung wurde daher erneut ein quadratisch gewichtetes Kappa für die Untersuchungen im Realverkehr und Krippendorff's Alpha für die videografierten Prüfungen verwendet. Tabelle 10 enthält die gleichen Statistiken für die Gesamtbewertung der

<sup>34</sup> So sind die sensorischen Informationen in den Filmen – trotz aller Bemühungen zur vollständigen Wiedergabe – nur eingeschränkt vorhanden. Beispielsweise kann der Prüfer Ereignisse nicht im Augenwinkel erfassen und Verhaltensweisen des Fahrlehrers, die auf besondere Ereignisse hindeuten, nur eingeschränkt wahrnehmen. Außerdem sind die visuellen und auditiven Informationen in den Videos weniger präzise als in der Realität und nicht dreidimensional dargestellt. Darüber hinaus können die Beobachter Lenkbewegungen und Geschwindigkeitsveränderungen nicht fühlen. Alle diese Einschränkungen führen dazu, dass die Fahrerlaubnisprüfer für die Bewertung der Fahrleistungen relevante Ereignisse beim Betrachten von Videos leichter übersehen als in realen Prüfungen.

Simulierte Prüfungen	Kappa	n	p	Abweichungen
Verkehrsbeobachtung	0,83	15	0,00**	1(2)
Fahrzeugpositionierung	0,89	15	0,00***	1(1)
Geschwindigkeitsanpassung	0,89	15	0,00***	1(1)
Kommunikation	1,00	15	0,00***	0
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,60	15	0,01*	1(2)
Gesamt	0,87	75	0,00***	1(6)
<b>Reale Prüfungen der Fahrerlaubnisklasse B</b>				
Verkehrsbeobachtung	0,93	20	0,00***	1(2)
Fahrzeugpositionierung	1,00	20	0,00***	0
Geschwindigkeitsanpassung	0,61	20	0,01**	1(2)
Kommunikation	0,64	20	0,00**	1(1)
Fahrzeugbedienung/Umweltbewusste Fahrweise	1,00	20	0,00***	0
Gesamt	0,92	100	0,00***	1(4)
<b>Reale Prüfungen verschiedener Fahrerlaubnisklassen</b>				
Verkehrsbeobachtung	0,66	18	0,00**	1(5); 2(1)
Fahrzeugpositionierung	0,24	18	0,25	1(5); 2(1)
Geschwindigkeitsanpassung	0,70	19	0,00**	1(7)
Kommunikation	0,35	19	0,10	1(10)
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,48	19	0,00**	1(6)
Gesamt	0,54	93	0,00***	1(33); 2(2)

Tab. 10: Übereinstimmung (quadratisch gewichtetes Kappa) zwischen den Fahrerlaubnisprüfern bei der Gesamtbewertung der Kompetenzbereiche bzw. Beobachtungskategorien

Kompetenzbereiche bzw. Beobachtungskategorien wie die Tabelle 6 für die Gesamtbewertung der Fahraufgaben.

Für die ersten beiden Untersuchungen im Realverkehr liegt die Beobachterübereinstimmung bei der Gesamtbewertung der Kompetenzbereiche bzw. Beobachtungskategorien in vier bzw. drei von fünf Fällen über dem geforderten Wert von 0,80. Für alle Beobachtungskategorien zusammen liegt der Wert für beide Untersuchungen mit 0,87 bzw. 0,92 sogar deutlich über diesem Niveau. Bei der dritten Untersuchung hingegen liegen alle Übereinstimmungswerte deutlich unter 0,80.

Die relativ niedrigen Kappa-Werte im Kompetenzbereich „Fahrzeugbedienung/Umweltbewusste Fahrweise“ in der ersten Untersuchung sowie in den Kompetenzbereichen „Geschwindigkeitsanpassung“ und „Kommunikation“ in der zweiten Untersuchung basieren auf einseitig verteilten Daten und könnten somit der bereits diskutierten Konstruktionsschwäche des Kappa-Koeffizienten geschuldet sein (s. o.). SCHUSTER (2010) empfiehlt in solchen Fällen, zusätzliche Kriterien zur Beurteilung der „wirklichen“ Übereinstimmung hinzuzuziehen. Im vorliegenden Fall erscheint es – wie auch bei den Gesamtbewertungen der Fahraufgaben – sinnvoll, die tatsächliche (prozentuale) Übereinstimmung zu berücksichtigen: Bezüglich des Kompetenzbereiches „Fahrzeugbedienung/Umweltbewusste Fahrweise“ stimmten die Fahrerlaubnisprüfer bei der ersten Untersuchung in 18 von 20 Fällen (90 %) überein. Bei der zweiten Untersuchung stimmten die Fahrerlaubnisprüfer bei der Bewertung der „Geschwindigkeitsanpassung“ in 18 von 20 (90 %) und bezüglich der „Kommunikation“ in 19 von 20 Fällen (95 %) überein. Demnach sind die relativ niedrigen Kappa-Werte bei der Bewertung dieser Beobachtungskategorien einer einseitigen

Verteilung der Daten und nicht einer mangelnden Objektivität und Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung geschuldet.

Bei der Beurteilung videografiertter Prüfungen (s. Tabelle 11) steigen die Übereinstimmungswerte bei den Gesamtbewertungen der Kompetenzbereiche kontinuierlich zwischen dem ersten und dritten Messzeitpunkt an: So liegt die Übereinstimmung zum ersten Messzeitpunkt bei 0,46, zum zweiten Messzeitpunkt bei 0,52 und zum dritten Messzeitpunkt bei 0,57. Mit Ausnahme des Kompetenzbereiches „Geschwindigkeitsanpassung“ liegen zum dritten Messzeitpunkte alle Werte über denen vom ersten Messzeitpunkt. Auffällig ist, dass zum ersten Messzeitpunkt die Werte für alle einzelnen Kompetenzbereiche unter dem Gesamtwert für alle Kompetenzbereiche liegen. Das erklärt sich wiederum durch die geringere Varianz der den einzelnen Kompetenzbereichen zugrunde liegenden Bewertungen gegenüber der größeren Varianz der Bewertungen, auf denen der Gesamtwert basiert. Wie oben bereits mehrfach erläutert wurde, steigt bei geringer Varianz die statistisch erwartete Übereinstimmung, wodurch der Alpha-Koeffizient sinkt.

Im Ergebnis kann festgehalten werden, dass die Befundlage hinsichtlich der Beobachterübereinstimmung bei der Gesamtbewertung der Kompetenzbereiche bzw. Beobachtungskategorien uneinheitlich ist. Während die Daten der ersten beiden Untersuchungen im Realverkehr ausreichende Übereinstimmungswerte aufweisen, wurde der anvisierte Wert von 0,80 in der dritten Untersuchung im Realverkehr deutlich verfehlt. Auf Basis dieser Befunde lassen sich – wie schon bei der Gesamtbewertung der Fahraufgaben – keine eindeutigen Aussagen über die Objektivität und Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung ableiten. Die vom ersten bis zum dritten Messzeit-

	Krippendorff's Alpha		
	1. Messzeitpunkt (n = 38)	2. Messzeitpunkt (n = 28)	3. Messzeitpunkt (n = 31)
Verkehrsbeobachtung	0,31	0,36	0,56
Fahrzeugpositionierung	0,42	0,56	0,72
Geschwindigkeitsanpassung	0,43	0,42	0,38
Kommunikation	0,18	0,25	0,19
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,20	0,39	0,29
Gesamt	0,46	0,52	0,57

**Tab. 11:** Übereinstimmung zwischen den Fahrerlaubnisprüfern bei den Gesamtbewertungen der Kompetenzbereiche bzw. Beobachtungskategorien – Videografierte Prüfungen

punkt kontinuierlich steigende Übereinstimmung bei der Gesamtbewertung der Kompetenzbereiche deutet auf eine wünschenswerte Verbesserung der Objektivität und Reliabilität bei zunehmender Übung im Umgang mit der optimierten Praktischen Fahrerlaubnisprüfung hin.

### Ereignisbezogene Bewertungen

Die Berechnung der Beobachterübereinstimmung bei der Bewertung spezifischer entscheidungsrelevanter Prüfungsereignisse ist mit methodischen Hindernissen verbunden: Erstens können im Straßenverkehr eine Vielzahl potenziell bewertungsrelevanter Ereignisse in kurzer Zeit auftreten, die von verschiedenen Fahrerlaubnisprüfern nicht immer in gleicher Reihenfolge bewertet werden. Beispielsweise könnte einerseits das zweite bewertete Ereignis des einen Fahrerlaubnisprüfers dem fünften des anderen Fahrerlaubnisprüfers entsprechen. Andererseits könnten zwei Fahrerlaubnisprüfer im Rahmen einer Prüfung das gleiche Kriterium ausgewählt haben, aber bezogen auf verschiedene Ereignisse. Zweitens erscheint es vorstellbar, dass ein Fahrerlaubnisprüfer bei einem leichten Fehler eines Bewerbers wegen einer Bewertungsunsicherheit zunächst nichts notiert und eine ähnliche Situation erneut herbeiführt, um eine abschließende Bewertung vorzunehmen („Adaptive Prüfstrategie“), während ein anderer Fahrerlaubnisprüfer den Fehler sofort dokumentiert. Hier liegen – obwohl dieser Anschein in der Datenstruktur erweckt wird – keine grundsätzlichen Abweichungen in der Fehlerwahrnehmung zwischen den Fahrerlaubnisprüfern vor, sondern nur Abweichungen im Dokumentationsverhalten.

Zur Ermittlung der Beobachterübereinstimmung bei den Untersuchungen im Realverkehr wurden die Bewertungen des Fahrerlaubnisprüfers, der die jeweiligen Prüfungen geleitet hat, als maßgeblich betrachtet. Sämtliche von ihm vorgenommenen ereignisbezogenen Bewertungen in einer bestimmten

Situation wurden daraufhin verglichen, ob der zweite Fahrerlaubnisprüfer diese Bewertungen ebenfalls vorgenommen hat. So ergab sich für jede vom leitenden Fahrerlaubnisprüfer vorgenommene Bewertung eine Übereinstimmung oder nicht. Zur Quantifizierung der Übereinstimmungen kann in diesem Fall die intuitiv verständliche prozentuale Übereinstimmung verwendet werden. Der Schwachpunkt der prozentualen Übereinstimmung besteht darin, dass das Ergebnis nicht um die zufällig erwartete Übereinstimmung bereinigt wird (s. o.). Aufgrund der sehr großen Anzahl möglicher bewertungsrelevanter Prüfungsereignisse ist die Wahrscheinlichkeit, dass zwei Fahrerlaubnisprüfer dieselbe Bewertung zufällig vornehmen, aber so gering, dass die Korrektur um die zufällig erwartete Übereinstimmung an dieser Stelle vernachlässigt werden kann.

Hinsichtlich der Beobachterübereinstimmung bei der Dokumentation ereignisbezogener Bewertungen wurde in der ersten Untersuchung im Realverkehr (simulierte Prüfungen) und in der zweiten Untersuchung im Realverkehr (reale Prüfungen der Fahrerlaubnisklasse B) der geforderte Wert von 0,70 bzw. 70 Prozent erreicht (s. Tabelle 12). Dies gilt auch bei einer Beschränkung der Auswertung auf die für die Prüfungsentscheidung besonders relevanten schweren Fehler, „Todsünden“ (d. h. schwere Fehler, die zum sofortigen Nichtbestehen der Prüfung führen) und Prüfungsabbrüche wegen Gefährdung bzw. Schädigung.

Im Rahmen der dritten Untersuchung im Realverkehr (reale Prüfungen verschiedener Fahrerlaubnisklassen) wurde der geforderte Wert hingegen nur bei einer Beschränkung der Auswertung auf die schweren Fehler, „Todsünden“ und Prüfungsabbrüche erreicht. Bei Berücksichtigung aller ereignisbezogenen Bewertungen – also auch der überdurchschnittlichen Leistungen und leichten Fehler – liegt die prozentuale Übereinstimmung mit 45 Prozent deutlich unter den geforderten 70 Prozent. Auf

	Alle ereignisbezogenen Bewertungen		Schwere Fehler, „Todsünden“, Prüfungsabbrüche	
	Übereinstimmungen	Prozentuale Übereinstimmung	Übereinstimmungen	Prozentuale Übereinstimmung
Simulierte Prüfungen	46 von 57	81 %	16 von 21	76 %
Reale Prüfungen der Fahrerlaubnisklasse B	35 von 42	83 %	16 von 17	94 %
Reale Prüfungen verschiedener Fahrerlaubnisklassen	20 von 56	45 %	4 von 5	80 %

**Tab. 12:** Prozentuale Übereinstimmung zwischen den Fahrerlaubnisprüfern bei der Dokumentation ereignisbezogener Bewertungen – Prüfungen im Realverkehr



Basis der vorliegenden Daten kann folglich auch hinsichtlich der Dokumentation ereignisbezogener Bewertungen nicht eindeutig von einer objektiven und reliablen optimierten Praktischen Fahrerlaubnisprüfung gesprochen werden.

Eine Berechnung der Beobachterübereinstimmung bei der Bewertung videografiert Prüfungen wurde für die Ebene der ereignisbezogenen Bewertungen aus zwei Gründen nicht durchgeführt: Erstens ist die Vergabe einzelner ereignisbezogener Bewertungen äußerst selten, wodurch sich eine extreme Gleichverteilung der den Analysen zugrunde liegenden Daten ergeben würde.<sup>35</sup> Bei einer starken Gleichverteilung der Daten werden die zu verwendenden Übereinstimmungsmaße aber erheblich verzerrt (s. o.). Zweitens ist die Auswertung bei mehreren tausend ereignisbezogenen Bewertungsmöglichkeiten und 38 Beobachtern aus technischer Sicht mit erheblichen Problemen verbunden.

Zusammenfassend bleibt festzuhalten, dass die Ergebnisse der ersten beiden Untersuchungen im Realverkehr durchweg eine hohe Beobachterübereinstimmung belegen, während in der dritten Untersuchung im Realverkehr niedrige Übereinstimmungswerte bei der Bewertung der Fahraufgaben und Beobachtungskategorien sowie der Dokumentation überdurchschnittlicher Leistungen und leichter Fehler gefunden wurden. Hierfür sind folgende Gründe in Betracht zu ziehen:

- Erstens könnte die Einbeziehung zusätzlicher Fahrerlaubnisklassen – im Gegensatz zu den auf die Fahrerlaubnisklasse B beschränkten ersten beiden Untersuchungen – zu einer Reduktion der Übereinstimmungswerte geführt haben (z. B. weil die Fahrerlaubnisprüfer für den schnellen Wechsel zwischen den Fahraufgabekatalogen der verschiedenen Fahrerlaubnisklassen noch keine Routinen ausgebildet hatten).
- Zweitens muss berücksichtigt werden, dass die beiden Fahrerlaubnisprüfer vor der Schulung zur vorliegenden Untersuchung noch keine Gelegenheit gehabt hatten, den Umgang mit dem

Fahraufgabekatalog und dem e-Prüfprotokoll zu trainieren. Hingegen waren drei der vier Fahrerlaubnisprüfer der ersten beiden Untersuchungen in verschiedener Form am Entwicklungsprozess der optimierten Praktischen Fahrerlaubnisprüfung direkt beteiligt und besaßen bereits Vorkenntnisse hinsichtlich der Handhabung des Fahraufgabekatalogs bzw. des e-Prüfprotokolls. Dieser „Trainingsrückstand“ der erstgenannten Gruppe dürfte sich – insbesondere vor dem Hintergrund des ersten Arguments, also der erhöhten Anforderung durch den Wechsel verschiedener Fahrerlaubnisklassen – negativ auf die Bewertungsübereinstimmungen ausgewirkt haben.

Empirisch bestätigt wird die Hypothese zur Verzerrung von Prüfungsbewertungen durch fehlendes Beobachtertraining durch die Ergebnisse der Untersuchung zur Beobachterübereinstimmung anhand videografiert Prüfungen. Hierbei zeigt sich über die Messzeitpunkte hinweg ein Anstieg der Beobachterübereinstimmung für die Prüfungsentscheidungen und die Gesamtbewertungen der Kompetenzbereiche bzw. Beobachtungskategorien, was einen starken fördernden Einfluss des Beobachtertrainings auf die Prüfungsgüte belegt und auf die Erzielbarkeit zufriedenstellender Güteparameter hindeutet. Dieser wünschenswerte Anstieg konnte zwar für die Gesamtbewertungen der Fahraufgaben nicht nachgewiesen werden, was aber durch das geringere Abstraktionsniveau der situationsbezogenen Fahraufgaben gegenüber den situationsübergreifenden Kompetenzbereichen und der Prüfungsentscheidung erklärbar ist (s. o.).

### 5.1.3 Überprüfung der Reliabilität anhand der Bewertungsstabilität

Bei der Berechnung der Bewertungsstabilität werden die Bewertungen der Prüfer – z. B. hinsichtlich eines Kompetenzbereiches bzw. einer Beobachtungskategorie – zwischen zwei Messzeitpunkten miteinander verglichen. Die Stabilität der Messungen kann über Korrelations- bzw. Zusammenhangsmaße quantifiziert werden. Ausgewählt wird das angemessene Zusammenhangsmaß dabei anhand des Skalenniveaus der zugrunde liegenden Messreihen bzw. Variablen.

Die Prüfungsentscheidung ist eine dichotome Variable. Im Falle von zwei dichotomen Variablen ist Pearson's Produktmomentkorrelation  $R$  das geeignete Zusammenhangsmaß (BORTZ, 1999). Die

<sup>35</sup> Da die große Mehrheit der im e-Prüfprotokoll dokumentierbaren Fehler oder überdurchschnittlichen Leistungen in den meisten Prüfungen nicht vorkommt, enthalten die betroffenen Datenzellen fast immer eine „0“. Eine „0“ indiziert, dass ein Ereignis nicht eingetreten ist, eine „1“ indiziert, dass ein Ereignis eingetreten ist.

	Pearson's R	n	p
Stabilität zwischen Messzeitpunkt 1 und 2	0,59	24	0,00***
Stabilität zwischen Messzeitpunkt 2 und 3	0,56	21	0,00***

Tab. 13: Bewertungsstabilität der Fahrerlaubnisprüfer beim Treffen der Prüfungsentscheidung – Videografierte Prüfungen

Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche erfolgen dagegen anhand einer vierstufigen Ordinalskala (s. o.). Bei zwei ordinalskalierten Variablen ist Spearman's Rho das am häufigsten verwendete Korrelationsmaß. Der Vorteil von Spearman's Rho besteht vor allem darin, dass die Ergebnisse direkt mit der o. g. Produktmomentkorrelation R vergleichbar sind. Bei einer großen Zahl von Rangbindungen – d. h. viele Beobachtungen einer Variablen haben die gleiche Ausprägung und damit den gleichen Rang – ist aber Kendall's Tau, ein anderer Rangkorrelationskoeffizient, Spearman's Rho vorzuziehen (WIRTZ & CASPAR, 2002). Bei gleicher Datengrundlage fällt Kendall's Tau allgemein niedriger aus als Spearman's Rho, d. h. der Zusammenhang zwischen den untersuchten Variablen wird schwächer geschätzt.

Wie fast alle Zusammenhangsmaße haben Pearson's R und Kendall's Tau einen Wertebereich von „-1“ bis „1“, wobei „-1“ für einen perfekten negativen Zusammenhang und „1“ für einen perfekten positiven Zusammenhang zwischen den Variablen steht; bei „0“ besteht keine Zusammenhang. Bei den folgenden Berechnungen der genannten Zusammenhangsmaße für Ordinalskalen wird die Angabe „Fahraufgabe bzw. Beobachtungskategorie nicht vorgekommen“ stets als fehlender Wert behandelt, da sie sich nicht in ein sinnvolles Verhältnis zu den anderen Ausprägungen der verwendeten Ordinalskala setzen lässt: So kann weder behauptet werden, dass „Nicht vorgekommen“ kleiner sei noch dass es größer sei als beispielsweise „Ungenügend“ oder „Gut“. Die beschriebenen Zusammenhangsmaße können daher nicht unter Einbeziehung der Kategorie „Nicht vorgekommen“ berechnet werden.

Im Folgenden werden die Ergebnisse der Untersuchungen zur Bewertungsstabilität anhand videografiert Prüfungen zunächst für die Prüfungsentscheidung und danach für die Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien vorgestellt und diskutiert. Die Berechnungen erfolgten dabei immer für alle drei bewerteten Filme gemeinsam (s. o.). Für die

ereignisbezogenen Bewertungen wurden wegen der oben erläuterten Gründe – hoher technischer Aufwand und Verzerrung der Ergebnisse durch extreme Gleichverteilung der Daten – keine Stabilitätswerte berechnet.

Wie der zweiten Spalte der Tabelle 13 entnommen werden kann, liegt die Bewertungsstabilität für die Prüfungsentscheidung – quantifiziert durch Pearson's R – zwischen der ersten und der zweiten Messung bei 0,59 und zwischen der zweiten und der dritten Messung bei 0,56. Dies ist nach FIELD (2009) als starker Effekt zu bezeichnen.<sup>36</sup> Der Wert in der dritten Spalte (n) der Tabelle gibt an, auf wie vielen Fahrerlaubnisprüfern die Berechnungen basieren, d. h. wie viele Prüfer an beiden Messungen teilgenommen haben.

Bezüglich der Gesamtbewertungen der Fahraufgaben (s. Tabelle 14) liegt die Bewertungsstabilität in fast allen Fällen um 0,30 oder 0,50; sie kann somit als mittelhoch oder hoch bezeichnet werden (s. Fußnote 36). Einzige Ausnahme ist die Fahraufgabe „Schienenverkehr“ mit einem nur schwachen Effekt, der allerdings statistisch nicht signifikant ist. Ursächlich hierfür ist, dass diese Fahraufgabe in den drei Filmen insgesamt nur einmal vorgekommen ist und daher in den meisten Fällen gar nicht bewertet wurde

Eine Berechnung der Bewertungsstabilität für die Gesamtheit aller Fahraufgaben wäre zwar technisch gesehen möglich, aber inhaltlich nicht sinnvoll: Als justiertes Maß berechnet sich Kendall's Tau anhand der Relationen einzelner Bewertungen zur Gesamtheit der restlichen Bewertungen. Bei einer Berechnung von Kendall's Tau für die Gesamtheit aller Fahraufgaben würde sich der Bezugsmaßstab (alle restlichen Bewertungen) für die einzelnen Bewertungen verändern. Insofern würde eine Berechnung von Kendall's Tau für die Gesamtheit aller Fahraufgaben ggf. zu einer unerwünschten Verzerrung der Ergebnisse führen.

<sup>36</sup> FIELD (2009) klassifiziert Effektgrößen von „+0,1“ bis unter „+0,3“ als „schwach“, von „+0,3“ bis unter „+0,5“ als „mittelstark“ und ab „+0,5“ als „stark“.

	Messung 1 und 2 (n = 24)		Messung 2 und 3 (n = 21)	
	Tau	p	Tau	p
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	0,47	0,00***	0,53	0,00***
Kurven	0,57	0,00***	0,46	0,00***
Vorbeifahren, Überholen	0,32	0,00**	0,36	0,00**
Kreuzungen, Einmündungen	0,45	0,00***	0,37	0,00**
Kreisverkehr	0,56	0,00***	0,52	0,00***
Schienenverkehr	0,12	0,28 <sup>ns</sup>	0,01	0,94 <sup>ns</sup>
Haltestellen, Fußgänger, Radfahrer	0,31	0,01**	0,45	0,00***
Geradeausfahren	0,51	0,00***	0,57	0,00***

Tab. 14: Bewertungsstabilität der Fahrerlaubnisprüfer bei den Gesamtbewertungen der Fahraufgaben – Videografierte Prüfungen

	Messung 1 und 2 (n = 24)		Messung 2 und 3 (n = 21)	
	Tau	p	Tau	p
Verkehrsbeobachtung	0,32	0,00**	0,50	0,00***
Fahrzeugpositionierung	0,65	0,00***	0,79	0,00***
Geschwindigkeitsanpassung	0,63	0,00***	0,48	0,00***
Kommunikation	0,36	0,00**	0,39	0,00**
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,38	0,00**	0,53	0,00***

Tab. 15: Bewertungsstabilität der Fahrerlaubnisprüfer bei den Gesamtbewertungen der Kompetenzbereiche – Videografierte Prüfungen

Hinsichtlich der Gesamtbewertungen der Kompetenzbereiche (s. Tabelle 15) fällt die Bewertungsstabilität in allen Fällen mittelhoch oder hoch aus. Darüber hinaus ist bei fast allen Kompetenzbereichen eine Zunahme der Bewertungsstabilität zwischen der zweiten und dritten Messung gegenüber der Bewertungsstabilität zwischen der ersten und zweiten Messung erkennbar. Die Bewertungsstabilität für die Gesamtheit aller Fahraufgaben wurde aus den oben genannten Gründen nicht berechnet. Zusammenfassend kann in Bezug zu allen Auswertungsebenen – Prüfungsentscheidungen sowie Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche – von einer mittelhohen oder hohen Bewertungsstabilität bzw. Retest-Reliabilität gesprochen werden. Dies stützt genauso wie der Anstieg der Bewertungsstabilität im Erprobungszeitraum die Annahme einer zuverlässigen bzw. reliablen Fahrkompetenzmessung durch die optimierte Praktische Fahrerlaubnisprüfung. Bei der Interpretation der Stabilitätswerte ist auch zu berücksichtigen, dass der im Erprobungszeitraum stattgefundenene Zugewinn an Prüfkompetenz bei den Fahrerlaubnisprüfern die Stabilitätswerte schwächt. Vor

diesem Hintergrund sind die erreichten Stabilitätswerte als Erfolg anzusehen. Im Normalfall sind die Stabilitätsuntersuchungen zu allen Messpunkten – also sowohl die Erstuntersuchungen als auch die Wiederholungsuntersuchungen – mit routinierten Prüfern durchzuführen, um die maximal erreichbare Stabilität zu ermitteln. Dies ist noch nicht erfolgt und bleibt künftigen Untersuchungen vorbehalten.

#### 5.1.4 Fazit

Zur Sicherstellung der Verfahrensgüte der optimierten Praktischen Fahrerlaubnisprüfung wurde die Beobachterübereinstimmung als bedeutendster Indikator für die Objektivität und Reliabilität des Messverfahrens bei 15 simulierten Prüfungen der Fahrerlaubnisklasse B, bei 20 realen Prüfungen der Fahrerlaubnisklasse B, bei 20 realen Prüfungen verschiedener Fahrerlaubnisklassen und bei 3 videografierten Prüfungen untersucht. Die videografierten Prüfungen wurden den Fahrerlaubnisprüfern jeweils zu Beginn, zur Mitte und zum Ende der viereinhalbmonatigen Felderprobung zur systema-

tischen Beobachtung und Bewertung vorgeführt. Durch einen Vergleich der Bewertungen bei diesen drei Messzeitpunkten lässt sich die Bewertungsstabilität bestimmen, die zusätzliche Hinweise auf die Reliabilität der optimierten Praktischen Fahrerlaubnisprüfung bietet.

Im Ergebnis der drei Untersuchungen im Realverkehr kann festgehalten werden, dass bezüglich der Prüfungsentscheidungen und der Dokumentation der für die Prüfungsentscheidung besonders bedeutsamen schweren Fehler die vorab festgelegten Mindestbeobachterübereinstimmungswerte bei allen drei Untersuchungen erreicht bzw. – teilweise deutlich – übertroffen wurden. Hinsichtlich der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien sowie der Dokumentation der leichten Fehler wurden nur in den beiden erstgenannten Untersuchungen (simulierte Prüfungen, reale Prüfungen der Fahrerlaubnisklasse B) zufriedenstellende Übereinstimmungswerte erzielt. Bei der dritten Untersuchung (reale Prüfungen verschiedener Fahrerlaubnisklassen) konnte ein solcher belastbarer Nachweis nicht erbracht werden. Dies könnte durch die erhöhten, noch ungewohnten Anforderungen eines ständigen Wechsels der zu prüfenden Fahrerlaubnisklasse und einen „Trainingsrückstand“ der Prüfer im Vergleich mit den ersten beiden Untersuchungen begründet sein. Bestätigt wird diese Annahme durch die kontinuierliche Zunahme der Beobachterübereinstimmungswerte bei der Beurteilung videografiert Prüfungen vom ersten bis zum dritten Messzeitpunkt. Demnach gelangten die an dieser Untersuchung teilnehmenden Prüfer mit zunehmender Übung im Umgang mit dem e-Prüfprotokoll zu immer objektiveren und reliableren Ergebnissen, was auf die Erreichbarkeit einer hohen Verfahrensgüte bei ausreichendem Beobachtertraining schließen lässt.

Auch die Befunde zur Bewertungsstabilität zwischen dem ersten und zweiten Messzeitpunkt sowie dem zweiten und dritten Messzeitpunkt bei der Beurteilung videografiert Prüfungen belegen eine zuverlässige bzw. reliable Messung der Fahrkompetenz mittels der optimierten Praktischen Fahrerlaubnisprüfung. Die Bewertungsstabilität bzw. Retest-Reliabilität liegt für die Prüfungsentscheidung sowie für alle Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche in einem mittelhohen oder hohen Bereich.

## 5.2 Validität

### 5.2.1 Überblick

Ein psychometrisches Messverfahren gilt als valide, wenn es tatsächlich das misst, was es zu messen vorgibt. Im vorliegenden Fall ist daher zu überprüfen, ob mit der optimierten Praktischen Fahrerlaubnisprüfung auch tatsächlich (ausschließlich) die Fahrkompetenz der Bewerber erfasst wird.

Die Validität ist das wichtigste und zugleich am schwierigsten zu erfassende Gütekriterium psychometrischer Messverfahren; sie setzt die Objektivität und Reliabilität des Verfahrens voraus (BORTZ & DÖRING, 2006). Im Revisionsprojekt zur optimierten Praktischen Fahrerlaubnisprüfung wurden daher in verschiedenen Untersuchungen Daten gesammelt, die Rückschlüsse auf die Validität der Prüfung zulassen. Dazu zählen eine Befragung der Fahrlehrer zur Fahrkompetenz ihrer Fahrschüler und eine Anwendung der Technik der bekannten Gruppen im Rahmen der simulierten Prüfungen (s. Kapitel 4.1.1), eine Analyse der Lernverläufe im Rahmen der Beurteilung videografiert Prüfungen (s. Kapitel 4.1.2) und eine Untersuchung der populationsspezifischen Äquivalenz im Rahmen der Felderprobung (s. Kapitel 4.5). Im Folgenden sollen die Ergebnisse dieser Untersuchungen erläutert werden.

### 5.2.2 Überprüfung der Validität anhand einer Fahrlehrerbefragung

Im Rahmen der ersten Untersuchung zur Beobachterübereinstimmung (simulierte Prüfungen der Fahrerlaubnisklasse B) wurden die Fahrlehrer der teilnehmenden Fahrschüler mittels eines Fragebogens zur Fahrkompetenz ihrer Schüler befragt. Der Fragebogen enthielt – analog zu den Bewertungen, welche die Fahrerlaubnisprüfer im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung vornehmen – Fragen zur situationsübergreifenden Gesamtbewertung der (Teil-)Fahraufgaben<sup>37</sup> und der Beobachtungskategorien bzw. Fahrkompetenzbereiche der Fahrschüler. Diese Bewertungen wurden von den Fahrlehrern vor der Prüfung (er-

<sup>37</sup> Einige Fahraufgaben setzen sich aus mehreren Teilfahraufgaben zusammen. Um eine Gesamtbewertung der Fahraufgaben zu erhalten, wurde der Durchschnitt der Bewertungen der dazugehörigen Teilfahraufgaben berechnet und zur nächstliegenden ganzen Zahl gerundet.



fahrbasiert) und nach der Prüfung (prüfungsbasiert) vorgenommen. Die Fragebogenversion für die prüfungsbasierten Einschätzungen enthielt zusätzlich eine Frage zur erwarteten Prüfungsentcheidung.

Die Einschätzungen der Fahrleistungen durch die Fahrlehrer erfolgten völlig unabhängig von den Bewertungen der Fahrerlaubnisprüfer und stellen ein Außenkriterium zur Validierung der Fahrkompetenzmessung mittels der optimierten Praktischen Fahrerlaubnisprüfung dar: Würden die Bewertungen der Fahrlehrer mit den Bewertungen der Fahrerlaubnisprüfer korrespondieren, könnte dies als ein Indiz für eine valide Fahrkompetenzmessung betrachtet werden. Im Vergleich zur Beobachterübereinstimmung zwischen den Fahrerlaubnisprüfern (s. o.) wären hierbei allerdings geringere Übereinstimmungen als ausreichend anzusehen, da die Fahrlehrer bei der Beurteilung der Fahrkompetenz ihrer Fahrschüler wahrscheinlich zwar ähnliche, aber nicht die gleichen Maßstäbe anlegen wie die Fahrerlaubnisprüfer. So ist davon auszugehen, dass Fahrlehrer in erster Linie pädagogisch motiviert handeln und Fahrschüler, die sich selbst überschätzen, kritischer bewerten als unsichere Schüler, die in stärkerem Maße emotionale Unterstützung durch positive Leistungsrückmeldungen benötigen. Darüber hinaus haben die Fahrlehrer im Verlauf der Ausbildung ein umfassendes Leistungsbild vom Bewerber und von seinem Leistungsfortschritt gewonnen, was ihnen eine eher situationsunabhängige, differenzierte Einschätzung ermöglicht. Die Fahrerlaubnisprüfer hingegen sind beauftragt, nur solchen Personen ausreichende Leistungen zu bescheinigen, die im Rahmen der zeitlich befristeten und nicht selten Prüfungsangst auslösenden Prüfungssituation nachgewiesen haben,

dass sie zur sicheren Teilnahme am motorisierten Straßenverkehr in der Lage sind. Hierbei müssen die Prüfer kriteriengeleitet beurteilen, ob der Bewerber bestimmte, vom Ordnungsgeber festgelegte Anforderungen erfüllt hat; pädagogische Überlegungen spielen im Kontext dieser Leistungsbewertung nur eine untergeordnete Rolle.

### Vergleich von erfahrungsbasierten und prüfungsbasierten Einschätzungen der Fahrlehrer

Die erfahrungsbasierten Einschätzungen der Fahrlehrer beruhen auf der Kenntnis des Fahrkompetenzstatus und der Lernverläufe ihrer Fahrschüler in der fahrpraktischen Ausbildung („aggregiertes Urteil“). Die prüfungsbasierten Einschätzungen der Fahrlehrer fußen hingegen ausschließlich auf den Leistungen der Fahrschüler im Rahmen der Prüfung. Ein Vergleich der erfahrungsbasierten mit den prüfungsbasierten Einschätzungen der Fahrlehrer gibt daher Hinweise darauf, inwieweit die Prüfungsleistungen mit den Ausbildungsleistungen korrespondieren (s. Tabelle 16).

Bezüglich der Gesamtbewertung der Fahraufgaben<sup>38</sup> zeigten sich – erneut quantifiziert anhand des quadratisch gewichteten Kappa-Koeffizienten – ausschließlich geringe und nicht signifikante Übereinstimmungen zwischen den erfahrungsbasierten und den prüfungsbasierten Einschätzungen der Fahrlehrer. Für die Gesamtheit aller Fahraufgaben (ohne Geradeausfahren) liegt der Kappa-Wert bei 0,30.

<sup>38</sup> Auf die Auswertung der Fahraufgabe „Geradeausfahren“ wurde verzichtet, weil diese Fahraufgabe eine „Restkategorie“ mit vielfältigen Anforderungen darstellt, deren Bewertungsregeln den Fahrlehrern nicht bekannt sind.

Fahraufgaben	Kappa	n	p	Abweichungen
Einfädelungs-/Ausfädelungsstreifen, Fahrstreifenwechsel	0,15	15	0,53	1(5); 2(1)
Kurven	0,34	14	0,14	1(5); 2(1)
Vorbeifahren, Überholen	0,25	15	0,28	1(6)
Kreuzungen, Einmündungen	0,26	15	0,25	1(4); 2(1)
Kreisverkehr	-	-	-	-
Schienenverkehr	-	-	-	-
Haltestellen, Fußgänger, Radfahrer	0,35	15	0,13	1(6)
Gesamt	0,30	78	0,01**	1(31)

Tab. 16: Übereinstimmung zwischen erfahrungsbasierten und prüfungsbasierten Fahrkompetenzeinschätzungen der Fahrlehrer bei der Gesamtbewertung der Fahraufgaben

Beobachtungskategorien	Kappa	N	p	Abweichungen
Verkehrsbeobachtung	-0,13	15	0,34	1(7); 2(2)
Fahrzeugpositionierung	0,69	15	0,01**	1(4)
Geschwindigkeitsanpassung	-0,10	15	0,70	1(8); 2(1)
Kommunikation	0,20	15	0,31	1(8)
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,15	15	0,53	1(5); 2(1)
Gesamt	0,14	75	0,23	1(32); 2(3)

Tab. 17: Übereinstimmung (quadratisch gewichtetes Kappa) zwischen erfahrungsbasierten und prüfungsbasierten Fahrkompetenzeinschätzungen der Fahrlehrer bei der Gesamtbewertung der Beobachtungskategorien

Zwar ist dieses Ergebnis aufgrund der größeren Fallzahl signifikant, die Stärke des Zusammenhangs muss aber als unzureichend betrachtet werden.

Auch bei den Gesamtwertungen der Kompetenzbereiche bzw. Beobachtungskategorien (s. Tabelle 17) zeigten sich, abgesehen von der Kategorie „Fahrzeugpositionierung“, keine nennenswerten Übereinstimmungswerte zwischen den erfahrungsbasierten und den prüfungsbasierten Einschätzungen der Fahrlehrer. Ob und ggf. wie sich der signifikante und relativ hohe Kappa-Wert für die „Fahrzeugpositionierung“ von 0,69 inhaltlich erklären lässt, müsste genauer untersucht werden. Die Wahrscheinlichkeit, dass diese Übereinstimmung im Rahmen der relativ kleinen Stichprobe zufällig entstanden ist, liegt bei 0,8 Prozent.

Zusammenfassend kann festgehalten werden, dass sich keine nennenswerten Übereinstimmungen zwischen den erfahrungsbasierten und den prüfungsbasierten Einschätzungen der Fahrkompetenz der Fahrschüler durch die Fahrlehrer finden lassen. Dies bedeutet, dass die Prüfungsreifeinschätzungen der Fahrlehrer der Stichprobe im Hinblick auf die Prüfungsleistungen prognostisch nicht belastbar waren.

### Vergleich von erfahrungsbasierten Einschätzungen der Fahrlehrer und Prüfungsbewertungen der Fahrerlaubnisprüfer

Hinsichtlich des Vergleichs zwischen den (erfahrungsbasierten) Fahrkompetenzeinschätzungen durch die Fahrlehrer und den Prüfungsbewertungen durch die Fahrerlaubnisprüfer<sup>39</sup> wurde bereits vermutet, dass die Bewertungsmaßstäbe bei der Beurteilung der Fahrkompetenz voneinander abweichen könnten. Tatsächlich haben die Fahrlehrer ihre Schüler (auf einer vierstufigen Bewertungsskala von „0“ bis „3“) bei den Fahraufgaben um durchschnitt-

lich ca. 0,84 Skalenstufen und bei den Beobachtungskategorien um ca. 0,93 Skalenstufen besser bewertet als die Fahrerlaubnisprüfer. Statt exakter Übereinstimmungen wurden daher relative Übereinstimmungen untersucht. Kendall's Tau ist ein geeignetes justiertes Maß für Zusammenhänge zwischen ordinalskalierten Variablen bei Fallzahlen, die unter 30 liegen<sup>40</sup> (WIRTZ & CASPAR, 2002). Im Ergebnis zeigten sich auch beim Vergleich der erfahrungsbasierten Einschätzungen der Fahrlehrer mit den Prüfungsbewertungen der Fahrerlaubnisprüfer weder bezüglich der Fahraufgaben noch bezüglich der Kompetenzbereiche bzw. Beobachtungskategorien nennenswerte Übereinstimmungen.

