

VOM FAHRSIMULATOR BIS ZU NATURALISTIC DRIVING STUDIES

TEIL 1

WISSENSCHAFTLICHE QUALITÄTSKRITERIEN FÜR FAHRVERHALTENSBEOBSACHTUNGEN.

Autor: Dr. Martin Gründl, Ergoneers GmbH

Whitepaper, 02.2016



ERGONEERS
FROM SCIENCE TO INNOVATION

ABSTRACT

In den letzten Jahren lässt sich ein vermehrtes Interesse an alltagsnahen Fahrverhaltensstudien beobachten, die nicht durch zu starke experimentelle Vorgaben beeinflusst sind. Dazu werden zunehmend so genannte *Naturalistic Driving Studies (NDS)* durchgeführt, die das Verhalten des Fahrers in einer natürlichen Fahrumgebung erfassen sollen. Zusätzlich gibt es stärker kontrollierte experimentelle Realfahrzeugstudien (= experimentelle Feldstudien, *experimental Field Operational Tests, eFOT*), deren Zweck in der Regel die Untersuchung der Effekte von Fahrerassistenz- oder Fahrerinformationssystemen ist. Die häufigste Methode der Fahrverhaltensbeobachtung ist jedoch nach wie vor das Experiment im Fahrsimulator.

Doch egal für welche Methode zur Fahrverhaltensbeobachtung man sich entscheidet – am Ende müssen die Ergebnisse jedoch immer denselben wissenschaftlichen Gütekriterien standhalten, nämlich *Objektivität, Reliabilität, interne Validität, Inhaltsvalidität, externe Validität* und *Replizierbarkeit*. Werden diese Qualitätsstandards nicht ausreichend berücksichtigt, ist eine Studie in ihrer Aussagekraft eingeschränkt – im Extremfall sogar wertlos. In diesem Teil 1 des Whitepapers zu den Methoden der Fahrverhaltensbeobachtung werden daher die wichtigsten Qualitätsstandards einer wissenschaftlichen Untersuchung beschrieben und mit Beispielen aus der Praxis der Fahrverhaltensbeobachtung illustriert.

INHALT

ABSTRACT	2
1. ÜBERBLICK WISSENSCHAFTLICHER GÜTEKRITERIEN.....	4
2. OBJEKTIVITÄT	5
3. RELIABILITÄT	5
4. INTERNE VALIDITÄT	6
5. INHALTSVALIDITÄT	7
6. EXTERNE VALIDITÄT	8
7. REPLIZIERBARKEIT	9
8. FAZIT.....	10
9. LITERATUR.....	11

1. ÜBERBLICK WISSENSCHAFTLICHER GÜTEKRITERIEN

Bevor neue Assistenz- und Informationssysteme Einzug ins Auto finden, müssen Automobilhersteller aufgrund einer Selbstverpflichtung (Alliance of Automobile Manufacturers, 2006) durch Studien nachweisen, dass diese neuen Systeme sicher sind. Das heißt vor allem, dass sie auch in seltenen, kritischen Situationen korrekt funktionieren, dass der Fahrer sie jederzeit unter Kontrolle halten kann und er durch ihre Benutzung nicht zu stark von der Fahraufgabe abgelenkt wird. Für die Durchführung solcher Untersuchungen gibt es unterschiedliche Methoden, die sich vor allem in ihrem technischen Aufwand, in ihrer Realitätsnähe und im Ausmaß ihrer methodischen Kontrolle unterscheiden (Näheres dazu siehe [Teil 2](#)).

Bei diesen Evaluierungsstudien lässt sich in letzter Zeit ein Trend beobachten: Weg von standardisierten Laborexperimenten mit einem kontrollierten Versuchsdesign und hin zur Erhebung von Fahrdaten im realen Straßenverkehr unter möglichst alltagsnahen Bedingungen (Lietz et al., 2001). Dazu werden zunehmend so genannte *Naturalistic Driving Studies (NDS)* durchgeführt, die das Verhalten des Fahrers in einer natürlichen Fahrumgebung erfassen sollen. Hinzu kommen zahlreiche experimentelle Realfahrzeugstudien (= experimentelle Feldstudien, *experimental Field Operational Tests, eFOT*), in deren Mittelpunkt in der Regel die Untersuchung der Effekte eines oder mehrerer Fahrerassistenz- oder -informationssysteme steht. Auch hier wird das Fahrverhalten in einer natürlichen Fahrumgebung beleuchtet, und es werden Unterschiede zwischen dem Fahren mit und ohne Assistenzsystem (bzw. Informationssystem) herausgearbeitet. Trotz dieses Trends werden jedoch nach wie vor immer noch die meisten Evaluationsstudien zu Fahrerassistenz- und Fahrerinformationssystemen im Fahrsimulator unter Laborbedingungen durchgeführt.