Festzuhalten bleibt, dass die ausbildungsbezogenen Fahrkompetenzeinschätzungen der Fahrlehrer in der vorliegenden Studie anscheinend keine belastbare Prognose der Prüfungsleistungen in der optimierten Praktischen Fahrerlaubnisprüfung erlaubt haben. Ein erster Erklärungsansatz hierfür wäre, dass die Anforderungen der Praktischen Fahrerlaubnisprüfung zwar im Grundsatz für jeden Bewerber bzw. in jeder Prüfung gleich sind, sich in der konkreten Umsetzung jedoch von Fall zu Fall unterscheiden: Die Prüfung findet im realen Straßenverkehr unter variierenden

<sup>39</sup> Hierbei waren die Bewertungen des (nicht „stillen“) Fahrerlaubnisprüfers maßgeblich, der die Prüfung geleitet hat. Da die Übereinstimmungen zwischen den beiden Fahrerlaubnisprüfern aber sehr hoch sind (s. o.), würden sich die Ergebnisse auch nicht grundlegend ändern, wenn der „stille“ Fahrerlaubnisprüfer bzw. der Durchschnitt beider Fahrerlaubnisprüfer als Vergleichsmaßstab herangezogen würde.

<sup>40</sup> Kendall's Tau verwendet die gleiche Skala („-1“ bis „+1“) und ist ähnlich zu interpretieren wie Cohen's Kappa, d. h. der Wert „-1“ entspricht absolut keiner Übereinstimmung, „0“ entspricht der zufällig erwarteten Übereinstimmung, und „1“ entspricht einer perfekten Übereinstimmung.

Witterungs-, Straßen und Verkehrsverhältnissen statt. Einzelne, für das Bestehen oder Nichtbestehen des Bewerbers unter Umständen entscheidende Ereignisse sind nicht vorhersehbar. Dementsprechend können sie vom Fahrlehrer bei der erfahrungsbasierten Einschätzung der Fahrkompetenz des Bewerbers nicht berücksichtigt werden, während sie die prüfungsbasierten Einschätzungen des Fahrlehrers und des Fahrerlaubnisprüfers maßgeblich beeinflussen. Ein zweiter Grund für die fehlenden Zusammenhänge könnte in der Prüfungssituation zu finden sein: Durch Versagensängste verhalten sich Menschen in Stresssituationen – wie einer Praktischen Fahrerlaubnisprüfung – oftmals anders als unter „gewöhnlichen“ Umständen (z. B. bei der fahrpraktischen Ausbildung). Dies kann einerseits zu einer erhöhten Aufmerksamkeit der Bewerber und weniger Fahrfehlern führen; andererseits kann Prüfungsangst den Bewerber verunsichern und Fahrfehler verursachen. Drittens schließlich mangelt es in der Fahrschulausbildung an methodisch anspruchsvollen lernstandsdiagnostischen Verfahren zur Prüfungsreifefeststellung; den Fahrlehrern fehlen also objektive Einschätzungskriterien. Daher bleibt zu hoffen, dass der methodische Aufschwung der optimierten Praktischen Fahrerlaubnisprüfung auch dazu führt, dass sich die erfolgreich erprobten Durchführungs-, Anforderungs- und Bewertungsstandards der Prüfung in den Verfahren der Fahrlehrer zur Prüfungsreifefeststellung widerspiegeln.

### Vergleich von prüfungsbasierten Einschätzungen der Fahrlehrer und Fahrerlaubnisprüfer

Die Übereinstimmung zwischen den prüfungsbasierten Einschätzungen der Fahrlehrer und Fahr-

erlaubnisprüfer kann als wichtiges Kriterium für die Validität der Fahrkompetenzmessung mittels der optimierten Praktischen Fahrerlaubnisprüfung betrachtet werden: Sowohl die Fahrlehrer als auch die Fahrerlaubnisprüfer können als Experten für Fahrleistungseinschätzungen gelten; bewerten sie dieselben Prüfungsleistungen, sollten sie auch zum gleichen Bewertungsergebnis kommen. Unterschiede könnten allerdings daraus resultieren, dass die einheitliche Anwendung der inhaltlichen und methodischen Standards zur Prüfungsbeobachtung und Prüfungsbewertung vom Fahrerlaubnisprüfer in weitaus stärkerem Maße geübt und häufiger praktiziert wird.

Die Prüfungsentscheidung stellt für den Bewerber die wichtigste Ebene der Prüfungsbewertung dar. Sie ist eine dichotome Variable mit nur zwei Ausprägungen. Eine justierte Analyse der Zusammenhänge ist dabei nicht sinnvoll, weshalb nicht Kendall's Tau, sondern Cohen's Kappa als Quantifizierungsmaß verwendet wurde. Die Übereinstimmung zwischen den vom jeweiligen Fahrlehrer erwarteten Prüfungsentscheidungen und den tatsächlichen Prüfungsentscheidungen der Fahrerlaubnisprüfer beträgt 0,87 ( $n = 15$ ;  $p = 0,00^{***}$ ; Abweichungen = 1) und kann somit als hoch bezeichnet werden (s. o.). Die Irrtumswahrscheinlichkeit liegt dabei unter 0,1 Prozent; das Ergebnis ist hochsignifikant.

Weniger stark fallen die Zusammenhänge zwischen den prüfungsbasierten Fahrkompetenzeinschätzungen der Fahrlehrer und Fahrerlaubnisprüfer bei der Gesamtbewertung der Fahraufgaben aus. Wie der Tabelle 18 entnommen werden kann, zeigten sich teilweise schwache und teilweise gerade noch akzeptable Zusammenhänge, wobei nicht alle Ergebnisse statistisch signifikant sind. Besonders niedrig ist der Zusammenhang beim „Geradeausfahren“, was vermutlich mit den

Fahraufgaben	Tau	n	p
Einfädlungs-/Ausfädungsstreifen, Fahrstreifenwechsel	0,41	15	0,06
Kurven	0,59	14	0,01**
Vorbeifahren, Überholen	0,48	15	0,03*
Kreuzungen, Einmündungen	0,39	15	0,06
Kreisverkehr	-	-	-
Schienenverkehr	-	-	-
Haltestellen, Fußgänger, Radfahrer	0,40	15	0,62
Geradeausfahren	0,30	15	0,11

Tab. 18: Zusammenhänge zwischen den prüfungsbasierten Fahrkompetenzeinschätzungen der Fahrlehrer und Fahrerlaubnisprüfer bei der Gesamtbewertung der Fahraufgaben

Beobachtungskategorien	Tau	n	p
Verkehrsbeobachtung	0,47	15	0,03*
Fahrzeugpositionierung	0,64	15	0,01**
Geschwindigkeitsanpassung	0,53	15	0,02*
Kommunikation	0,00	15	0,50
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,61	15	0,01**

**Tab. 19:** Zusammenhänge zwischen den prüfungsbasierten Fahrkompetenzeinschätzungen der Fahrlehrer und Fahrerlaubnisprüfer bei der Gesamtbewertung der Beobachtungskategorien

bereits erwähnten Interpretationsschwierigkeiten der Fahrlehrer bei der Bewertung dieser Fahraufgabe in Zusammenhang steht.

Neben dem Vergleich der fahraufgabenbezogenen Prüfungseinschätzungen von Fahrlehrern und Fahrerlaubnisprüfern besitzt auch der Vergleich der kompetenzbezogenen Prüfungseinschätzungen eine hohe Relevanz für die Validitätsbeurteilung: Aufgrund des höheren Generalisierungsgrads von Kompetenzeinschätzungen ist diesem Vergleich wahrscheinlich sogar eine höhere Bedeutung beizumessen. Der Tabelle 19 ist zu entnehmen, dass sich für fast alle Kompetenzbereiche bzw. Beobachtungskategorien signifikante Zusammenhänge mit akzeptablen Zusammenhangsstärken zeigten. Einzige Ausnahme ist der Kompetenzbereich „Kommunikation“, für den sich in der Stichprobe kein Zusammenhang nachweisen ließ. Ein vertiefender Blick in die Struktur der Daten zeigt, dass dieser Kompetenzbereich von den Fahrerlaubnisprüfern in 14 von 15 Fällen mit „Ausreichend“ bewertet wurde. Bei der Berechnung von Kendall’s Tau werden die Ränge von Beobachtungen zwischen den beiden Beobachtern verglichen. Bewertet der erste Beobachter 14 von 15 Fällen identisch, kann überhaupt nur in einem Fall eine Rangabstufung festgestellt und verglichen werden. Diese 15. Beobachtung liegt für die Fahrerlaubnisprüfer unter den 14 anderen und für den betroffenen Fahrlehrer darüber („Inversion“). Daraus ergibt sich ein Kendall’s Tau von 0 für die Beobachtungskategorie „Kommunikation“, was allerdings weniger inhaltlichen Abweichungen zwischen den Einschätzungen der Fahrerlaubnisprüfer und Fahrlehrer geschuldet ist als einer ausgeprägten Gleichverteilung der Daten.

Zusammenfassend bleibt festzuhalten, dass sich akzeptable Zusammenhänge zwischen den prüfungsbasierten Fahrlehrereinschätzungen und Prüferereinschätzungen nachweisen ließen: Fast

alle Koeffizienten fallen signifikant aus und liegen deutlich über dem Wert von 0,40. Dieser Befund kann auch deshalb als zufriedenstellend gelten, weil die Fahrlehrer nicht in gleicher Weise wie die Fahrerlaubnisprüfer in der Beurteilung von Fahrkompetenz geschult wurden (s. o.). Die Ergebnisse der vergleichenden Validitätsuntersuchungen von prüfungsbasierten Fahrlehrer- und Prüferereinschätzungen stellen damit ein Indiz für eine valide Messung der Fahrkompetenz bei der optimierten Praktischen Fahrerlaubnisprüfung dar.

### 5.2.3 Überprüfung der Validität anhand der „Technik der bekannten Gruppen“

Bei der „Technik der bekannten Gruppen“ geht es darum, zwei Gruppen zu unterscheiden, die sich augenscheinlich durch unterschiedliche Kompetenzniveaus im Gegenstandsbereich auszeichnen (SCHNELL et al., 2005). Im Anschluss wird die Hypothese geprüft, dass das Kompetenzniveau der Bewerber – operationalisiert über die Einschätzung der Fahrkompetenz (abhängige Variable) – mit den Prüfungsergebnissen (unabhängige Variablen) zusammenhängt. Fänden sich signifikante Zusammenhänge in fachlich akzeptabler Stärke, würde dies die Annahme einer validen Fahrkompetenzmessung bei der optimierten Praktischen Fahrerlaubnisprüfung stützen. Die an der ersten Untersuchung im Realverkehr (simulierte Prüfungen) teilnehmenden Fahrlehrer wurden im Vorfeld gebeten, jeweils zur Hälfte Fahrschüler für die Prüfungen vorzustellen, die gemäß festgelegter Kriterien als „Sehr kompetent“ bzw. als „Wenig kompetent“ einzustufen sind. Die Fahrkompetenzniveaus nach STURZBECHER et al. (2014) bildeten dabei die wissenschaftlich begründete Basis für die in der Tabelle 20 dargestellten Kategorien.

Während die als „Sehr kompetent“ eingestuften Fahrschüler – nach Einschätzung ihrer Fahrlehrer – kurz vor der Praktischen Fahrerlaubnisprüfung



Kompetenzniveau	Bewerbermerkmale
Sehr kompetent	Der Fahrschüler verhält sich in (fast) allen Verkehrssituationen vorausschauend, richtig und effektiv.
Wenig kompetent	Der Fahrschüler verhält sich in standardmäßigen Verkehrssituationen (d. h. ohne besondere Anforderungen) überwiegend vorausschauend, richtig und effektiv. In ungewohnten – d. h. in der Ausbildung eher selten vorkommenden – oder komplexen Situationen treten leichte bzw. schwere Fehler auf. Für die selbstständige motorisierte Verkehrsteilnahme sind Verbesserungen notwendig.

Tab. 20: Kriterien zur Auswahl von „Sehr kompetenten Fahrschülern“ und „Weniger kompetenten Fahrschülern“

	Fahraufgaben gesamt	Beobachtungskategorien gesamt
Median „Sehr kompetent“	1	1
Median „Wenig kompetent“	1	0

Tab. 21: Median der Gesamtbewertung der Fahraufgaben und Beobachtungskategorien durch die Fahrerlaubnisprüfer gegliedert nach dem Kompetenzniveau der Fahrschüler (Skala: 0 = „Ungenügend“, 1 = „Ausreichend“, 2 = „Gut“, 3 = „Sehr gut“)

stehen (Kategorie „Prüfungsreif“), sind die als „Wenig kompetent“ eingestuftten Fahrschüler nur unter Aufsicht ihrer Fahrlehrer zur sicheren Teilnahme am Straßenverkehr in der Lage („Nicht prüfungsreif“).<sup>41</sup> Für die zweite Gruppe von Fahrschülern bzw. für ihre Fahrlehrer war die Teilnahme an den simulierten Prüfungen anscheinend deutlich weniger attraktiv, weswegen 13 von 15 simulierten Prüfungen mit als „Sehr kompetent“ eingeschätzten Fahrschülern durchgeführt wurden. Auf Basis zweier so ungleich besetzter Gruppen, von denen eine nur zwei Probanden enthält, lassen sich keine statistisch belastbaren Ergebnisse berechnen. Im Folgenden werden daher – zur Erhöhung der Fallzahlen – vorwiegend aggregierte Ergebnisse dargestellt. Die zwei als „Wenig kompetent“ eingestuftten Fahrschüler haben die simulierte Prüfung nicht bestanden (0 %). In der Gruppe der als „Sehr kompetent“ eingeschätzten Schüler lag die Bestehensquote bei 46 Prozent. Somit besteht ein erheblicher Unterschied zwischen den beiden Gruppen.

Möchte man die Leistungen der beiden Gruppen bezüglich der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien vergleichen, erscheint es sinnvoll, ein Maß der zentralen Tendenz für beide Gruppen zu berechnen. Die Gesamtbewertung der Fahraufgaben und Kompetenzbereiche nahmen die Fahrerlaubnisprüfer anhand einer vierstufigen Ordinalskala mit den Stufen „Sehr gut“, „Gut“, „Ausreichend“ und „Ungenügend“ vor. Für ordinalskalierte Variablen ist das geeignete Maß der zentralen Tendenz der Median.<sup>42</sup>

Wie die Tabelle 21 zeigt, unterscheiden sich auf der Grundlage des Medians die als „Sehr kompetent“

und die als „Wenig kompetent“ eingestuftten Fahrschüler nicht in Bezug auf die Gesamtbewertung der Fahraufgaben: Beide Gruppen tendieren im Mittel zu ausreichenden Leistungen. Unterschiede hingegen gibt es bei der Gesamtbewertung der Kompetenzbereiche. Hier liegt der Median für die als „Sehr kompetent“ eingeschätzten Schüler bei „Ausreichend“; der Median für die als „Wenig kompetent“ eingestuftten Schüler dagegen bei „Ungenügend“. Der Unterschied beträgt somit eine Kategorienstufe. Geht man davon aus, dass die Fahrkompetenzbewertungen einen höheren Generalisierungsgrad aufweisen und sich über alle Fahraufgaben erstrecken, deuten die Befunde darauf hin, dass sich die Gruppen in ihren Fahrleistungen bzw. Fahrkompetenzen zugunsten der als „Sehr kompetent“ eingestuftten Fahrschüler unterscheiden.

Verknüpft man die Stufen der Ordinalskala mit numerischen Werten – z. B. von „0“ bis „3“ – lässt sich das arithmetische Mittel berechnen, welches differenziertere Aussagen als der Median ermöglicht. Dieses Vorgehen ist in den Sozialwissenschaften für Ordinalskalen durchaus gebräuch-

<sup>41</sup> Gemäß § 6 der Fahrschüler-Ausbildungsordnung dürfen nur Personen zur Prüfung zugelassen werden, die den Unterricht im gesetzlich vorgeschriebenen Umfang absolviert haben sowie zur sicheren, verantwortungsvollen und umweltbewussten Verkehrsteilnahme befähigt sind.

<sup>42</sup> Der Median ist diejenige Bewertungskategorie, ober- und unterhalb welcher jeweils die Hälfte der Beobachtungen liegt. Der Nachteil des Medians liegt darin, dass er bei einer vierstufigen Skala nur vier Ausprägungen annehmen kann und eine feinstufige Differenzierung zwischen zwei Gruppen nicht möglich ist.

	Fahraufgaben gesamt	Beobachtungskategorien gesamt
Arithmetisches Mittel „Sehr kompetent“	0,85	0,85
Arithmetisches Mittel „Wenig kompetent“	0,57	0,40

Tab. 22: Arithmetisches Mittel der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche durch die Fahrerlaubnisprüfer (gegliedert nach dem Kompetenzniveau der Fahrschüler; Skala: 0 = „Ungenügend“, 1 = „Ausreichend“, 2 = „Gut“, 3 = „Sehr gut“)

lich.<sup>43</sup> Wie der Tabelle 22 zu entnehmen ist, schneiden die als „Sehr kompetent“ eingeschätzten Fahrschüler bei der Gesamtbewertung der Fahraufgaben um 0,28 Skalenstufen und bei der Gesamtbewertung der Kompetenzbereiche um 0,45 Skalenstufen besser ab als die „Wenig kompetenten Fahrschüler“. Dies muss vor dem Hintergrund der Tatsache, dass die Fahrerlaubnisprüfer die Bewertungsstufen „Gut“ und „Sehr gut“ nur selten ausgewählt haben, die meisten Bewertungen also – numerisch ausgedrückt – bei „0“ oder „1“ lagen, als deutlicher Unterschied bezeichnet werden.

Zusammenfassend bleibt festzuhalten, dass die Fahrkompetenz der (vorab) als „Wenig kompetent“ eingestuftten Fahrschüler im Vergleich zur Fahrkompetenz der „Sehr kompetenten Fahrschüler“ von den Fahrerlaubnisprüfern erheblich schlechter eingeschätzt wurde. Dies betrifft sowohl die Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche als auch die Prüfungsentscheidungen. Vor dem Hintergrund der geringen Fallzahlen sind diese statistischen Befunde aber mit einer gewissen Unsicherheit behaftet. Die dargestellten Befunde aus der Anwendung der „Technik der bekannten Gruppen“ im Rahmen der simulierten Prüfungen stellen demzufolge weitere empirische Indizien, aber keinen belastbaren Nachweis für die Validität der optimierten Praktischen Fahrerlaubnisprüfung dar.

#### 5.2.4 Lernverläufe im Rahmen der Beurteilung videografiert Prüfungen

Im Rahmen der Untersuchung zur Ermittlung der Beobachterübereinstimmung anhand videografiert Prüfungen wurden die teilnehmenden Fahr-

erlaubnisprüfer gebeten, die Prüfungsleistungen aus jeweils drei Prüfungsfahrtvideos zu insgesamt drei verschiedenen Messzeitpunkten mittels e-Prüfprotokoll zu dokumentieren und zu bewerten. Zwischen den Messungen lagen Zeiträume von jeweils ca. acht Wochen, in denen die Prüfer im Rahmen der Felderprobung Praktische Fahrerlaubnisprüfungen mittels e-Prüfprotokoll abgenommen haben (s. Kapitel 4.1.2). Dabei ist anzunehmen, dass die Prüfer zwischen den Messzeitpunkten durch die umfassende praktische Anwendung des e-Prüfprotokolls Lernerfolge hinsichtlich der korrekten Anwendung des optimierten Prüfkonzepts erzielen konnten. Testpsychologisch gesehen bedeutet dies, dass der Einfluss von – durch mangelnde Erfahrung bedingten – Störfaktoren auf die Messung abnehmen und der Einfluss des eigentlich zu erhebenden Merkmals – der Fahrkompetenz – zunehmen sollte: Auf Erfahrungsmangel basierende Störfaktoren wirken sich in der Regel interindividuell verschieden aus, d. h. die einzelnen Prüfer begehen nicht die gleiche Art von Messfehlern. Die Fahrkompetenz des zu beurteilenden Bewerbers ist hingegen für alle bewertenden Prüfer gleich. Daher sollte die Beobachterübereinstimmung zwischen den Prüfern mit zunehmender Übung im Umgang mit der optimierten Prüfung steigen.

Wie in Kapitel 5.1.2 dargestellt wurde, hat die Beobachterübereinstimmung bei der Bewertung videografiert Prüfungen über die drei Messzeitpunkte hinweg insgesamt deutlich zugenommen. Dies kann folglich als ein weiteres Indiz für eine valide Fahrkompetenzmessung mittels der optimierten Praktischen Fahrerlaubnisprüfung gewertet werden.

<sup>43</sup> Das arithmetische Mittel darf streng genommen nur für metrisch skalierte Variablen berechnet werden. In der sozialwissenschaftlichen Praxis ist es aber schon länger Usus, parametrische Verfahren auch auf ordinalskalierte Variablen anzuwenden, wenn für die Zielstellung keine vergleichbaren nichtparametrischen Verfahren verfügbar sind oder diese weniger aussagekräftige Ergebnisse liefern. LABOVITZ (1967) und KIM (1975) zeigen die Robustheit vieler parametrischer Verfahren gegenüber der Verletzung dieser Anwendungsvoraussetzung.

#### 5.2.5 Untersuchung der populations-spezifischen Äquivalenz im Rahmen der Felderprobung

Die optimierte Praktische Fahrerlaubnisprüfung besitzt populations-spezifische Äquivalenz, wenn Merkmale, die für die Aneignung von Fahrkompetenz inhaltlich irrelevant sind, keinen Einfluss auf

die Prüfungsergebnisse haben. Solche Merkmale dürften beispielsweise das Geschlecht oder die Haarfarbe des Bewerbers sein. Fänden sich derartige Zusammenhänge, würde dies die Validität der Prüfung einschränken, da sie in diesem Fall nicht nur die Fahrkompetenz der Bewerber, sondern auch Beurteilungsfehler der Fahrerlaubnisprüfer erfasst. Umgekehrt existieren Bewerbermerkmale wie die Fahrerfahrung oder der Vorbesitz einer anderen Fahrerlaubnis, welche die Aneignung von Fahrkompetenz begünstigen sollten. Zusammenhänge zwischen solchen förderlichen Bewerbermerkmalen und den Prüfungsergebnissen könnten ggf. als Indiz für eine valide Prüfung gedeutet werden. Die Ergebnisse der Untersuchungen zur populations-spezifischen Äquivalenz finden sich im Kapitel 5.3.

### 5.3 Populationsspezifische Äquivalenz

#### 5.3.1 Überblick

Bei der Gütebeurteilung psychometrischer Messverfahren spielen – neben den drei Hauptkriterien „Objektivität“, „Reliabilität“ und „Validität“ – auch bestimmte Nebengütekriterien eine wichtige Rolle. Wie STURZBECHER et al. (2014) begründet haben, ist die populationsspezifische Äquivalenz im Kontext der Praktischen Fahrerlaubnisprüfung von hoher Bedeutung. Auf den engen inhaltlichen Zusammenhang zwischen der Validität und der populationsspezifischen Äquivalenz eines Verfahrens wurde bereits hingewiesen.

Im vorliegenden Revisionsprojekt wurden für die Analysen der populationsspezifischen Äquivalenz folgende Bewerbermerkmale ausgewählt und im Kapitel 3.5 begründet: Bildungsniveau, Prüfungsangst, Alter, Geschlecht und Migrationshintergrund. Fände sich ein (unerwarteter) Zusammenhang zwischen einem dieser Bewerbermerkmale und den Prüfungsergebnissen, wäre durch weiterführende Untersuchungen zu erhellen, ob sich dies inhaltlich erklären lässt. So könnte beispielsweise ein alters-typisch vermindertes Konzentrationsvermögen ggf. zu schlechteren Prüfungsergebnissen älterer Bewerber führen. Falls für gefundene Zusammenhänge zwischen Bewerbermerkmalen und Prüfungsergebnissen jedoch Diskriminierungsprozesse ursächlich wären, läge eine unsachgemäße Anwendung der optimierten Praktischen Fahrerlaubnisprüfung vor. Das würde die Validität der Prüfung einschränken, da sie in diesem Fall offenkundig

nicht nur die Fahrkompetenz der Bewerber, sondern auch Beurteilungsfehler der Fahrerlaubnisprüfer erfassen würde (s. o.).

Bestehen keine Zusammenhänge zwischen den Prüfungsergebnissen und Merkmalen, die keinen inhaltlich begründbaren Einfluss auf die Fahrkompetenz haben, ist eine wichtige Voraussetzung für die Validität des Verfahrens erfüllt. Die Untersuchung des Zusammenhangs der Testergebnisse mit theoretisch irrelevanten Merkmalen bezeichnet man als „diskriminante Validierung“, die einen wichtigen Bestandteil der Konstruktvalidierung darstellt (s. Kapitel 3.4). Die zugehörigen Bewerbermerkmale werden daher im Folgenden als „diskriminante Merkmale“ bezeichnet. Im Umkehrschluss sind aber auch verifizierende Befunde möglich, wenn Bewerbermerkmale, welche die Aneignung von Fahrkompetenz begünstigen sollten, in einem positiven Zusammenhang mit den Prüfungsergebnissen stehen. Dies wird als „konvergente Validierung“ bezeichnet; die entsprechenden Variablen heißen „konvergente Merkmale“. Hierzu zählen im vorliegenden Fall der Vorbesitz einer Fahrerlaubnis, die Vorfahrerfahrung und der regelmäßige Konsum straßenverkehrsbezogener Computerspiele. In Kapitel 3.4 wurde begründet, warum sich diese Merkmale positiv auf die Aneignung von Fahrkompetenz auswirken sollten.

Im vorliegenden Kapitel werden die Ergebnisse der statistischen Berechnungen zur populationsspezifischen Äquivalenz der optimierten Praktischen Fahrerlaubnisprüfung im Hinblick auf die oben genannten Bewerbermerkmale dargestellt und interpretiert. Dabei werden zunächst jene Merkmale betrachtet, für die kein Zusammenhang mit den Prüfungsergebnissen erwartet wird (diskriminante Merkmale), und anschließend jene, für die ein positiver Zusammenhang erwartet wird (konvergente Merkmale).

#### 5.3.2 Diskriminante Bewerbermerkmale

Bei der Analyse der Zusammenhänge von Bewerbermerkmalen und Prüfungsergebnissen bestimmt das Skalenniveau der einbezogenen Variablen über das anzuwendende statistische Verfahren. Die Prüfungsentscheidung als wichtigste Ebene der Prüfungsbewertung besitzt ein nominales Skalenniveau. Die Bewerbermerkmale weisen im vorliegenden Fall ein metrisches (Alter), ein ordinales (Prüfungsangst, Bildungsniveau) oder ein nomina-

les Skalenniveau (Migrationshintergrund, Geschlecht) auf. Insofern sind zur Analyse der (bivariaten) Zusammenhänge mit der Prüfungsentscheidung verschiedene statistische Verfahren angezeigt. Im vorliegenden Fall wären das Mittelwertvergleiche mit T-Tests (bei metrischem Skalenniveau), Kreuztabellen mit U-Tests (bei ordinalem Skalenniveau) und Kreuztabellen mit Chi-Quadrat-Tests (bei nominalem Skalenniveau). Der Nachteil an diesem Vorgehen ist, dass sich die Ergebnisse verschiedenartiger Verfahren nur begrenzt vergleichen lassen. Zur besseren Vergleichbarkeit werden daher im Folgenden alle nicht nominalen Bewerbermerkmale dichotomisiert und somit auf nominales Skalenniveau reduziert. Bevor die Zusammenhänge der diskriminanten Bewerbermerkmale mit den Prüfungsergebnissen analysiert werden, soll die Verteilung der Merkmale zunächst (univariat) beschrieben und das Vorgehen bei der Dichotomisierung begründet werden.

Im Fall des Bildungsniveaus wurde die vierstufige Einteilung in „Hauptschulabschluss“, „Mittlere Reife“, „Abitur/Fachabitur“ und „Anderer Abschluss“ aufgehoben und nur zwischen „Abitur“ (55,4 %) und „Kein Abitur“ (44,6 %) unterschieden. Die Grenze wurde an dieser Stelle gezogen, da zum einen der Unterschied zwischen „Hauptschulabschluss“ (9 Bildungsjahre) und „Realschulabschluss“ (10 Bildungsjahre) als geringer einzustufen ist als zwischen „Realschulabschluss“ und „Abitur/Fachabitur“ (12 bzw. 13 Bildungsjahre). Zum anderen wurde die Gruppe der Bewerber so in zwei etwa gleich große Subpopulationen geteilt, was einen Vorteil für die statistische Analyse bedeutet.

Hinsichtlich der Prüfungsangst wurde ursprünglich zwischen „Nicht aufgeregt“, „Kaum aufgeregt“, „Etwas aufgeregt“ und „Sehr aufgeregt“ unterschieden. Da der Großteil der Menschen in Prüfungssituationen zumindest „Etwas aufgeregt“ sein dürfte, erschien in diesem Fall die Unterteilung in „Sehr aufgeregt“ (41,6 %) und „Nicht sehr aufgeregt“

(58,4 %) sinnvoll, was auch wieder den Vorteil zweier ungefähr gleichgroßer Subgruppen mit sich bringt.

Die Variablen „Migrationshintergrund“ und „Geschlecht“ lagen bereits in dichotomisierter Form vor, wobei 56,1 Prozent der Teilnehmenden männlich und 43,9 Prozent weiblich waren. Das Merkmal „Migrationshintergrund“ traf auf 16,5 Prozent der Bewerber zu und auf 83,5 Prozent nicht zu.

Das Bewerbermerkmal „Alter“ lag ursprünglich als Zahlenwert vor. Als Grenze zwischen einem „Alten Bewerber“ und einem „Nicht alten Bewerber“ wurde aus verschiedenen Gründen das vollendete 40. Lebensjahr festgelegt: Zum einen gilt der 40. Geburtstag gemeinhin als Eintritt in die zweite Lebenshälfte. Zum anderen würde ein höherer Grenzwert wie beispielsweise 50 Jahre den Anteil „alter“ Bewerber so drastisch reduzieren (in diesem Fall nur noch ca. 3 Prozent), dass keine relevante Teilgruppe mehr betrachtet werden könnte. Bei einer Grenze von 40 Jahren befinden sich 7,5 Prozent „Alte Bewerber“ und 92,5 Prozent „Nicht alte Bewerber“ in der Stichprobe (s. Tabelle 23)

Zur Analyse des Zusammenhangs zweier dichotomer Merkmale sind Kreuztabellen geeignet. Bei Kreuztabellen kann die statistische Unabhängigkeit zweier Populationen (z. B. Bewerber mit bzw. ohne bestandene Prüfung) im Hinblick auf ein zweites Merkmal – z. B. das Geschlecht – mittels eines Chi-Quadrat-Tests untersucht werden. Aus dem resultierenden Chi-Quadrat-Wert lässt sich berechnen, wie groß die Wahrscheinlichkeit ist, irrtümlicherweise anzunehmen, die beiden Populationen würden nicht der gleichen Grundgesamtheit entstammen (d. h. sie würden sich also im Hinblick auf das zweite Merkmal unterscheiden). Im vorliegenden Beispiel wäre das die Wahrscheinlichkeit, fälschlicherweise anzunehmen, Männer und Frauen unterschieden sich hinsichtlich ihrer Bestehensquote systematisch.

Bewerbermerkmale	Anteil der Bewerber mit „Ja“	Anteil der Bewerber mit „Nein“
Abitur angestrebt oder erreicht	55,4 %	44,6 %
Bei Prüfungen sehr aufgeregt	41,6 %	58,4 %
Migrationshintergrund	16,5 %	83,5 %
Weiblich	43,9 %	56,1 %
Über 40 Jahre alt	7,5 %	92,5 %

Tab. 23: Verteilung der „diskriminanten“ Bewerbermerkmale



Neben der Irrtumswahrscheinlichkeit ( $p$ ), die häufig auch als „statistische Signifikanz“ bezeichnet wird, ist auch die Effektstärke für die inhaltliche Interpretation der Befunde bedeutsam. So kann ein Unterschied zwischen zwei Populationen in Bezug auf ein Merkmal zwar statistisch signifikant sein, aber von der Effektstärke her so gering, dass er keine inhaltliche Bedeutung besitzt. Dieses Phänomen tritt insbesondere bei sehr großen Stichproben auf, da die Stichprobengröße eine der Determinanten der statistischen Signifikanz ist. Mit 4.821 Personen kann die vorliegende Stichprobe, die für die Äquivalenzuntersuchungen zur Verfügung stand, als sehr groß bezeichnet werden; das Phänomen ist deshalb nicht generell auszuschließen. Zur Feststellung der Effektstärke bzw. der inhaltlichen Bedeutsamkeit der Zusammenhänge wird im Fall von 2x2-Kreuztabellen gewöhnlich der Phi-Koeffizient berechnet, der direkt Cohen's  $W$  entspricht. Nach BORTZ (1999) kann die Effektgröße Phi bzw. Cohen's  $W$  wie folgt interpretiert werden: ab „0,10“ bis unter „0,30“ = „Schwacher Effekt“; ab „0,30“ bis unter „0,50“ = „Mittlerer Effekt“; ab „0,50“ = „Starker Effekt“.

Wie der Tabelle 24 zu entnehmen ist, finden sich in der vorliegenden Stichprobe zwar signifikante Zusammenhänge zwischen den Prüfungsergebnissen einerseits sowie dem Bildungsniveau, der Prüfungsangst, dem Alter und dem Geschlecht der Bewerber andererseits; allerdings erreichen alle diese Zusammenhänge keine nennenswerte Effektstärke. Lediglich das Vorhandensein eines Migrationshintergrundes scheint die Aussichten auf die Erteilung der angestrebten Fahrerlaubnis substantiell zu schmälern; jedoch findet sich auch hier nur eine sehr geringe Effektstärke.

Bei der vergleichenden Betrachtung der Prozentsätze fällt auf, dass sich die jeweiligen Teilpopula-

tionen – mit Ausnahme beim „Geschlecht“ – um mindestens fünf Prozent hinsichtlich der Bestehensquoten voneinander unterscheiden. Insbesondere Bewerber mit Migrationshintergrund bzw. mit einem Alter über 40 Jahren weisen eine über zehn Prozent niedrigere Bestehensquote auf als ihre jeweilige Referenzgruppe. Eine um fünf bis zehn Prozent reduzierte Bestehensquote kann – obgleich die statistische Maßzahl dies suggeriert – wohl nicht als nur marginale Einschränkung der populationsspezifischen Äquivalenz betrachtet werden; sie stellt auf jeden Fall einen Anlass für vertiefte Analysen dar.<sup>44</sup>

Zur Aufklärung der inhaltlichen Ursachen der in Tabelle 24 dargestellten bivariaten Zusammenhänge können multivariate Modelle beitragen. Multivariate Modelle erlauben es, bei der Berechnung des Zusammenhangs zweier Variablen den Einfluss einer dritten – gegebenenfalls vermittelnden – Variablen zu kontrollieren. Im Folgenden soll dies beispielhaft demonstriert werden.

Im Kapitel 3.5 wurde erläutert, dass ein Einfluss des Migrationshintergrundes des Bewerbers auf die Prüfungsbewertungen bzw. die Prüfungsentcheidung des Fahrerlaubnisprüfers gegebenenfalls – neben vielen anderen möglichen Ursachen – über die Sprachkompetenz bzw. die Fähigkeit, Fahrhinweise zu verstehen, vermittelt sein könnte. Die Sprachkompetenz der Bewerber wurde im Rahmen der Untersuchungen zur populationsspezifischen Äquivalenz zwar nicht erhoben;

<sup>44</sup> Entsprechende Untersuchungen können im Rahmen der kontinuierlichen Evaluation durchgeführt werden, da diese sich ggf. auf eine breitere, bundesweite Datengrundlage stützen wird.

Bewerbermerkmale	Bestehensquote bei „Ja“	Bestehensquote bei „Nein“	Effektstärke (Cohen's $W$ )	Irrtumswahrscheinlichkeit ( $p$ )
Abitur angestrebt oder erreicht	88,5 %	83,5 %	0,07	0,00***
Bei Prüfungen sehr aufgeregt	83,1 %	88,4 %	-0,08	0,00***
Migrationshintergrund	76,3 %	88,2 %	-0,13	0,00***
Weiblich	84,9 %	87,3 %	-0,03	0,02*
Über 40 Jahre alt	68,8 %	77,8 %	-0,06	0,00***

\*\*\* signifikant auf dem 0,1-Prozent-Niveau,  
 \*\* signifikant auf dem 1-Prozent-Niveau,  
 \* signifikant auf dem 5-Prozent-Niveau, keine Sternchen = Koeffizient ist nicht signifikant.

Tab. 24: Bestehensquoten nach „diskriminanten“ Bewerbermerkmalen

allerdings ist die Sprachkompetenz nach ESSER (2006) eine entscheidende Voraussetzung für den Bildungserfolg von Migranten. Ein hohes Bildungsniveau kann also als Indiz für eine ausgeprägte Sprachkompetenz von Migranten interpretiert werden. Migranten erreichen im Durchschnitt niedrigere Bildungsabschlüsse als Deutsche ohne Migrationshintergrund (Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2012). In der vorliegenden Stichprobe könnte also die niedrigere Bestehensquote von Migranten über die niedrigere Bestehensquote von Bewerbern ohne Abitur vermittelt sein. Träfe dies zu, so müsste der Einfluss des Migrationshintergrundes auf die Prüfungsergebnisse bei statistischer Kontrolle des Bildungsniveaus verschwinden oder zumindest abnehmen.

Die statistische Kontrolle von möglicherweise vermittelnden Drittvariablen erfolgt meist durch die Konstanthaltung der Drittvariablen bei der Berechnung des Zusammenhangs zwischen den zwei interessierenden Merkmalen (EID, GOLLWITZER & SCHMITT, 2010). Im vorliegenden Fall wurde der Einfluss des Migrationshintergrundes auf die Prüfungsentscheidung bei konstantem Bildungsniveau berechnet. Wie der Tabelle 25 entnommen werden kann, unterscheidet sich die Bestehensquote zwischen Bewerbern mit und ohne Migrationshintergrund in Abhängigkeit davon, ob das Abitur angestrebt wird bzw. bereits erreicht wurde oder ob das Abitur nicht angestrebt wird bzw. erreicht wurde. So beträgt dieser Unterschied 11,0 Prozent in der Gruppe der Personen mit angestrebtem oder erreichtem Abitur und 11,5 Prozent in der Gruppe der Personen ohne angestrebtes oder erreichtes Abitur. Somit deutet sich an, dass sich zumindest ein kleiner Teil der Benachteiligung von Bewerbern mit Migrationshintergrund über das Bildungsniveau erklären lässt. Absolut gesehen ist dieser Vermittlungseffekt aber als sehr gering zu betrachten, wie sich auch in der Effektgröße Cohen's *W* ausdrückt, die lediglich um eine

Hundertstel zwischen den Gruppen differiert. Insgesamt wirkt sich das Bildungsniveau also nur sehr geringfügig auf den Zusammenhang von Migrationshintergrund und Prüfungsentscheidungen aus.

Das obige Beispiel des multivariaten Zusammenspiels von Migrationshintergrund, Bildungsniveau und Prüfungsentscheidungen lässt sich anschaulich anhand von Kreuztabellen illustrieren. Möchte man die Zahl der Prädiktoren jedoch von zwei auf drei oder mehr erweitern, stößt man mit diesem Vorgehen schnell an Grenzen. Im vorliegenden Fall fungieren die fünf Bewerbermerkmale „Bildungsniveau“, „Prüfungsangst“, „Migrationshintergrund“, „Geschlecht“ und „Alter“ als mögliche Prädiktoren für die Prüfungsentscheidung. Möchte man ihren Einfluss auf die Prüfungsentscheidung miteinander vergleichen, benötigt man ein statistisches Modell, das in der Lage ist, alle Prädiktoren gleichzeitig einzubeziehen und dabei auch die Korrelationen der Prädiktoren untereinander zu berücksichtigen. Durch Letzteres wird der Einfluss eines jeden Prädiktors auf die Prüfungsentscheidung unter Bereinigung des Einflusses der anderen Prädiktoren berechnet, sodass der isolierte Einzelbeitrag jedes Prädiktors zur Prüfungsentscheidung ermittelt wird. Dies leistet eine multiple Regressionsanalyse.

In der vorliegenden Untersuchung stellt die Prüfungsentscheidung als vorherzusagendes Kriterium eine dichotome Variable dar. Daher ist eine logistische Regressionsanalyse anzuwenden, bei der mithilfe der logistischen Funktion ein nicht-linearer Zusammenhang zwischen den Prädiktoren und dem Kriterium modelliert wird. Dabei werden die Regressionskoeffizienten als Wahrscheinlichkeit interpretiert, dass dieses Kriterium bzw. Ereignis eintritt (BACKHAUS et al., 2006). Bezogen auf die vorliegende Analyse wird also anhand der Prädiktoren die Wahrscheinlichkeit geschätzt, dass der Bewerber die Prüfung besteht.

Bewerbermerkmale	Bestehensquote bei „Ja“	Bestehensquote bei „Nein“	Effektgröße (Cohen's <i>W</i> )	Irrtumswahrscheinlichkeit ( <i>p</i> )
<b>Bewerber mit angestrebtem oder erreichtem Abitur</b>				
Migrationshintergrund	89,9 %	78,9 %	-0,12	0,00***
<b>Bewerber ohne angestrebtes oder erreichtes Abitur</b>				
Migrationshintergrund	86,0 %	74,5 %	-0,13	0,00***

Tab. 25: Bestehensquoten nach Migrationshintergrund unter Konstanthaltung des Bildungsniveaus

Eine wichtige Voraussetzung für die Anwendung eines logistischen Regressionsmodells ist das Nichtvorhandensein von Multikollinearität (d. h. die einzelnen Prädiktoren lassen sich nicht als Linearkombination der anderen Prädiktoren darstellen). Im vorliegenden Fall – und in allen weiteren dargestellten logistischen Regressionsmodellen – sind sowohl die bivariaten Korrelationen zwischen den Prädiktoren als auch die multiplen Korrelationen zwischen den Prädiktoren durchgängig gering (kleiner als 0,2), weswegen Multikollinearität ausgeschlossen werden kann.

Die Tabelle 26<sup>45</sup> enthält die Ergebnisse einer multiplen logistischen Regressionsanalyse mit den diskriminanten Bewerbermerkmalen als Prädiktoren und der Prüfungsentscheidung als Kriterium. Der Omnibus-Test mit den Modellkoeffizienten ergibt einen Chi-Quadrat-Wert von 117,25, wonach das Modell die Zielvariable mit einer Irrtumswahrscheinlichkeit von  $p = 0,00$  besser vorhersagt als ein „leeres“ Modell ohne Prädiktoren. Das Modell erklärt gemäß Nagelkerkes R-Quadrat 4,4 Prozent der Varianz des Kriteriums. Bei der Interpretation dieses Wertes ist zu berücksichtigen, dass eine hohe Varianzaufklärung aufgrund der bereits dargelegten methodischen Zusammenhänge unerwünscht ist.

Als Nachteil von logistischen Regressionsanalysen ist die erschwerte Interpretation der Koeffizienten aufgrund des nichtlinearen Zusammenhangs zwi-

schen den Prädiktoren und dem Kriterium anzusehen. Aus den B-Koeffizienten lässt sich daher lediglich die Richtung und – da im vorliegenden Fall alle Prädiktoren als dichotome Variablen den gleichen Wertebereich („0“ und „1“) aufweisen – die relative Stärke des Einflusses der jeweiligen diskriminanten Variablen auf die Prüfungsentscheidung vergleichend ablesen. Im vorliegenden Modell senken ein Migrationshintergrund, ein Alter über 40 Jahren und starke Prüfungsangst des Bewerbers die Wahrscheinlichkeit, die Prüfung zu bestehen. Das Erreichen oder Anstreben eines Abiturs erhöht diese Wahrscheinlichkeit hingegen. Der Einfluss des Geschlechts ist unter Bereinigung der Einflüsse der anderen Prädiktoren statistisch nicht signifikant. Aus den B-Koeffizienten kann abgelesen werden, dass der Migrationshintergrund den stärksten bereinigten Einfluss auf die Prüfungsentscheidung besitzt, gefolgt vom Alter, der Prüfungsangst und dem Bildungsniveau. Allerdings stehen die Koeffizienten in keinem linearen Verhältnis zueinander, sodass z. B. der Einfluss des Migrationshintergrundes nicht als doppelt so groß wie der Einfluss der Prüfungsangst bezeichnet werden kann.

Zusammenfassend lässt sich auf der Grundlage der dargestellten Analysen festhalten, dass Bewerber mit Migrationshintergrund im Erprobungsprojekt eine deutlich geringere Chance besaßen, die optimierte Praktische Fahrerlaubnisprüfung zu bestehen, als andere Bewerber. Dagegen spielte das Geschlecht der Bewerber keine nennenswerte Rolle im Hinblick auf die Prüfungsentscheidung. Bestätigt wurde die bekannte Erfahrung, dass Bewerber in hohem Lebensalter oder mit Prüfungsangst (HAMPEL, 1977) die Prüfung seltener als andere bestehen. Es sind jedoch weiterführende Studien nötig, um die Effektstärken derartiger Einflüsse näher zu bestimmen und ggf. vermittelnde Variablen zu identifizieren. Damit könnten aus wissenschaftlicher Sicht wertvolle Beiträge zur Optimierung der Fahrschulausbildung und Fahrerlaubnisprüfung geleistet werden.

<sup>45</sup> Der Omnibus-Test ergibt mit den Modellkoeffizienten einen Chi-Quadrat-Wert von 117,25, wonach das Modell die Zielvariable mit einer Irrtumswahrscheinlichkeit von  $p = 0,00$  besser vorhersagt als ein „leeres“ Modell ohne Prädiktoren. Das Modell erklärt gemäß Nagelkerkes R-Quadrat 4,4 Prozent der Varianz des Kriteriums. Dieser Wert ist vor dem Hintergrund zu interpretieren, dass eine hohe Varianzaufklärung an dieser Stelle aufgrund der bereits dargelegten methodischen Zusammenhänge zur Validitätsbeurteilung nicht wünschenswert erscheint.

Bewerbermerkmale	B-Koeffizient
Migrationshintergrund (nein -> ja)	-0,75***
Geschlecht (männlich -> weiblich )	-0,15
Alter in Jahren (unter 40 J. -> über 40 J.)	-0,51**
Bildungsniveau (Abitur nicht erreicht oder nicht angestrebt -> Abitur erreicht oder angestrebt )	0,31***
Prüfungsangst (nicht sehr aufgeregt -> sehr aufgeregt)	-0,38***

**Tab. 26:** Einfluss diskriminanter Bewerbermerkmale auf die Prüfungsentscheidung (B-Koeffizienten einer logistischen multiplen Regressionsanalyse)

Bis dahin muss sogar offen bleiben, ob die gefundenen Zusammenhänge zwischen diskriminanten Bewerbermerkmalen und der Prüfungsentscheidung überhaupt als bedeutsam zu interpretieren sind. Dies kann nämlich auf der Basis des vorliegenden Datensatzes nicht abschließend beurteilt werden; hierfür wären bundesweit repräsentative Daten notwendig. Wie im Kapitel 4.4 dargelegt wurde, stammen die vorliegenden Felderprobungsdaten aus insgesamt vier Modellregionen Deutschlands und formen somit eine Klumpenstichprobe. Das Verhalten der Bewerber und Prüfer in den Modellregionen bzw. Klumpen kann – beispielsweise aufgrund regionaler Besonderheiten – von dem Verhalten der Bewerber und Prüfer in anderen Teilen Deutschlands abweichen. Dies könnte zu Bewertungsbesonderheiten geführt haben. Auf jeden Fall ist die Repräsentativität der Daten aus der Felderprobung nicht sichergestellt, d. h. es ist beispielsweise unklar, ob die gefundenen Effekte der Auswahl der Modellregionen geschuldet sind.

Bundesweit repräsentative Ergebnisse zur populationspezifischen Äquivalenz sollten im Rahmen einer kontinuierlichen Prüfungsevaluation generiert werden, bei der Daten aus allen Prüferten Deutschlands über einen längeren Zeitraum erhoben werden. Die kontinuierliche testpsychologische Evaluation der optimierten Praktischen Fahrerlaubnisprüfung erscheint – ähnlich wie die kontinuierliche Evaluation der theoretischen Fahrerlaubnisprüfung – aus wissenschaftlicher Sicht dringend geboten, denn sie würde empirisch begründete Aussagen und Empfehlungen zur Weiterentwicklung der Prüfung ermöglichen. Darüber hinaus wäre die kontinuierliche Evaluation der Prüfung – im Falle einer bundesweiten Einführung der optimierten Prüfung – aufgrund der ohnehin geplanten elektronischen Prüfungsdokumentation und -bewertung mittels e-Prüfprotokoll sowie der IT-gestützten Archivierung der Prüfungsdaten mit tragbarem technischen und finanziellen Aufwand möglich.

### 5.3.3 Konvergente Bewerbermerkmale

Im Rahmen der Untersuchungen zur populationspezifischen Äquivalenz der optimierten Praktischen Fahrerlaubnisprüfung wurden einige Bewerbermerkmale erhoben, welche die Aneignung von Fahrkompetenz begünstigen sollten (s. Kapitel 3.4) und von denen daher ein positiver Einfluss auf die

Prüfungsentscheidung erwartet wird. Diese Merkmale werden als konvergente Merkmale bezeichnet (s. o.). Zu den konvergenten Merkmalen zählen der Vorbesitz einer (anderen) Fahrerlaubnis, die Fahrerfahrung und der Konsum straßenverkehrsbezogener Computerspiele. Zur besseren Vergleichbarkeit der Einflussstärken dieser Merkmale auf die Prüfungsentscheidung wurden die zugrundeliegenden Variablen wiederum dichotomisiert. Bevor im Folgenden die (bivariaten) Zusammenhängen zwischen den konvergenten Merkmalen und der Prüfungsentscheidung untersucht werden, sollen zunächst das Vorgehen bei der Dichotomisierung der Variablen begründet und eine univariate Beschreibung der Variablen vorgenommen werden.

Der Vorbesitz einer (anderen) Fahrerlaubnis ist bereits eine dichotome Variable. 26,3 Prozent der an der Befragung teilnehmenden Fahrerlaubnisbewerber verfügten bereits über eine Fahrerlaubnis in einer anderen Fahrerlaubnisklasse; 73,7 Prozent besaßen noch keine andere Fahrerlaubnis.

Die Fahrerfahrung wurde bei den Bewerbern mit vorheriger Fahrerlaubnis über die bereits im Straßenverkehr gefahrenen Kilometer bestimmt. Als Grenzwert zwischen einem Bewerber mit „Viel Fahrerfahrung“ und „Nicht viel Fahrerfahrung“ wurde der Median der Verteilung gewählt. Der Median ist derjenige Wert, oberhalb und unterhalb dessen je 50 Prozent der befragten Personen liegen. Der Median der gefahrenen Kilometer lag bei 40.000. Dieser Wert entspricht auch ungefähr dem Wert, den Wissenschaftler als notwendig erachten, um von einem erfahrenen Fahrer sprechen zu können.<sup>46</sup> Bei den Bewerbern ohne Fahrerlaubnissvorbesitz wurde – statt der gefahrenen Kilometer – die Anzahl absolvierter Fahrstunden erhoben. Dabei wurde unterschieden zwischen „20 Stunden oder weniger“, „21-30 Stunden“, „31-40 Stunden“, „41-50 Stunden“ und „Mehr als 50 Stunden“. Bundesweit benötigen Bewerber zum Erlangen der Fahrerlaubnisklasse B ungefähr 30 Ausbildungs-

<sup>46</sup> Der Abschluss der Fahrschulausbildung ist nicht mit dem Abschluss des Fahrkompetenzerwerbs gleichzusetzen. Stattdessen ist der Fahrkompetenzerwerb als ein komplexer Lernprozess zu verstehen, der sich mindestens über einen zwei- bis dreijährigen Zeitraum erstreckt (MAYCOCK, LOCKWOOD & LESTER, 1991; GREGERSEN & NYBERG, 2002; SCHADE, 2001) und Fahrleistungen von etwa 50.000 Kilometern erfordert (SUMMALA, 1987; CENTRAAL BUREAU RIJVAARDIGHEIDSBEWIJZEN, 2012).



stunden.<sup>47</sup> Daher wurden alle Bewerber mit 31 oder mehr Fahrausbildungsstunden der Kategorie „Viele Fahrstunden“ zugeordnet. Dies entspricht 58,4 Prozent der Befragten in der vorliegenden Stichprobe, wohingegen 41,6 Prozent der Kategorie „Nicht viele Fahrstunden“ zugeordnet wurden.

Der Konsum straßenverkehrsbezogener Computerspiele wurde ursprünglich über eine fünfstufige Skala mit den Ausprägungen „Mehrere Male pro Woche“, „Etwa einmal pro Woche“, „Etwa einmal in 14 Tagen“, „Einmal im Monat oder seltener“ und „Nie“ erhoben. Da 75 Prozent der befragten Bewerber die Kategorie „Nie“ ausgewählt haben, ist eine Unterteilung in zwei ungefähr gleich große Subgruppen nicht möglich. Daher wurde die Gruppeneinteilung nicht nach statistischen Parametern, sondern anhand inhaltlicher Kriterien vorgenommen. Die in diesem Kontext zu überprüfende Hy-

pothese besagt, dass Handlungsskripte, welche die Aneignung von Fahrkompetenz begünstigen, auch im Rahmen straßenverkehrsbezogener Computerspiele trainiert werden können (s. Kapitel 3.4). Ein relevanter Trainingseffekt ist aber nur zu erwarten, wenn das Training mindestens einmal pro Woche durchgeführt wird; nur einmal Training in zwei Wochen oder seltener erscheint hierfür nicht ausreichend. Daher wurden Bewerber, die mindestens einmal pro Woche Computerspiele mit Bezug zum Straßenverkehr spielen, als „Regelmäßige Spieler“ eingestuft. Dies betrifft 10,2 Prozent der befragten Bewerber. 89,8 Prozent der Bewerber spielen diese Spiele hingegen seltener oder nie und gelten damit als „Nicht regelmäßige Spieler“. Bei den „Regelmäßigen Spielern“ wurde ferner gefragt, wie viele Stunden sie diese straßenverkehrsbezogenen Computerspiele durchschnittlich pro Woche spielen. Der Median der erhobenen Verteilung lag bei drei Stunden pro Woche. Bewerber, die drei Stunden pro Woche oder häufiger spielen, gelten demnach als „Vielspieler“; Bewerber, die weniger als drei Stunden pro Woche spielen, gelten als „Nicht Vielspieler“ (s. Tabelle 27).

Die Auswertung der bivariaten Zusammenhänge zwischen den konvergenten Bewerbermerkmalen einerseits und der Prüfungsentscheidung andererseits in Form von Kreuztabellen (s. Tabelle 28)

<sup>47</sup> Die Anzahl der Fahrstunden liegt im Ermessen des Fahrlehrers; lediglich sogenannte „Besondere Ausbildungsfahrten“ (z. B. Fahrten auf Bundesstraßen), die in der fortgeschrittenen Fahrpraktischen Ausbildung absolviert werden, sind mit einem Umfang von mindestens 12 Fahrstunden zu je 45 Minuten festgelegt. Üblicherweise nehmen Fahranfänger darüber hinaus etwa 10 bis 25 weitere Fahrstunden in Anspruch (BAHR, 2013).

Bewerbermerkmale	Anteil der Bewerber mit „Ja“	Anteil der Bewerber mit „Nein“
Vorbesitz einer anderen Fahrerlaubnis	26,3 %	73,7 %
Mehr als 40.000 Kilometer Fahrerfahrung	51,3 %	48,7 %
Mehr als 30 Ausbildungsfahrstunden	58,4 %	41,6 %
Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele	10,2 %	89,8 %
Drei oder mehr Stunden durchschnittlicher Konsum straßenverkehrsbezogener Computerspiele pro Woche	45,4 %	54,6 %

Tab. 27: Verteilung der „konvergenten“ Bewerbermerkmale

Bewerbermerkmale	Bestehensquote bei „Ja“	Bestehensquote bei „Nein“	Effektgröße (Cohen's W)	Irrtumswahrscheinlichkeit (p)
Vorbesitz einer anderen Fahrerlaubnis	89,2 %	84,8 %	0,06	0,00***
Mehr als 40.000 Kilometer Fahrerfahrung	92,8 %	85,7 %	0,11	0,00***
Mehr als 30 Ausbildungsfahrstunden	85,4 %	84,9 %	0,01	0,68
Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele	89,7 %	85,8 %	0,04	0,02*
Drei oder mehr Stunden durchschnittlicher Konsum straßenverkehrsbezogener Computerspiele pro Woche	91,0 %	89,0 %	0,03	0,47

Tab. 28: Bestehensquoten nach „konvergenten“ Bewerbermerkmalen

ergibt, dass lediglich die Variable „Mehr als 40.000 Kilometer Fahrerfahrung“ einen signifikanten und – von der Effektstärke her – nennenswerten Einfluss auf die Prüfungsentscheidung besitzt. Sehr geringfügige (signifikante) Effekte lassen sich allenfalls noch im Hinblick auf den Fahrerlaubnisvorbesitz und den wöchentlichen Konsum straßenverkehrsbezogener Computerspiele finden.

Insbesondere eine Fahrerfahrung von mehr als 40.000 Kilometern wirkt sich also förderlich auf die Wahrscheinlichkeit aus, die Prüfung zu bestehen; dies ist ein starkes Validitätsindiz für die optimierte Praktische Fahrerlaubnisprüfung! „Mehr als 30 Ausbildungsfahrstunden“ oder „Drei oder mehr Stunden durchschnittlicher Konsum straßenverkehrsbezogener Computerspiele pro Woche“ haben in der vorliegenden Stichprobe dagegen keinen signifikanten Einfluss auf die Prüfungsentscheidung.<sup>48</sup> Der erstgenannte Befund dürfte nicht zuletzt darin begründet sein, dass das Absolvieren vieler kostenpflichtiger Fahrstunden insbesondere bei Bewerbern zu finden ist, die besondere Schwierigkeiten bei der Aneignung von fahrpraktischen Fähigkeiten und demzufolge auch beim Bestehen der Fahrprüfung haben. Der zweitgenannte Befund deutet auf die Grenzen des Übens in virtuellen Welten für das Erlernen von Manövrierfähigkeiten im realen Verkehrsgeschehen hin.

Ein Vergleich der bereinigten Einflussstärken der konvergenten Bewerbermerkmale auf die Prüfungsentscheidung mittels einer multiplen logistischen Regressionsanalyse (s. o.) ist an dieser Stelle nur bedingt möglich. Da die Merkmale „Mehr als 40.000 Kilometer Fahrerfahrung“ und „Mehr als 30 Ausbildungsfahrstunden“ in Abhängigkeit vom Merkmal „Vorbesitz einer anderen Fahrerlaubnis“ erhoben wurden, stehen diese Variablen in einem

<sup>48</sup> Dieser Befund ändert sich auch nicht, wenn der Zusammenhang zwischen den Ausbildungsfahrstunden bzw. dem durchschnittlichen Konsum straßenverkehrsbezogener Computerspiele pro Woche und der Prüfungsentscheidung vor der Dichotomisierung der beiden Bewerbermerkmale mittels U-Test bzw. T-Test berechnet wird.

direkten Verhältnis zueinander, was die Schätzungen des Regressionsmodells verzerrt. Gleiches gilt für die Bewerbermerkmale „Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele“ und „Drei oder mehr Stunden durchschnittlicher Konsum straßenverkehrsbezogener Computerspiele pro Woche“. Die Einbeziehung der beiden Filtervariablen „Vorbesitz einer anderen Fahrerlaubnis“ und „Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele“ in ein Regressionsmodell stellt hingegen kein Problem dar, da diese Variablen in keiner logischen Abhängigkeit zueinander stehen und jeweils von allen Befragten zu beantworten waren. Wie der Tabelle 29 zu entnehmen ist, fällt der bereinigte Einfluss dieser beiden Bewerbermerkmale auf die Prüfungsentscheidung positiv und ähnlich stark aus. Demnach wirken sich praktische Fahrerfahrungen und der regelmäßige Konsum straßenverkehrsbezogener Computerspiele positiv auf die Aneignung von Fahrkompetenz aus, was im Sinne der in Kapitel 3.4 dargelegten Argumentation klar auf die Validität der optimierten Praktischen Fahrerlaubnisprüfung hindeutet.

Die gleichwertige Gewichtung der beiden Faktoren erscheint allerdings verwunderlich, da das konkrete Fahren im Straßenverkehr die zur Aneignung von Fahrkompetenz nötigen Handlungsskripte besser trainieren sollte als das simulierte Fahren in virtuellen Welten.

Eine denkbare statistische Erklärung dieses unerwarteten und aus psychologischer Sicht sehr wenig plausiblen Befundes wäre die Möglichkeit, dass beide Merkmale untereinander stark korrelieren. Die Korrelation zwischen den Merkmalen beträgt aber nur 0,004, was als extrem niedrig zu bezeichnen ist. Ein weiterer Erklärungsansatz wäre die Existenz eines weiteren unbekanntem Faktors, der die Aneignung von Fahrkompetenz und das Interesse an straßenverkehrsbezogenen Computerspielen gleichermaßen begünstigt. Dies könnte beispielsweise eine generelle Affinität zu Autos und Autofahren sein. Diese Hypothese kann auf Basis der vorliegenden Daten allerdings nicht überprüft werden. In-

Bewerbermerkmale	B-Koeffizient
Vorbesitz einer anderen Fahrerlaubnis (nein -> ja)	0,37***
Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele (nein -> ja)	0,38*

**Tab. 29:** Einfluss konvergenter Bewerbermerkmale auf die Prüfungsentscheidung (B-Koeffizienten einer logistischen multiplen Regressionsanalyse; das Regressionsmodell ist gemäß Omnibus-Test signifikant)

wiefern sich reales und virtuelles Fahren gleichermaßen positiv auf die Aneignung von Fahrkompetenz auswirken, wäre also eine Frage für weiterführende Studien. Die Beantwortung dieser Frage hätte weitreichend Bedeutung für den Einsatz von Fahr-simulatoren in der Fahrausbildung und könnte die vorliegenden Befunde ergänzen (BREDOW, 2015).

### 5.3.4 Zusammenfassung der Befunde

Die Tabelle 30 beinhaltet die Ergebnisse einer multiplen logistischen Regressionsanalyse mit der Prüfungsentscheidung als Kriterium sowie den diskriminanten und konvergenten Bewerbermerkmalen als Prädiktoren. Dies ermöglicht einen interessanten Vergleich der relativen Einflussstärken der diskriminanten und konvergenten Bewerbermerkmale auf die Prüfungsentscheidung. Im Vergleich zu den beiden einzelnen Regressionsanalysen unter Einbeziehung der diskriminanten bzw. der konvergenten Merkmale bleiben die Wirkungsrichtungen und Stärken der einzelnen Merkmale im Wesentlichen bestehen. Beim Betrachten der Tabelle fällt auf, dass die diskriminanten Merkmale „Migrationshintergrund“ und „Alter über 40 Jahre“ den stärksten Einfluss auf die Prüfungsentscheidung haben. Sie rangieren damit vor dem konvergenten Merkmal „Vorbesitz einer anderen Fahrerlaubnis“, hinter dem sich das Ausmaß der Fahrerfahrung verbirgt (s. o.). Dahinter folgen die diskriminanten Merkmale „Bildungsniveau“ und „Prüfungsangst“ etwa gleichauf.

Weiterhin erscheint interessant, dass das Merkmal „Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele“ unter Berücksichtigung des Einflusses der diskriminanten Merkmale keinen signifikanten Einfluss auf die Prüfungsentscheidung mehr hat. Dies unterstützt die oben formulierte Hypothese, dass der positive Ein-

fluss des Konsums straßenverkehrsbezogener Computerspiele auf die Aneignung von Fahrkompetenz zumindest teilweise anderen, vermittelnden Faktoren geschuldet ist.

Zusammenfassend kann festgehalten werden, dass die populationsspezifische Äquivalenz der optimierten Praktischen Fahrerlaubnisprüfung sehr wahrscheinlich durch einen Einfluss des (diskriminanten) Bewerbermerkmals „Migrationshintergrund“ eingeschränkt wird. Dieser Einfluss ist hinsichtlich seiner Stärke und psychologischen Vermittlungsmechanismen weiter aufzuklären. Der festgestellte förderliche Einfluss der Merkmale „Vorbesitz einer anderen Fahrerlaubnis“, „Fahrerfahrung“ und „Konsum straßenverkehrsbezogener Computerspiele“ auf das Bestehen der Prüfung spricht für eine valide Fahrkompetenzmessung bei der optimierten Praktischen Fahrerlaubnisprüfung, da diese Merkmale die Aneignung von Fahrkompetenz nach vorliegenden empirischen Befunden tatsächlich begünstigen. Für eine abschließende Bestätigung dieser Befunde und ein genaueres Verständnis der ihnen zugrunde liegenden Mechanismen ist aber aus wissenschaftlicher Sicht die Erhebung eines inhaltlich erweiterten und bundesweit repräsentativen Datensatzes im Rahmen einer kontinuierlichen Prüfungsevaluation notwendig. Darauf aufbauend könnten dann empirisch begründete Empfehlungen für die Weiterentwicklung der Praktischen Fahrerlaubnisprüfung abgeleitet werden.

## 5.4 Auswertung von Prüfungsergebnissen

### 5.4.1 Überblick

Die Ergebnisdarstellungen der vorhergehenden Kapitel betrafen die instrumentelle Evaluation und somit die methodische Verfahrensgüte der opti-

Bewerbermerkmale	B-Koeffizient
Migrationshintergrund (nein -> ja)	-0,74***
Geschlecht (männlich -> weiblich )	-0,04
Alter in Jahren (unter 40 J. -> über 40 J.)	-0,64***
Bildungsniveau (Abitur nicht erreicht oder nicht angestrebt -> Abitur erreicht oder angestrebt )	0,35***
Prüfungsangst (nicht sehr aufgeregt -> sehr aufgeregt)	-0,34***
Vorbesitz einer anderen Fahrerlaubnis (nein -> ja)	0,47***
Mindestens einmal pro Woche Konsum straßenverkehrsbezogener Computerspiele (nein -> ja)	0,27

Tab. 30: Einfluss diskriminanter und konvergenter Bewerbermerkmale auf die Prüfungsentscheidung (B-Koeffizienten einer logistischen multiplen Regressionsanalyse; das Regressionsmodell ist gemäß Omnibus-Test signifikant)

mierten Praktischen Fahrerlaubnisprüfung. Die beiden folgenden Kapitel sind hingegen den Ergebnissen der prozessualen Evaluation gewidmet und beschreiben die Qualität der Durchführung der optimierten Praktischen Fahrerlaubnisprüfung unter den Arbeitsbedingungen der technischen Prüfstellen in der Alltagspraxis.

Eine wichtige Methode zur Ermittlung der Durchführungsqualität ist die Analyse der Prüfungsergebnisse, die im Rahmen der Felderprobung (s. Kapitel 4.4) erhoben wurden. Das vorliegende Kapitel beinhaltet die Auswertung dieser Prüfungsergebnisse. Dabei wird zwischen einer

- deskriptiven Auswertung der Prüfungsanforderungen, Prüfungsbewertungen und Prüfungsentscheidungen, einer
- Analyse der Zusammenhänge zwischen den Prüfungsrahmenbedingungen, Prüfungsbewertungen und Prüfungsentscheidungen sowie einer
- Analyse von Prüfungsbewertungsmustern und Prüfertypen

unterschieden (s. Kapitel 2.3).

Da die Dokumentation der Prüfungsanforderungen und -bewertungen mittels des e-Prüfprotokolls detailliert erfolgt, ist eine Fülle potenzieller qualitätsrelevanter Auswertungsmöglichkeiten gegeben. Die Auswertung der Prüfungsergebnisse muss sich daher im Folgenden auf exemplarisch herausgegriffene Aspekte beschränken, die zum jetzigen Zeitpunkt als besonders relevant für den Nachweis der Durchführungsqualität der optimierten Praktischen Fahrerlaubnisprüfung betrachtet wurden. Ferner sollte stets bedacht werden, dass das vorliegende Revisionsprojekt in erster Linie der Erprobung des methodischen Auswertungsinstrumentariums dient und die Ergebnisse keine Gültigkeit für das gesamte Bundesgebiet besitzen.<sup>49</sup> Eine Verifizierung und vertiefende Analyse der dabei entdeckten Zusammenhänge könnte

erst anhand einer kontinuierlichen und bundesweiten Prüfungsevaluation erfolgen.

#### 5.4.2 Deskriptive Auswertung der Prüfungsanforderungen, Prüfungsbewertungen und Prüfungsentscheidungen

##### Prüfungsanforderungen

Die Praktische Fahrerlaubnisprüfung stellt – methodisch gesehen – eine Arbeitsprobe zur Diagnose der Fahrkompetenz dar, die im realen Straßenverkehr erbracht wird und daher nur begrenzt standardisierbar ist (STURZBECHER et al., 2014). Dem Fahrerlaubnisbewerber obliegt es dabei nachzuweisen, dass er sicherheitsrelevante Anforderungen der Fahrzeugführung weitestgehend beherrscht. Welche Anforderungen im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung mit welcher Begründung als sicherheitsrelevant betrachtet werden, wurde im Kapitel 2.1 dargelegt. Im Ergebnis dieser Überlegungen entstand der Fahraufgabenkatalog zur optimierten Praktischen Fahrerlaubnisprüfung, der auf der Ebene der Prüfungsanforderungen acht Fahraufgaben definiert. Ein Ziel im Zuge der Prüfungsoptimierung ist die Erhöhung der Verkehrssicherheit von Fahranfängern. Ein umfassender Nachweis sicherheitsrelevanter Fahrkompetenzen setzt daher die weitestgehend fehlerfreie Durchführung aller Fahraufgaben voraus. Zeigt die Analyse der Prüfungsergebnisse jedoch, dass bestimmte Fahraufgaben häufig nicht geprüft werden, müssen die Gründe dafür analysiert und bei der Weiterentwicklung der Prüfungsmethodik berücksichtigt werden.

Wie der Tabelle 31 entnommen werden kann, wurden die Fahraufgaben „Kreisverkehr“ und „Schienenverkehr“ nur in zirka der Hälfte der Fälle geprüft; dabei traten erhebliche Unterschiede zwischen den Standorten auf. Hintergrund dieses Befundes ist vielleicht der Umstand, dass Kreisverkehre und Straßenbahnen bzw. Bahnübergänge in manchen Prüforten nicht (mehr) vorkommen. In diesem Fall würde es sich um eine Prüfortproblematik handeln, die wahrscheinlich nur durch eine Anpassung der Anforderungen an Prüforte gemildert werden könnte.

Weniger leicht zu erklären ist hingegen, dass die Fahraufgabe „Ein-, Ausfädeln, Fahrstreifenwechsel“ im Standort 1 in 21,4 Prozent der Fälle „Nicht vorgekommen“ sein soll. Aufgrund der Häufigkeit von Fahrstreifenwechseln beim Führen eines Kraft-

<sup>49</sup> Die im Rahmen der Felderprobung erhobene Klumpenstichprobe ist statistisch gesehen nicht repräsentativ für die Grundgesamtheit aller Praktischen Fahrerlaubnisprüfungen in Deutschland im Erhebungszeitraum. Nichtsdestotrotz dürften die Stichprobenergebnisse nicht grundsätzlich von den wahren Parametern in der Grundgesamtheit abweichen. Daher werden die Ergebnisse im Folgenden als Hinweise bzw. Indizien der „tatsächlichen“ Sachlage behandelt.



	Gesamt	Standort 1	Standort 2	Standort 3	Standort 4
Ein-, Ausfädeln, Fahrstreifenwechsel	8,9 %	21,4 %	4,3 %	7,1 %	8,4 %
Kurven	2,6 %	3,5 %	1,4 %	3,6 %	2,3 %
Vorbeifahren, Überholen	4,0 %	7,7 %	4,3 %	3,3 %	2,2 %
Kreuzungen, Einmündungen	0,4 %	0,4 %	0,4 %	0,7 %	0,3 %
Kreisverkehr	48,5 %	28,0 %	48,6 %	74,2 %	28,8 %
Schieneverkehr	51,3 %	37,4 %	41,1 %	58,4 %	64,2 %
Fußgänger, Radfahrer, Haltestellen	8,7 %	14,2 %	9,9 %	5,2 %	8,3 %
Geradeausfahren	0,7 %	1,4 %	0,5 %	0,7 %	0,5 %

Tab. 31: Relative Häufigkeiten des Nichtvorkommens der einzelnen Fahraufgaben

	Sehr gut	Gut	Ausreichend	Ungenügend
Ein-, Ausfädeln, Fahrstreifenwechsel	3,0 %	72,6 %	19,7 %	4,7 %
Kurven	1,9 %	83,9 %	13,7 %	0,5 %
Vorbeifahren, Überholen	2,5 %	76,3 %	19,3 %	1,9 %
Kreuzungen, Einmündungen	2,1 %	57,6 %	28,0 %	12,4 %
Kreisverkehr	1,5 %	80,0 %	17,5 %	1,0 %
Schieneverkehr	3,4 %	82,5 %	13,0 %	1,1 %
Fußgänger, Radfahrer, Haltestellen	2,6 %	81,3 %	13,9 %	2,2 %
Geradeausfahren	1,9 %	62,2 %	28,2 %	7,7 %

Tab. 32: Verteilungen der Antwortkategorien bei den Gesamtbewertungen der Fahraufgaben

fahrzeuges erscheint dieser Befund nicht plausibel und spricht für einen systematischen Dokumentationsfehler bei der Anwendung der optimierten Praktischen Fahrerlaubnisprüfung.

Abgesehen von diesen Ausnahmen, deuten die Ergebnisse darauf hin, dass an allen Standorten alle Fahraufgaben bei nahezu jeder Prüfung durchgeführt wurden.<sup>50</sup> Auf dieser Grundlage kann die Qualität der Prüfungsdurchführung insgesamt als zufriedenstellend – obgleich ausbaufähig – eingeschätzt werden.

### Prüfungsbewertungen

Die Bewertung der vom Bewerber im Rahmen der Prüfungsfahrt gezeigten Fahrkompetenz erfolgt einerseits anhand situationsbezogener Bewertungskriterien (Fehler und überdurchschnittliche Leistungen) und andererseits mittels situationsübergreifender Gesamtbewertungen der Fahraufgabenbewältigung und der Kompetenzbereiche bzw. Beobachtungskategorien (s. Kapitel 2.1). Die Gesamtbewertungen werden anhand einer vierstufigen Skala mit den Ausprägungen „Sehr gut“, „Gut“, „Ausreichend“ und „Ungenügend“ vorgenommen.

In der Tabelle 32 werden die Verteilungen der Bewertungskategorien für die Gesamtbewertungen der Fahraufgaben gezeigt. Der Großteil der Leistungen wurde mit „Sehr gut“ oder „Gut“ bewertet. Bei einem weitaus kleineren Teil der Leistungen haben die Prüfer hingegen noch Defizite festgestellt (Bewertungen mit „Ausreichend“ oder sogar mit „Ungenügend“). Dies betrifft insbesondere die Fahraufgaben „Kreuzungen, Abbiegen“ und „Gerade-

<sup>50</sup> Es ist darauf hinzuweisen, dass grundsätzlich – d. h. auch in Fällen, in denen die Daten keine augenscheinlichen Widersprüche aufweisen – eine gewisse Unsicherheit hinsichtlich der Richtigkeit der Ereignisdokumentationen verbleibt: Einerseits besteht die Möglichkeit, dass bestimmte Ereignisse stattfinden, aber nicht dokumentiert werden. Andererseits können Ereignisse dokumentiert werden, die nicht stattgefunden haben. Derartige Fehldokumentationen entstehen in der Regel unbewusst. Bei der zukünftigen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung sind deshalb zum einen Untersuchungen zu Beobachtungs- und Dokumentationsfehlern durchzuführen. Zum anderen müssen unbedingt wirksame Strategien zur Plausibilitätsprüfung entwickelt werden. Diese Strategien können sowohl der Selbstreflexion der Fahrerlaubnisprüfer als auch dem Qualitätsmanagement der technischen Prüfstellen dienen.

ausfahren“; hier liegt der Anteil der Leistungen mit Defiziten bei 40,4 bzw. 35,9 Prozent.

Bezüglich der Fahrkompetenzbereiche wurden ebenfalls vorrangig „Gute“ oder „Sehr gute“ Leistungen festgestellt (s. Tabelle 33). Dies gilt vor allem für die Bereiche „Kommunikation“ und „Fahrzeugbedienung/Umweltbewusste Fahrweise“. In den Bereichen „Fahrzeugpositionierung“ und „Geschwindigkeitsanpassung“ wurden mit 64,8 bzw. 63,8 Prozent etwas weniger gute und sehr gute Leistungen erzielt. Lediglich hinsichtlich der „Verkehrsbeobachtung“ fanden sich bei den Prüferbewertungen annähernd gleich viele tendenziell gute (55,4 %) und schlechte (44,6 %) Leistungen. Die „Verkehrsbeobachtung“ ist folglich – gefolgt von „Geschwindigkeitsanpassung“ und „Fahrzeugpositionierung“ – der Fahrkompetenzbereich mit den meisten Leistungsmängeln der Bewerber.

Betrachtet man die Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche differenziert nach Gruppen von Fahrerlaubnisklassen, bestätigt sich das dargestellte Bild für alle diese Gruppen: Bei den Fahraufgaben bereiten „Kreuzungen, Einmündungen“ und „Geradeausfahren“ die meisten Schwierigkeiten und bei den Kompetenzbereichen die „Verkehrsbeobachtung“, „Geschwindigkeitsanpassung“ und „Fahrzeugpositionierung“. Erwartungsgemäß verursacht die „Verkehrsbeobachtung“ bei den Nutzfahrzeugklassen (C- und D-Klassen) nicht mehr die meisten Probleme, sondern sie liegt etwa gleichauf mit der „Geschwindigkeitsanpassung“ und „Fahr-

zeugpositionierung“ bzw. sogar leicht hinter diesen. Die Fahrerlaubnis für Nutzfahrzeuge wird meist im beruflichen Kontext angestrebt und ist außerdem mit recht hohen Kosten verbunden, sodass die Bewerber dieser Klassen im Allgemeinen zum Prüfungszeitpunkt besonders gut ausgebildet und relativ fahrfertig sind. Gleichzeitig sind Mängel in der Verkehrsbeobachtung ein fahranfängertypisches Problem, da die motorischen und koordinativen Prozesse der Fahrzeugführung in diesem Ausbildungsstadium noch zu viele kognitive Ressourcen binden, um eine umfassende und ständige Verkehrsbeobachtung zuzulassen (GENSCHOW & STURZBECHER, 2015). Es erscheint daher plausibel, dass dieser Umstand auf die besonders gut ausgebildeten und relativ erfahrenen Bewerber der Nutzfahrzeugklassen in geringerem Maße zutrifft als auf die Bewerber anderer Fahrerlaubnisklassen. Der niedrigere Anteil der mit „Ausreichend“ oder „Ungenügend“ bewerteten Leistungen im Bereich der „Verkehrsbeobachtung“ in den Nutzfahrzeugklassen stellt damit ein weiteres Indiz für die Validität der optimierten Praktischen Fahrerlaubnisprüfung dar.