Egal für welche Methode zur Fahrverhaltensbeobachtung man sich entscheidet – am Ende müssen die Ergebnisse jedoch immer denselben wissenschaftlichen Gütekriterien standhalten, nämlich

- Objektivität
- Reliabilität
- Interne Validität
- Inhaltsvalidität
- Externe Validität
- Replizierbarkeit

Werden diese Qualitätsstandards nicht ausreichend berücksichtigt, ist eine Studie in ihrer Aussagekraft eingeschränkt – im schlimmsten Fall sogar völlig wertlos. In diesem Teil 1 des Whitepapers zu den Methoden der Fahrverhaltensbeobachtung werden daher die wichtigsten Qualitätsstandards einer wissenschaftlichen Untersuchung kurz beschrieben und mit Beispielen aus der Praxis der Fahrverhaltensbeobachtung illustriert. In [Teil 2](#) werden die verschiedenen Methoden vorgestellt und ihre Vor- und Nachteile anhand der hier beschriebenen Gütekriterien erläutert.

2. OBJEKTIVITÄT

Objektivität ist bei einer Methode dann gegeben, wenn die mit dieser Methode erzielten Ergebnisse nicht davon abhängen, *wer* die Methode anwendet (Klein et al., 2012). Gibt es beispielsweise zwei verschiedene Versuchsleiter, dann muss sichergestellt werden, dass beide sich gegenüber den Versuchspersonen gleich verhalten und z. B. identische Instruktionen geben. Ansonsten könnte es sein, dass durch die Art, wie ein Versuchsleiter die Aufgabenstellung erklärt, manche Probanden indirekt nützliche Tipps bekommen und andere nicht. Um eine möglichst hohe *Durchführungsobjektivität* zu gewährleisten, ist es daher empfehlenswert, Instruktionen schriftlich zu geben. Bei einer Realfahrt kann beispielsweise die zu fahrende Strecke von einem Navigationssystem angesagt werden, da somit Inhalt und Timing der Anweisungen für alle Probanden immer exakt gleich sind.

Bei der Analyse der aufgenommenen Daten muss zudem *Auswertungsobjektivität* gegeben sein. Wird beispielsweise ausgewertet, ob der Fahrer einen Fehler begangen hat, darf das Ergebnis nicht davon abhängen, welcher Auswerter die Beurteilung vornimmt. Ob z. B. der Fahrer vor einer Rechts-vor-Links-Kreuzung stark genug abgebremst hat, ob ein Schulterblick stattgefunden hat oder nicht, muss unabhängig vom jeweiligen Beurteiler sein. Entsprechend präzise müssen daher auch die Regeln zur Codierung des beobachteten Verhaltens sein, um *Objektivität* sicherzustellen.

Ein anderes Beispiel, bei dem die *Auswertungsobjektivität* kritisch hinterfragt werden muss, ist die Messung des Fahrerblickverhaltens *ohne* Blickerfassungssystem (= Eye-Tracker). Dabei wird mit einer Kamera im Auto lediglich das Gesicht des Fahrers gefilmt. Hinterher muss dann ein Beurteiler (= *Rater*) bei der Analyse des Videos einschätzen und codieren, wo der Fahrer zu welchem Zeitpunkt hingesehen hat (z. B. auf die Straße, Display eines Systems, Tacho, Rückspiegel, Außenspiegel). Unterschiedliche Beurteiler kommen dabei häufig zu ganz unterschiedlichen Resultaten. Man kann das Ausmaß der Urteilsübereinstimmung zwischen zwei Urteilern leicht berechnen (= *Interrater-Reliabilität*). Bei Blickzuwendungen zu bestimmten Bereichen im Auto (= *Areas of Interest, AOs*) handelt es um eine Übereinstimmung bei kategorialen Daten. Als statistisches Maß verwendet man hierfür das sogenannte *Cohen's Kappa* (Cohen, 1960; Landis & Koch, 1977; Bortz, Lienert & Boehnke, 1990), das die Urteilerübereinstimmung als Zahl zwischen 0 und 1 ausgibt. Blickzuwendungen auf *Areas of Interest*, die lediglich mit einem einfachen Video (also ohne Eye-Tracker) gemessen wurden, sind ein Spezialfall von videobasierten Verhaltenscodierungen. Anders als beim Einsatz eines Eye-Trackers schwebt dabei immer eine gehörige Portion Skepsis mit, wie objektiv eigentlich solche Ergebnisse sind und ob diese nicht stark vom Auswerter abhängen. Durch Berechnung einer Urteilerübereinstimmung ließen sich solche Zweifel zerstreuen. Dennoch fehlen sehr vielen Publikationen zur Fahrerhaltensbeobachtung diese statistischen Kennwerte zur *Objektivität* des verwendeten Messverfahrens.

3. RELIABILITÄT

Die *Reliabilität* (= Zuverlässigkeit) gibt den Grad der Messgenauigkeit (= Präzision) einer Messmethode an (Bortz & Döring, 2006). Jede Messung ist auch immer mit einem Messfehler behaftet. Je kleiner dieser Messfehler, desto höher ist die *Reliabilität*. Das Angenehme an der *Reliabilität* ist, dass man sie – anders als *interne* oder *externe Validität* – quantifizieren kann. Denn wenn man beispielsweise zweimal dasselbe misst, muss auch zweimal dasselbe Ergebnis rauskommen. Durchfährt also eine Gruppe von Probanden zweimal ein identisches Verkehrsszenario unter identischen Voraussetzungen, dann müssen die gemessenen Fahrkennwerte auch in beiden Fällen etwa gleich sein. Sind sie

deutlich verschiedenen, stimmt etwas nicht. Wie gut die Daten in den beiden Bedingungen übereinstimmen, lässt sich mit statistischen Verfahren berechnen, je nach Art der Daten beispielsweise mit einer Korrelation. Obwohl sich die *Reliabilität* eines Messverfahrens relativ leicht berechnen lässt – Fachbegriffe sind hier *Retest-Reliabilität*, *Paralleltest-Reliabilität* und *Split-Half-Reliabilität* – wird es bei Fahrverhaltensbeobachtungen viel zu selten gemacht.

4. INTERNE VALIDITÄT

Validität bedeutet allgemein Gültigkeit. Sie gibt an, ob ein Messverfahren das misst, was es messen soll. Die *interne Validität* betrifft die Frage, wie gut es in einer Untersuchung gelingt, bedeutsame Störvariablen zu kontrollieren. Eine Untersuchung ist also dann *intern valide*, wenn ihre Ergebnisse eindeutig interpretierbar sind (Campbell & Stanley, 1963; Cook & Campbell, 1979).

Bei Fahrerverhaltensmessungen sind beispielsweise wichtige Störvariablen das Verhalten anderer Verkehrsteilnehmer, Witterungs- oder Sichtverhältnisse. Möchte man z. B. messen, wie sich ein bestimmtes Assistenzsystem im Auto auf den eingehaltenen Sicherheitsabstand zum Vordermann auswirkt, und findet man Unterschiede in den beiden Versuchsbedingungen „mit System“ und „ohne System“, dann sollten die Unterschiede auch eindeutig auf die Versuchsbedingungen zurückzuführen sein. Hat jedoch in einer Bedingung ein geringerer Sicherheitsabstand in Wirklichkeit seine Ursache in einem dichteren Verkehrsaufkommen, dann ist die Messung *intern nicht valide*. Probleme mit geringer *interner Validität* aufgrund zahlreicher Störvariablen sind v. a. für Studien im realen Straßenverkehr typisch. Im Fahrsimulator hingegen hat man die Möglichkeit, Fremdverkehr und Umweltbedingungen zu kontrollieren und konstant zu halten.