In der Tabelle 34 sind die Anteile der mangelhaften Leistungen – d. h. der nur mit „Ausreichend“ oder „Ungenügend“ bewerteten Leistungen – hinsichtlich der einzelnen Kompetenzbereiche nach Modellstandorten aufgeführt. Dabei zeigt sich, dass die „Geschwindigkeitsanpassung“, die „Fahrzeugpositionierung“ und insbesondere die „Verkehrsbeobachtung“ standortübergreifend am schlechtes-

	Sehr gut	Gut	Ausreichend	Ungenügend
Verkehrsbeobachtung	3,5 %	51,9 %	29,3 %	15,3 %
Fahrzeugpositionierung	2,2 %	62,6 %	27,2 %	8,0 %
Geschwindigkeitsanpassung	2,3 %	61,5 %	30,1 %	6,2 %
Kommunikation	2,1 %	77,3 %	20,2 %	0,5 %
Fahrzeugbedienung/Umweltbewusste Fahrweise	2,1 %	67,9 %	28,5 %	1,4 %

Tab. 33: Verteilungen der Gesamtbewertungen der Kompetenzbereiche

	Standort 1	Standort 2	Standort 3	Standort 4
Verkehrsbeobachtung	41,4 %	32,9 %	47,8 %	57,7 %
Fahrzeugpositionierung	26,3 %	24,7 %	40,3 %	48,1 %
Geschwindigkeitsanpassung	26,1 %	28,4 %	36,9 %	52,2 %
Kommunikation	17,5 %	9,9 %	20,7 %	36,4 %
Fahrzeugbedienung/Umweltbewusste Fahrweise	23,2 %	23,0 %	27,8 %	46,0 %
<b>Durchschnitt aller Kompetenzbereiche</b>	<b>26,9%</b>	<b>23,8 %</b>	<b>34,7 %</b>	<b>48,1 %</b>

Tab. 34: Anteil der mit „Ausreichend“ oder „Ungenügend“ bewerteten Kompetenzbereiche nach Standorten

ten bewertet wurden. Die Prüfer am Standort 4 scheinen in diesem Zusammenhang besonders streng zu bewerten, was in einem gewissen Widerspruch zu der vergleichsweise hohen Bestehensquote an diesem Standort steht. Dieses Bild (d. h. strenge Bewertung und milde Prüfungsentscheidung) bestätigt sich bei einer Analyse der Gesamtbewertungen der Fahraufgaben nach Standorten.

Wie bereits mehrfach erwähnt wurde, erfolgt im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung – neben der situationsübergreifenden Gesamtbewertung der Fahraufgaben und Kompetenzbereiche – eine situationsbezogene Dokumentation von Fehlern und überdurchschnittlichen Leistungen während der Prüfungsfahrt. Diese Fehler und überdurchschnittlichen Leistungen werden mit Bezug zu den Teilfahraufgaben und Situationsunterklassen (s. Kapitel 2.1) erhoben, sodass – im Gegensatz zu den situationsübergreifenden Gesamtbewertungen, die nur bezogen auf die (Haupt-)Fahraufgaben und ohne Bezug zu den Situationsunterklassen vorgenommen werden – ergänzende Rückschlüsse auf die Teilfahraufgaben und Situationsunterklassen möglich sind.

Aus der Analyse der Gesamtbewertungen der Fahraufgaben ist bereits bekannt, dass die Bewältigung der Fahraufgabe „Kreuzungen, Einmündungen“ den Bewerbern die meisten Schwierigkeiten bereitet (s. o.). Dies spiegelt sich auch in der Anzahl der dokumentierten „Todsünden“ wider, welche für die Fahraufgabe „Kreuzungen, Einmündungen“ mit deutlichem Abstand am größten ausfällt. Innerhalb dieser Fahraufgabe entfallen die meisten „Todsünden“ auf das „Linksabbiegen an Kreuzungen und Einmündungen“. Für keine andere der insgesamt 17 Teilfahraufgaben sind annähernd so viele „Todsünden“ vermerkt worden, sodass von allen Teilfahraufgaben „Linksabbiegen an Kreuzungen und Einmündungen“ als die schwierigste Teilfahraufgabe mit dem größten Schulungs- bzw. Übungsbedarf betrachtet werden kann. Bestätigt sich dieser Befund auf Basis eines bundesweit repräsentativen Datensatzes, könnte dies ein Beispiel dafür sein, wie die detaillierte Dokumentation und Auswertung aller Prüfungsergebnisse wertvolle Entscheidungsgrundlagen für die weitere Optimierung der Fahrschulung und Fahrerlaubnisprüfung erbringen kann.

### **Prüfungsentscheidungen**

Die Prüfungsentscheidung ist für den Zugang zum motorisierten Straßenverkehr maßgeblich und

spielt daher eine besonders wichtige Rolle nicht nur für den Fahrerlaubnisbewerber, sondern auch bei der Auswertung von Prüfungsergebnissen und insbesondere bei öffentlichen Diskussionen der Bestehensquoten. In diesem Zusammenhang wird in der Öffentlichkeit unter dem Stichwort „Prüfungsgerechtigkeit“ in der Regel gefordert, dass alle Fahrerlaubnisbewerber gleich schwierige Prüfungsanforderungen bewältigen sollen; gleichzeitig wird oft eine einheitliche Ausbildungsqualität an den Fahrschulen vorausgesetzt. Schon die letztgenannte Voraussetzung ist nicht gegeben: In der Realität variiert die Ausbildungsqualität an den Fahrschulen beträchtlich (STURZBECHER, SCHELLHAS & HERMANN, 2004), wofür nicht zuletzt unterschiedliche Ausbildungstraditionen, Wettbewerbsgegebenheiten und Lernvoraussetzungen der Fahrschüler verantwortlich sein dürften. Über das Zusammenspiel dieser Gegebenheiten, die alle auch einer soziokulturellen Prägung unterliegen, oder gar ihre regionale Verteilung ist auf wissenschaftlicher Ebene nichts bekannt.

Der wichtigen Frage, inwieweit unter den Bedingungen der Praktischen Fahrerlaubnisprüfung das Ideal der „Prüfungsgerechtigkeit“ verwirklicht werden sollte und kann, wurde bereits an anderer Stelle ausführlich nachgegangen (STURZBECHER et al., 2010; STURZBECHER et al., 2014). Selbst wenn man die Verwirklichung dieses Ideals für wünschenswert hält, wären einem solchen Bestreben durch die Prüfungsdurchführung in der „lebensweltlichen Domäne“ (STURZBECHER, 2010) des Straßenverkehrs und – daraus resultierend – die geringe Planbarkeit der Verkehrs- bzw. Prüfungsbedingungen Grenzen gesetzt. Zweifellos stellen die überarbeiteten Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung einen Schritt zur Vereinheitlichung der Prüfungsbedingungen und damit auch zu mehr „Prüfungsgerechtigkeit“ dar; die Unterschiedlichkeit der Prüfungsbedingungen beispielsweise in Großstädten und ländlichen Regionen lässt sich damit aber nicht überwinden, sondern nur verringern. Insofern gilt es auch zu akzeptieren, dass die Bestehensquoten in Großstädten aufgrund der erhöhten Verkehrsdichte und der oftmals komplexeren Verkehrsführungen tendenziell unter denen in ländlichen Regionen liegen. Dementsprechend streuen auch die Bestehensquoten in den 4 Modellstandorten des Revisionsprojekts: Sie betragen an zwei Standorten ca. 80 Prozent und an zwei Standorten ca. 74 Prozent.

	Gesamt	A	B	BE	C	D
Bestehensquote	77,0 %	87,4 %	73,9 %	87,8 %	86,4 %	84,1 %

Tab. 35: Bestehensquoten nach Gruppen von Fahrerlaubnisklassen

Anzumerken bleibt, dass die unterschiedliche Prägung der Modellstandorte durch urbane und rurale Gegebenheiten die gefundenen Differenzen bei den Bestehensquoten nicht völlig erklären kann. Sofern sich der Befund regional bedeutsam abweichender Bestehensquoten auf Basis einer bundesweit repräsentativen Stichprobe replizieren lässt, sollten im Rahmen einer kontinuierlichen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung die Ursachen dafür erhellt werden.

Neben der Analyse der regionalen Bestehensquoten erscheint die Ermittlung der Bestehensquoten nach Fahrerlaubnisklassen von Interesse. In der Tabelle 35 werden die Bestehensquoten nach Gruppen von Fahrerlaubnisklassen<sup>51</sup> dargestellt. Motorradprüfungen (A-Klassen), Pkw-Anhängerprüfungen (Klasse BE) sowie Lastkraftwagenprüfungen (C-Klassen) und Omnibusprüfungen (D-Klassen) weisen demnach deutlich höhere Bestehensquoten auf als Prüfungen für Personenkraftwagen (Klasse B).

Da Fahrerlaubnisbewerber für die Klasse B zum Zeitpunkt der Prüfung – verglichen mit Bewerbern für andere Fahrerlaubnisklassen – zu einem wesentlich größeren Anteil noch Fahranfänger sind und über keine andere Fahrerlaubnis verfügen, lassen sich ihre deutlich niedrigeren Bestehensquoten wahrscheinlich durch ihre geringeren Fahrerfahrungen und die entsprechend schlechtere Fahrkompetenz zum Prüfungszeitpunkt erklären. Dies stellt ein Indiz für die Validität der optimierten Praktischen Fahrerlaubnisprüfung dar und steht im Einklang mit den im Kapitel 5.3 beschriebenen Ergebnissen. Somit ergeben sich an dieser Stelle keine Hinweise auf eine Beeinträchtigung der all-

täglichen Durchführungsqualität der Prüfung im Revisionsprojekt.

### Zwischenfazit

Im Ergebnis der deskriptiven Auswertung der Prüfungsanforderungen, Prüfungsbewertungen und Prüfungsentscheidungen kann festgehalten werden, dass Fragen im Hinblick auf mögliche Einschränkungen der alltäglichen Durchführungsqualität der optimierten Praktischen Fahrerlaubnisprüfung offen geblieben sind: Zum einen zeigte sich bei den Bestehensquoten eine deutliche regionale Varianz, deren Ursachen unklar sind. Zum anderen können die Fahraufgaben „Kreisverkehr“ und „Schienenverkehr“ in vielen Prüforten nicht geprüft werden, was die Selektionsfunktion der Prüfung unterminiert.

Darüber hinaus fand sich eine Reihe von plausiblen Zusammenhängen, die als Validitätsindizien interpretiert werden können. Entsprechend vorliegender wissenschaftlicher Erkenntnisse hatten beispielsweise die Bewerber der Fahrerlaubnisklasse B aufgrund ihrer vergleichsweise geringen Fahrerfahrung die niedrigsten Bestehensquoten und die größten Leistungsmängel bei der „Verkehrsbeobachtung“; die für Fahranfänger besonders unfallträchtige Fahraufgabe „Kreuzungen, Einmündungen“ wurde auch besonders häufig fehlerhaft bewältigt.

### 5.4.3 Ausgewählte Zusammenhänge zwischen Prüfungsrahmenbedingungen, Prüfungsbewertungen und Prüfungsentscheidungen

Von den Fahrerlaubnisprüfern wird gefordert, die Prüfungsentscheidungen auf der Basis der dokumentierten Fehler und überdurchschnittlichen Leistungen sowie vor allem der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche zu treffen. Abgesehen davon, dass der Gesetzgeber im Fall einer dokumentierten „Todsünde“ das Nichtbestehen der Prüfung vorschreibt, hat der Fahrerlaubnisprüfer Ermessensspielräume<sup>52</sup> bei der Gewichtung der Art und Häufigkeit von Fehlern im Zuge des Treffens der Prüfungsentscheidung

<sup>51</sup> Aufgrund der großen Zahl von Fahrerlaubnisklassen – die sich untereinander in ihren Anforderungen teilweise stark ähneln – wurden zur besseren Übersichtlichkeit der Ergebnisdarstellung Gruppen von Fahrerlaubnisklassen gebildet.

<sup>52</sup> Ermessensspielräume für Fahrerlaubnisprüfer sind auch aus methodischer Sicht notwendig und müssen von den Prüfern mit fachlicher und pädagogisch-psychologischer Expertise sowie im Einklang mit den grundlegenden gesetzlichen und methodischen Vorgaben ausgestaltet werden (STURZBECHER, 2010).



(STURZBECHER et al., 2014): Die Prüfungsrichtlinie schreibt lediglich vor, dass Fehler, die keine „Todsünden“ darstellen, bei Wiederholung oder Häufung zum Nichtbestehen führen. Was dies konkret bedeutet und inwieweit überdurchschnittliche Leistungen Fehler kompensieren können, wird nicht konkretisiert. Auch gibt es keine Vorschriften über die Gewichtung der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche beim Treffen der Prüfungsentscheidung.<sup>53</sup> Selbst im Fall einer mit „Ungenügend“ bewerteten Fahraufgabe oder eines mit „Ungenügend“ bewerteten Kompetenzbereichs kann der Prüfer im Ausnahmefall die Prüfung noch als bestanden werten, sofern dafür triftige Gründe vorliegen.<sup>54</sup> Nachfolgend soll daher exemplarisch beschrieben werden, wie die Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche empirisch mit den Prüfungsentscheidungen zusammenhängen. Weiterhin soll exemplarisch am Beispiel der Tageszeit untersucht werden, wie sich Prüfungsrahmenbedingungen auf die Wahrscheinlichkeit auswirken, die Prüfung zu bestehen. Aus solchen Erkenntnissen könnten sich wertvolle Ansatzpunkte für die Optimierung der Fahrschulausbildung und der Praktischen Fahrerlaubnisprüfung ergeben.

### **Gesamtbewertung der Fahraufgaben und Prüfungsentscheidung**

Die Gesamtbewertungen der Fahraufgaben werden auf einer vierstufigen Ordinalskala vorgenommen (s. o.); die Prüfungsentscheidung stellt eine dichotome Variable mit den Ausprägungen „Bestanden“ und „Nicht bestanden“ dar. Der Zusammenhang zwischen ordinalskalierten und dichotomisierten Variablen kann anhand des rangbasierten Mann-Whitney-U-Tests untersucht werden. Dieser belegt hochsignifikante Zusammenhänge zwischen den Gesamtbewertungen der Fahraufgaben und der Prüfungsentscheidung, d. h. positive Gesamtbewertungen erhöhen die Wahrscheinlichkeit, die Prüfung zu bestehen. Dies gilt für alle Fahraufgaben.

Der Nachteil an dieser nichtparametrischen Zusammenhangsüberprüfung ist, dass der Zusammenhang zwischen den Gesamtbewertungen und der Prüfungsentscheidung nicht anschaulich interpretiert werden kann. Darüber hinaus kann der Einfluss der einzelnen Gesamtbewertungen der Fahraufgaben auf die Prüfungsentscheidung nicht in einem multivariaten Modell unter Bereinigung des Einflusses der anderen Gesamtbewertungen zusammenfassend dargestellt werden. Zur Vermeidung dieses Nachteils rangbasierter Verfahren werden ordinal skalierte Variablen häufig dichotomisiert, um die Anwendung einer Regressionsanalyse zu ermöglichen. Die Regressionsanalyse bietet im vorliegenden Fall einen anschaulichen Vergleich der bereinigten Einflussstärken aller Gesamtbewertungen der Fahraufgaben auf die Prüfungsentscheidung.

Wie ist bei der Dichotomisierung vorzugehen? Die möglichen Ausprägungen bei der Gesamtbewertung der Fahraufgaben lauten „Sehr gut“, „Gut“, „Ausreichend“ und „Ungenügend“. Es ist anzunehmen, dass „Sehr gute“ und „Gute“ Leistungen einen positiven Einfluss auf die Prüfungsentscheidung haben und „Ungenügende“ Leistungen einen negativen. „Ausreichende“ Leistungen sind vermutlich eher negativ konnotiert; sie sind aber per Definition ausreichend für das Bestehen der Prüfung. Folglich ist die entscheidende Frage hinsichtlich des Einflusses einer Gesamtbewertung auf die Prüfungsentscheidung, ob die Gesamtbewertung „Ungenügend“ oder „Nicht ungenügend“ ist. Daher werden die Gesamtbewertungen der Fahraufgaben anhand dieses Kriteriums für die Regressionsanalyse dichotomisiert. Da auch das Kriterium des Modells – die Prüfungsentscheidung – eine dichotome Variable darstellt, kommt wiederum eine logistische Regressionsanalyse zum Einsatz (s. Kapitel 5.3). In der Tabelle 36 werden die Ergebnisse dieser multiplen logistischen Regressionsanalyse mit den dichotomisierten Gesamtbewertungen der Fahraufgaben als Prädiktoren und der Prüfungsentscheidung als Kriterium dargestellt.<sup>55</sup>

<sup>53</sup> Dies ist auch gar nicht möglich, da die Prüfungsrichtlinie (noch) nicht an die optimierte Praktische Fahrerlaubnisprüfung angepasst wurde und es im bestehenden Prüfungssystem keine Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche gibt. Hier entscheiden lediglich die dokumentierten Fehler über das Bestehen oder Nichtbestehen der Prüfung.

<sup>54</sup> In der vorliegenden Stichprobe haben 10,8 Prozent der Bewerber mit einer als „Ungenügend“ bewerteten Fahraufgabe die Prüfung bestanden. Bezogen auf die Kompetenzbereiche liegt dieser Wert bei 10,0 Prozent.

<sup>55</sup> Der Omnibus-Test der Modellkoeffizienten ergibt einen Chi-Quadrat-Wert von 423,66, wonach das Modell die Zielvariable mit einer Irrtumswahrscheinlichkeit von  $p = 0,00$  besser vorhersagt als ein „leeres“ Modell ohne Prädiktoren. Auch das R-Quadrat (Nagelkerkes) von 37,3 Prozent erklärter Varianz indiziert einen guten Modellfit; die Tests auf Multikollinearität erbrachten keine kritischen Ergebnisse.

	B-Koeffizient
Ein-, Ausfädeln, Fahrstreifenwechsel	1,47***
Kurven	2,45
Vorbeifahren, Überholen	0,85*
Kreuzungen, Einmündungen	2,83***
Kreisverkehr	0,11
Schienenverkehr	0,95
Fußgänger, Radfahrer, Haltestellen	1,38***
Geradeausfahren	1,04***

**Tab. 36:** Einfluss der Gesamtbewertungen der Fahraufgaben auf die Prüfungsentscheidung (B-Koeffizienten einer logistischen multiplen Regressionsanalyse)

	B-Koeffizient
Verkehrsbeobachtung	4,37***
Fahrzeugpositionierung	2,88***
Geschwindigkeitsanpassung	2,09***
Kommunikation	2,42**
Fahrzeugbedienung/Umweltbewusste Fahrweise	4,17***

**Tab. 37:** Einfluss der Gesamtbewertungen der Kompetenzbereiche auf die Prüfungsentscheidung (B-Koeffizienten einer logistischen multiplen Regressionsanalyse)

Den stärksten Einfluss auf die Prüfungsentscheidung hat nach diesen Berechnungen die Fahraufgabe „Kreuzungen, Einmündungen“. Dies deckt sich mit den Erkenntnissen aus Kapitel 5.4.1, wonach diese Fahraufgabe bei den Bewerbern besonders viele Schwierigkeiten verursacht. Danach folgen die Fahraufgaben „Ein-, Ausfädeln, Fahrstreifenwechsel“, „Fußgänger, Radfahrer, Haltestellen“ und „Geradeausfahren“. „Vorbeifahren, Überholen“ hat einen nochmals geringeren Einfluss.

Für die verbleibenden Fahraufgaben lässt sich – unter Bereinigung des Einflusses der anderen Fahraufgaben auf die Prüfungsentscheidung – kein statistisch abgesicherter Einfluss auf die Prüfungsentscheidung nachweisen. Dies hängt auch mit der Tatsache zusammen, dass „Ungenügende“ Leistungen bei diesen Fahraufgaben relativ selten vorkamen.

### Gesamtbewertung der Kompetenzbereiche und Prüfungsentscheidung

Die Gesamtbewertung der Kompetenzbereiche erfolgt auf derselben Skala wie die oben beschriebenen Gesamtbewertungen der Fahraufgaben. Daher wurden zur Analyse der Zusammenhänge mit der Prüfungsentscheidung zunächst wieder

Mann-Whitney-U-Tests durchgeführt, die hochsignifikante positive Korrelationen zwischen allen Gesamtbewertungen der Kompetenzbereiche und der Prüfungsentscheidung ergeben. Für einen zusammenfassenden Vergleich der bereinigten Einflussstärken der einzelnen Gesamtbewertungen der Fahrkompetenzbereiche auf die Prüfungsentscheidung wurden die Gesamtbewertungen wiederum nach dem oben beschriebenen Kriterium dichotomisiert und als Prädiktoren für eine multiple logistische Regression auf die Prüfungsentscheidung verwendet. Entsprechend den Ergebnissen aus Kapitel 5.4.1, hat die „Verkehrsbeobachtung“ die größte Bedeutung für die Prüfungsentscheidung (s. Tabelle 37).<sup>56</sup> Danach folgen allerdings nicht die „Fahrzeugpositionierung“ und die „Geschwindigkeitsanpassung“, sondern der Kompetenzbereich „Fahrzeugbedienung/Umweltbewuss-

<sup>56</sup> Der Omnibus-Test der Modellkoeffizienten ergibt einen Chi-Quadrat-Wert von 3.764,61, wonach das Modell die Zielvariable mit einer Irrtumswahrscheinlichkeit von  $p = 0,00$  besser vorhersagt als ein „leeres“ Modell ohne Prädiktoren. Gemäß der Berechnung von Nagelkerkes R-Quadrat beträgt der Anteil der durch das Modell erklärten Varianz 64,4 Prozent, was als sehr hoch bezeichnet werden kann. Die Tests auf Multikollinearität ergaben keine kritischen Ergebnisse.

te Fahrweise“. Auch die „Kommunikation“ hat einen – verglichen mit den univariaten Analysen aus Kapitel 5.4.1 – unerwartet starken Einfluss auf die Prüfungsentscheidung. Dies könnte dem Umstand geschuldet sein, dass „Ungenügende“ Leistungen in den Bereichen „Fahrzeugbedienung/ Umweltbewusste Fahrweise“ und „Kommunikation“ nur in 1,4 bzw. 0,5 Prozent der Fälle auftreten. Die äußerst seltene Verwendung dieser Bewertungsstufe für diese beiden Fahraufgaben hat eine geringere Varianz bei der Bewertung dieser Fahraufgaben zur Folge. Faktisch beschränkt sich die Gesamtbewertung der Fahraufgaben „Fahrzeugbedienung/Umweltbewusste Fahrweise“ und „Kommunikation“ fast vollständig auf nur drei statt vier Stufen, wodurch sich der zugehörige B-Koeffizient im Vergleich zu den Fahraufgaben mit höherer Bewertungsvarianz erhöht.

### **Prüfungsrahmenbedingungen und Prüfungsentscheidung**

Aus ökonomischen Erwägungen wurden im Revisionsprojekt bzw. bei den Bewerberbefragungen zur populationsspezifischen Äquivalenz (s. Kapitel 5.3) nicht speziell Prüfungsrahmenbedingungen erhoben, um deren Einfluss auf die Prüfungsentscheidung zu untersuchen. Im vorliegenden Datensatz stehen daher lediglich Daten zur Tageszeit der Prüfung für diesbezügliche Analysen zur Verfügung. So bleibt es künftigen Untersuchungen – ggf. im Rahmen der kontinuierlichen Evaluation – vorbehalten, die Einflüsse von Prüfungsbedingungen wie beispielsweise Witterungs-, Straßen- und Verkehrsbedingungen auf die Prüfungsentscheidung zu erhellen.

Bezüglich des Einflusses der Tageszeit auf die Prüfungsentscheidungen lässt sich erstens vermuten, dass Prüfungen bei reduzierter Helligkeit aufgrund der eingeschränkten Sicht und der damit verbundenen erhöhten Anforderungen an die Bewerber mit schlechteren Prüfungsleistungen einhergehen. Zur Überprüfung dieser Hypothese wurden mithilfe astronomischer Daten alle Prüfungen identifiziert, die vollständig unter reduzierter Helligkeit stattfanden. Als Beginn bzw. Ende der reduzierten Helligkeit wurden dabei Sonnenaufgangs- und Sonnenuntergangszeiten der jeweiligen Modellregion verwendet. Alle Prüfungen, die vor Sonnenaufgang endeten oder nach Sonnenuntergang begannen, galten als Prüfungen unter reduzierter Helligkeit. Dabei zeigte es sich, dass der Anteil dieser Prüfungen an der Gesamtheit aller Prüfungen äußerst gering ausfiel und

sich im Rahmen der vorliegenden Stichprobe keine statistisch signifikanten Zusammenhänge zwischen der Helligkeit während der Prüfungen einerseits und den Prüfungsbewertungen bzw. Prüfungsentscheidungen andererseits fanden.

Als zweiter tageszeitabhängiger Einflussfaktor auf die Prüfungsleistungen ist die Vigilanz<sup>57</sup> der Bewerber vorstellbar. Diese ist aufgrund circadianer Rhythmen bei den meisten Menschen am Vormittag höher als am Nachmittag (DINGES & POWELL, 1985). HORNE und REYNER (1995) haben nachgewiesen, dass die Wahrscheinlichkeit, mit einem Kraftfahrzeug aufgrund von Müdigkeit zu verunfallen, nachts sowie nachmittags gegen 15 Uhr am höchsten ist. Nachts finden keine Fahrerlaubnisprüfungen statt. Die verminderte Vigilanz am Nachmittag könnte sich aber auch auf die Prüfungsleistungen der Fahrerlaubnisbewerber zu dieser Tageszeit auswirken.

Zur Überprüfung dieser Hypothese wurden Prüfungen, die zwischen 14 und 16 Uhr absolviert wurden, mit Prüfungen verglichen, die zwischen 9 und 12 Uhr stattfanden. Im Ergebnis zeigt sich kein signifikanter Zusammenhang zwischen den genannten Zeitfenstern der Prüfungsdurchführung und den Prüfungsbewertungen bzw. Prüfungsentscheidungen. Dies könnte dadurch erklärt werden, dass die Praktische Fahrerlaubnisprüfung eine Ausnahmesituation für die Bewerber darstellt. Der damit verbundene erhöhte Erregungsgrad und die relativ kurze Dauer der Prüfung begünstigen ein hohes durchgängiges Aufmerksamkeitsniveau, auch wenn die Prüfung am Nachmittag stattfindet. Allerdings fanden sich geschlechtsspezifische Besonderheiten: Während sich bei Frauen kein signifikanter Zusammenhang fand, hatten Männer zwischen 9 und 12 Uhr eine um mehr als 10 Prozent erhöhte Bestehensquote gegenüber dem Referenzzeitraum (14 bis 16 Uhr). Hierfür sind verschiedene Erklärungen vorstellbar. So ist beispielsweise bekannt, dass Frauen sich in der Regel intensiver auf Prüfungen vorbereiten als Männer (CHETCUTI, 2009), womit sich Aufmerksamkeitsschwächen u. U. leichter vermeiden oder kompensieren lassen. Ob diese oder andere Erklärungen zutreffen, lässt sich nicht auf der Grundlage der vorliegenden Daten, sondern ggf. erst im Rahmen der kontinuierlichen Prüfungsevaluation überprüfen.

<sup>57</sup> Unter Vigilanz versteht man in der Psychologie Zustände andauernder Aufmerksamkeit.

## Zwischenfazit

Im Ergebnis der Analysen bleibt festzuhalten, dass sich in der vorliegenden Stichprobe schlechte Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche sowie eine hohe Zahl dokumentierter Fehler negativ auf die Wahrscheinlichkeit auswirken, die Prüfung zu bestehen. Hinsichtlich des Zusammenhangs zwischen den Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche einerseits sowie der Prüfungsentscheidung andererseits bestätigen sich vorliegende wissenschaftliche Erkenntnisse, wonach die Bewältigung der Fahraufgabe „Kreuzungen, Einmündungen“ und der Kompetenzbereich „Verkehrsbeobachtung“ besonders große Herausforderungen für Fahranfänger darstellen. Beide Befunde stützen die Annahme, dass die Fahrerlaubnisprüfer die Durchführungsstandards sowie die Bewertungs- und Entscheidungskriterien korrekt anwenden.

### 5.4.4 Prüfungsbewertungsmuster und Prüfertypen

Für die Untersuchung multivariater Prüfungsbewertungsmuster (s. Kapitel 3.5) sollen im vorliegenden Kapitel zunächst Systematiken der Prüfungsbewertung ermittelt werden. In einem zweiten Schritt sollen diese Systematiken dann in einen Zusammenhang mit bestimmten (pseudonymisierten) Beobachtern gebracht werden. Dadurch lassen sich gegebenenfalls Prüfertypen mit Bewertungsbesonderheiten identifizieren, deren Kenntnis für das Qualitätsmanagement sowie die Aus- und Fortbildung der Fahrerlaubnisprüfer der technischen Prüfstellen von großem Interesse sein könnte.

#### Identifikation von Prüfungsbewertungsmustern

Die Analyse von Prüfungsbewertungsmustern im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung könnte grundsätzlich auf der Basis der situationsbezogenen und der situationsübergreifenden Bewertungskriterien erfolgen. Die situa-

tionsbezogenen Bewertungskriterien – d. h. die konkreten Fehler und überdurchschnittlichen Leistungen – treten zwar sehr vielfältig, aber jeweils für sich gesehen relativ selten auf. Dadurch ergibt sich eine extrem große Zahl möglicher Kombinationen, die sich nicht sinnvoll miteinander vergleichen lassen. Die Analyse der Prüfungsbewertungsmuster beschränkt sich daher im Folgenden auf die situationsübergreifenden Bewertungen der Fahraufgaben und Kompetenzbereiche.

Die Bewertung der Fahraufgaben und Kompetenzbereiche erfolgt auf einer vierstufigen Ordinalskala (s. o.). Daraus ergeben sich allein schon 65.536 Kombinationsmöglichkeiten bezüglich der Gesamtbewertungen der 8 Fahraufgaben. Diese große Anzahl zeigt bereits, dass eine detaillierte Auswertung und Diskussion aller Kombinationsmöglichkeiten auch auf dieser Bewertungsebene nicht zielführend ist. Zusammenfassend kann jedoch festgehalten werden, dass die häufigsten Kombinationen ausschließlich die Bewertungsstufen „Gut“ und „Ausreichend“ enthalten. Die häufigste Kombination enthält sowohl für die Gesamtbewertungen der Fahraufgaben als auch für die Gesamtbewertungen der Kompetenzbereiche ausschließlich die Bewertung „Gut“. Gemäß der im e-Prüfprotokoll hinterlegten Definitionen der Bewertungsstufen ist die Kategorie „Gut“ auszuwählen, wenn weder Fehler noch überdurchschnittliche Leistungen dokumentiert wurden, sondern der Bewerber das erwartungsgemäße, weil fahranfängertypische Fahrleistungsniveau demonstriert hat.

Auf der Ebene der situationsübergreifenden Bewertungen der Fahraufgaben und Kompetenzbereiche kann man das ordinale Skalenniveau nutzen, um – anstelle einer einfachen Auszählung und des Vergleichs von Häufigkeiten bestimmter Bewertungskombinationen – Korrelationen zwischen den Gesamtbewertungen zu berechnen und dadurch Bewertungsmuster zu ermitteln. Für die Bestimmung des Zusammenhangs zweier Ordinalskalen ist Spearman's Rho das geeignete Maß (BORTZ, 1999). Tabelle 38 zeigt die Korrelations-

	VB	FP	GA	KO	FBU
Verkehrsbeobachtung (VB)	1,00***	0,49***	0,48***	0,44***	0,42***
Fahrzeugpositionierung (FP)		1,00***	0,48***	0,46***	0,45***
Geschwindigkeitsanpassung (GA)			1,00***	0,47***	0,47***
Kommunikation (KO)				1,00***	0,48***
Fahrzeugbedienung/Umweltbewusste Fahrweise (FBU)					1,00***

Tab. 38: Korrelationsmatrix (Spearman's Rho) für die Gesamtbewertungen der Kompetenzbereich



koeffizienten für die Gesamtbewertungen der Kompetenzbereiche als Matrix.

Aus Tabelle 38 ist ersichtlich, dass alle Kompetenzbereiche auf einem hohen und relativ einheitlichen Niveau miteinander korrelieren. Empirisch gesehen, stellen die fünf Kompetenzbereiche damit eigenständige und nicht redundante Faktoren der Fahrkompetenz dar. Dies bestätigt das theoretische Fahrkompetenzkonzept, das der optimierten Praktischen Fahrerlaubnisprüfung zugrunde liegt, und spricht für eine korrekte Prüfungsdurchführung im Prüfungstag. Die Vermutung aus Kapitel 3.5, wonach „Fahrzeugpositionierung“ und „Geschwindigkeitsanpassung“ sowie „Verkehrsbeobachtung“ und „Kommunikation“ erhöhte Zusammenhänge untereinander aufweisen könnten, haben sich damit nicht bestätigt.

Zur Prüfung des genannten Befundes soll mit der Faktorenanalyse ein multivariates Analyseverfahren herangezogen werden, mit dem die korrelativen Beziehungen der Gesamtbewertungen der Kompetenzbereiche nicht paarweise, sondern in ihrer Gesamtheit verglichen werden können. Die Faktorenanalyse ist ein Verfahren zur Reduktion einer größeren Zahl von Dimensionen bzw. Variablen auf wenige Einflussfaktoren, indem künstliche Variablen berechnet werden, die mit den vorhandenen Variablen möglichst hoch korrelieren. Diese künstlichen Variablen werden als „Faktoren“ bezeichnet. Auf diesem Wege können Variablen anhand ihrer korrelativen Beziehungen in Gruppen klassifiziert werden (BORTZ, 1999).

„Faktorenanalyse“ ist ein Sammelbegriff für eine Reihe dimensionsreduzierender Verfahren, wobei vor allem die Hauptachsenmethode große Verbreitung gefunden hat. Die Anwendung der Hauptachsenmethode auf die Gesamtbewertungen der Kompetenzbereiche<sup>58</sup> ergibt eine eindeutige Einfaktorenlösung.

Alle Variablen korrelieren hoch mit einem Faktor.<sup>59</sup> eitere Faktoren, die berechnet wurden, korrelieren viel niedriger mit den Gesamtbewertungen; sie

haben einen Eigenwert kleiner 1 und tragen somit nach dem Kaiser-Kriterium nicht zur Dimensionsreduktion bei (BACKHAUS et al., 2005).

Die Tabelle 39 zeigt die Ladungen der einzelnen Kompetenzbereiche auf den ersten Faktor. Das dargestellte Ergebnis stützt den Befund, dass die fünf Kompetenzbereiche gleichermaßen unabhängige und nicht redundante Facetten der Fahrkompetenz widerspiegeln, die zwar hoch miteinander korrelieren, aber alle in einem ausgewogenen Verhältnis zueinander und zum Fahrkompetenzkonstrukt stehen.

In Tabelle 40 ist die Matrix der Korrelationskoeffizienten nach Spearman für die Gesamtbewertungen der Fahraufgaben dargestellt. Dabei zeigen sich in vielen Fällen paarweise erhöhte Korrelationen zwischen den Fahraufgaben „Kurven“, „Kreisverkehr“, „Schienenverkehr“ und „Fußgänger, Radfahrer, Haltestellen“. Zur Erklärung dieses Befundes könnten inhaltliche Überlegungen herangezogen werden, wonach es beispielsweise gewisse offensichtliche Überschneidungen hinsichtlich der Anforderungen bei der „Fahrzeugpositionierung“ zwischen den Fahraufgaben „Kurven“ und „Kreisverkehr“ gibt. Allerdings wären demnach auch Überschneidungen hinsichtlich der Anforderungen bei der „Verkehrsbeobachtung“ zwischen den Fahraufgaben „Kreuzungen, Einmündungen“ und „Kreisverkehr“ zu erwarten (s. Kapitel 3.5), die sich aber nicht in überdurchschnittlichen Korrelationen

<sup>58</sup> Das Verfahren der Faktorenanalyse setzt eigentlich metrisch skalierte Variablen voraus. In der sozialwissenschaftlichen Praxis werden allerdings auch parametrische Verfahren auf ordinalskalierte Variablen angewendet, wenn für die Zielstellung keine vergleichbaren nichtparametrischen Verfahren verfügbar sind (s. dazu auch Fußnote 43).

<sup>59</sup> Die zugrunde liegende Stichprobe ist nach dem Test von KAISER-MEYER-OLKIN mit einem Wert von 0,85 gut für eine Faktorenanalyse geeignet. Das Modell ist nach dem Test von BARTLETT mit einer Irrtumswahrscheinlichkeit von 0,00 hochsignifikant. Der extrahierte Faktor erklärt 46,8 Prozent der Gesamtvarianz aller einbezogenen Variablen.

	Ladungen Faktor 1
Verkehrsbeobachtung	0,67
Fahrzeugpositionierung	0,70
Geschwindigkeitsanpassung	0,71
Kommunikation	0,68
Fahrzeugbedienung/Umweltbewusste Fahrweise	0,66

Tab. 39: Ergebnisse einer Hauptachsen-Faktorenanalyse auf die Gesamtbewertungen der Kompetenzbereiche

	FA1	FA2	FA3	FA4	FA5	FA6	FA7	FA8
FA1	1,00***	0,54***	0,53***	0,47***	0,63***	0,57***	0,51***	0,50***
FA2		1,00***	0,57***	0,44***	0,71***	0,67***	0,62***	0,48***
FA3			1,00***	0,46***	0,62***	0,60***	0,59***	0,49***
FA4				1,00***	0,51***	0,48***	0,47***	0,52***
FA5					1,00***	0,73***	0,68***	0,54***
FA6						1,00***	0,70***	0,50***
FA7							1,00***	0,48***
FA8								1,00***

Tab. 40: Korrelationsmatrix (Spearman's Rho) für die Gesamtbewertungen der Fahraufgaben-FA

zwischen den Gesamtbewertungen dieser Fahraufgaben niederschlagen. Auch ist inhaltlich kaum zu erklären, warum ausgerechnet die Fahraufgaben „Kreisverkehr“ und „Schienenverkehr“ die höchste paarweise Korrelation aller Fahraufgaben<sup>60</sup> aufweisen, wo deren Anforderungen an die Bewerber doch recht unterschiedlich sind.

Eine mögliche – nicht inhaltliche, sondern statistische – Erklärung für die erhöhten paarweisen Korrelationen zwischen den Fahraufgaben „Kurven“, „Kreisverkehr“, „Schienenverkehr“ und „Fußgänger, Radfahrer, Haltestellen“ bietet die univariate Verteilung der Antwortkategorien dieser Variablen. Wie bereits dargestellt (s. Kapitel 5.4.1), wurden alle diese Fahraufgaben in über 80 Prozent der Fälle mit „Gut“ bewertet, wohingegen die Bewertungen „Sehr gut“ und „Ungenügend“ fast nie ausgewählt wurden.

Folglich ist die Varianz bei diesen Fahraufgabenbewertungen geringer als bei den anderen, was eine erhöhte Korrelation zur Folge haben könnte. Insofern sind die erhöhten paarweisen Zusammenhänge zwischen den Gesamtbewertungen der oben genannten Fahraufgaben nicht in inhaltlichen Überschneidungen ihrer Anforderungen an die Bewerber begründet, sondern in der hohen Zahl „guter“ Leistungen bei der Durchführung dieser Fahraufgaben. Somit stellen auf Grundlage der gegebenen

empirischen Basis auch alle Fahraufgaben eigenständige und nicht redundante Anforderungskategorien dar. Dies spricht wiederum für eine gute theoretisch-konzeptionelle Fundierung und eine korrekte Anwendung der optimierten Praktischen Fahrerlaubnisprüfung. Eine Hauptachsen-Faktorenanalyse mit den Gesamtbewertungen der Fahraufgaben<sup>61</sup> bestätigt diesen Befund (s. Tabelle 41). Alle Fahraufgabenbewertungen laden auf einen Faktor hoch; alle weiteren Faktoren haben einen Eigenwert kleiner 1 und tragen nicht zur Dimensionsreduzierung bei. Die Fahraufgaben bilden also eine homogene Gruppe von Anforderungskategorien, die zwar untereinander hoch korrelieren, dies aber gleichmäßig tun und sich daher nicht zu Gruppen zusammenfassen lassen.

### Identifikation von Prüfertypen

Für die Identifikation von Prüfertypen werden im folgenden Kapitel die Prüfungsbewertungen und die Prüfungsentscheidungen herangezogen. Eine detaillierte Auswertung hinsichtlich der situationsbezogenen Bewertungskriterien erscheint dabei aufgrund der oben genannten Argumente (große Anzahl und seltenes Auftreten der einzelnen Kombinationen) erneut nicht sinnvoll. Daher wurde – statt der konkreten Fehler und überdurchschnittlichen Leistungen – lediglich deren durchschnittliche Anzahl pro Fahrerlaubnisprüfer berücksichtigt. Bei der Berechnung dieser Durchschnittswerte wurden die Prüfer mit weniger als 10 Prüfungen im Erprobungszeitraum nicht berücksichtigt, da bei solch kleinen Fallzahlen die Durchschnittswerte zufallsbedingt stark nach oben oder unten „ausreißen“ können. Damit verblieben 39 Prüfer in der Stichprobe. Um den Einfluss von „Ausreißerfällen“ bei der Interpretation der Daten noch weiter zu minimieren, wurden weiterhin das Zehnprozent-Quartil und das Neunzigprozent-Quartil berechnet und Mittelwerte darunter bzw. da-

<sup>60</sup> FA1 = „Ein-, Ausfädeln, Fahrstreifenwechsel“; FA2 = „Kurven“; FA3 = „Vorbeifahren, Überholen“; FA4 = „Kreuzungen, Einmündungen“; FA5 = „Kreisverkehr“; FA6 = „Schienenverkehr“; FA7 = „Fußgänger, Radfahrer, Haltestellen“; FA8 = „Geradausfahren“

<sup>61</sup> Der Test nach KAISER-MEYER-OLKIN ergibt einen Wert von 0,95, d. h. die Stichprobe ist gut für eine Faktorenanalyse geeignet. Nach dem Test von Bartlett ist das Modell mit einer Irrtumswahrscheinlichkeit von 0,00 hochsignifikant. Der extrahierte Faktor erklärt 46,8 Prozent der Gesamtvarianz der einbezogenen Variablen.

	Ladungen Faktor 1
Ein-, Ausfädeln, Fahrstreifenwechsel	0,79
Kurven	0,88
Vorbeifahren, Überholen	0,81
Kreuzungen, Einmündungen	0,70
Kreisverkehr	0,83
Schieneverkehr	0,83
Fußgänger, Radfahrer, Haltestellen	0,83
Geradeausfahren	0,74

Tab. 41: Ergebnisse einer Hauptachsen-Faktorenanalyse mit den Gesamtbewertungen der Fahraufgaben

rüber bei der Datenanalyse außer Acht gelassen.<sup>62</sup> Dadurch gehen sehr große und sehr kleine („Ausreißer“-)Werte nicht in die Dateninterpretation ein.

Wie bereits erwähnt, sollen bei der Identifikation von möglichen Prüfertypen auch die situationsübergreifenden Bewertungskriterien, also die Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche herangezogen werden. Dabei ermöglicht das ordinale Skalenniveau der Gesamtbewertungen eine breitere Auswahl möglicher Analyseverfahren; für den vorliegenden Zweck besonders geeignet erscheint die Clusteranalyse. Hinter dem Begriff „Clusteranalyse“ verbirgt sich eine Reihe von Verfahren, die eine bestehende Anzahl von Objekten anhand ausgewählter Merkmale in Gruppen einteilen. Das Ziel besteht darin, die Gruppen so einzuteilen, dass diese in Bezug auf die ausgewählten Merkmale in sich möglichst homogen und untereinander möglichst heterogen sind (LOHSE, LUDWIG & RÖHR, 1986). Im Hinblick auf die vorliegende Fragestellung interessiert uns, ob sich die Fahrerlaubnisprüfer im Hinblick auf ihre Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche systematisch voneinander unterscheiden: ob sich also Prüfergruppen finden lassen, die sich im Gruppenvergleich stark unterscheiden, deren Mitglieder sich aber jeweils hinsichtlich ihres Bewertungsverhaltens stark ähneln.

Eingangs wurde bereits erwähnt, dass sich hinter dem Begriff der Clusteranalyse verschiedenen Methoden zur Gruppierung von Objekten stehen. Unterschiede zwischen den Methoden bestehen hinsichtlich des Proximitätsmaßes,<sup>63</sup> mit dem Ähnlichkeiten bzw. Unähnlichkeiten zwischen den Objekten quantifiziert werden, und bezüglich des Fusionierungsalgorithmus,<sup>64</sup> auf dessen Basis kleine Gruppen schrittweise zu größeren Gruppen zusammengefasst werden.

Nach der Wahl des geeigneten Proximitätsmaßes und des Fusionierungsalgorithmus<sup>64</sup> muss (bei hierarchischen Verfahren) festgelegt werden, an welcher Stelle die Clusterfusionierung stoppen soll, welche Anzahl von Clustern also die beste Lösung

<sup>62</sup> Ordnet man alle prüferbezogenen Mittelwerte ihrer Größe nach, dann ist das Zehnprozent-Quartil der Wert, der die 10 Prozent der kleinsten Prüfermittelwerte vom Rest der Verteilung trennt. Entsprechend „schneidet“ das Neunzigprozent-Quartil die 10 Prozent der größten Prüfermittelwerte vom Rest der Verteilung ab.

<sup>63</sup> Für das Zusammenfassen der einzelnen Beobachtungen zu Fällen („Clustering“) muss zunächst die Ähnlichkeit der Objekte zueinander quantifiziert werden. Dies geschieht über ein sog. Proximitätsmaß (BACKHAUS et al., 2006). Das am häufigsten verwendete Proximitätsmaß ist die euklidische Distanz (BORTZ, 1999).

<sup>64</sup> Nachdem die Distanz zwischen den Objekten bestimmt wurde, müssen diese zu Clustern zusammengefasst werden. Dieser Prozess wird als „Fusionierung“ bezeichnet. Ähnlich wie bei der Distanzermittlung der Objekte steht auch bei deren Fusionierung eine Vielzahl möglicher Verfahren zur Verfügung. Grundlegend kann dabei zwischen partitionierenden und hierarchischen Verfahren unterschieden werden. Partitionierende Verfahren weisen die Objekte zu Beginn zufällig einer vordefinierten Anzahl von Gruppen zu. Bei den wichtigsten hierarchischen Verfahren stellt jedes Objekt zunächst ein eigenes Cluster dar. Die Objekte, welche die kleinste Distanz zueinander aufweisen, werden als erste fusioniert (BORTZ, 1999). Anschließend werden die Ähnlichkeiten zwischen den Objekten bzw. Clustern neu berechnet, und es werden wieder die zwei Cluster mit der geringsten Distanz zueinander zusammengefasst. Dieser Prozess wird so lange fortgeführt, bis alle Objekte einem Cluster angehören. Mittels eines Dendrogramms kann dieser Fusionierungsprozess grafisch veranschaulicht werden. Für die Fusionierung zweier Cluster wurden verschiedene Kriterien – z. B. „Single Linkage“, „Complete Linkage“, „Average Linkage“, „Medianverfahren“ – entwickelt, wobei das Ward-Verfahren nach BACKHAUS et al. (2006) in den meisten Fällen sehr gute Einteilungen findet. Bevor das Ward-Verfahren angewendet wird, empfiehlt sich aber die Identifikation und Eliminierung von „Ausreißerobjekten“ – also von Fällen mit sehr hohen bzw. niedrigen Werten – weil diese die Gruppenbildung ansonsten verzerren können (ebd.).

darstellt. Da die Clusteranalyse im Allgemeinen explorativ zur Identifikation bisher unbekannter Gruppen von Objekten angewendet wird, können sachlogische Überlegungen zur optimalen Anzahl der Cluster meist nicht als Entscheidungsgrundlage herangezogen werden. In diesem Fall sollte die Bestimmung der optimalen Clusteranzahl ausschließlich anhand statistischer Kriterien erfolgen.

Das Dendrogramm bereitet den schrittweisen Zusammenschluss der Clusterfusionierung bei hierarchischen Verfahren grafisch auf und gibt Hinweise auf optisch sinnvolle Gruppentrennungen. Beim Ward-Verfahren kann zusätzlich die Zunahme der Fehlerquadratsumme bei der Fusionierung der Cluster als Kriterium für die Bestimmung der optimalen Clusterzahl herangezogen werden. Zeigt sich im Prozess der schrittweisen Clusterfusionierung an einer Stelle eine sprunghafte Zunahme der Fehlerquadratsumme, so gibt das einen Hinweis darauf, dass die Objekte des neu gebildeten Clusters deutlich weniger homogen in Bezug auf die Gruppierungsvariablen sind, als dies bei den vor diesem Schritt existierenden Clustern der Fall war. Zur Identifikation einer solchen sprunghaften Heterogenitätszunahme ist es hilfreich, die Entwicklung der Fehlerquadratsumme gegen die Zahl der Cluster in einem Koordinatensystem abzutragen und die benachbarten Punkte jeweils mit einer

Linie zu verbinden. Zeigt sich in diesem Diagramm ein „Ellenbogen“, so kann dies als Entscheidungskriterium für die Auswahl der optimalen Clusteranzahl benutzt werden („Elbow-Kriterium“).

Neben dem Dendrogramm und dem „Elbow-Kriterium“ erscheint es sinnvoll, für eine gegebene Clusterlösung die Mittelwerte, Standardabweichungen, Minima und Maxima pro Cluster und Gruppierungsvariable in einer Kreuztabelle darzustellen. So entstehen „Steckbriefe“ der einzelnen Cluster, und es besteht die Möglichkeit Variablen zu identifizieren, die sich in den Clustern nicht unterscheiden und somit wenig Trennkraft besitzen. Die Analyse solcher Steckbriefe für verschiedene Clusterlösungen gibt zusätzliche Hinweise auf die Güte der berechneten Clusterlösungen.

Die Clusteranalyse nach dem Ward-Verfahren wurde zwei Mal durchgeführt – auf Basis der prüferbezogenen Mittelwerte der Gesamtbewertungen der Fahraufgaben ( $n = 41$ ) und auf Basis der prüferbezogenen Mittelwerte der Gesamtbewertungen der Kompetenzbereiche ( $n = 42$ ). Anhand des „Elbow-Kriteriums“ und der Entwicklung der Fehlerquadratsumme wurde in beiden Fällen eine Drei-Cluster-Lösung als die beste Lösung identifiziert. Die Tabelle 42 enthält pro Cluster die Mittelwerte, Minima und Maxima bezogen auf die durchschnittlichen Be-

		Verkehrsbeobachtung	Fahrzeugpositionierung	Geschwindigkeitsanpassung	Kommunikation	Fahrzeugbedienung/ Umweltbewusste Fahrweise
Streng	Mittelwert*	3,01	2,87	2,87	2,78	2,79
	Minimum*	2,91	2,77	2,76	2,61	2,59
	Maximum*	3,13	3,00	3,03	2,98	3,00
	N	3	3	3	3	3
Mittelstreng	Mittelwert*	2,75	2,50	2,49	2,25	2,38
	Minimum*	2,45	2,30	2,31	2,10	2,23
	Maximum*	2,95	2,73	2,65	2,46	2,58
	N	13	13	13	13	13
Milde	Mittelwert*	2,39	2,29	2,25	2,09	2,20
	Minimum*	2,08	2,10	2,06	2,00	2,05
	Maximum*	2,64	2,57	2,48	2,24	2,36
	N	26	26	26	26	26
Gesamt	Mittelwert*	2,55	2,40	2,37	2,19	2,30
	Minimum*	2,08	2,10	2,06	2,00	2,05
	Maximum*	3,13	3,00	3,03	2,98	3,00
	N	42	42	42	42	42

Tab. 42: Clusteranalyse mit den prüferbezogenen Mittelwerten der Gesamtbewertungen der Kompetenzbereiche („Steckbrief“ der Drei-Cluster-Lösung)



wertungen der zugehörigen Prüfer in den fünf Kompetenzbereichen.<sup>65</sup> Zusätzlich ist die Anzahl der Prüfer (N) vermerkt, die dem jeweiligen Cluster zugeordnet wurden. Die Prüfer im ersten Cluster weisen die nominal höchsten durchschnittlichen Bewertungen auf, was in diesem Fall schlechte Bewertungen indiziert. Sie werden daher als „Streng“ bezeichnet. Die Prüfer im dritten Cluster weisen die nominal niedrigsten Bewertungen auf und werden im Umkehrschluss als „Milde“ interpretiert. Die durchschnittlichen Bewertungen in der zweiten Gruppe liegen nominal zwischen denen der „Strengen“ und der „Milden“ Gruppe; die zugehörigen Prüfer werden daher als „Mittelstreng“ bezeichnet.

Ein sehr ähnliches Bild ergibt die Drei-Cluster-Lösung anhand der Gesamtbewertungen der Fahraufgaben. Die Übereinstimmung zwischen den beiden Clusterlösungen hinsichtlich der Zuordnung der Prüfer zu den Clustern liegt bei 78 %. Folglich hängt die „Strenge“ bei der Bewertung der Fahraufgaben eng mit der „Strenge“ bei der Bewertung der Kompetenzbereiche zusammen. Das Cluster mit den „strengen“ Prüfern ist in beiden Fällen mit (den gleichen) drei Prüfern vergleichsweise schwach besetzt. Es gibt also in der vorliegenden Stichprobe einen Anteil von ca. 7 % der Prüfer, die Prüfungsleistungen besonders „streng“ bewerten und daher auch besonders niedrige Bestehensquoten aufweisen.

Zur Identifikation weiterer Prüfertypen wurden weitere Clusteranalysen anhand verschiedener Kriterien durchgeführt. Interessante Ergebnisse ergaben sich bei Verwendung der prüferbezogenen Korrelationskoeffizienten zwischen den Kompetenzbereichen als Gruppierungskriterien. Prüfer mit weniger als 50 Prüfungen wurden dabei aus Gründen der statistischen Signifikanz nicht berücksichtigt; die Clusteranalyse basierte auf 38 von 44 Prüfern. Nach der Begutachtung des Dendrogramms und der Kontrolle des „Elbow-Kriteriums“ hat sich wiederum eine Drei-Cluster-Lösung als ideal erwiesen.

Die Tabelle 43 enthält pro Cluster die Mediane, Minima und Maxima der prüferbezogenen Korrelationen zwischen den Gesamtbewertungen der Kompetenzbereiche,<sup>66</sup> Zusätzlich ist die Anzahl der Prüfer vermerkt, die dem jeweiligen Cluster zugeordnet wurden (N). Das erste Cluster weist im Mittel (Median) die höchsten Korrelationen zwischen den Gesamtbewertungen der Kompetenzbereiche auf und wird daher als „Homogen“ hinsichtlich der Prüfungsbewertungen bezeichnet. Im zweiten

65 Skalenstufen: 1 = „Sehr gut“; 2 = „Gut“; 3 = „Ausreichend“; 4 = „Ungenügend“.

66 Skalenstufen: 1 = „Sehr gut“; 2 = „Gut“; 3 = „Ausreichend“; 4 = „Ungenügend“.

		VB~FP	VB~GA	VB~KO	VB~FBU	FP~GA	FP~KO	FP~FBU	GA~KO	GA~FBU	KO~FBU
Homogen	Median	0,82	0,81	0,72	0,70	0,83	0,84	0,83	0,78	0,82	0,87
	Minimum	0,63	0,73	0,68	0,63	0,71	0,79	0,73	0,71	0,67	0,71
	Maximum	0,85	10,00	0,85	0,86	0,85	10,00	10,00	0,93	0,92	10,00
	N	5	5	5	5	5	5	5	5	5	5
Mittel-homogen	Median	0,45	0,39	0,33	0,28	0,40	0,33	0,37	0,34	0,40	0,38
	Minimum	0,27	0,15	0,21	0,17	0,28	0,22	0,20	0,21	0,16	0,00
	Maximum	0,66	0,56	0,46	0,53	0,56	0,64	0,51	0,48	0,54	0,75
	N	15	15	15	15	15	15	15	15	15	15
Heterogen	Median	0,22	0,25	0,17	0,18	0,17	0,16	0,19	0,15	0,16	0,15
	Minimum	-0,02	0,06	-0,04	0,05	-0,00	-0,06	-0,04	-0,24	-0,02	-0,04
	Maximum	0,36	0,52	0,32	0,38	0,32	0,34	0,44	0,27	0,36	0,28
	N	18	18	18	18	18	18	18	18	18	18
Gesamt	Median	0,33	0,32	0,29	0,26	0,31	0,26	0,29	0,26	0,30	0,26
	Minimum	-0,02	0,06	-0,04	0,05	-0,00	-0,06	-0,04	-0,24	-0,02	-0,04
	Maximum	0,85	10,00	0,85	0,86	0,85	10,00	10,00	0,93	0,92	10,00
	N	38	38	38	38	38	38	38	38	38	38

Tab. 43: Clusteranalyse mit den prüferbezogenen Korrelationen zwischen den Gesamtbewertungen der Kompetenzbereiche

Cluster nehmen die Mediane der Korrelationen die zweithöchsten Werte an; dieses Cluster trägt daher die Bezeichnung „Mittelhomogen“. Das dritte Cluster weist schließlich die niedrigsten Korrelationen zwischen den Gesamtbewertungen der Kompetenzbereiche auf und wird folglich „Heterogen“ benannt.

Das Cluster mit den „homogenen“ Prüfern ist mit fünf Prüfern (13 %) vergleichsweise schwach besetzt. Diese Prüfer bewerten die Prüfungsleistungen ihrer Bewerber hinsichtlich aller fünf Kompetenzbereiche besonders ähnlich. Vor dem Hintergrund der Tatsache, dass alle fünf Kompetenzbereiche eigenständige Anforderungskategorien darstellen (s. o.), lässt sich dabei eine Verzerrung der Kompetenzbewertungen im Sinne eines Halo-Effektes (auch „Hofeffekt“) nicht ausschließen. Darunter wird in der Sozial- bzw. Testpsychologie die Tendenz verstanden, faktisch unabhängige oder nur mäßig korrelierte Eigenschaften bzw. Verhaltensmuster von Personen fälschlicherweise als zusammenhängend wahrzunehmen und zu bewerten. Dieser Effekt tritt auf, wenn bestimmte Verhaltensweisen eines Beobachteten (z. B. mangelnde Verkehrsbeobachtung) besonders hervorstechen und (oder) vom Beobachter als besonders wichtig empfunden werden. In diesem Falle können dann die Beobachtungs- und Bewertungsergebnisse zu anderen Verhaltensweisen (z. B. zur Geschwindigkeitsanpassung) „überstrahlt“ bzw. „vereinheitlicht“ werden.

## Fazit

Zusammenfassend bleibt erstens festzuhalten, dass die Analyse der Prüfungsergebnisse das theoretische und methodische Konzept der optimierten Praktischen Fahrerlaubnisprüfung bestätigt hat: Die Inhalte und die Struktur des Fahraufgabenkatalogs und der Kompetenzbereiche sowie das System der Bewertungskriterien haben sich als tragfähig erwiesen; alle Fahraufgaben und Kompetenzbereiche stellen eigenständige und nicht redundante Prüfungsanforderungen bzw. Facetten der Fahrkompetenz dar.

Zweitens fanden sich bei der Auswertung der Prüfungsleistungen im Hinblick auf die Fahraufgaben und Kompetenzbereiche Ergebnisse, die aus der Verkehrsforschung und Unfallstatistik bekannt sind (z. B. fahranfängertypische Kompetenzdefizite beim Bewältigen der Fahraufgabe „Kreuzungen, Einmündungen“ und im Kompetenzbereich „Ver-

kehrerbeobachtung“) oder zumindest fachlich plausibel erscheinen (z. B. Häufigkeiten bestimmter Bewertungskombinationen). Diese Befunde sprechen für die Validität des Prüfungsverfahrens und eine hohe Durchführungsqualität durch die beteiligten Fahrerlaubnisprüfer unter den Bedingungen des Prüfungsalltags.

Drittens zeigten sich bei den Prüfungsbewertungen und Prüfungsentscheidungen Standortunterschiede, deren Ursachen noch nicht abschließend geklärt werden konnten. Neben regionalen Besonderheiten lassen sich auch Unterschiede zwischen den technischen Prüfstellen bei der Umsetzung des methodischen Prüfungsrahmens nicht ausschließen. Daher erscheint es außerordentlich wichtig, an der prüfstellenübergreifenden gemeinsamen Erarbeitung, Evaluation und Weiterentwicklung der inhaltlichen und methodischen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung strikt festzuhalten und den fachlichen Austausch über die praktische Umsetzung dieser Grundlagen auszubauen.

Viertens schließlich lassen sich bei einer relativ kleinen Zahl der beteiligten Prüfer Besonderheiten beim Bewertungsverhalten (z. B. eine besondere Strenge) oder beim Dokumentationsverhalten (nur wenige negative Bewertungen, die sich aber homogen über verschiedene Kompetenzbereiche erstrecken) nicht ausschließen. Derartige Auffälligkeiten im Prüfverhalten würden, sofern sie sich in repräsentativen Stichproben bestätigen sollten, dem Ziel einer einheitlichen und anspruchsvollen Prüfungsdurchführung ebenfalls entgegenstehen. Derzeit ist aber eher davon auszugehen, dass es sich um Einzelfälle und Unsicherheiten beim Praktizieren der neuen Prüfungsstandards handelt. Derartige Unsicherheiten unterstreichen jedoch ggf. die zwingende Notwendigkeit, die Fahrerlaubnisprüfer frühzeitig und ausreichend auf die neuen Herausforderungen der optimierten Praktischen Fahrerlaubnisprüfung vorzubereiten und die Einhaltung der Standards durch eine kontinuierliche wissenschaftliche Evaluation der Prüfung zu kontrollieren.

Zum Abschluss sei nochmals explizit darauf hingewiesen, dass die vorliegenden Daten keine Repräsentativität für das gesamte Bundesgebiet besitzen und die Auswertung der Prüfungsergebnisse lediglich exemplarisch erfolgte. Es ging darum, die Auswertungsmöglichkeiten hinsichtlich der im Rahmen der optimierten Praktischen Fahrerlaubnisprüfung dokumentierten Prüfungsanforderungen und Prü-

fungsleistungen exemplarisch zu erproben und zu demonstrieren. Alle Befunde zur alltäglichen Durchführungsgüte der Prüfung besitzen damit nur den Charakter von Indizien, die auf der Basis bundesweit repräsentativer Stichproben im Rahmen einer kontinuierlichen Prüfungsévaluation vertieft analysiert werden sollten.

## 5.5 Kundenzufriedenheit

### 5.5.1 Überblick

Die Kundenzufriedenheit stellt ein wichtiges Qualitätsmerkmal von Dienstleistungen dar (s. Kapitel 3.4). Dies gilt auch für die Praktische Fahrerlaubnisprüfung, mit deren Durchführung die technischen Prüfstellen beliehen sind. Zu den „Kunden“ zählen in erster Linie die Fahrerlaubnisbewerber und die Fahrschulen. Im Evaluationssystem zur optimierten Praktischen Fahrerlaubnisprüfung ist vorgesehen, die Zufriedenheit dieser Kunden mit der (alltäglichen) Prüfungsdurchführung in einem Fünfjahresturnus zu erfassen, um empirisch begründete Anregungen für die Optimierung und Weiterentwicklung der Prüfung zu gewinnen (s. Kapitel 2.3).

Zu den Bestandteilen des Revisionsprojekts zur optimierten Praktischen Fahrerlaubnisprüfung zählt die Erprobung einer Bewerberbefragung und einer Fahrlehrerbefragung. Im Mittelpunkt der Erprobungsuntersuchungen standen die Verfahrensweisen der beiden Befragungen und ihre Umsetzung. Die Anlage der beiden Untersuchungen wurde bereits im Kapitel 4 beschrieben; nun sollen die Untersuchungsergebnisse der Bewerberbefragung (s. Kapitel 5.5.2) und der Fahrlehrerbefragung (s. Kapitel 5.5.3) dargestellt werden.

### 5.5.2 Bewerberbefragung

#### Allgemeine Aussagen und Befunde zur Stichprobe

Die Grundgesamtheit der Bewerberbefragung ergab sich aus allen Fahrerlaubnisbewerbern, die im Erprobungszeitraum (September 2014 bis Februar 2015) in den Modellregionen der Erprobungsuntersuchung (s. Kapitel 4.3) eine optimierte Praktische Fahrerlaubnisprüfung absolviert haben, also mittels e-Prüfprotokoll geprüft wurden und eine optimierte Rückmeldung zu ihren Prüfungsleistungen erhalten haben. Es wurde eine Vollerhebung<sup>67</sup> angestrebt.

Grundsätzlich muss sichergestellt sein, dass mit der ausgewählten Befragungsmethode alle potenziell zur Grundgesamtheit (bzw. zu einer geplanten Stichprobe) gehörenden Befragten erreicht werden können. Sofern dies nicht der Fall ist, gilt die Stichprobe nicht als repräsentativ (MAURER & JANDURA, 2009). Die Mehrheit der Bewerber gehört der Altersgruppe der 17- bis 19-Jährigen an. Laut der aktuellen JIM-Studie (2014) ist in den deutschen Haushalten bezüglich der Internetabdeckung mit 94 Prozent fast Vollausstattung erreicht. 92 Prozent der Jugendlichen besitzen sogar einen eigenen Internetzugang in ihren Zimmern. Somit sollte das Problem der Vorselektion der Befragten durch das Merkmal „Internetnutzung“ (TADDICKEN, 2013) bei dieser Zielgruppe nicht gegeben sein. Aus diesem Grund wurde eine Online-Befragung für die Erreichung der Zielgruppe als angemessen erachtet. Zur Rekrutierung der Stichprobe erhielten alle Bewerber von ihren Fahrerlaubnisprüfern eine Einladung zur Online-Befragung. Ausführliche Bemerkungen zur Untersuchungsanlage wurden bereits im Kapitel 4.6 dargestellt.

Bei Online-Befragungen gelten Ausschöpfungsquoten zwischen zwei und zwanzig Prozent als üblich (HOFTE-FRANKHAUSER & WÄLTJ, 2011), da die Quoten von einer Reihe unterschiedlicher Faktoren abhängen, auf die nur ein begrenzter Einfluss besteht (MAURER & JANDURA, 2009). BATINIC und MOSER (2005) wie auch TREIBLMAIER (2010) berichten von zunehmenden Schwierigkeiten bei der Stichprobengewinnung für Online-Studien; als Grund für diese Entwicklung nennen sie die stetig wachsende Zahl an internetbasierten Befragungen. Dieser Effekt fällt umso stärker aus, je geringer das Interesse und der persönliche Nutzen an der Befragung sind, je größer die Bedenken hinsichtlich der Vertraulichkeit der Ergebnisse sind und je länger die Befragung dauert (COLLINS, SYKES, WILSON, & BLACKSHAW, 1988; SINGER & PRESSER, 2008; HOLBROOK, KROSNICK & PFENT, 2008). Grundsätzlich gilt, dass spezielle, homogene Gruppen von Befragten, die vom Befra-

<sup>67</sup> Eine Vollerhebung bietet gegenüber der Ziehung einer Stichprobe den Vorteil, einen Stichprobenfehler bei der Generierung von Aussagen auf Basis statistischer Analysen ausschließen zu können. Davon unbenommen ist eine mögliche Ergebnisverzerrung aufgrund von Selbstselektion durch Teilnahmeverweigerer (EID, GOLLWITZER & SCHMITT, 2010).

gungsthema unmittelbar betroffen sind und das Befragungsthema als wichtig einstufen – wie im vorliegenden Fall die Fahrerlaubnisbewerber, aber auch die Fahrlehrer – eher zur Teilnahme bereit sind. Schließlich ist hinsichtlich der Ausschöpfung insbesondere dann mit Schwierigkeiten zu rechnen, wenn sowohl die Erhebung als auch die Kontaktaufnahme über das Internet erfolgen. Auch dieses Problem war im vorliegenden Fall nicht gegeben, da die Kontaktaufnahme nicht über das Internet, sondern persönlich über die Prüfer erfolgte. Trotz dieser persönlichen Ansprache konnte mit sechs Prozent ( $n = 544$ ) lediglich eine wenig befriedigende Ausschöpfungsquote erreicht werden.

Die Befragung erstreckte sich auch auf ausgewählte persönliche Merkmale der Bewerber (Alter, Geschlecht, Bildungsniveau), um zu überprüfen, ob bei bestimmten Teilgruppen der Befragten Unterschiede in ihrer Zufriedenheit mit bestimmten Prüfungsaspekten auftreten (s. u.). 70 Prozent der befragten Bewerber gehörten der Altersgruppe „24 Jahre und jünger“ an; 30 Prozent waren älter. 47 Prozent der befragten Bewerber waren weiblich und 53 Prozent männlich. Das Bildungsniveau der Befragten wurde in gleicher Weise wie bei der Untersuchung zur populationsspezifischen Äquivalenz erfasst (s. Kapitel 5.3); die Befragten wurden entsprechend in zwei Gruppen zusammengefasst („Abitur angestrebt oder erreicht“ vs. „Abitur nicht angestrebt oder erreicht“).

### Auswertungsstrategie und Darstellung der Befunde

Zunächst werden die erhobenen Daten für jeden Befragungskomplex in Form von Mittelwerten oder prozentualen Häufigkeiten beschrieben. Im Mittelpunkt steht die Auswertung der einzelnen Zufriedenheitsbereiche und der Gesamtzufriedenheit. Weiterhin wird hinsichtlich einzelner Zufriedenheitsaspekte untersucht, ob in Abhängigkeit von bestimmten personenbezogenen Merkmalen, Prüfungsergebnis und der Anzahl bereits abgelegter Praktischer Fahrerlaubnisprüfungen (Wiederholungsprüfung vs. Erstprüfung) Unterschiede bestehen. Dabei wurde mithilfe von T-Tests die Hypothese geprüft, dass die Zufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung von einem oder mehreren der genannten Bewerbermerkmale abhängt.<sup>68</sup> In einigen Fällen wurden zusätzlich Effektstärken berechnet, um die Bedeutsamkeit von Mittelwertunterschieden zu beurteilen.<sup>69</sup>

Um die Bedeutsamkeit einzelner Zufriedenheitsaspekte für die Gesamtzufriedenheit bestimmen zu können, wurden darüber hinaus Regressionsanalysen (s. Kapitel 5.3) durchgeführt. Mithilfe von Regressionsanalysen kann man beispielsweise den speziellen Einfluss eines Zufriedenheitsaspektes auf die Gesamtzufriedenheit unter Berücksichtigung des Einflusses aller anderen Zufriedenheitsaspekte bestimmen und ermitteln, welche Zufriedenheitsaspekte die Gesamtzufriedenheit am stärksten beeinflussen.<sup>70</sup>

An dieser Stelle sei abschließend angemerkt, dass häufig ordinalskalierte Variablen dichotomisiert werden, um die Aufnahme von ordinalskalierten Prädiktoren in ein Regressionsmodell zu ermöglichen (s. Kapitel 5.4). Alternativ dazu können die Ausprägungen der Ordinalskalen mit numerischen Werten hinterlegt und diese als (quasi-)metrische Variablen interpretiert werden. Auch dann ist die Anwendung parametrischer Verfahren wie der Regressionsanalyse möglich. Die Entscheidung, ob die unabhängigen ordinalskalierten Variablen dichotomisiert oder als metrisch behandelt werden, hängt im vorliegenden Fall von inhaltlichen Überlegungen ab. Während der Einfluss der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche auf die Prüfungsentscheidung sich v. a. anhand der Frage entscheidet, ob eine Leistung „Ungenügend“ oder „Nicht ungenügend“ war (s. Kapitel 5.4), lassen sich die Bewertungsstufen der einzelnen Zufriedenheitsaspekte hinsichtlich ihres

<sup>68</sup> Der T-Test erlaubt zu beurteilen, mit welcher Wahrscheinlichkeit Mittelwertunterschiede zwischen verschiedenen Teilgruppen der Bewerberstichprobe auch zwischen Teilgruppen der Grundgesamtheit aller Fahrerlaubnisbewerber bestehen oder ob die gefundenen Unterschiede lediglich einen zufälligen Auswahlfehler der Stichprobe abbilden. Liegt die Wahrscheinlichkeit für einen zufallsbasierten Mittelwertunterschied unter 5 Prozent, spricht man von einem signifikanten Effekt des Bewerbermerkmals (BORTZ, 2005). Sofern nicht anders angegeben, sind alle in den nachfolgenden Kapiteln berichteten Mittelwertunterschiede statistisch signifikant.

<sup>69</sup> Als Kennwert für die Bewertung von Mittelwertunterschieden wurde Cohen's  $d$  berechnet. Nach COHEN (1988) bezeichnen Werte im Bereich von  $d > 0,2$  „kleine“ Effekte, Werte von  $d > 0,5$  „mittlere“ und Werte von  $d > 0,8$  „große“ Effekte (BORTZ & DÖRING, 2002).

<sup>70</sup> Für jeden Aspekt bzw. Prädiktor werden Gewichte in Form von B-Koeffizienten berechnet, welche die Einflussstärke der einzelnen Prädiktoren auf die Gesamtzufriedenheit widerspiegeln. Die B-Koeffizienten der verschiedenen Prädiktoren können bei allen Regressionsmodellen in diesem Kapitel direkt miteinander verglichen werden, da sie alle auf derselben Skala beruhen. Somit erübrigt sich die Berechnung standardisierter Beta-Koeffizienten.



Einflusses auf die Gesamtzufriedenheit weniger eindeutig in zwei Gruppen aufteilen. An dieser Stelle erscheint es daher sinnvoller, die unabhängigen Variablen als metrisch zu interpretieren, auch wenn damit gewisse statistische Risiken einhergehen. Dies scheint vor dem Hintergrund der Erkenntnisse von LABOVITZ (1967) und KIM (1975) (s. Kapitel 5.4.3) gerechtfertigt.<sup>71</sup>

Für die methodenkritische Betrachtung der verwendeten Fragebogen wurden als weitere Formen der multivariaten Datenauswertung Faktorenanalysen durchgeführt. Das Ziel dieser Analysen besteht darin, miteinander stark korrelierende Einzelschätzungen zur Zufriedenheit mit unterschiedlichen Merkmalen der Fahrerlaubnisprüfung auf einer höheren Abstraktionsebene zu sogenannten „Faktoren“ zusammenzufassen (s. Kapitel 5.4).<sup>72</sup>

### Methodenkritische Betrachtung der Fragebogenstruktur

Bei Faktorenanalysen stellt sich stets die Frage, wie viele Faktoren in einem Modell berücksichtigt werden sollen. Im vorliegenden Fall bedeutet dies: Auf wie viele Faktoren lassen sich die verschiedenen Bewertungseinschätzungen zur Zufriedenheit mit unterschiedlichen Merkmalen der Praktischen Fahrerlaubnisprüfung verdichten? Auf Basis der in Kapitel 5.4 diskutierten Kriterien hat sich eine Drei-

Faktor-Lösung als ideal erwiesen. In der Tabelle 44 werden die Strukturen der Faktoren bzw. die Zugehörigkeit der verschiedenen Bewertungseinschätzungen zu den drei gefundenen Faktoren aufgezeigt. Weiterhin findet sich in dieser Tabelle ein Überblick über die von den Faktoren erklärte Varianz in den Befragungsergebnissen.

Es zeigte sich, dass sich die einzelnen Zufriedenheitseinschätzungen aufgrund ihrer statistischen Verteilungen zu drei inhaltlich voneinander abgrenzbaren Faktoren zusammenfassen lassen. Der stärkste Faktor wird durch Einschätzungen gebildet, die sich auf die Merkmale der schriftlichen

<sup>71</sup> Neben dem Skalenniveau der einbezogenen Variablen ist die lineare Regression an einige Voraussetzungen geknüpft. Hierzu zählen nach BACKHAUS et al. (2006): (1) lineare Zusammenhänge zwischen den Prädiktoren und dem Kriterium; (2) Residuen haben den Erwartungswert Null; (3) korrekte Auswahl der Prädiktoren; (4) Homoskedastizität; (5) keine Autokorrelation der Residuen; (6) keine perfekte Multikollinearität; (7) Normalverteilung der Residuen. Die Voraussetzungen wurden für alle im Verlauf des Kapitels dargestellten Regressionsmodelle überprüft; die Ergebnisse können als akzeptabel im Sinne der Anwendbarkeit des Regressionsmodells interpretiert werden.

<sup>72</sup> Die Faktorenanalyse setzt eigentlich metrisch skalierte Variable voraus. Aufgrund der in Kapitel 5.4 genannten Gründe erscheint aber auch im vorliegenden Fall eine Anwendung auf die ordinalskalierten Zufriedenheitsbewertungen als akzeptabel und zielführend.

Zufriedenheitsmerkmale	Faktoren		
	1	2	3
Prüfer-Einführungsgespräch		,61	
Prüfer-Rechtzeitigkeit der Fahrhinweise		,70	
Prüfer-Verständlichkeit der Fahrhinweise		,78	
Prüfer-Bemühungen, Prüfungsstress abzubauen		,74	
Prüfer-Prüfungsbewertung		,55	
Auswertungsgespräch-Erläuterungen (gute Leistungen, Fehler)			,80
Auswertungsgespräch-Hinweise des Prüfers zum Weiterlernen und Verbesserungsvorschläge			,74
Auswertungsgespräch-Beantwortung von Nachfragen			,77
Schriftliche Rückmeldung-Gestaltung	,82		
Schriftliche Rückmeldung-Übersichtlichkeit	,87		
Schriftliche Rückmeldung-Inhalte	,89		
Schriftliche Rückmeldung-Verständlichkeit	,86		
Schriftliche Rückmeldung-Hinweise zur Verbesserung der Fahrkompetenz	,83		
Gesamtzufriedenheit mit Prüfer		,76	
Gesamtzufriedenheit mit Auswertungsgespräch			,76
Gesamtzufriedenheit mit schriftlicher Rückmeldung	,89		
<b>Erklärte Gesamtvarianz</b>	<b>29,6 %</b>	<b>21,5 %</b>	<b>20,1 %</b>

Tab. 44: Rotierte Komponentenmatrix

Rückmeldung bei der optimierten Praktischen Fahrerlaubnisprüfung beziehen (Faktor 1 „Schriftliche Rückmeldung“). Dieser Faktor erklärt 29,6 % der Varianz in den Befragungsergebnissen.