Abbildung 1: Ein Verrutschen der Blickerfassungsbrille während einer längeren Fahrt kann die Blickmessungen der gesamten Fahrt unbrauchbar machen, da im Blickvideo dadurch nicht mehr die tatsächlichen Blicke angezeigt werden. Ein Eye-Tracker mit Möglichkeit zur Nachkalibrierung (hier Dikablis von Ergoneers) kann diesen Fehler jedoch korrigieren und erhöht dadurch die interne Validität gemessener Blickdaten.

Auch bei der Verwendung eines Eye-Trackers für Blickverhaltensmessungen im Auto ist *interne Validität* ein Thema. Wird damit auch tatsächlich gemessen, wohin eine Person schaut, oder sind die Ergebnisse teilweise durch Störeinflüsse verfälscht? Ein typischer Fehler ist beispielsweise, dass während einer längeren Fahrt die Blickerfassungsbrille etwas nach unten verrutscht. Dadurch stimmt die anfängliche Kalibrierung nicht mehr und im Blickvideo werden die Blicke des Probanden falsch angezeigt. Verwendet man einen Eye-Tracker, der bei der Blickdatenanalyse die Möglichkeit einer Nachkalibrierung bietet, kann dieser Fehler behoben werden. Ohne diese Korrekturmöglichkeit hingegen sind ab dem Verrutschen der Brille alle gemessenen Blicke falsch und nicht mehr *intern valide*.

5. INHALTSVALIDITÄT

Inhaltsvalidität (auch *Face Validity*, Augenscheinvalidität oder Logische Validität) ist gegeben, wenn das, was man eigentlich messen möchte, von der Messmethode in seinen wichtigsten Aspekten erschöpfend erfasst wird (Bortz & Döring, 2006; Rosnow & Rosenthal, 2007).

Beispiel: Ein Spurhalteassistent ist ein Assistenzsystem, das dafür gemacht ist, Unfälle durch versehentliches Abkommen von der Fahrbahn zu verhindern. Kurz bevor ein Fahrer die Fahrbahnmarkierung überfahren würde (gemessen als sogenannte *Time to Line Crossing, TLC*), warnt das System den Fahrer oder greift mit einem kurzen, schwachen Lenkimpuls in das Fahren ein. Durch eine frühe Warnung, also das Einstellen einer langen *TLC*, kann ein Fahrer vom System dazu angehalten werden, frühzeitig kleine Lenkkorrekturen durchzuführen und so möglichst in der Mitte seiner Fahrspur zu fahren. Fährt ein Fahrer mit diesem Assistenzsystem, wird er schnell lernen so zu fahren, dass er möglichst keine Warnungen oder Lenkeingriffe bekommt, er wird also brav in der Mitte seiner Spur bleiben. Ohne Assistenzsystem dagegen fährt er wie immer, kommt also auch gelegentlich in die Nähe der Fahrbahnmarkierung (unterschreitet also die als kritisch definierte *TLC*). Vergleicht man nun statistisch die beiden Versuchsbedingungen miteinander, wird man feststellen, dass Fahrer mit Spurhalteassistent mehr in der Mitte ihrer Fahrspur fahren und seltener die als kritisch definierte *Time to Line Crossing* unterschreiten als Fahrer ohne System. Die Schlussfolgerung, dass sie deswegen auch sicherer fahren und ein geringeres Risiko haben, vor der Fahrbahn abzukommen, ist aber trotzdem nicht zwingend. Denn ein Unterschreiten einer *TLC* von z. B. 1,4 Sekunden ist noch lange nicht gleichbedeutend mit einem Beinahe-Abkommen von der Fahrbahn, und das perfekte Fahren in der Fahrbahnmitte erst recht nicht.

Der Operationalisierung fehlt hier es also an *Inhaltsvalidität*: Man will eigentlich das Risiko eines Unfalls durch Abkommen von der Fahrbahn messen, misst aber in Wirklichkeit etwas Anderes, nämlich wie genau und vorbildlich ein Fahrer in der Mitte der Fahrspur fahren kann bzw. fahren will. Die *Inhaltsvalidität* wäre größer, würde man in beiden Versuchsbedingungen nur diejenigen Ereignisse zählen, in denen der Fahrer *tatsächlich* die Fahrbahnmarkierung überfährt. Dies wäre für eine Abschätzung des Unfallvermeidungspotenzials eines solchen Systems viel aussagekräftiger. Doch man tut es in der Regel nicht, weil solche Ereignisse beim normalen Autofahren so selten sind, dass man sie statistisch schlecht auswerten kann. Denn derartige Fahrversuche würden ein Vielfaches an Zeit, Probanden und damit Kosten bedeuten.