Der zweite Faktor umfasst die Einschätzungen der Bewerber zu verschiedenen Merkmalen des Prüferverhaltens (Faktor 2 „Prüfer“) und erklärt weitere 21,5 % der Ergebnisvarianz. Der dritte Faktor beinhaltet schließlich Einschätzungen zum mündlichen Auswertungsgespräch nach der Prüfungsfahrt (Faktor 3 „Auswertungsgespräch“) und trägt 20,1 % zur Erklärung der Gesamtvarianz bei. Alle Faktoren erscheinen inhaltlich sinnvoll und klar konturiert; die klaren Konturen der gefundenen Faktoren werden nicht durch formale Itemmerkmale (z. B. unterschiedliche Frage- oder Antwortformate) verursacht und sind daher auf inhaltliche Zusammenhänge zurückzuführen. Insofern erscheint das Ergebnis der Faktorenanalyse außerordentlich befriedigend und als Beleg für eine hohe Fragebogengüte. Zur methodischen Zufriedenheit trägt auch bei, dass die drei gefundenen Faktoren zusammen immerhin 71,2 % der Gesamtvarianz erklären.

Empirisch gefundene faktorielle Merkmalsgruppen eignen sich gut für die Strukturierung von Ergebnisdarstellungen. Daher werden die im Folgenden dargestellten Ergebnisse nach der Zugehörigkeit der Einschätzungen zu den oben beschriebenen drei Faktoren gegliedert. Die Auswertung erfolgt zunächst für jeden Zufriedenheitsbereich und schließlich für die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung.

### Befunde zur Zufriedenheit mit der schriftlichen Rückmeldung

Im Rahmen des Revisionsprojekts wurde neben dem e-Prüfprotokoll (s. Kapitel 2.2) auch ein optimiertes Rückmeldesystem (s. Kapitel 2.4) entwickelt und erprobt. Einen Bestandteil des Rückmeldesystems stellt die schriftliche Rückmeldung an

die Fahrerlaubnisbewerber dar, die – entgegen der bisherigen Prüfungspraxis – während der Erprobung auch bei bestandener Prüfung an die Fahrerlaubnisbewerber ausgehändigt wurde. Hinsichtlich der Aushändigung ist festzuhalten, dass diese auf unterschiedlichem Wege verlief: Je nach Technischer Prüfstelle wurde die schriftliche Rückmeldung entweder direkt ausgehändigt oder konnte im Nachgang der Prüfung im Internet über einen Zugangscode abgerufen werden.

Zunächst wurden die Bewerber gefragt, ob sie im Besitz einer schriftlichen Rückmeldung seien. Insgesamt gaben 87 % der Befragten an, eine schriftliche Rückmeldung zu haben. Davon hatten 76 % der Bewerber die Rückmeldung über das Internet bezogen, 22 % erhielten die schriftliche Rückmeldung direkt im Anschluss an die Prüfung als Ausdruck, und zwei Prozent hatten die Rückmeldung bei der technischen Prüfstelle persönlich abgeholt.

Interessant erscheint, warum 13 % der Bewerber keine schriftliche Rückmeldung bekommen haben. Auf die entsprechende Frage antworteten 26 Befragte, dass es technische Gründe gab (entweder wurde die Rückmeldung nicht hochgeladen oder der entsprechende Link funktionierte nicht). Das könnte daran liegen, dass die technischen Prüfstellen aus organisatorischen Gründen die Rückmeldungen in der Regel erst am nächsten Tag hochgeladen haben. Alle Bewerber, die am Prüfungstag ihre Rückmeldung abrufen wollten, blieben daher erfolglos. In welchem Umfang die Bewerber über diese Gegebenheiten informiert waren bzw. die Informationen aufgrund der emotionalen Prüfungsumstände nicht wahrgenommen oder verdrängt haben, ist unklar. Künftig sollten Informationsdefizite ggf. ausgeschlossen werden.

In der Tabelle 45 sind die Untersuchungsergebnisse zur Gesamtzufriedenheit mit der schriftlichen Rückmeldung und zur Zufriedenheit mit einzelnen Aspekten dargestellt. Aus den Befunden wird deutlich,

Wie zufrieden waren Sie bei der schriftlichen Rückmeldung ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden
... mit der Gestaltung?	32	50	13	5
... mit der Übersichtlichkeit?	42	42	13	3
... mit den Inhalten?	39	44	11	6
... mit der Verständlichkeit?	45	40	10	5
... mit den Hinweisen zur Verbesserung ihrer Fahrkompetenz?	35	41	17	7
<b>Wie zufrieden waren Sie mit der schriftlichen Rückmeldung insgesamt?</b>	<b>38</b>	<b>44</b>	<b>12</b>	<b>6</b>

Tab. 45: Zufriedenheit der Bewerber mit der schriftlichen Rückmeldung (in Prozent)

dass die Mehrheit der Befragten mit der schriftlichen Rückmeldung zufrieden war (82 %). Das größte Verbesserungspotenzial besteht im Hinblick auf die Gestaltung (einschließlich Layout) der Rückmeldung.

Um Zusammenhänge zwischen der Gesamtzufriedenheit mit der schriftlichen Rückmeldung und den einzelnen Zufriedenheitsaspekten zu bestimmen, wurde eine Regressionsanalyse durchgeführt (s. o.), die 79 % der Ergebnisvarianz erklärt. Die einzelnen Zufriedenheitsmerkmale haben also einen erheblichen Einfluss auf die Zufriedenheit mit der schriftlichen Rückmeldung und sollten daher auch im Fokus ihrer Weiterentwicklung stehen. An den B-Koeffizienten lässt sich ablesen, dass die Inhalte der Rückmeldung und speziell die Hinweise zur Verbesserung der Fahrkompetenz den stärksten Einfluss auf die Gesamtzufriedenheit mit den schriftlichen Rückmeldungen haben (s. Tabelle 46).

Es bestehen keine signifikanten Zusammenhänge zwischen der Gesamtzufriedenheit mit der schriftlichen Rückmeldung einerseits und den bewerberbezogenen Merkmalen „Wiederholungsprüfung“, „Geschlecht“, „Bildungsniveau“ und „Alter“<sup>73</sup> andererseits. Erwartungsgemäß zeigten sich diejenigen Bewerber, welche die Prüfung bestanden hatten,

deutlich zufriedener als die Bewerber, die nicht bestanden hatten (Cohen's  $d = 0,98$ ; großer Effekt). Dieses Ergebnis könnte darin begründet sein, dass die Bewerber bei nicht bestandener Prüfung in der Rückmeldung mehr nützliche Inhalte und ein aussagekräftigeres Feedback erwarten als bei bestandener Prüfung.

Damit deuten sich Herausforderungen bei der Weiterentwicklung der Rückmeldung an: Erfolgreiche und damit hinsichtlich ihrer Kompetenzerwartung enttäuschte Bewerber haben meist relativ viele Fahrfehler begangen, die in selbstwertdienlicher und lernfördernder Weise zu erläutern und mit verständlichen Lernhinweisen zu versehen sind.

70 % der Befragten gaben an, dass sie die schriftliche Rückmeldung als hilfreich für das weitere Üben und Lernen im Straßenverkehr empfunden haben. Dieses Empfinden fiel stärker aus, wenn die Bewerber die Rückmeldung vom Prüfer erklärt bekommen hatten. 71 % der Prüfer verbanden die Aushändigung der Rückmeldung bzw. des Zugangscodes mit vertiefenden Erläuterungen. Ebenfalls statistisch überprüft wurde, ob der Bereitstellungsweg einen Einfluss auf das Nützlichkeitsempfinden hatte. Unter denjenigen Bewerbern, die sich die Rückmeldung im Internet heruntergeladen hatten, fand sich der größte Anteil von Befragten, welche die Rückmeldung als nützlich empfanden (s. Tabelle 47).

Die bei den offenen Fragen geäußerten Verbesserungsvorschläge lassen sich wie folgt zusammenfassen: Die Rückmeldungen und insbesondere die Hinweise zur Verbesserung der Fahrkompetenz werden – wie

<sup>73</sup> Zur Ermittlung des Zusammenhangs zwischen den einzelnen Zufriedenheitsaspekten und dem Bewerberalter wurden die Probanden anhand des Medians des Alters in zwei Gruppen unterteilt. So wurde die jüngere mit der älteren Hälfte der Bewerber bezüglich ihrer durchschnittlichen Zufriedenheit mittels eines T-Tests verglichen. Der Median des Alters betrug 23 Jahre.

	B-Koeffizienten
Gestaltung	0,19***
Übersichtlichkeit	0,20***
Inhalte	0,25***
Verständlichkeit	0,16***
Hinweise zur Verbesserung der Fahrkompetenz	0,25***

Tab. 46: Einflussfaktoren auf die Gesamtzufriedenheit mit der schriftlichen Rückmeldung (B-Koeffizienten einer linearen Regressionsanalyse)

		Empfinden Sie die schriftliche Rückmeldung als hilfreich für das weitere Lernen im Straßenverkehr? (in Prozent)		
		Ja	Nein	Gesamt
Auf welchem Wege haben Sie Ihre schriftliche Rückmeldung erhalten?	Download	73	27	n = 318
	Ausdruck	62	38	n = 93
	Selbstabholung	50	50	n = 8

Tab. 47: Einfluss des Bereitstellungsweges auf das Nützlichkeitsempfinden

einige Bewerber explizit anmerkten – grundsätzlich als sehr nützlich erachtet, sollten aber sprachlich vereinfacht und individualisiert werden, um an Verständlichkeit zu gewinnen. Vor allem die Fahrfehler sollten eindeutiger und stärker auf die persönliche Prüfungsfahrt bezogen beschrieben werden.

### Befunde zur Zufriedenheit mit dem Prüferverhalten

Um valide Aussagen zur Zufriedenheit mit dem Rückmeldesystem der optimierten Praktischen Fahrerlaubnisprüfung treffen zu können, wurden auch Faktoren erfasst, die möglicherweise die Zufriedenheitseinschätzungen beeinflussen. Wie bereits im Kapitel 4.7 erläutert, spielen bei der Prüfungsdurchführung im Allgemeinen und bei der Leistungsrückmeldung im Besonderen die Fahrerlaubnisprüfer eine entscheidende Rolle. Aus diesem Grund wurde auch die Zufriedenheit mit Prüferverhaltensmerkmalen erfasst (s. Tabelle 48). Würde der Prüfer beispielsweise vom Bewerber als unfair oder unsympathisch empfunden, könnte das unter Umständen die Zufriedenheitseinschätzungen in anderen Bereichen verzerren. Die Befunde zeigen, dass die befragten Fahrerlaubnisbewerber insgesamt „Sehr zufrieden“ (68 %) oder „Eher zufrieden“ (26 %) mit dem Fahrerlaubnisprüfer waren. Das größte Optimierungspotenzial fand sich hinsichtlich der Bemühungen des Prüfers, Prüfungs-

stress abzubauen und eine angenehme Prüfungsatmosphäre zu schaffen.

Auch für diesen Zufriedenheitsbereich wurde anhand einer Regressionsanalyse versucht, mögliche Zusammenhänge zwischen der Gesamtzufriedenheit mit dem Fahrerlaubnisprüfer und einzelnen Zufriedenheitsaspekten aufzudecken. Die einzelnen Zufriedenheitsmerkmale erklären 62 Prozent der Ergebnisvarianz und besitzen somit einen erheblichen Einfluss auf die Gesamtzufriedenheit mit dem Prüfer. Am stärksten beeinflussen das Einführungsgespräch und die Bemühungen des Prüfers, Prüfungsstress beim Bewerber abzubauen, die Gesamtzufriedenheit. Nimmt man die Variable „Prüfungsentscheidung“ in das Regressionsmodell auf, zeigt sich, dass das Prüfungsergebnis bei der Gesamtbewertung des Prüfers nur eine untergeordnete Rolle spielt; Bewerber, welche die Prüfung bestanden hatten, waren aber mit dem Prüfer etwas zufriedener als nicht erfolgreiche Bewerber (s. Tabelle 49).

Keine Unterschiede bei der Gesamtzufriedenheit mit dem Fahrerlaubnisprüfer zeigen sich in Abhängigkeit vom Geschlecht, Alter und Bildungsniveau der Befragten. Darüber hinaus scheint es keinen Einfluss auf die Gesamtzufriedenheit mit dem Prüfer zu haben, ob der Bewerber eine Erstprüfung oder eine Wiederholungsprüfung absolvierte.

Wie zufrieden waren Sie beim Prüfer ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden	Nicht vorgekommen
... mit seinem Einführungsgespräch?	67	27	3	0	3
... mit der Rechtzeitigkeit seiner Fahrhinweisungen?	71	24	4	1	0
... mit der Verständlichkeit seiner Fahrhinweisungen?	78	19	3	0	0
... mit seinen Bemühungen, Prüfungsstress abzubauen?	57	27	7	3	6
... mit seiner Prüfungsbewertung?	66	25	6	3	0
<b>Wie zufrieden waren Sie mit dem Prüfer insgesamt?</b>	<b>68</b>	<b>26</b>	<b>5</b>	<b>1</b>	<b>0</b>

Tab. 48: Zufriedenheit der Bewerber mit dem Prüferverhalten (in Prozent)

Zufriedenheitsmerkmale	B-Koeffizienten
Einführungsgespräch	0,28***
Rechtzeitigkeit der Fahrhinweisungen	0,12**
Verständlichkeit der Fahrhinweisungen	0,13**
Bemühungen, Prüfungsstress abzubauen	0,27***
Prüfungsentscheidung (Bestanden = 1; Nicht bestanden = 0)	0,18***

Tab. 49: Einflussfaktoren auf die Gesamtzufriedenheit mit dem Prüferverhalten (B-Koeffizienten einer linearen Regressionsanalyse)



### Befunde zur Zufriedenheit mit dem mündlichen Auswertungsgespräch

Jede Praktische Fahrerlaubnisprüfung wird mit einer mündlichen Leistungsrückmeldung durch den Fahrerlaubnisprüfer abgeschlossen. In diesem Auswertungsgespräch gibt der Prüfer das Ergebnis bekannt und erläutert anhand einiger beispielhafter Verkehrssituationen die guten Leistungen sowie ggf. aufgetretene Fehler und Verbesserungsmöglichkeiten. Dabei hat der Bewerber auch die Möglichkeit für Nachfragen. Die Befunde zur Gesamtzufriedenheit mit dem Auswertungsgespräch und einzelnen Zufriedenheitsaspekten finden sich in der Tabelle 50. Insgesamt waren 87 Prozent der Befragten mit dem Auswertungsgespräch „Sehr zufrieden“ oder „Eher zufrieden“. Die größte Zufriedenheit zeigte sich hinsichtlich der Erläuterungen des Prüfers, das größte Optimierungspotenzial bei der Beantwortung von Nachfragen. Bei der Interpretation dieser Ergebnisse ist zu berücksichtigen, dass bei diesen Fragen auch die Möglichkeit bestand, „Nicht vorgekommen“ anzukreuzen: Sofern also keine Nachfragen notwendig waren, konnten auch keine Bewertungen vorgenommen werden.

Bei der regressionsanalytischen Untersuchung der Zusammenhänge zwischen der Gesamtzufriedenheit mit dem Auswertungsgespräch und den einzelnen Zufriedenheitsaspekten erklären die verschiedenen Prädiktoren 70 Prozent der Varianz der Gesamtzufriedenheit. Am weitaus stärksten beeinflussen die Erläuterungen des Prüfers im Anschluss an

die Prüfungsfahrt die Gesamtzufriedenheit mit dem Auswertungsgespräch (s. Tabelle 51).

Bewerber, welche die Prüfung nicht bestanden hatten, waren insgesamt mit dem mündlichen Auswertungsgespräch deutlich unzufriedener als erfolgreiche Bewerber (Cohen's  $d = 1,21$ ; großer Effekt). Darüber hinaus zeigten sich Zufriedenheitsunterschiede in Abhängigkeit vom Bildungsstand und Alter der Befragten. Bewerber mit höherem Bildungsniveau bzw. jüngere Bewerber (unter 23 Jahre) waren insgesamt etwas unzufriedener mit dem mündlichen Auswertungsgespräch als Bewerber mit einem niedrigeren Bildungsniveau bzw. ältere Bewerber (jeweils Cohen's  $d = 0,21$ ; kleiner Effekt). Vermutlich sind mit einem höheren Bildungsniveau auch höhere Erwartungen an den Prüfer und daher kritischere Bewertungen verbunden. Die höhere Zufriedenheit älterer Bewerber könnte daran liegen, dass ältere Bewerber (in den Augen der Fahrerlaubnisprüfer) meist erkennbar über mehr Lebensreife und Selbstreflexionskompetenz verfügen, was die Diskussion ihrer Prüfungsleistungen begünstigt. Keine Mittelwertunterschiede fanden sich hingegen in Abhängigkeit vom Geschlecht der Bewerber und dem Umstand, ob es sich um eine Erstprüfung oder Wiederholungsprüfung handelte.

### Befunde zur Gesamtzufriedenheit mit der optimierten Praktischen Fahrerlaubnisprüfung

60 Prozent der Befragten waren insgesamt mit der (letzten) Praktischen Fahrerlaubnisprüfung „Sehr

Wie zufrieden waren Sie beim Auswertungsgespräch ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden	Nicht vorgekommen
... mit den Erläuterungen (gute Leistungen, Fehler) des Prüfers?	50	32	9	3	6
... mit den Hinweisen des Prüfers zum Weiterlernen und seinen Verbesserungsvorschlägen?	42	29	9	2	18
... mit der Beantwortung von Nachfragen?	40	16	3	1	40
<b>Wie zufrieden waren Sie mit dem Auswertungsgespräch insgesamt?</b>	<b>55</b>	<b>32</b>	<b>11</b>	<b>2</b>	<b>0</b>

Tab. 50: Zufriedenheit der Bewerber mit dem mündlichen Auswertungsgespräch (in Prozent)

Zufriedenheitsmerkmale	B-Koeffizienten
Erläuterungen (gute Leistungen, Fehler)	0,41***
Hinweise zum Weiterlernen	0,21***
Beantwortung von Nachfragen	0,27***

Tab. 51: Einflussfaktoren auf die Gesamtzufriedenheit mit dem mündlichen Auswertungsgespräch (B-Koeffizienten einer linearen Regressionsanalyse)

Zufriedenheitsmerkmale	B-Koeffizienten
Gesamtzufriedenheit Prüfer	0,54***
Gesamtzufriedenheit Auswertungsgespräch	0,16**
Gesamtzufriedenheit schriftliche Rückmeldung	0,17***

Tab. 52: Einfluss einzelner Zufriedenheitsbereiche auf die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung (B-Koeffizienten einer linearen Regressionsanalyse)

zufrieden“ und 32 Prozent „Eher zufrieden“. Lediglich acht Prozent zeigten sich „Eher unzufrieden“ (6 %) oder „Sehr unzufrieden“ (2 %). Am stärksten beeinflusst die Zufriedenheit mit dem Prüfer die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung (s. Ergebnisse der Regressionsanalyse in Tabelle 52). Die einzelnen Prädiktoren („Gesamtzufriedenheit Prüfer“, „Gesamtzufriedenheit Auswertungsgespräch“, „Gesamtzufriedenheit schriftliche Rückmeldung“) erklären 46 Prozent der Ergebnisvarianz.

Nimmt man die Variable „Prüfungsentscheidung“ ins Regressionsmodell auf, so zeigt sich, dass die Zufriedenheit mit dem Prüfer zwar den stärksten Einfluss auf die Gesamtzufriedenheit mit der Prüfung behält, aber – entgegen aller anderen Zufriedenheitsbereiche – das Prüfungsergebnis den zweitstärksten Einfluss auf die Gesamtzufriedenheit mit der Prüfung ausübt.<sup>74</sup> Dies spiegelt sich auch in der erklärten Varianz wider: Bei Hinzunahme der Variablen „Prüfungsentscheidung“ steigt die erklärte Varianz von 46 Prozent auf immerhin 53 Prozent. Aus Sicht der Bewerber kommt dem Prüfer also die entscheidende Rolle bei der Gestaltung der Prüfung zu, und insbesondere die Zufriedenheit mit dem Prüfer entscheidet über die Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung. Zufriedenheitsunterschiede in Abhängigkeit vom Alter, Geschlecht, Bildungsniveau oder dem Umstand, ob es sich um eine Erstprüfung oder Wiederholungsprüfung handelte, fanden sich nicht. Erwartungsgemäß zeigten sich Bewerber, die bestanden hatten, insgesamt deutlich zufriedener als erfolglose Bewerber (Cohen's  $d = 2,03$ ; großer Effekt).

Die Auswertung der offenen Frage nach Vorschlägen zur Verbesserung der Praktischen Fahrerlaub-

nisprüfung deutete auf Optimierungspotenzial im Hinblick auf den Abbau von Prüfungsangst und Prüfungsstress durch den Prüfer hin. Einige Bewerber wünschten sich mehr Einfühlungsvermögen, auch positives Feedback zu ihrer Prüfungsfahrt sowie Hinweise für das weitere selbstständige Fahren. Vereinzelt wurden zu späte Fahrhinweise des Prüfers und störende Gespräche zwischen Prüfer und Fahrlehrer kritisiert.

#### Zwischenfazit zur Praktikabilität der Bewerberbefragung

Insgesamt bleibt festzuhalten, dass sich die gewählte Verfahrensweise zur Befragung der Fahrerlaubnisbewerber grundsätzlich als funktions-tüchtig erwiesen hat und im Rahmen der kontinuierlichen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung in der Zukunft angewendet werden kann. Allerdings empfiehlt sich bei künftigen Befragungen eine bessere Ausgestaltung und Kontrolle der Einladungssituation, um die Ausschöpfungsquote zu erhöhen und damit u. a. die Gefahr stichprobenbedingter Verzerrungen der Ergebnisse zu minimieren. Es erscheint erfolgversprechender, die Einladungen künftig per E-Mail zu verschicken, damit sowohl die Einladung zur Befragung als auch ggf. notwendige Nachfassaktionen direkt über das mit der Befragung beauftragte Institut abgewickelt werden können. MAURER und JANDURA (2009) empfehlen zur Steigerung der Ausschöpfungsquote prinzipiell einen Vorabkontakt mit den Befragten. Weiterhin sollte dieser Kontakt wie auch ggf. notwendige Nachfassaktionen grundsätzlich mit einer persönlichen Anrede der Bewerber erfolgen. Zudem sollte künftig die Teilnahmebereitschaft durch Anreize wie kleine Geschenke oder die Teilnahme an Verlosungen sowie durch die Zurverfügungstellung der Befragungsergebnisse erhöht werden (TREIBLEMAIER, 2010; THEOBALD, 2000).

<sup>74</sup> Da es sich bei der Variable „Prüfungsentscheidung“ um ein zweistufiges Antwortformat handelt (den Zufriedenheitsmerkmalen liegt ein vierstufiges Antwortformat zugrunde), wurden in diesem Fall die Beta-Koeffizienten berechnet.

### 5.5.3 Fahrlehrerbefragung

#### Allgemeine Befunde zur Stichprobe

Für die Rekrutierung der Fahrschulstichprobe wurden alle Fahrschulen aus den Modellregionen (s. Kapitel 4.3) von ihren zuständigen technischen Prüfstellen zur Teilnahme an der Befragung eingeladen. Die daraus resultierenden anteiligen Grundgesamtheiten betragen bei DEKRA 85 Fahrschulen, beim TÜV NORD 238 Fahrschulen, beim TÜV Rheinland 226 Fahrschulen und beim TÜV SÜD – aufgrund der dort gewählten dezentralen Standortwahl und einer Standortbeschränkung auf Regensburg in Bayern – 110 Fahrschulen.

Beim TÜV NORD fand die Online-Datenerfassung im April 2015 statt. Bei den anderen technischen Prüfstellen wurden die ursprünglichen Untersuchungsplanungen nicht umgesetzt, um eine Ergänzung der Befragungsinhalte zu ermöglichen. Nach entsprechenden konzeptionellen Arbeiten und einer erfolgreichen Abstimmung des weiteren Vorgehens mit der BAST begann die Datenerhebung in den verbliebenen drei Prüfstellen am 15.06.2015. Sie dauerte – einschließlich der durch den Poststreik notwendigen Nachfassaktionen – bis zum 15.07.2015. Insgesamt erbrachte der Rücklauf 234 Datensätze; die Rücklaufquote beträgt demzufolge 35 Prozent.

Bei der Fahrlehrerbefragung wurden auch das Alter, das Geschlecht und der Berufsstatus der Fahrlehrer sowie die Größe der Fahrschule erfragt, um zu prüfen, ob Zufriedenheitsunterschiede zwischen bestimmten Teilgruppen der befragten Fahrlehrer bzw. Fahrschulen bestehen. 20 Prozent der Fahrlehrer waren jünger als 46 Jahre; 80 Prozent waren 46 Jahre alt oder älter. Lediglich 7 Prozent der Befragten waren weiblich. Im Hinblick auf den Berufsstatus gaben 88 Prozent der Befragten an, Fahrschulinhaber zu sein. Lediglich sechs Prozent angestellte Leiter und nochmals sechs Prozent Fahrschulangestellte ohne Leitungsfunktion nahmen an der Befragung teil. Die prozentuale Verteilung der teilnehmenden Fahrschulen hinsichtlich ihrer Größe lässt sich dem Bild 7 entnehmen.

#### Auswertungsstrategie und Darstellung der Befunde

Die Auswertung der erhobenen Daten und die Darstellung der Befunde der Fahrlehrerbefragung erfolgten – soweit sinnvoll und möglich – analog zu den verwendeten Auswertungs- und Darstellungs-

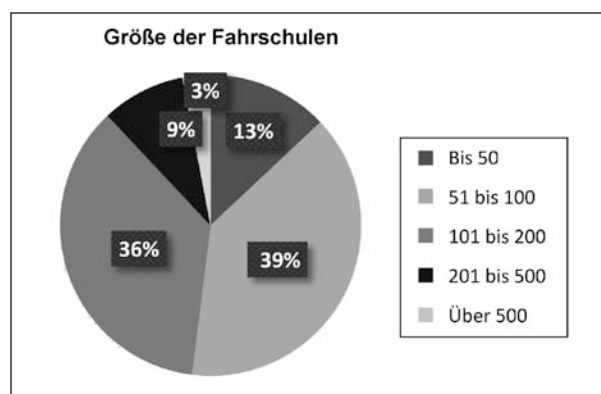


Bild 7: Anzahl der jährlich ausgebildeten Fahrschüler pro Fahrschule

strategien bei der Bewerberbefragung (s. Kapitel 5.5.1). Um die Mittelwerte mehrerer Fallgruppen – wie es bei den Variablen „Berufsstatus“ und „Fahrschulgröße“ der Fall ist – miteinander vergleichen zu können, wurde zusätzlich eine Varianzanalyse durchgeführt.

#### Methodenkritische Betrachtung der Fragebogenstruktur

Im Ergebnis der Faktorenanalyse (s. Tabelle 53) zeigen sich zwei Faktoren; sie betreffen den „Prüfer“ (Faktor 1) und die organisatorischen „Rahmenbedingungen“ der Prüfung (Faktor 2). Im Gegensatz zum ursprünglichen Fragebenaufbau wird die Variable „Zufriedenheit mit dem Treffpunkt der Prüfungsfahrt“ statistisch nicht zum Prüferverhalten, sondern zu den organisatorischen Rahmenbedingungen zugeordnet. Diese Zuordnung erscheint auch inhaltlich plausibel, da die Wahl des Treffpunkts nicht als ein individueller Verhaltensaspekt des Prüfers zu betrachten ist, sondern als eine institutionelle organisatorische Vorgabe. Aus diesem Grund sollte diese Frage künftig auch im Fragebogen zu den organisatorischen Rahmenbedingungen gezählt werden.

Der stärkste Faktor wird durch Zufriedenheitseinschätzungen der Fahrlehrer gebildet, die sich auf das Prüferverhalten bei der zuletzt erlebten Praktischen Fahrerlaubnisprüfung beziehen (Faktor „Prüfer“). Dieser erste Faktor erklärt 41,7 Prozent der Varianz in den Befragungsergebnissen. Der zweite Faktor umfasst die Einschätzungen zu verschiedenen Aspekten der Zufriedenheit mit den organisatorischen Rahmenbedingungen (Faktor „Rahmenbedingungen“) in den letzten 12 Monaten und erklärt weitere 21,5 Prozent der Ergebnisvarianz. Die klaren Konturen der gefundenen Faktoren werden

Zufriedenheitsmerkmale	Faktoren	
	1	2
Organisatorisches-Meldefristen		.85
Organisatorisches-Bearbeitungszeiten und Rücklaufinformationen		.71
Organisatorisches-Terminvergabe		.84
Organisatorisches-Stornierungsmöglichkeiten		.56
Gesamtzufriedenheit-Organisatorische Rahmenbedingungen		.76
Erreichbarkeit Treffpunkt Prüfungsfahrt		.50
Prüfer-Bemühungen, Prüfungsstress abzubauen	.82	
Prüfer-Einführungsgespräch	.82	
Prüfer-Eindeutigkeit Fahrweisungen	.77	
Prüfer-Rechzeitigkeit Fahrweisungen	.74	
Prüfer-Streckenwahl	.78	
Prüfer-Prüfungsbewertung	.79	
Auswertungsgespräch-(Fehler-)Erläuterungen	.81	
Auswertungsgespräch-Erwähnung guter Leistungen	.74	
Auswertungsgespräch-Hinweise und Verbesserungsvorschläge	.77	
Auswertungsgespräch-Beantwortung von Nachfragen	.50	
Gesamtzufriedenheit mit dem Prüfer bei der letzten Prüfung	.85	
Gesamtzufriedenheit mit der letzten Prüfung	.80	
<b>Erklärte Gesamtvarianz</b>	<b>41,7 %</b>	<b>21,5 %</b>

Tab. 53: Rotierte Komponentenmatrix

nicht durch formale Itemmerkmale (z. B. unterschiedliche Frage- oder Antwortformate) verursacht und sind daher auf inhaltliche Zusammenhänge zurückzuführen. Insofern erscheint das Ergebnis der Faktorenanalyse außerordentlich befriedigend und als Beleg für eine hohe Fragebogengüte. Zur methodischen Zufriedenheit trägt auch bei, dass die zwei gefundenen Faktoren zusammen 63,2 Prozent der Gesamtvarianz erklären.

### Befunde zur Zufriedenheit mit dem Prüferverhalten

Die Zufriedenheitseinschätzungen der Fahrlehrer zum Prüferverhalten des Fahrerlaubnisprüfers bei der Durchführung der letzten Praktischen Prüfung finden sich in der Tabelle 54. Mit dem Prüferverhalten insgesamt waren 84 Prozent der befragten Fahrlehrer „Sehr zufrieden“ oder „Eher zufrieden“. Sowohl bei der Gesamteinschätzung des Prüferverhaltens als auch bei der Bewertung der einzelnen Zufriedenheitsaspekte zeigten sich die Fahrlehrer deutlich kritischer als die Fahrerlaubnisbewerber. In ähnlicher Weise zeigte sich hingegen in beiden Befragungen das größte Optimierungspotenzial hin-

sichtlich des Prüferverhaltens bei den Bemühungen, das Prüfungsklima zu entspannen und Prüfungsstress abzubauen. Dazu gehört – inhaltlich gesehen – aus Fahrlehrersicht auch die allmähliche Steigerung der Anforderungen im Prüfungsverlauf „vom Leichten zum Schweren“. Die größte Zufriedenheit äußerten die Fahrlehrer – wie auch die Bewerber – in Bezug auf die Eindeutigkeit und Rechzeitigkeit der Fahrweisungen.

Bei der Einschätzung der Zufriedenheit mit dem Auswertungsgespräch (s. Tabelle 55) zeigte sich, dass die größte Zufriedenheit bei der Beantwortung von Nachfragen (sofern gegeben) und bei den (Fehler-)Erläuterungen des Prüfers besteht. Weniger zufrieden waren die Fahrlehrer hingegen mit der Erwähnung guter Leistungen und den Verbesserungshinweisen an den Bewerber nach der Prüfungsfahrt. Diese Bewertungen der Fahrlehrer ähneln stark den entsprechenden Einschätzungen der Bewerber (s. o.). 10 Prozent der Befragten gaben an, dass gute Leistungen gar nicht erwähnt wurden. Weiterhin hat sich bei 28 Prozent der Bewerber keine Beantwortung von Nachfragen ergeben, da keine gestellt wurden.



Wie zufrieden waren Sie bei der letzten Praktischen Fahrerlaubnisprüfung mit dem Prüferverhalten hinsichtlich ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden
... den Bemühungen, das Prüfungsklima zu entspannen und Prüfungsstress abzubauen (z. B. Einfühlsamkeit, Fairness)?	42	36	17	5
... den Einweisungen/dem Einführungsgespräch in den Prüfungsablauf?	45	42	11	2
... der Eindeutigkeit der Fahrhinweisungen?	50	42	8	0
... der Rechtzeitigkeit der Fahrhinweisungen?	48	46	6	0
... der Wahl der Strecke (z. B. vom Leichten zum Schweren, i. g. O/ a. g. O)?	31	45	18	6
... der Prüfungsbewertung insgesamt?	44	46	8	2
<b>Wie zufrieden waren Sie mit dem Prüfer insgesamt?</b>	<b>38</b>	<b>46</b>	<b>13</b>	<b>3</b>

Tab. 54: Zufriedenheit der Fahrlehrer mit dem Prüferverhalten (in Prozent)

Wie zufrieden waren Sie beim Auswertungsgespräch mit ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden
... den (Fehler-)Erläuterungen?	36	46	15	3
... der Erwähnung von guten Leistungen?	24	37	30	9
... den Hinweisen zur Verbesserung der Fahrkompetenz an den Bewerber?	22	44	26	8
... der Beantwortung von Nachfragen?	41	45	10	4

Tab. 55: Zufriedenheit der Fahrlehrer mit dem Auswertungsgespräch (in Prozent)

Zufriedenheitsmerkmale	B-Koeffizienten
Bemühungen, Prüfungsstress abzubauen	0,22***
Eindeutigkeit der Fahrhinweisungen	0,24***
Streckenwahl	0,19***
Auswertungsgespräch (Fehler-)Erläuterungen	0,13**
Einführungsgespräch	0,15**
Auswertungsgespräch Erwähnung guter Leistungen	0,07**

Tab. 56: Einflussfaktoren auf die Gesamtzufriedenheit mit dem Prüferverhalten (B-Koeffizienten einer linearen Regressionsanalyse)

Weder bei der Zufriedenheit mit dem Prüferverhalten noch bei der Zufriedenheit mit dem Auswertungsgespräch zeigten sich Zusammenhänge mit dem Alter, Geschlecht oder Berufsstatus der befragten Fahrlehrer oder der Größe ihrer Fahrschule.

Wie bei der Auswertung der Bewerberdaten, wurde auch bei der Fahrlehrerbefragung regressionsanalytisch überprüft, welche Zufriedenheitsaspekte den größten Einfluss auf die Gesamtzufriedenheit mit dem Prüfer hatten. Das entsprechende Regressionsmodell (s. Tabelle 56) erklärt 76 Prozent der Ergebnisvarianz: Die einzelnen Zufriedenheitsaspekte besitzen also einen erheblichen Einfluss auf die Gesamtzufriedenheit mit dem Fahrerlaubnisprüfer. Der Vergleich der B-Koeffizienten verdeutlicht, dass – neben der Qualität (Eindeutigkeit,

Verständlichkeit, Rechtzeitigkeit) der Fahrhinweisungen des Prüfers – vor allem seine Bemühungen, Prüfungsstress beim Bewerber abzubauen, und die damit zusammenhängende Streckenwahl („vom Leichten zum Schweren“, s. o.) den stärksten Einfluss auf die Gesamtzufriedenheit mit dem Prüfer ausüben. Zur hohen Zufriedenheitsbedeutung der Streckenwahl mag beitragen, dass dieser Aspekt – im Gegensatz zur Kontrolle von Stress und Angst – überhaupt nicht vom Bewerber oder Fahrlehrer beeinflussbar ist.

### Befunde zur Zufriedenheit mit den organisatorischen Rahmenbedingungen

Die Frage nach der Zufriedenheit mit der Erreichbarkeit des Treffpunkts bei der letzten Praktischen

Fahrerlaubnisprüfung wurde bei der Faktorenanalyse eindeutig zum Faktor „Organisatorische Rahmenbedingungen“ zugeordnet. Es zeigte sich, dass die überwältigende Mehrheit der befragten Fahrlehrer mit der Erreichbarkeit des Treffpunkts ihrer letzten Prüfungsfahrt zufrieden war: 94 Prozent gaben „Sehr zufrieden“ oder „Eher zufrieden“ an.

Die Fragen nach der Gesamtzufriedenheit mit den organisatorischen Rahmenbedingungen und nach einzelnen Zufriedenheitsaspekten bei der Zusammenarbeit mit den technischen Prüfstellen (Meldefristen, Terminvergabe, Bearbeitungszeiten und Stornierungsmöglichkeiten) bezogen sich auf den Zeitraum der letzten 12 Monate. Der Tabelle 57 ist die prozentuale Verteilung der Antworten zu entnehmen. Es zeigt sich, dass die Fahrlehrer auch in diesem Zufriedenheitsbereich mehrheitlich „Sehr zufrieden“ oder „Eher zufrieden“ waren, wenn auch die Zufriedenheit hier nicht so positiv ausfiel wie bei den Einschätzungen zum Prüfer. Am wenigsten zufrieden waren die Befragten mit den Möglichkeiten zur Stornierung von Prüfungsterminen. Insgesamt gab fast die Hälfte der Fahrlehrer an, mit den Stornierungsmöglichkeiten „Eher unzufrieden“ oder „Sehr unzufrieden“ zu sein (48 %).

Bei der regressionsanalytischen Auswertung der Bedeutung der einzelnen Zufriedenheitsaspekte

für die Gesamtzufriedenheit der Fahrlehrer mit den organisatorischen Rahmenbedingungen zeigte sich, dass die einzelnen Zufriedenheitsmerkmale lediglich 54 Prozent der Ergebnisvarianz erklären. Es gibt also noch andere, in der vorliegenden Analyse nicht berücksichtigte wichtige Einflussfaktoren auf die Gesamtzufriedenheit der Fahrlehrer mit den organisatorischen Rahmenbedingungen. Nimmt man die Variable „Zufriedenheit mit der Erreichbarkeit des Treffpunkts zur Praktischen Fahrerlaubnisprüfung“ in das Regressionsmodell auf, steigt die erklärte Varianz nur unwesentlich auf 55 Prozent; diese Variable spielt also offensichtlich in dieser Hinsicht keine wichtige Rolle. Der Tabelle 58 ist zu entnehmen, dass die Zufriedenheit mit den Meldefristen am stärksten die Gesamtzufriedenheit mit den organisatorischen Rahmenbedingungen beeinflusst. Die Stornierungsmöglichkeiten haben hingegen den geringsten Einfluss auf die Gesamtzufriedenheit. Dies erklärt auch, warum trotz relativ großer Unzufriedenheit mit den Stornierungsmöglichkeiten die Gesamtzufriedenheit mit den organisatorischen Rahmenbedingungen trotzdem so positiv ausfällt (s. o.).

Auch in diesem Zufriedenheitsbereich existieren keine Zusammenhänge der Zufriedenheitseinschätzungen mit dem Alter, Geschlecht oder Berufsstatus der befragten Fahrlehrer oder der Größe ihrer Fahrschule.

Wie zufrieden waren Sie mit der Prüfauftragsverwaltung bzw. dem Terminmanagement hinsichtlich ...	Sehr zufrieden	Eher zufrieden	Eher unzufrieden	Sehr unzufrieden
... Meldefristen (z. B. die Möglichkeit, Termine frühzeitig beantragen zu können)?	30	46	20	4
... Bearbeitungszeiten und Rücklaufinformationen (z. B. Terminbestätigung)?	40	43	16	1
... Terminvergabe?	26	46	24	4
... Stornierungsmöglichkeiten (z. B. Kurzfristigkeit der Terminstornierung)?	18	34	33	15
<b>Wie zufrieden waren Sie mit den organisatorischen Rahmenbedingungen insgesamt?</b>	<b>26</b>	<b>56</b>	<b>18</b>	<b>0</b>

Tab. 57: Zufriedenheit der Fahrlehrer mit den organisatorischen Rahmenbedingungen (in Prozent)

Zufriedenheitsmerkmale	B-Koeffizienten
Meldefristen	0,30***
Bearbeitungszeiten und Rücklaufinformationen	0,20***
Stornierungsmöglichkeiten	0,17***
Erreichbarkeit des Treffpunkts zur Prüfungsfahrt	0,20***

Tab. 58: Einflussfaktoren auf die Gesamtzufriedenheit mit den organisatorischen Rahmenbedingungen (B-Koeffizienten einer linearen Regressionsanalyse)

### **Befunde zur Gesamtzufriedenheit mit der Praktischen Fahrerlaubnisprüfung**

Auf die Frage zur Gesamtzufriedenheit mit der (letzten) Praktischen Fahrerlaubnisprüfung gaben 60 Prozent der befragten Fahrlehrer an, „Sehr zufrieden“ gewesen zu sein; 32 Prozent waren „Eher zufrieden“. Lediglich acht Prozent der Befragten waren „Eher unzufrieden“ (6 %) oder „Sehr unzufrieden“ (2 %). Im Ergebnis einer Regressionsanalyse mit zwei Prädiktoren zeigte sich, dass allein die Gesamtzufriedenheit mit dem Prüferverhalten – nicht aber die Gesamtzufriedenheit mit den organisatorischen Rahmenbedingungen – einen nennenswerten Erklärungsbeitrag zur Gesamtzufriedenheit mit der (letzten) Praktischen Fahrerlaubnisprüfung leistet. Dieser Erklärungsbeitrag fällt mit 74 Prozent sehr hoch aus.

Am Ende der Befragung hatten die Fahrlehrer die Möglichkeit, Verbesserungsvorschläge zur Praktischen Fahrerlaubnisprüfung zu äußern. Die Auswertung dieser abschließenden Vorschläge führte zu ähnlichen Ergebnissen wie die Analyse der einzelnen Zufriedenheitsaspekte (s. o.). Die Mehrheit der Wünsche betraf das Prüferverhalten gegenüber den Bewerbern: Dabei ging es insbesondere um die Schaffung einer angenehmen Prüfungsatmosphäre sowie um ein stärkeres Einfühlungsvermögen und den Abbau von Prüfungsangst. Weiterhin wurde mehrfach erwähnt, dass die Bewerber ausführlichere Rückmeldungen zu ihren Prüfungsleistungen erhalten sollten. Dabei sollten unbedingt auch positive Leistungen erwähnt und bei der Prüfungsentscheidung berücksichtigt werden. Weiterhin wurde mehr Einheitlichkeit bei der Durchführung und Bewertung der Prüfung gefordert. Den letztgenannten

Wünschen wird die optimierte Praktische Fahrerlaubnisprüfung mit ihren verbesserten Anforderungs- und Dokumentationsstandards sowie Bewertungskriterien weit entgegenkommen. Ziel sollte – wie einige Einträge betonen – eine höhere Prüfungstransparenz und der Ausschluss von Benachteiligungen einzelner Bewerber sein. Darüber hinaus wurde mehrfach die Wahl der Prüfungstrecke kritisiert; dies betraf sowohl die Aufteilung von innerörtlichen und außerörtlichen Strecken als auch den Aufbau der Streckenanforderungen vom Leichten zum Schweren. Schließlich wurden Wünsche nach einer schnelleren Bearbeitung von Terminanfragen und nach Möglichkeiten für kurzfristige Terminänderungen geäußert, um die Flexibilität der Fahrschulen zu erhöhen.

### **Befunde zur optimierten Praktischen Fahrerlaubnisprüfung**

Wie bereits im Kapitel 4.7 erwähnt, wurden bei der zeitversetzt stattfindenden Befragung der Fahrlehrer aus den Regionen der drei technischen Prüfstellen DEKRA, TÜV Rheinland und TÜV SÜD drei zusätzliche Fragen zur optimierten Praktischen Fahrerlaubnisprüfung gestellt. Diese Fragen werden nachfolgend ausgewertet. Die entsprechenden Befunde sind für die Fahrlehrerschaft der Modellregionen des Revisionsprojektes nicht repräsentativ, sollen aber wegen ihres Anregungsgehalts für die Weiterentwicklung des Rückmeldesystems für die optimierte Praktische Fahrerlaubnisprüfung skizziert werden.

Durch die elektronische Dokumentation der Prüfungsleistungen mittels e-Prüfprotokoll wurden die Voraussetzungen für eine differenzierte Leistungsrückmeldung an die Bewerber ausgebaut und optimiert. Diese Rückmeldung soll Aufschluss über die jeweiligen Stärken und Schwächen – differenziert nach Fahraufgaben und Fahrkompetenzbereiche – geben und künftig allen Bewerbern unabhängig von ihrem Prüfungsergebnis zur Verfügung gestellt werden. Da die Rückmeldung auch hilfreiche Informationen für die Fahrlehrer und ihre Ausbildungsplanung bietet, wurden die Fahrlehrer gefragt, ob sie sich bei Einführung der optimierten Prüfung auch die Bereitstellung der schriftlichen Rückmeldung für die jeweilige Fahrschule des Bewerbers wünschen. Sofern diese Frage mit „Ja“ beantwortet wurde, sollten die befragten Fahrlehrer weiterhin angeben, auf welchem Bereitstellungsweg sie diese Rückmeldung gern erhalten würden (s. Tabelle 59).

Die Mehrheit der Befragten gab an, dass auch die Fahrlehrer künftig die schriftliche Rückmeldung erhalten sollten. Bisher liegt es allein im Ermessen des Bewerbers, ob er das bei Nichtbestehen ausgehändigte Protokoll auch seinem Fahrlehrer zeigt. Allerdings erscheint es wahrscheinlich, dass bei anstehender Wiederholungsprüfung die begangenen Fehler vom Fahrlehrer und seinem Fahrschüler gemeinsam erörtert werden. Da nach geltendem Prüfungsrecht bislang lediglich ein Protokoll mit exemplarisch aufgelisteten Fehlern erstellt und ausgehändigt wird, würde eine regelmäßige Bereitstellung dieses Protokolls für die Fahrschulen vermutlich bei den Fahrlehrern kaum zu Erkenntnisgewinnen führen. Da künftig aber alle Bewerber eine ausführliche schriftliche Rückmel-

<b>Sollte – neben dem Bewerber und mit seiner Zustimmung – auch der Fahrlehrer eine schriftliche Rückmeldung zu den Prüfungsleistungen des Bewerbers erhalten?</b>	
Ja	63 %
Nein	37 %
<b>Falls ja, auf welchem Wege wünschen Sie die Bereitstellung der schriftlichen Rückmeldung?</b>	
Per E-Mail	76 %
Per Download auf einem Online-Portal (Homepage)	24 %

Tab. 59: Fragen zur optimierten Rückmeldung und ihre Bereitstellung

dung erhalten sollen, könnten die Rückmeldungen auch im Fall bestandener Prüfungen hilfreiche und nützliche Optimierungshinweise für die Ausbildung im Allgemeinen und für den ausbildenden Fahrlehrer im Besonderen beinhalten.

Bei der Auswertung der offenen Frage zu den Wünschen der Fahrlehrer, sofern die optimierte Prüfung eingeführt wird, wurden vor allem Ängste der Fahrlehrer deutlich. Dies betraf im Fall einer möglichen Einführung des e-Prüfprotokolls vor allem den Entscheidungsaspekt: Es wurde mehrfach darauf hingewiesen, dass nicht ein Computer, sondern der Prüfer über das Ergebnis entscheiden soll. Diesem Anliegen entspricht das methodische Konzept der optimierten Praktischen Fahrerlaubnisprüfung als kriteriengeleitetes Beurteilungsverfahren (STURZBECHER et al., 2014). Die Erreichung des Ziels, dass der Fahrerlaubnisprüfer am Prüfungsende auf der Grundlage einer adaptiven Prüfstrategie und der dokumentierten Prüfungsleistungen eine reflektierte und wohlabgewogene fachkundige Prüfungsentscheidung trifft, gehört zu den erklärten Prioritäten der vorliegenden Entwicklungsarbeiten von Anfang an (STURZBECHER et al., 2008 bzw. 2010).

Schließlich gaben einige Fahrlehrer wichtige Anregungen zur inhaltlichen Optimierung des e-Prüfprotokolls und betonten dabei, dass die Dokumentation und Rückmeldung aller Prüfungsleistungen des jeweiligen Bewerbers gewährleistet werden müssen. An dieser Stelle sei noch einmal darauf hingewiesen, dass – auch mit Blick auf die Akzeptanz bei der Verkehrsverwaltung und den Fahrerlaubnisprüfern – der fachlichen Aufklärung der Fachöffentlichkeit eine ganz entscheidende Bedeutung zukommt. Nur wenn alle am Prüfungswesen Beteiligten ausreichend informiert sind, kann die Einführung einer optimierten Praktischen Fahrerlaubnisprüfung zum angestrebten Erfolg führen.

### **Zusammenfassung der Befunde zur Kundenzufriedenheit**

Im Evaluationssystem zur optimierten Praktischen Fahrerlaubnisprüfung ist vorgesehen, die Zufriedenheit von Fahrerlaubnisbewerbern und Fahrlehrern mit der (alltäglichen) Durchführung der optimierten Praktischen Fahrerlaubnisprüfung in einem Fünfjahresturnus anhand einer bundesweit repräsentativen Studie zu erfassen, um empirisch begründete Anregungen für die kontinuierliche Optimierung und Weiterentwicklung der Prüfung zu gewinnen. Im Rahmen des Revisionsprojekts sollten mögliche Umsetzungsstrategien und Verfahrensweisen für eine Bewerberbefragung und eine Fahrlehrerbefragung erprobt werden.

Im Verlauf des Revisionsprojektes wurden die Rahmenbedingungen und Ziele der beiden Erprobungsuntersuchungen modifiziert. Erstens wurden die Inhalte der Bewerberbefragung auf die Zufriedenheit der Fahrerlaubnisbewerber mit dem neuen Rückmeldesystem für die Prüfungsleistungen und die Erfassung von Anregungen zu seiner Optimierung fokussiert. Eine verbesserte Rückmeldung der Prüfungsleistungen ist ein wichtiges gemeinsames Anliegen von Fahrerlaubnisbewerbern und Fahrlehrern, mit dem die Lernwirksamkeit der optimierten Praktischen Fahrerlaubnisprüfung erhöht und ihre Sicherheitswirksamkeit gestärkt werden könnte. Deshalb musste das erarbeitete Rückmeldesystem dringend erprobt werden, um den verkehrspolitischen Entscheidungsträgern empirisch begründete Informationen für die Entscheidung über die Systemimplementierung bereitzustellen. Zweitens wurden beide Befragungen – nicht zuletzt ausgelöst durch diese inhaltliche Fokussierung – räumlich auf die Modellregionen des Revisionsprojekts beschränkt. Dies stellt, verglichen mit einer bundesweiten Untersuchungsanlage, keinen substantiellen Verlust an Erkenntnismöglichkeiten dar, weil sich die erarbeiteten Erhebungsprozeduren mit ihren im Revisionsprojekt gefundenen Verbesse-



rungsmöglichkeiten bei künftigen Untersuchungen problemlos auf das Bundesgebiet ausweiten lassen.

Beide Befragungen haben – mittelbar oder unmittelbar – Ergebnisse erbracht, die eine empirisch begründete Abwägung der Frage gestatten, ob ein verbessertes Leistungsrückmeldesystem eingeführt werden sollte und wie es ggf. gestaltet werden könnte. Bei der Fahrlehrerbefragung wurde vor allem anhand der offenen Einträge zu wünschenswerten Verbesserungen des derzeitigen Prüfungssystems deutlich, dass die Optimierung der Leistungsrückmeldungen eine überfällige Maßnahme darstellt, die Fahrlehrern und Fahrerlaubnisbewerbern nützlich sein könnte. Eine direkte Bestätigung erfuhren diese Anregungen durch die Bewerberbefragung: 82 Prozent der Befragten waren mit der schriftlichen Rückmeldung „Sehr zufrieden“ (38 %) oder „Eher zufrieden“ (44 %), 76 Prozent waren mit den Hinweisen zur Verbesserung ihrer Fahrkompetenz „Sehr zufrieden“ (35 %) oder „Eher zufrieden“ (41 %), und 70 Prozent gaben an, dass sie die schriftliche Rückmeldung als hilfreich für das weitere Üben und Lernen im Straßenverkehr empfunden haben. Dies stellt ein überzeugendes Votum für die Einführung des neuen Rückmeldesystems dar; die Wertschätzung des Systems ist unabhängig vom Geschlecht, Alter und Bildungsniveau der Befragten.

Bei der Fahrlehrerbefragung hat sich gezeigt, dass der eingesetzte Fragebogen alle Themen anspricht, welche die Fahrlehrerschaft im Hinblick auf die (optimierte) Praktische Fahrerlaubnisprüfung beschäftigen. Um die Stärken und Schwächen der Prüfung sowie vor allem Optimierungsmöglichkeiten zu erfahren, bietet sich mit diesem Fragebogen demnach ein funktionstüchtiges Werkzeug: Die Stellschrauben zur Steigerung der Prüfungsqualität und der Fahrlehrerzufriedenheit werden umfassend abgebildet.

Unter methodischen Aspekten bleibt festzuhalten, dass sich – wie bereits erwähnt – die gewählte Verfahrensweise zur Befragung der Fahrerlaubnisbewerber grundsätzlich als funktionstüchtig erwiesen hat und im Rahmen der kontinuierlichen Evaluation der optimierten Praktischen Fahrerlaubnisprüfung in der Zukunft angewendet werden kann. Einige Optimierungsmöglichkeiten (z. B. eine bessere Ausgestaltung der Einladungssituation und die Verwendung von Anreizsystemen) haben sich gezeigt, die dafür sorgen dürften, dass die Ausschöpfungs-

quote bei künftigen Untersuchungen besser ausfällt. Bei der Fahrlehrerbefragung ist dagegen davon auszugehen, dass mit der postalischen Befragung die besseren Ausschöpfungsquoten gegenüber der Online-Befragung erreicht wurden. Dies könnte daran liegen, dass es sich bei den Fahrlehrern um eine Berufsgruppe mit einem hohen Durchschnittsalter handelt, die sich mit einem Fragebogen in Papierform wesentlich vertrauter fühlen dürfte als mit einem Onlineportal im Internet. Allerdings handelt es sich bei der postalischen Befragung um eine sehr kosten- und zeitintensive Befragungsmethode, was auch in Zukunft gegen den Einsatz dieses Verfahrens sprechen wird. Daher wird in Anlehnung an STURZBECHER et al. (2014) empfohlen, bei künftigen Fahrlehrerbefragungen auf eine telefonische Befragung zurückzugreifen: Es ist zu erwarten, dass der persönliche Kontakt im Telefongespräch zu einem wesentlich besseren Rücklauf führt.

## 6 Implementierungsvoraussetzungen

### 6.1 Schaffung der fahrerlaubnisrechtlichen Voraussetzungen

Für die bundesweite Einführung der optimierten Praktischen Fahrerlaubnisprüfung ist es erforderlich, die rechtlichen Rahmenbedingungen anzupassen. JAGOW (2010) prüfte auf der Grundlage der von STURZBECHER et al. (2010) vorgeschlagenen Weiterentwicklungen der Praktischen Fahrerlaubnisprüfung, welche fahrerlaubnisrechtlichen Anpassungen bei der Umsetzung der Vorschläge notwendig wären. Dabei stellte JAGOW einerseits fest, dass die Vorschläge zu einer inhaltlichen und methodischen Bereicherung der Praktischen Fahrerlaubnisprüfung führen würden: „Das neue Konzept ist zu begrüßen, denn die Beobachtungskategorien werden sich künftig deutlicher von den Prüfungsaufgaben abheben, und sie erscheinen gestrafter und übersichtlicher“. Andererseits konstatierte er, dass die Prüfungszeiten „knapp bemessen sind, um sämtliche Erfordernisse einer fachgerechten Prüfung zu berücksichtigen“ (JAGOW, 2010, S. 146 ff.).

Hinsichtlich der notwendigen Anpassung der fahrerlaubnisrechtlichen Grundlagen empfahl JAGOW (2010) allgemein die Änderung der Fahrerlaubnisverordnung bzw. Prüfungsrichtlinie hinsichtlich fol-

gender Inhalte: stärkere Strukturierung der Prüfungsaufgaben und Beobachtungskategorien, Überarbeitung der Bewertungs- und Entscheidungsregelungen, bessere Verankerung der Prüfungsdokumentation. Dabei sollten sich die Vorgaben auf die wichtigsten Rahmenbedingungen beschränken und insbesondere von Detailvorgaben „entschlackt“ werden, mit deren Veränderung im Verlaufe der evaluationsbasierten Weiterentwicklung der Praktischen Fahrerlaubnisprüfung zu rechnen ist. Um diesen Forderungen zu genügen, müssten folgende Änderungen in der Fahrerlaubnis-Verordnung (Anlage 7, Stand 16.04.2014) vorgenommen werden:

1. Zu Pkt. 2.1.5 „Prüfungsfahrt“: Anstelle der Aufzählung von einzelnen Punkten, bei welchen auf richtige Verhaltensweisen zu achten ist, sollte auf den Fahraufgabenkatalog und die darin beschriebenen Handlungsanforderungen bezüglich der einzelnen Fahraufgaben verwiesen werden.
  2. Zu Pkt. 2.3 „Prüfungsdauer und Mindestfahrzeit“: Die klassenspezifische Prüfungsdauer nebst der reinen Fahrzeiten sollten auf Basis der in der Aufwandsabschätzung zur optimierten Praktischen Fahrerlaubnisprüfung (s. u.) ermittelten Durchführungszeiten überarbeitet werden.
  3. Zu Pkt. 2.5.2 „Fehlerklassifikation“: Hier sollte die Klassifikation der Fehler vom Fahraufgabenkatalog übernommen werden. Anstelle der „erheblichen Fehler“ sollte auf „schwere Fehler, die zur sofortigen Beendigung der Prüfung führen“ verwiesen werden. Die Festlegungen zur Wiederholung oder Häufung von verschiedenen Fehlern könnte mit der Ergänzung konkretisiert werden, dass es sich dabei sowohl um „schwere Fehler“ als auch um „leichte Fehler“ handeln kann.
  4. Zu Pkt. 2.6 „Rückmeldung an den Bewerber“: Bei Beendigung der nicht bestandenen Prüfung sollte anstelle der kurzen Benennung der wesentlichen Fehler eine kompetenzbezogene Rückmeldung an den Fahrerlaubnisbewerber auf Basis der absolvierten Fahraufgaben und übergeordneten Fahrkompetenzbereiche erfolgen.
1. Zu Pkt. 1.3.8 „Prüfungsfahrt“: Die angeführten Anforderungen (ab Pkt. 1.3.8.3) an das Bewerberverhalten sollten ggf. durch einen Verweis auf den Fahraufgabenkatalog ersetzt werden. Im Zuge dessen erübrigt sich die „Tabelle der Fahraufgaben“, welche die geforderte Häufigkeit der Anforderungssituationen bei 5 Fahrprüfungen festlegt.
  2. Zu Pkt. 1.5 „Bewertung der „Prüfung“: Die aufgeführten „erheblichen Fehler“ sollten durch den Wortlaut der im Fahraufgabenkatalog definierten „schweren Fehler, die zur sofortigen Beendigung der Prüfung führen“ ersetzt werden. Die Festlegungen zur Wiederholung oder Häufung von verschiedenen Fehlern müssten mit der Ergänzung konkretisiert werden, dass es sich dabei sowohl um „schwere Fehler“ als auch um „leichte Fehler“ handeln kann. Darüber hinaus sollte – ohne einzelne Beispiele zu nennen – auf den Fahraufgabenkatalog verwiesen werden.
  3. Zu Pkt. 1.6 „Ergebnis der Prüfung“: Die Dokumentation der Prüfung im Prüfungsteil „Fahrtechnische Vorbereitung der Fahrt, Grundfahraufgaben und Prüfungsfahrt“ sollte näher geregelt werden. Es wird empfohlen festzuschreiben, dass im Prüfungselement „Prüfungsfahrt“ sowohl die Dokumentation des konkreten Bewerberverhaltens bei der Bewältigung der Fahraufgaben als auch die explizite kompetenzbezogene Bewertung der Fahraufgaben und Fahrkompetenzbereiche mittels eines elektronischen Prüfprotokolls erfolgen sollten.
  4. Zu Pkt. 3 „Muster für Prüfprotokoll“: Es ist festzuschreiben, dass das Muster für die schriftliche Rückmeldung an den Fahrerlaubnisbewerber dem für die optimierte Praktische Fahrerlaubnisprüfung entwickelten Rückmeldesystem entsprechen sollte.

Die vorstehenden, überblicksartig aufgeführten Änderungsempfehlungen zu den fahrerlaubnisrechtlichen Grundlagen werden im Zusammenhang mit dem Revisionsprojekt von der TÜV | DEKRA arge tp 21 konkretisiert und dem für den Straßenverkehr zuständigen Bundesministerium als Änderungsvorschlag unterbreitet.

Korrespondierend mit den genannten Änderungen in der Fahrerlaubnis-Verordnung müssten zusätzlich Änderungen in der Prüfungsrichtlinie (Stand April 2014) erfolgen:

Im Zuge der erforderlichen rechtlichen Anpassungen muss auch entschieden werden, wie der Fahraufgabenkatalog als zentrale inhaltliche Grundlage der optimierten Praktischen Fahrerlaubnisprüfung

rechtlich verankert werden soll. Es wird entsprechend den oben genannten Vorschlägen empfohlen, den Fahraufgabenkatalog verbindlich in den Rechtsrahmen der Praktischen Fahrerlaubnisprüfung aufzunehmen. Hierdurch sollen die im Fahraufgabenkatalog beschriebenen Inhalte einen verbindlichen Charakter erhalten, um die mit der Optimierung der Praktischen Fahrerlaubnisprüfung angestrebte Erhöhung der inhaltlichen und methodischen Güte der Prüfung bundesweit gleichermaßen zu gewährleisten. Aufgrund der Komplexität des Fahraufgabenkataloges wird allerdings empfohlen, diesen nicht detailliert als Teil der Fahrerlaubnis-Verordnung und Prüfungsrichtlinie zu verankern, sondern innerhalb der Verwaltungs- bzw. Richtlinientexte lediglich auf diesen zu verweisen (wie oben exemplarisch aufgeführt). Somit ließen sich Änderungen des Fahraufgabenkataloges unabhängig von einer Anpassung der Fahrerlaubnis-Verordnung bzw. der Prüfungsrichtlinie umsetzen.

Für die Integration des Fahraufgabenkataloges in das rechtliche Regelwerk der Fahrerlaubnisprüfung gibt es zwei Möglichkeiten:

1. Vorgehen entsprechend der theoretischen Fahrerlaubnisprüfung: Bei der theoretischen Fahrerlaubnisprüfung sind die Grundlagen der Prüfungsdurchführung in der Fahrerlaubnis-Verordnung und Prüfungsrichtlinie beschrieben. Die konkreten Prüfungsinhalte (Prüfungsfragen) erscheinen in Form eines „Amtlichen Fragenkataloges“ zwar als Teil der rechtlichen Grundlagen; gleichwohl kann der „Amtliche Fragenkatalog“ als separates Dokument unabhängig von der Fahrerlaubnis-Verordnung und Prüfungsrichtlinie geändert werden. Übertragen auf die optimierte Praktische Fahrerlaubnisprüfung hieße dies, dass der Fahraufgabenkatalog als rechtsverbindliches Dokument firmiert und notwendige Änderungen in diesem Katalog ebenfalls in einem amtlichen Freigabeverfahren umgesetzt werden müssten. Die Dokumentation des Verfahrens würde – analog der theoretischen Fahrerlaubnisprüfung – in einem web-basierten Administrations- und Dokumentationsprogramm der TÜV | DEKRA arge tp 21 erfolgen.
2. Vorgehen entsprechend der Beurteilungskriterien der Medizinisch-Psychologischen Fahreignungsdiagnostik: Seit 2006 sind die von der Deutschen Gesellschaft für Verkehrspsychologie e. V. und der Deutschen Gesellschaft für Verkehrsmedizin e. V. erarbeiteten Beurtei-

lungskriterien der Medizinisch-Psychologischen Fahreignungsdiagnostik in den Anforderungen an amtlich anerkannte Träger von Begutachtungsstellen für Fahreignung enthalten. Seither müssen diese Beurteilungskriterien von den durchführenden Stellen als verbindliche Verfahrensweisen angewendet werden. Änderungen an diesen Standards (beispielsweise aufgrund neuer empirischer Befunde) werden nicht mittels eines amtlichen Freigabeverfahrens vorgenommen, sondern aufgrund wissenschaftlich begründeter Expertenempfehlungen. Zur Inkraftsetzung dieser Neuerungen genügt die Veröffentlichung. Übertragen auf die optimierte Praktische Fahrerlaubnisprüfung hieße dies, dass der Fahraufgabenkatalog als von Wissenschaftlern und Fachexperten definierter und verbindlicher Standard für die Prüfungsinhalte nebst der Bewertungs- und Entscheidungskriterien existiert und Änderungen ohne ein amtliches Freigabeverfahren umgesetzt werden könnten.

Bei der Abwägung der Vor- und Nachteile der beiden Möglichkeiten sprechen die Analogie der Vorgehensweisen bei der Theoretischen und Praktischen Fahrerlaubnisprüfung und die daraus resultierende Vertrautheit der zuständigen Bundes- und Länderbehörden wie auch der TÜV | DEKRA arge tp 21 mit den notwendigen Verfahrensweisen für die erstgenannte Lösung. Weiterhin wären die höhere Rechtsverbindlichkeit der Durchführungs-, Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung und eine stärkere staatliche Kontrolle als mögliche Vorzüge zu nennen. Gleichzeitig würden sich aber auch die Arbeitsbelastungen der zuständigen Behörden in einem nicht unerheblichen Umfang erhöhen. Außerdem würde der mit Rechtsänderungen verbundene Zeitverlauf zu einer gewissen Anpassungsträgheit des Rechtsrahmens bei notwendigen Rechtsänderungen führen. Diese Trägheit könnte bei den zu erwartenden Anpassungsnotwendigkeiten infolge der evaluationsbasierten Weiterentwicklung des Systems vor allem in den nächsten Jahren hinderlich werden. Die zweitgenannte Lösung würde dagegen zwar weniger Rechtsverbindlichkeit der Standards und weniger staatliche Kontrolle mit sich bringen; als wahrscheinliche Vorzüge wären aber eine starke Arbeitsentlastung der zuständigen Behörden und eine mögliche höhere Anpassungsdynamik der inhaltlichen und methodischen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung zu nen-

nen, die der zügigen Qualitätsoptimierung dienen könnte.

Die künftigen Funktionen und Inhalte des Fahraufgabenkatalogs sowie die Erwartungen an seine fahrerlaubnisrechtliche Verbindlichkeit wurden bereits auf der 22. Sitzung der Entwicklungs- und Evaluierungsgruppe (EEG) am 02.12.2014 in Dresden diskutiert. Dieser Gruppe gehören – neben der TÜV|DEKRA arge tp 21 und den technischen Prüfstellen – auch Vertreter von zuständigen Bundes- und Landesbehörden, des VdTÜV, der Bundeswehr und der Bundesvereinigung der Fahrlehrerverbände an. Insbesondere wurde auf der Sitzung die grundsätzliche Frage erörtert, ob der Fahraufgabenkatalog ein fahrerlaubnisrechtlich verbindliches Regelwerk unter dem Dach der Fahrerlaubnis-Verordnung sein sollte oder ob er den Charakter einer fachwissenschaftlichen Empfehlung ähnlich wie die Begutachtungsleitlinien zur Kraffahreignung tragen sollte.

Sowohl der Vertreter des BMVI als auch die anwesenden Ländervertreter empfahlen, dass zwar die inhaltlichen und methodischen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung Prüfung rechtlich verankert werden sollten, die detaillierteren Inhaltsebenen aber nicht. Als vorbildliches Beispiel für eine solche Mehrstufigkeit der verkehrsrechtlichen Verankerung wurden die bereits erwähnten Begutachtungsleitlinien zur Kraffahreignung angeführt (Anlage 4a der FeV): Hier würden wissenschaftliche Grundlagen benannt bzw. vorgegeben, aber nicht inhaltlich konkretisiert. Die Konkretisierung erfolge außerhalb der rechtlichen Grundlagen, was einer fortlaufenden Anpassung der Maßnahme an den Stand der Wissenschaft dienen würde. Dagegen würde eine rechtliche Verankerung des Fahraufgabenkataloges in der Fahrerlaubnis-Verordnung bzw. Prüfungsrichtlinie zur Folge haben, dass jede inhaltliche Änderung im Bund-Länder-Fachausschuss „Fahrerlaubnisrecht/Fahrlehrerrecht“ (BLFA) abgestimmt werden müsse. Dies spräche – auch nach Meinung anderer Sitzungsteilnehmer – eher dafür, zwar die Fahrkompetenzbereiche fahrerlaubnisrechtlich zu verankern, nicht jedoch die Fahraufgaben (EEG-PROTOKOLL vom 02.12.2014, TÜV|DEKRA arge tp 21, 2014).

Derzeit besteht immer noch Klärungsbedarf bezüglich der künftigen fahrerlaubnisrechtlichen Verbindlichkeit des Fahraufgabenkataloges. Eine diesbezügliche Entscheidung muss kurzfristig im Bund-

Länder-Fachausschuss „Fahrerlaubnisrecht/Fahrlehrerrecht“ (BLFA) vorbereitet bzw. herbeigeführt werden, denn fahrerlaubnisrechtliche Änderungen wie die Verankerung des Fahraufgabenkatalogs in der Anlage 7 der Fahrerlaubnis-Verordnung wären – sofern gewünscht – nicht kurzfristig umzusetzen. Weiterhin werden zurzeit Detailvorschläge zu notwendigen Rechtsanpassungen von der TÜV|DEKRA arge tp 21 erarbeitet, die in Kürze den für den Straßenverkehr zuständigen Bundes- und Länderbehörden übermittelt werden.

## 6.2 Schaffung der technischen und organisatorischen Voraussetzungen

Die Einführung der optimierten Praktischen Fahrerlaubnisprüfung erfordert – neben den oben aufgezeigten notwendigen Anpassungen der rechtlichen Grundlagen – auch materielle und personelle Aufwände seitens der technischen Prüfstellen. Obgleich im Rahmen des vorliegenden Revisionsprojektes in den Modellregionen der Erprobung bereits Vorbereitungen für den Praxiseinsatz der Prüfung getroffen wurden, handelte es sich dabei lediglich um lokal und temporär begrenzte Insellösungen.<sup>75</sup> Für die bundesweite Einführung des optimierten Prüfungssystems können die technischen Prüfstellen damit nunmehr zwar auf die Erprobungserfahrungen zurückgreifen; allerdings ist damit der technische und organisatorische Aufwand für den flächendeckenden Einsatz noch längst nicht abgedeckt. Nachfolgend soll daher bezüglich unterschiedlicher Aufwandskategorien überblicksartig konkretisiert werden, welche Aufwände bzw. Kostenpositionen noch zu erwarten sind:

<sup>75</sup> Für die Bereitstellung dieser Lösungen wie auch für die großzügige unentgeltliche Unterstützung des Revisionsprojektes mit materiellen und personellen Ressourcen insgesamt gebührt den technischen Prüfstellen großer Dank. Zu den Unterstützungsleistungen zählen ca. 9.000 Prüfungen mit verlängerter Dispositionszeit, die zweitägige Schulung der Fahrerlaubnisprüfer, der ebenfalls zwei Tage umfassende Erfahrungsaustausch mit ihnen, die Anschaffung der Erprobungshardware, die Schaffung aller notwendigen IT-Schnittstellen für die Erprobungsstandorte (Modellregionen) und nicht zuletzt die zeitintensive Erarbeitung des Fahraufgabenkataloges.

Die Erarbeitung des Fahraufgabenkataloges und die Durchführung des Revisionsprojektes wurden auch maßgeblich durch die Bundesvereinigung der Fahrlehrerverbände unterstützt; auch dafür sei herzlich gedankt.



## 1. Aufbau und Betrieb des Prüfungssystems

Für eine bundesweite Einführung der optimierten Praktischen Fahrerlaubnisprüfung müssen die IT-Systeme der technischen Prüfstellen flächendeckend vorbereitet werden. Hierzu zählen die Anschaffung mobiler Endgeräte (für das e-Prüfprotokoll) nebst Zubehör, die Schaffung einer entsprechenden Server-Architektur sowie die Anpassung bestehender Verwaltungssysteme. Nach der Implementierung der optimierten Praktischen Fahrerlaubnisprüfung müssen die Systeme durch die technischen Prüfstellen gewartet und Unterstützungsleistungen für ihre Nutzung gewährleistet werden. Hierzu zählen auch Maßnahmen zur Einhaltung des Datenschutzes im Rahmen der elektronischen Datenverarbeitung. Darüber hinaus ist es erforderlich, alle mit der Durchführung der Fahrerlaubnisprüfung betrauten Fahrerlaubnisprüfer hinsichtlich der optimierten Prüfungsverfahren und -abläufe in den Bildungseinrichtungen der technischen Prüfstellen zu schulen (im Revisionsprojekt erfolgte die Schulung der teilnehmenden Fahrerlaubnisprüfer durch den Projektnehmer bzw. dessen Kooperationspartner). Hierfür müssen zunächst einheitliche Aus- und Fortbildungsmaterialien erarbeitet werden (s. Kapitel 2.5).

## 2. Verlängerte Prüfungsdurchführung

Obwohl sich der allgemeine Prüfungsablauf in Form von Begrüßung, Formalien, Einführungsgepräch, Durchführung klassenspezifischer Prüfungselemente, Leistungsrückmeldung und Verabschiedung grundsätzlich nicht ändert, führen die Inhalte der optimierten Praktischen Fahrerlaubnisprüfung zum Teil zu veränderten Prozessschritten. Hierzu zählen technisch bedingte Prozesse wie etwa das Übertragen von Bewerberdaten zwischen den Verwaltungssystemen der technischen Prüfstellen und dem e-Prüfprotokoll und umfangreichere Dokumentationsaufgaben des Fahrerlaubnisprüfers – nicht nur im Rahmen der Prüfungsfahrt, sondern auch in den anderen klassenspezifischen Prüfungselementen.

Die Erprobung der optimierten Praktischen Fahrerlaubnisprüfung zeigte, dass insbesondere im Rahmen der Prüfungsfahrt ein höherer Zeitbedarf notwendig ist. Die Fahraufgaben und Kompetenzbereiche werden im e-Prüfprotokoll in Form einer Matrix angeordnet und geben dem Fahrerlaubnisprüfer zu jedem Zeitpunkt der Prüfung einen anschaulichen Überblick über die während der Prü-

fahrungsfahrt dokumentierten Prüfungsleistungen (s. Kapitel 2.2). Durch die kontinuierliche Anzeige der vorgegebenen Mindestanforderungen und deren aktuellem Bewältigungsstand – verbunden mit der Notwendigkeit, jede Fahraufgabe nach der Prüfungsfahrt explizit zu beurteilen – projektieren die Fahrerlaubnisprüfer die Prüfungsfahrt im Sinne der adaptiven Prüfstrategie sorgfältiger. Die Fahraufgaben werden für die gezielte Überprüfung der Anforderungsstandards aus dem Fahraufgabenkatalog konsequenter durchgeführt als bisher. Dies – so zeigen die Erfahrungen der Erprobung – erfordert eine längere Prüfungsfahrt.

Auch die daran anknüpfende Prüfungsbewertung bzw. Prüfungsentscheidung erfordert nach den Erprobungserfahrungen zusätzliche Zeit im Vergleich zur gegenwärtigen Praxis. Die Prüfungsleistungen werden bei der optimierten Praktischen Fahrerlaubnisprüfung umfangreicher und detaillierter dokumentiert als bisher. Für die Beurteilung der acht Fahraufgaben und fünf Kompetenzbereiche muss der Fahrerlaubnisprüfer die einzelnen Prüfungsleistungen des Bewerbers differenziert reflektieren und abwägen (s. Kapitel 2.1). Dabei handelt es sich um einen neuen Prozessschritt. Auf Basis dieser kompetenzbezogenen Bewertungen trifft der Fahrerlaubnisprüfer dann begründeter und transparenter als bisher seine Prüfungsentscheidung.

Wie im Kapitel 2.4 ausgeführt, stellt die ausführliche Rückmeldung der Prüfungsleistungen ein wesentliches Element der optimierten Praktischen Fahrerlaubnisprüfung dar. Diese Rückmeldung erfolgt für jeden Bewerber – ob bestanden oder nicht bestanden – zum einen durch ein Auswertungsgespräch im Anschluss an die Prüfung und zum anderen in Form einer zeitnahen schriftlichen Rückmeldung. Im Auswertungsgespräch erläutert der Fahrerlaubnisprüfer das Prüfungsergebnis und die zusammenfassenden Kompetenzbeurteilungen mit beispielhafter Benennung der beobachteten Fehler bzw. überdurchschnittlichen Leistungen; weiterhin gibt er Empfehlungen zum weiteren Fahrkompetenzerwerb. Dieses im Vergleich zu heute ausführlichere Rückmeldegespräch erfordert zusätzliche Zeit.

Aus dem e-Prüfprotokoll wird eine schriftliche Rückmeldung für den Bewerber generiert. Diese Rückmeldung wird jedem Bewerber in Form eines mehrseitigen Dokuments bereitgestellt. Gegenwärtig erhalten nur Bewerber, welche nicht bestanden haben, ein einseitiges Fehlerprotokoll ausgehändigt. In diesem sind gemäß den Anforderungen aus

der Prüfungsrichtlinie lediglich die Fehler aufgelistet, die zum Nichtbestehen geführt haben. Bei DEKRA ist geplant, die schriftliche Rückmeldung mittels Ausdruck unmittelbar am Ende der Prüfung (im Fahrzeug) bereitzustellen. Die TÜVs planen die Bereitstellung über ein Web-Portal. Hierzu generiert das e-Prüfprotokoll einen individuellen bewerberbezogenen Code, welcher vom Fahrerlaubnisprüfer auf einem Abreibblock notiert und dem Bewerber ausgehändigt wird. Beide Umsetzungsmöglichkeiten sind mit zeitlichem Aufwand verbunden.

Im Zuge der Erprobung der optimierten Praktischen Fahrerlaubnisprüfung wurde der Zeitbedarf für die optimierte Prüfungsdurchführung von der TÜV | DEKRA arge tp 21 und den technischen Prüfstellen untersucht. Die Auswertung der Untersuchungen war bis zum Abgabetermin des vorliegenden Berichts noch nicht abgeschlossen. Daher wird empfohlen, dass die Bundesanstalt für Straßenwesen eine direkte Abstimmung mit der TÜV | DEKRA arge tp 21 und den technischen Prüfstellen vornimmt.

### **3. Weiterentwicklung und Evaluation der Prüfung**

Um den Beitrag der (Praktischen) Fahrerlaubnisprüfung zur Verkehrssicherheit angesichts sich wandelnder und komplexer werdender Verkehrsbedingungen auch künftig angemessen zu gewährleisten, müssen die eingesetzten technischen Mittel ebenso wie die angewandten Verfahren und Abläufe des Fahrerlaubnisprüfungssystems kontinuierlich weiterentwickelt werden. Insbesondere geht es darum, fahranfängertypische Unfallursachen und Kompetenzdefizite bei der Weiterentwicklung von Prüfungselementen und Fahraufgaben immer besser zu berücksichtigen und den technischen Fortschritt für eine prüfungsdidaktisch angemessene Prüfungsgestaltung und Prüfungsdokumentation zu nutzen. Hierzu ist die wissenschaftliche Abstützung und Weiterentwicklung der Prüfungsinhalte und Prüfungsmethoden dringend geboten.