6. EXTERNE VALIDITÄT

Eine Untersuchung ist *extern valide*, wenn ihre Ergebnisse über die besonderen Bedingungen der Untersuchungssituation hinaus generalisierbar sind. Man unterscheidet hierbei zwei Formen der Generalisierbarkeit, nämlich eine Generalisierung über *Personen* und *Situationen* (Hager & Westermann, 1983; Cook & Shadish, 1994).

Sollen die Ergebnisse über die konkret in der Studie untersuchten *Personen* hinaus verallgemeinert werden, muss die untersuchte Probandenstichprobe repräsentativ sein. Hat man beispielsweise wegen der leichteren Verfügbarkeit nur Studierende (= junge Fahrer) oder nur Fahrzeugingenieure eines Automobilherstellers (= Technik-Experten) untersucht, lassen sich die Ergebnisse kaum auf andere Populationen übertragen, z. B. auf ältere Fahrer (= typische Käufer von Neufahrzeugen der Oberklasse). Generell gilt: Je geringer die Repräsentativität der untersuchten Probanden-Stichprobe, desto geringer ist die *externe Validität*.

Sollen Ergebnisse über die konkret untersuchte *Situation* hinaus verallgemeinerbar sein, setzt dies beispielsweise voraus, dass sich Menschen in einem Fahrsimulator tatsächlich genauso verhalten wie in einem echten Auto im echten Straßenverkehr. Ebenso setzt es voraus, dass sie sich in einem hoch standardisierten Experiment in einem Simulator oder bei einer Realfahrtstudie mit einem mitfahrenden Versuchsleiter im Auto genauso verhalten, wie sie es tun, wenn sie unbeobachtet privat mit dem Auto fahren. Dies sind sehr starke Annahmen und abgesehen von *Naturalistic Driving Studies* sind sie bei typischen Studien zur Fahrverhaltensbeobachtung auch kaum erfüllt. Generell gilt: Je unnatürlicher die Untersuchungsbedingungen, desto geringer ist die *externe Validität*.

Wenn die Beobachtung oder Messung das zu beobachtende oder zu messende Verhalten und somit auch die erfassten Daten beeinflussen, spricht man auch von *Reaktivität*. Diese mindert die *externe Validität*. *Reaktivität* ist damit ein Negativ-Kriterium. Bei Beobachtungsstudien ist dieses Phänomen seit fast 100 Jahren unter dem Begriff *Hawthorne-Effekt* bekannt (Roethlisberger & Dickson, 1964). Dieser besagt, dass die Teilnehmer einer Studie ihr natürliches Verhalten ändern, weil sie wissen, dass sie an einer Studie teilnehmen und unter Beobachtung stehen. Sie strengen sich deswegen mehr an, bringen eine bessere Leistung und verhalten sich sozial erwünschter. Im Fahrbeobachtungsstudien heißt das: Die Probanden fahren konzentrierter, machen weniger Fahrfehler, unterlassen ablenkende Nebentätigkeiten und fahren rücksichtsvoller als sonst. All das geht zulasten der *externen Validität*.

Eine methodisch saubere Studie mit gleichzeitig hoher *externer Validität* durchzuführen, ist schwer. Denn es gibt es einen Zielkonflikt bei der Optimierung von *interner* und *externer Validität*: Optimiert man nämlich die *externe Validität*, wirkt sich dies meist nachteilig auf die *interne Validität* aus (und umgekehrt), so dass man sich in der Regel immer mit einem Kompromiss zufriedengeben muss (Bortz & Döring, 2006).