Darüber hinaus bildet die systematische Auswertung von durchgeführten Prüfungen die Voraussetzung für die Kontrolle der Wirksamkeit der Praktischen Fahrerlaubnisprüfung im Sinne der Zielsetzungen. Zu diesen Auswertungen zählen sowohl testpsychologische Untersuchungen zur instrumentellen Prüfungsgüte als auch statistische Analysen und Kundenbefragungen zur Qualität der Prüfungsprozesse. Ein methodisch anspruchsvolles und leicht handhabbares Prüfverfahren und Prüfproto-

koll bieten die Grundlage für die Evaluation der Praktischen Fahrerlaubnisprüfung (s. Kapitel 2.3) und ihrer Steuerungsfunktion für die Fahranfängervorbereitung.

Damit die optimierte Praktische Fahrerlaubnisprüfung – ebenso wie die theoretische Fahrerlaubnisprüfung – bundesweit einheitlich und kontinuierlich evaluiert und weiterentwickelt werden kann, haben die technischen Prüfstellen die dafür notwendigen Arbeiten der für diesen Zweck 1999 gegründeten Arbeitsgemeinschaft TÜV | DEKRA arge tp 21 übertragen. Die TÜV | DEKRA arge tp 21 bereitet die anonymisierten Prüfungsdaten auf, evaluiert diese mithilfe externer wissenschaftlicher Einrichtungen kontinuierlich und setzt auf dieser Grundlage empirisch begründete Weiterentwicklungen der Prüfungsinhalte und der Prüfungsmethodik um. Hierzu zählen auch Vorschläge zum Umgang mit fahrzeugtechnischen Entwicklungen im Rahmen der Prüfung und zur Weiterentwicklung der Anforderungen an die Prüferte. Die Erarbeitung dieser Vorschläge fällt ebenso wie die Erarbeitung von Änderungsempfehlungen bezüglich des Fahraufgabekataloges und von Entwürfen neuer Prüfungselemente in die Zuständigkeit der TÜV | DEKRA arge tp 21-AG „Praktische Prüfung und Fahraufgabenbeschreibung“.

Darüber hinaus stellt die TÜV | DEKRA arge tp 21 sicher, dass die technischen Rahmenbedingungen der optimierten Praktischen Fahrerlaubnisprüfung aufrechterhalten werden. Hierzu zählt auch die Entwicklung und Pflege des zentralen Dokumentationsinstruments der optimierten Praktischen Fahrerlaubnisprüfung – des e-Prüfprotokolls (s. Kapitel 2.2). Außerdem pflegt die TÜV | DEKRA arge tp 21 eine zentrale Datenbank, in welcher alle Prüfungsinhalte (Prüfungselemente, Fahraufgaben, Beobachtungskategorien, Bewertungskriterien und Entscheidungskriterien) hinterlegt sind. Von dieser Datenbank aus werden die Informationen über die spezifische IT-Struktur der technischen Prüfstellen an die Prüfungscomputer vor Ort verteilt, sodass die Prüfungsergebnisse in einheitlichem Format erfasst werden können. Dabei wird sichergestellt, dass alle Prüfungscomputer in Deutschland mit dem gleichen Datenbestand arbeiten. Mittels eines Administrations- und Dokumentationsprogramms werden die Prüfungsinhalte beschrieben; darüber hinaus werden damit die Erstellungsprozesse (einschließlich der Freigabe) und die Prozessergebnisse in Anlehnung an die bewährten Verfahren der theoretischen Fahrerlaubnisprüfung dokumentiert.

Eine weitere Aufgabe der TÜV | DEKRA arge tp 21 besteht darin, die Öffentlichkeit durch eine periodische Berichtslegung über den Stand und den Fortschritt der Optimierungsarbeiten zu informieren. Hervorzuheben ist in diesem Zusammenhang der Innovationsbericht, in dem alle mit der mittel- und langfristigen Weiterentwicklung des Fahrerlaubnisprüfungssystems zusammenhängenden wesentlichen Forschungs- und Entwicklungsprozesse sowie ihre Ergebnisse im Hinblick auf die Theoretische und Praktische Fahrerlaubnisprüfung dokumentiert werden. Dieser Bericht wird in einem festen Turnus den zuständigen Behörden vorgelegt.

Alle diese Aufgaben der TÜV | DEKRA arge tp 21 bzw. der technischen Prüfstellen sowie grundlegende Vorgaben zur Aufgabenerfüllung sind in einem „Handbuch zum Fahrerlaubnisprüfungssystem (Praxis)“ beschrieben. Dieses Handbuch liegt in einer Entwurfsfassung vom 04.02.2011 vor und muss auf der Grundlage der Erfahrungen des Revisionsprojekts überarbeitet werden, um dann dem Bund-Länder-Fachausschuss „Fahrerlaubnisrecht/Fahrlehrerrecht“ (BLFA FE/FL) zur Kenntnisnahme vorgelegt zu werden.

Die hier aufgeführten Aufwände für die Implementierung, den Betrieb und die Weiterentwicklung der optimierten Praktischen Fahrerlaubnisprüfung werden derzeit in einer Aufwandsabschätzung konkretisiert, welche alle Mehraufwände und Einsparungsmöglichkeiten im Vergleich zur gegenwärtigen Prüfungspraxis berücksichtigt. Diese Aufwandsabschätzung wird – nach Attestierung durch eine unabhängige Wirtschaftsprüfungsgesellschaft – den für den Straßenverkehr zuständigen Bundes- und Landesbehörden von der TÜV | DEKRA arge tp 21 vorgelegt. Wenn anhand der Aufwandsabschätzung ein Mehraufwand nachgewiesen werden kann, sollte dies auch in der Gebührenordnung für Maßnahmen im Straßenverkehr (GebOSt) seinen Niederschlag finden.

### 6.3 Zeitplan für die Implementierung

Bevor ein Zeitplan für die Implementierung der optimierten Praktischen Fahrerlaubnisprüfung erstellt und ein Einführungstermin genannt werden können, müssen die für die Implementierungsentscheidung zuständigen Gremien auf der Grundlage des vorliegenden Erprobungsberichtes und der o. g. Aufwandsabschätzung eine Richtungsentscheidung treffen, ob und ggf. in welcher Form

die neuen Durchführungs-, Anforderungs- und Bewertungsstandards eingeführt werden sollen. Dazu gehören auch Entscheidungen zur fahrerlaubnisrechtlichen Verankerung dieser Standards (v. a. des Fahraufgabenkatalogs) und zur mehrkostenadäquaten Anpassung der Gebührenordnung. Sobald diese Entscheidungen getroffen sind, lassen sich der organisatorische, technische und rechtliche Vorbereitungsaufwand und die zu seiner Bewältigung notwendigen Zeiträume kurzfristig und treffsicher kalkulieren sowie in einen Zeitplan für die Implementierung überführen.

## 7 Zusammenfassung und Ausblick

### Zusammenfassung der Ergebnisse

Die Durchführungs-, Anforderungs- und Bewertungsstandards der optimierten Praktischen Fahrerlaubnisprüfung wurden im Rahmen des Revisionsprojekts mit Untersuchungen zur instrumentellen Evaluation und mit ca. 9.000 realen Praktischen Fahrerlaubnisprüfungen in vier Modellregionen erprobt. Die nachfolgend aufgeführten Ergebnisse können zusammenfassend festgehalten werden und stellen die wichtigsten Befunde dar.

Zur Sicherstellung der Verfahrensgüte der optimierten Praktischen Fahrerlaubnisprüfung wurde die Beobachterübereinstimmung als bedeutendster Indikator für die Objektivität und Reliabilität des Messverfahrens bei 15 simulierten Prüfungen der Fahrerlaubnisklasse B, bei 20 realen Prüfungen der Fahrerlaubnisklasse B, bei 20 realen Prüfungen verschiedener Fahrerlaubnisklassen und bei drei videografierten Prüfungen untersucht. Die videografierten Prüfungen wurden den Fahrerlaubnisprüfern jeweils zu Beginn, zur Mitte und zum Ende der viereinhalbmonatigen Felderprobung zur systematischen Beobachtung und Bewertung vorgeführt. Im Ergebnis der drei Untersuchungen im Realverkehr kann festgehalten werden, dass bezüglich der Prüfungsentscheidungen und der Dokumentation der für die Prüfungsentscheidung besonders bedeutsamen schweren Fehler die vorab festgelegten Mindestbeobachterübereinstimmungswerte bei allen drei Untersuchungen erreicht bzw. – teilweise deutlich – übertroffen wurden. Hinsichtlich der Gesamtbewertungen der Fahraufgaben und Kompetenzbereiche bzw. Beobachtungskategorien sowie

der Dokumentation der leichten Fehler wurden nur in den beiden erstgenannten Untersuchungen (simulierte Prüfungen, reale Prüfungen der Fahrerlaubnisklasse B) zufriedenstellende Übereinstimmungswerte erzielt. Bei der dritten Untersuchung (reale Prüfungen verschiedener Fahrerlaubnisklassen) konnte ein solcher belastbarer Nachweis (noch) nicht erbracht werden. Dies dürfte durch die erhöhten, noch ungewohnten Anforderungen eines ständigen Wechsels zwischen den zu prüfenden Fahrerlaubnisklassen bzw. den unterschiedlichen Bestandteilen des Fahraufgabenkatalogs sowie einen „Trainingsrückstand“ der Prüfer im Vergleich mit den ersten beiden Untersuchungen begründet sein. Bestätigt wird diese Annahme durch die kontinuierliche Zunahme der Beobachterübereinstimmungswerte bei der Beurteilung videografiert Prüfungen vom ersten bis zum dritten Messzeitpunkt. Demnach gelangten die an dieser Untersuchung teilnehmenden Prüfer mit zunehmender Übung im Umgang mit dem e-Prüfprotokoll zu immer objektiveren und reliableren Ergebnissen, was auf die Erreichbarkeit einer hohen Verfahrensgüte bei ausreichendem Beobachtertraining schließen lässt.

Hinsichtlich der Validität der Prüfung fanden sich folgende Validitätsbelege:

- Die Analyse der Prüfungsergebnisse bestätigte das theoretische und methodische Konzept der optimierten Praktischen Fahrerlaubnisprüfung. Die Inhalte und die Struktur des Fahraufgabenkatalogs und der Kompetenzbereiche sowie das System der Bewertungskriterien haben sich als tragfähig erwiesen; alle Fahraufgaben und Kompetenzbereiche stellen eigenständige und nicht redundante Prüfungsanforderungen bzw. Facetten der Fahrkompetenz dar.
- Weiterhin fanden sich bei der Auswertung der Prüfungsleistungen im Hinblick auf die Fahraufgaben und Kompetenzbereiche Ergebnisse, die aus der Verkehrsforschung und Unfallstatistik bekannt sind (z. B. fahranfängertypische Kompetenzdefizite beim Bewältigen der Fahraufgabe „Kreuzungen, Einmündungen“ und im Kompetenzbereich „Verkehrsbeobachtung“) oder zumindest fachlich plausibel erscheinen (z. B. Häufigkeiten bestimmter Bewertungskombinationen).
- Die Bewerbermerkmale „Vorbesitz einer anderen Fahrerlaubnis“, „Fahrerfahrung“ und „Konsum straßenverkehrsbezogener Computerspie-

le“ wirkten sich erwartungsgemäß förderlich auf das Bestehen der Prüfung aus.

- Es existieren akzeptable Zusammenhänge zwischen den prüfungsbasierten Fahrlehrereinschätzungen und Prüferereinschätzungen.

Die Ergebnisse der Reliabilitäts- und Validitätsanalysen belegen also weitestgehend die Verfahrensgüte und damit die Einsatztauglichkeit der optimierten Praktischen Fahrerlaubnisprüfung. Die Implementierung des vorliegenden Verfahrens kann demnach aus wissenschaftlicher Sicht klar empfohlen werden. Dies gilt insbesondere, wenn einzelne bei der Erprobung auffällig gewordene Verbesserungsmöglichkeiten bis dahin noch verwirklicht werden. Im Rahmen des Revisionsprojekts zeigten sich bei den Prüfungsbewertungen und Prüfungsentscheidungen aber auch einige überzufällige Standort- und Prüferunterschiede, deren Ursachen noch nicht abschließend geklärt werden konnten, weil die vorliegenden Daten keine Repräsentativität für das gesamte Bundesgebiet besitzen und die Auswertung der Prüfungsergebnisse bislang lediglich exemplarisch erfolgte. Trotzdem unterstreichen die gefundenen Unterschiede die Notwendigkeit, an der prüfstellenübergreifenden wissenschaftlichen Erarbeitung, Evaluation und Weiterentwicklung der inhaltlichen und methodischen Grundlagen der optimierten Praktischen Fahrerlaubnisprüfung strikt festzuhalten und den fachlichen Austausch über die praktische Umsetzung dieser Grundlagen auszubauen.

Schließlich sollten im Rahmen des Revisionsprojekts mögliche Umsetzungsstrategien und Verfahrensweisen für eine künftige bundesweite Bewerberbefragung und Fahrlehrerbefragung erprobt werden. Die Erprobungsergebnisse zeigen, dass die im Fünfjahresturnus geplanten Befragungen mit aussagekräftigen Ergebnissen und vertretbarem Aufwand durchführbar sind. Darüber hinaus haben beide Befragungen – mittelbar oder unmittelbar – Ergebnisse erbracht, die auf die notwendige Einführung eines verbesserten Leistungsrückmeldesystems hindeuten. Das erprobte Rückmeldesystem wurde von der Mehrheit der befragten Bewerber akzeptiert: 82 Prozent der Befragten waren mit der schriftlichen Rückmeldung „Sehr zufrieden“ (38 %) oder „Eher zufrieden“ (44 %), 76 Prozent waren mit den Hinweisen zur Verbesserung ihrer Fahrkompetenz „Sehr zufrieden“ (35 %) oder „Eher zufrieden“ (41 %), und 70 Prozent gaben an, dass sie die schriftliche Rückmeldung als hilfreich für das



weitere Üben und Lernen im Straßenverkehr empfunden haben. Dies stellt ein überzeugendes Votum für die Einführung des neuen Rückmeldesystems dar; die Wertschätzung des Systems ist unabhängig vom Geschlecht, Alter und Bildungsniveau der Befragten.

### Ausblick

Die Praktische Fahrerlaubnisprüfung nimmt im Maßnahmensystem zur Fahranfängervorbereitung eine Schlüsselfunktion ein: Die Fahrerlaubnisbewerber sollen in der Prüfung genau das selbstständige und sichere Fahrverhalten zeigen, das nachfolgend beim selbstständigen Fahren im öffentlichen Straßenverkehr gefordert wird. Mit den neuen Durchführungs-, Anforderungs- und Bewertungsstandards sowie dem e-Prüfprotokoll können nun erstmalig in der Geschichte der Fahrerlaubnisprüfung alle besonders guten Fahrleistungen wie auch die Fahrfehler mit Bezug auf konkrete Fahraufgaben und Fahrkompetenzen verlaufsbezogen und vollständig dokumentiert werden. Damit sind eine Reihe von Nutzenaspekten verbunden, von denen nur drei genannt werden sollen: Erstens erlauben es die Standards und das e-Prüfprotokoll künftig, die instrumentelle Güte und die Sicherheitswirkung der Praktischen Fahrerlaubnisprüfung auf empirischer Basis zu beurteilen und zu erhöhen. Es lässt sich beispielsweise künftig erkennen, welche Fahraufgaben geprüft werden, welche Fahrfehler dabei besonders häufig auftreten und wie die Fahrschul Ausbildung darauf reagieren kann. Zweitens ermöglichen die Standards und das e-Prüfprotokoll künftig eine verbesserte Rückmeldung der Prüfungsergebnisse an alle Bewerber. Damit kann ein gezieltes Weiterlernen bei allen Bewerbern und nicht zuletzt bei denen, die nicht bestanden haben, gefördert werden. Drittens schließlich werden die Standards und das e-Prüfprotokoll die erfolgreiche Umsetzung der Bestrebungen nach einer bundesweit einheitlichen Prüfungsdurchführung bzw. nach Prüfungsgerechtigkeit vorantreiben.

Die vorliegenden Ergebnisse des Erprobungsprojekts weisen darauf hin, dass das differenzierte Bewertungsverfahren wie auch das Auswertungsgespräch von den meisten Bewerbern als förderlich für das weitere Lernen angesehen wird und sie davon profitieren (s. o.).

Im Rahmen der Forschungs- und Entwicklungsarbeiten zur Optimierung der Praktischen Fahr-

erlaubnisprüfung ist das zur sicheren Verkehrsteilnahme erforderliche Handlungswissen („Fahraufgaben“ einschließlich Bewertungskriterien) durch Fachexperten (Fahrlehrer, Fahrerlaubnisprüfer) und Wissenschaftler (Verkehrspädagogen und -psychologen) differenziert und anwendungsnah beschrieben worden. Im vorliegenden Revisionsprojekt haben sich diese Anforderungs- und Bewertungsstandards empirisch bewährt. Der damit erstmals vorliegende erprobte und öffentlich zugängliche Fahraufgabenkatalog enthält also gemeinsame Bildungsstandards für die Fahrschul Ausbildung und Fahrerlaubnisprüfung. Somit sind messbare einheitliche Leistungskriterien vorgegeben, die sowohl von Fahrlehrern bei der Fahrpraktischen Ausbildung und Prüfungsreifefeststellung als auch von Fahrerlaubnisprüfern bei der Einschätzung und Rückmeldung von Prüfungsleistungen anzuwenden sind. Die auf diesen wissenschaftlichen Standards beruhende differenzierte Leistungserfassung mittels des e-Prüfprotokolls eröffnet Chancen nicht nur für eine verbesserte Fahrerlaubnisprüfung, sondern nicht zuletzt auch für die wesentliche Verbesserung der Lernstandsdiagnostik in der Fahrschul Ausbildung, die nach Experteneinschätzung überfällig ist. Da nun die Erprobung des e-Prüfprotokolls erfolgreich abgeschlossen ist, sollten auf der Basis der Erprobungserfahrungen Möglichkeiten herausgearbeitet werden, Teile des entwickelten inhaltlichen und methodischen Know-hows in ausbildungsspezifischen lernstandsdiagnostischen Materialien zu nutzen. Damit entstünde eine geeignete Grundlage für die Feststellung von Lernfortschritten in der fahrpraktischen Ausbildung und für die individualisierte Förderung der Fahrschüler.

Im Ergebnis des Revisionsprojektes bleibt abschließend festzuhalten, dass sich die in den letzten zehn Jahren durchgeführten Forschungs- und Entwicklungsarbeiten zur Optimierung der Praktischen Fahrerlaubnisprüfung offensichtlich gelohnt haben und die neuen Prüfungsstandards als belastbares Fundament für die Stärkung der Fahranfängervorbereitung in Deutschland anzusehen sind.

## 8 Literatur

- ALBRECHT, F. (2005): Die rechtlichen Rahmenbedingungen bei der Implementierung von Fahrerassistenzsystemen zur Geschwindigkeitsbeeinflussung. *Deutsches Autorecht – DAR* 4, 186-198
- BACKHAUS, K., ERICHSON, B., PLINKE, W. & WEIBER, R. (2006): *Multivariate Analysemethoden*, 11. Aufl., Berlin u. a.
- BAMBERG, E. (2010): *Feedback – eine Klärung. Gruppendynamik und Organisationsberatung*. VS Verlag
- BAMF – Bundesamt für Migration und Flüchtlinge (Hrsg. 2014): *Migrationsbericht des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung (Migrationsbericht 2012)*.
- BANGERT-DROWNS, R. L., KULIK, C., KULIK, J. A. & MORGAN, M. T. (1991): The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238
- BATINIC, B. & MOSER, K. (2005): Determinanten der Rücklaufquote in Online-Panels. *Medienpsychologie*, 17, 64-74
- 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
- BÖNNINGER, J. & STURZBECHER, D. (2005): Optimierung der Fahrerlaubnisprüfung. Ein Reformvorschlag für die theoretische Fahrerlaubnisprüfung. *Berichte der Bundesanstalt für Straßenwesen*, Heft M 168. Bremerhaven: Wirtschaftsverlag NW
- BORTZ, J. (1999): *Statistik für Sozialwissenschaftler*. 5. Auflage. Berlin : Springer
- BORTZ, J. (2005): *Statistik für Human- und Sozialwissenschaftler* (6., vollst. überarb. und aktualisierte Aufl.). Heidelberg: Springer Medizin
- BORTZ, J. & DÖRING, N. (2002): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer
- BORTZ, J. & DÖRING, N. (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer
- BREDOW, B. (2015): Ausbildungskonzepte zur Schulung von Verkehrswahrnehmung und Gefahrenvermeidung. In: TÜV | DEKRA arge tp 21 (Hrsg.): *Verkehrswahrnehmung und Gefahrenvermeidung – Grundlagen und Umsetzungsmöglichkeiten in der Fahranfängervorbereitung*. Innovationsbericht zum Fahrerlaubnisprüfungssystem 2011-2014. Dresden: TÜV | DEKRA arge tp 21
- 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
- BUNGARD, W., HOLLING, H. & SCHULTZ-GAMBARD, J. (1996): *Methoden der Arbeits- und Organisationspsychologie*. Weinheim: Psychologie-Verlags Union
- CARVER, C. S. & SCHEIER, M. F. (2000): On the structure of behavioral selfregulation. In: M. BOEKAERTS, P. R. PINTRICH & M. ZEIDNER: *Handbook of selfregulation* (pp. 41-84). San Diego: Academic Press
- Centraal Bureau Rijvaardigheidsbewijzen (2012): *Leerboek voor de RIS-instructeur*
- CHETCUTI, D. (2009): Identifying a gender-inclusive pedagogy from Maltese teachers' personal practical knowledge. *International Journal of Science Education*, 31 (1), S. 81-99
- COHEN, J. (1988): *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates: Hillsdale

- COLLINS, M., CARNINE, D. & GERSTEN, R. (1987): Elaborated corrective feedback and the acquisition of reasoning skills: A study of computer-assisted instruction. *Exceptional Children*, 54, 254-262
- COLLINS, M., SYKES, W., WILSON, P. & BLACKSHAW, N. (1988): Nonresponse: the UK Experience. In: R. M. GROVES, P. P. BIEMER, L. E. LYBERG, T. J. MASSEY, L. W. NICHOLLS II & J. WAKSBERG (Hrsg.): *Telephone survey methodology*. New York, S. 213-231
- DECI, E. L., KOESTNER, R. & RYAN, R. M. (2001): Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71:1, 1-17
- DITTON, H. & MÜLLER, A. (Hrsg.) (2014): *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. Münster: Waxmann
- EID, M., GOLLWITZER M. & SCHMITT, M. (2010): *Statistik und Forschungsmethoden*. Weinheim: Beltz
- ESSER, H. (2006): *Migration, Sprache und Integration*, AKI-Forschungsbilanz 4, Berlin
- 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
- FEHM, L. & FYDRICH, T. (2011): *Prüfungsangst*. Göttingen: Hogrefe
- FIEGUTH, G. (1977): Die Entwicklung eines kategorialen Beobachtungsschemas. In: U. MEES & H. SELG (Hrsg.): *Verhaltensbeobachtung und Verhaltensmodifikation – Anwendungsmöglichkeiten im pädagogischen Bereich* (S. 33-42). Stuttgart: Klett
- FIELD, A. P. (2009): *Discovering statistics using SPSS: 3. Auflage*, London: Sage publications
- FISCHER, P. M. & MANDL, H. (1988): Improvement of the acquisition of knowledge by informing feedback. In: H. MANDL & A. LESGOLD (Eds.): *Learning issues for intelligent tutoring systems* (pp. 187-241). New York, Berlin, Heidelberg: Springer.
- FLEISS, J. L. (1971): Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5 pp. 378-382
- FRIEDEL, T., MÖRL, S. & RÜDEL, M. (2012): *Bericht zur Machbarkeitsstudie für die Entwicklung und Ersterprobung eines elektronischen Prüfprotokolls zur Optimierung der Praktischen Fahrerlaubnisprüfung*. Dresden: TÜV|DEKRA arge tp 21
- FRIEDRICH, J. (1973): *Methoden empirischer Sozialforschung*. Reinbeck: Rowohlt
- FRITZ, J. (2011): *Wie Computerspieler ins Spiel kommen. Theorien und Modelle zur Nutzung und Wirkung virtueller Spielwelten*. Schriftenreihe Medienforschung der Landesanstalt für Medien NRW (LfM), Band 67. Berlin
- GENSCHOW, J. & STURZBECHER, D. (2015): *Verkehrswahrnehmungstests als innovative Prüfungsform in der Fahranfängervorbereitung*. In: TÜV|DEKRA arge tp 21 (Hrsg.): *Verkehrswahrnehmung und Gefahrenvermeidung – Grundlagen und Umsetzungsmöglichkeiten in der Fahranfängervorbereitung*. Innovationsbericht zum Fahrerlaubnisprüfungssystem 2011-2014. Dresden: TÜV|DEKRA arge tp 21
- GENSCHOW, J., STURZBECHER, D. & WILLMES-LENZ, G. (2013): *Fahranfängervorbereitung im internationalen Vergleich. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit*. Bremerhaven: Wirtschaftsverlag NW
- GILMAN, D. A. (1969): Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology*, 60 (6), 503-508
- 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
- HAMPEL, B. (1977): *Möglichkeiten zur Standardisierung der Praktischen Fahrprüfung: Bericht zum Forschungsauftrag 7516 der Bundesanstalt*

- für Straßenwesen. Köln: Technischer Überwachungs-Verein Rheinland e. V
- HAMPEL, B. & KÜPPERS, F. (1982): Ermittlung der an Fahrprüfungsorte zu stellenden Anforderung. Bericht zum FA 7516 der Bundesanstalt für Straßenwesen. Köln: TÜV Rheinland
- 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 (S. 121-176). Bonn: Kirschbaum
- HANCOCK, T. E., THURMAN, R. A. & HUBBARD, D. C. (1995): An expanded control model for the use of instructional feedback. *Contemporary Educational Psychology*, 20, 410-425
- HASEMANN, K. (1983): Verhaltensbeobachtung und Ratingverfahren. In: K. J. Groffmann & L. MICHEL (Hrsg.): *Enzyklopädie der Psychologie (Serie II: Bd. 4)* (S. 434-488). Göttingen: Hogrefe
- HOFTE-FANKHAUSER, K. & WÄLTJ, H. F. (2011): *Marktforschung: Grundlagen mit zahlreichen Beispielen, Repetitionsfragen mit Antworten und Glossar*. Compendio Bildungsmedien, Zürich
- HOLBROOK, A. L., KROSNICK, J. A. & PFENT, A. (2008): The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms: 499-528. In: LEPKOWSKI, J. M., TUCKER, C., BRICK, J. M. et al. (Hrsg.): *Advances in Telephone Survey Methodology*. John Wiley & Sons, New Jersey
- HOLTZ, R. F. v. (1998): *Der Zusammenhang zwischen Mitarbeiterzufriedenheit und Kundenzufriedenheit*. München: FGM-Verlag
- HOMBURG, C. (Hrsg.) (2008): *Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen*. Wiesbaden: Gabler
- HORNE, J. A. & REYNER, L. A. (1995): Sleep-related vehicle accidents, *British Medicine Journal* 310 (6979); 565-567
- INGENKAMP, K.-H. & LISSMANN, U. (2008): *Lehrbuch der pädagogischen Diagnostik*. Weinheim: Beltz
- JACOBS, B. (2002): Aufgaben stellen und Feedback geben. Verfügbar unter: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/projekt.htm> [April, 2015]
- JACOBY, J., MAZURSKY, D., TROUTMAN, T. & KUSS, A. (1984): When Feedback is Ignored – Disutility of Outcome Feedback. In: *Journal of Applied Psychology*, 69 (3), S. 531-545.
- JUNG, H. (2003): *Personalwirtschaft*. 5., überarbeitete und erweiterte Auflage. München: Oldenbourg
- KIM, J.-O., (1975): Multivariate analysis of ordinal variables, *American Journal of Sociology* 81, 261-298
- KLUGER, A. N. & DeNISI, A. (1996): The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 199, 254-284
- KNIEBEL, T. (2012): *Feedbacks. Die Bedeutung von Rückmeldungen und Regeln zum konstruktiven Geben und (An)Nehmen*. Vodcast
- KRAUS, U.-M. (2007): *Feedback und kooperatives Lernen*. Münster: Waxmann
- KRAUSE, U.-M., STARK, R. & MANDL, H. (2003): *Förderung des computerbasierten Wissenserwerbs im Bereich empirischer Forschungsmethoden durch kooperatives Lernen und eine Feedbackmaßnahme (Forschungsbericht Nr. 160)*. München: Ludwig-Maximilians-Universität, Institut für Pädagogische Psychologie
- KRIPPENDORFF, K. (2011): Computing Krippendorff's Alpha-Reliability. Abgerufen am 19.08.2015 von [http://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc\\_papers](http://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc_papers)
- KULHAVY, R. W. (1977): Feedback in written instruction. *Review of Educational Research*, 47(1), 211-232.
- KULHAVY, R.W & STOCK, W. A. (1989): Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1 (4), 279-308
- KULHAVY, R. W., WHITE, M. T., TOPP, CHAN, A. L. & ADAMS, J. (1985): Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*, 10, 285-291



- LABOVITZ, S. (1967): Some observations on measurement and statistics. *Social Forces*, 46, 151-160
- LIENERT, G. A. & RAATZ, U. (1998): Testaufbau und Testanalyse. Weinheim: Beltz/PVU
- LOHSE, H., LUDWIG, R. & RÖHR, M. (1986): Statistische Verfahren für Psychologen, Pädagogen und Soziologen. Berlin: Verlag Volk und Wissen
- MAURER, M. & JANDURA, O. (2009): Masse statt Klasse? Einige kritische Anmerkungen zu Repräsentativität und Validität von Online-Befragungen. In: N. JACKOB, H. SCHOEN, & T. ZERBACK (Hrsg.): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. (S. 61-73). Wiesbaden: VS Verlag für Sozialwissenschaften
- MAYCOCK, G., LOCKWOOD, C. & LESTER, F. (1991): The accident liability of car drivers (TRRL RR 315, UMTRI-82404). Crowthorne: Transport and Road Research Laboratory
- McKNIGHT, A. J. & ADAMS, B. B. (1970): Driver Education Task Analysis. Volume I: Task Descriptions. 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. & ADAMS, A. G. (1971): Driver Education Task Analysis. Volume IV: The development of instructional objectives. Final Report. Alexandria, VA: Human Resources Research Organization
- MEES, U. (1977): Einführung in die systematische Verhaltensbeobachtung. In: U. MEES & H. SELG (Hrsg.): *Verhaltensbeobachtung und Verhaltensmodifikation – Anwendungsmöglichkeiten im pädagogischen Bereich* (S. 14-32). Stuttgart: Klett
- MEFFERT, H. & BRUHN, M. (2009): Dienstleistungsmarketing. Grundlagen – Konzepte – Methoden. Mit Fallstudien. Wiesbaden: Gabler
- MORY, E. H. (1996): Feedback research. In: D. H. JONASSEN (Ed.): *Handbook of research for educational communications and technology* (pp. 919-956). New York: Macmillan.OECD (2006). *Young Drivers: The Road to Safety*. Policy Brief, October, 2006, OECD Observer
- MÜLLER, K., DAVID, J., & STRAATMANN, T. (2011): Qualitative Beobachtungsverfahren. In: G. NADERER, E. & BALZER (Hrsg.): *Qualitative Marktforschung in Theorie und Praxis: Grundlagen, Methoden und Anwendungen*. 2. Auflage. Wiesbaden: Gabler
- NARCISS, S. (2006): Informatives tutorielles Feedback. Münster: Waxmann.
- PILLER, F. (2006): Kundenintegration im Innovationsprozess als Schlüssel zur Kundenzufriedenheit. In: C. HOMBURG (Hrsg.): *Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen* (S. 431-460). Wiesbaden: Gabler
- PRIDEMORE, D. A & KLEIN, J. D. (1991): Control of feedback in computer-assisted instruction. *Educational Technology, Research and Development*, 39 (4), 27-32
- 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 M 71. Unveröffentlichtes Manuskript. Flensburg: Kraftfahrt-Bundesamt
- SCHARNBACHER, K.; KIEFER, G. (1998): Kundenzufriedenheit: Analyse, Messbarkeit und Zertifizierung, 2. Auflage, Oldenburg.
- SCHNEIDER, W. & KORNMEIER, M. (2006): Kundenzufriedenheit: Konzept, Messung, Management. Bern: Haupt
- SCHNELL, R., HILL, P. B. & ESSER, E. (2005): *Methoden der empirischen Sozialforschung*, 7. Auflage, München (Oldenbourg)
- SCHULER, H. (2001): *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe

- SCHWENKMEZGER, P. & HANK, P. (1993): Papier-Bleistift versus computerunterstützte Darbietung von State-Trait-Fragebogen: eine Äquivalenzuntersuchung. *Diagnostica*, 39 (3), 189-210
- SINGER, E. & PRESSER, S. (2008): Privacy, confidentiality, and respondent burden as factors in telephone survey nonresponse. In: J. M. LEPKOWSKI, C. TUCKER, J. M. BRICK, E. D. DeLEEUW, L. JAPEC, P. J. LAVRAKAS, M. W. LINK & R. L. SANGSTER (Hrsg.): *Advances in telephone survey methodology* (S. 449-470). New York: Wiley
- SPRUNG, L. & SPRUNG, H. (1984): *Grundlagen der Methodologie und Methodik der Psychologie*. Berlin: Deutscher Verlag der Wissenschaften
- Statistisches Bundesamt (2014): *Verkehrsunfälle – Unfälle von 18- bis 24-Jährigen im Straßenverkehr*. Wiesbaden: Statistisches Bundesamt
- STURZBECHER, D. (Hrsg.) (2004): *Manual für die pädagogisch qualifizierte Fahrschulüberwachung*. Potsdam: Universität
- STURZBECHER, D. & MÖRL, S. (2008): *Methodensystem zur Erfassung der Zufriedenheit mit der Fahrerlaubnisprüfung*. Dresden: TÜV | DEKRA arge tp 21
- STURZBECHER, D., SCHELLHAS, B. & HERMANN, U. (2004): Beobachtungsinventar und Checkliste zur Erfassung der theoretischen Fahrschulbildung. In: D. STURZBECHER (Hrsg.): *Manual für die pädagogisch qualifizierte Fahrschulüberwachung (Modul 2)*. Potsdam: Universität
- 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., 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 (S. 70-144). Wirtschaftsverlag NW
- STURZBECHER, D., KASPER, D., BÖNNINGER, J. & RÜDEL, M. (2008): *Evaluation der theoretischen Fahrerlaubnisprüfung – Methodische Konzeption und Ergebnisse des Revisionsprojekts*. Dresden: TÜV | DEKRA arge tp 21
- STURZBECHER, D., MÖRL, S. & KALTENBAEK, J., (2014): *Optimierung der Praktischen Fahrerlaubnisprüfung*. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit. Bremerhaven: Wirtschaftsverlag NW
- STURZBECHER, D., KLEEBERG-NIEPAGE, A. & HOFFMANN, L. (Hrsg.) (2011): *Aufschwung Ost? Lebensperspektiven und Wertorientierungen ostdeutscher Jugendlicher*. Wiesbaden: VS-Verlag
- SUMMALA, H. (1987): *Young Driver Accidents: Risk Taking or Failure of Skills? Alcohol, Drugs, and Driving*, 3, 3-4, 79-91
- 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* (S. 13-25). Amsterdam: Elsevier
- TADDICKEN, M. (2013): *Online-Befragung*. In: MÖHRING, WIEBKE & SCHLÜTZ, D. (Hrsg.): *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden, S. 201-217
- THEOBALD, A. (2000): *Das World Wide Web als Befragungsinstrument*. Wiesbaden: Gabler
- TREIBLMAIER, H. (2011): *Datenqualität und Validität bei Online-Befragungen*. *der Markt – Journal für Marketing* 50 (1): 3-18
- VASILYEVA, E., PECHENIZKIY, M. & BRA, P. D. (2008): *Adaptation of elaborated feedback in e-learning*. *Lecture Notes in Computer Science*, 5149, 235-244
- VORDERER, P. & KLIMMT, C. (2006): *Rennspiele am Computer: Implikationen für die Verkehrssicherheitsarbeit*. Zum Einfluss von Computerspielen mit Fahrzeugbezug auf das Fahrverhalten junger Fahrer. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit. Bremerhaven: Wirtschaftsverlag NW

- WEIDENMANN, B. (1998): Erfolgreiche Kurse und Seminare – Professionelles Lernen mit Erwachsenen. Weinheim, Basel: Beltz
- WESTERMANN, R. (2002): Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. In: Zeitschrift für Psychologie, 210 (1), 4-26
- WIRTZ, M. & CASPAR, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe
- WOTTAWA, H. & THIERAU, H. (2003). Lehrbuch Evaluation. Bern: Hans Huber

## Schriftenreihe

### Berichte der Bundesanstalt für Straßenwesen

#### Unterreihe „Mensch und Sicherheit“

## 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, Rüdell 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 Fahrleistung 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, Leipzig € 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 punktauffällige Fahrer – Konzeptgrundlagen des Fahreignungsseminars  
Glitsch, Bornwasser, 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 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 alkoholauffällige 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, Mehli, 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 € 14,00

M 259: Alkoholkonsum und Verkehrsunfallgefahren bei Jugendlichen  
Hoppe, Tekaat € 16,50

M 260: Leistungen des Rettungsdienstes 2012/13  
Schmiedel, Behrendt € 16,50

M 261: Stand der Radfahrausbildung an Schulen und motorische Voraussetzungen bei Kindern  
Günther, Kraft € 18,50

M 262: Qualität in Fahreignungsberatung und fahreignungsfördernden Maßnahmen  
Klipp, Bischof, Born, DeVol, Dreyer, Ehlert, Hofstätter, Kalwitzki, Schattschneider, Veltgens € 13,50

M 263: Nachweis alkoholbedingter Leistungsveränderungen mit einer Fahrverhaltensprobe im Fahrsimulator der BAST  
Schumacher  
Dieser Bericht liegt nur in digitaler Form vor und kann unter <http://bast.opus.hbz-nrw.de/> heruntergeladen werden.

## 2016

M 264: Verkehrssicherheit von Radfahrern – Analyse sicherheitsrelevanter Motive, Einstellungen und Verhaltensweisen von Below € 17,50

M 265: Legalbewährung verkehrsauffälliger Kraftfahrer nach Neuerteilung der Fahrerlaubnis  
Kühne, Hundertmark € 15,00

M 266: Die Wirkung von Verkehrssicherheitsbotschaften im Fahrsimulator – eine Machbarkeitsstudie  
Wandtner  
Dieser Bericht liegt nur in digitaler Form vor und kann unter <http://bast.opus.hbz-nrw.de/> heruntergeladen werden.

267: Wahrnehmungspsychologische Analyse der Radfahraufgabe  
Platho, Paulenz, Kolrep in Vorbereitung

268: Revision zur optimierten Praktischen Fahrerlaubnisprüfung  
Sturzbecher, Luniak, Mörl € 20,50

Fordern Sie auch unser kostenloses Gesamtverzeichnis aller lieferbaren Titel an! Dieses sowie alle Titel der Schriftenreihe können Sie unter der folgenden Adresse bestellen:

Fachverlag NW in der Carl Schünemann Verlag GmbH  
Zweite Schlachtpforte 7 · 28195 Bremen  
Tel. +(0)421/3 69 03-53 · Fax +(0)421/3 69 03-63

Alternativ können Sie alle lieferbaren Titel auch auf unserer Website finden und bestellen.

[www.schuenemann-verlag.de](http://www.schuenemann-verlag.de)