7. REPLIZIERBARKEIT

Eine Replikation ist eine Wiederholung einer Untersuchung unter identischen oder sehr ähnlichen Bedingungen. Wiederholt man beispielsweise eine Studie zum Effekt eines Assistenzsystems auf das Fahrverhalten und kommen dabei dieselben oder sehr ähnliche Ergebnisse heraus wie beim ersten Mal, dann sind die Schlussfolgerungen aus der Studie deutlich besser abgesichert als ohne die Wiederholung. Gelingt es jedoch trotz gleicher Voraussetzungen nicht, die Ergebnisse zu reproduzieren, dann ist die Studie *nicht replizierbar*. Dies ist ein Hinweis darauf, dass in der Durchführung oder Auswertung der Studien – also auch schon beim ersten Mal – irgendwelche Störvariablen das Ergebnis verfälscht haben, z. B. Mängel bei der *Objektivität*, *nicht reliable* Messungen, unkontrollierte Störvariablen wie Fremdverkehr (also mangelnde *interne Validität*) oder auch Fehler bei der statistischen Datenanalyse.

Der Grundsatz, dass wissenschaftliche Untersuchungsergebnisse reproduzierbar sein müssen, ist ein fundamentales Prinzip der Wissenschaft. Können Ergebnisse von anderen, unabhängigen Forschergruppen nicht reproduziert werden, oder kommen diese sogar zu gegenteiligen Ergebnissen, bedeutet es auf lange Sicht das Aus für die Schlussfolgerungen aus solchen Studien.

Um die *Replizierbarkeit* sozialwissenschaftlicher Studien ist es generell nicht sonderlich gut bestellt, und gerade in den letzten Jahren ist eine hitzige Debatte unter Wissenschaftlern entstanden, woran das liegt und was sich dagegen tun lässt (Francis, 2012; Pashler & Harris, 2012; Giner-Sorolla, 2012; Ioannidis, 2012). Es gibt keinen Grund zur Annahme, dass verkehrspsychologische Untersuchungen mit Fahrverhaltensbeobachtungen von diesem Problem nicht betroffen seien.

Beispiele: Bei Realfahrstudien sind die Störeinflüsse durch den nicht beeinflussbaren Fremdverkehr beispielsweise so groß, dass zufällige Einflüsse einen bedeutenden Effekt auf die Ergebnisse haben. Und bei üblichen Stichprobengrößen von 30 bis 50 Probanden kann man auch nicht annehmen, dass sich sämtliche zufällig wirkenden Störungen am Ende schon irgendwie herausmitteln. Die Ergebnisse von *Naturalistic Driving Studies* sind wegen der fehlenden Kontrolle noch schlechter reproduzierbar. Bei Experimenten im Fahrsimulator sind die Voraussetzungen zwar deutlich besser, doch auch hier zeigt sich immer wieder, dass die Ergebnisse stark vom verwendeten Fahrsimulator und den jeweiligen konkreten technischen Einstellungen abhängen. Hier wird die Replikation eines Ergebnisses durch eine unabhängige Forschergruppe schon dadurch erschwert, dass im Methodenteil von Publikationen der jeweils verwendete Fahrsimulator mit all seinen Einstellungen nicht detailliert genug beschrieben ist, um eine Wiederholung der Studie unter gleichen Bedingungen zu gewährleisten.

Echte Replikationsstudien gibt es bei Fahrverhaltensbeobachtungen sehr selten. Gleich sind in der Regel höchstens Fragestellungen oder zu testende Systeme, aber nicht der gesamte Untersuchungskontext (Makel, Plucker & Hegarty, 2012). Zudem fehlt sowohl der Anreiz, die Studien anderer Forscher identisch zu wiederholen, als auch eigene Ergebnisse kritisch zu hinterfragen (Kooze & Lakens, 2012). Gerade im Automotive-Bereich ändert sich auch der Stand der Technik sehr schnell. Und selbst wenn ein Forscher ein ernstes Interesse an der Replikation eines schon bekannten Ergebnisses z. B. zum Nutzen eines bestimmten Fahrerassistenzsystems hätte – die Automobilindustrie, die diese Studien meist finanziert, hat es nicht. Viel lieber möchte sie den neuesten Stand des weiterentwickelten Systems evaluiert haben; und damit sind die Untersuchungsbedingungen schon nicht mehr gleich und eine echte Replikation ausgeschlossen.

8. FAZIT

Die beschriebenen Kriterien sind die Qualitätsstandards für jede Art von wissenschaftlichen Untersuchungen und gelten nicht nur für Untersuchungen zum Fahrverhalten. Alle Kriterien sind wichtig. Daher darf es nicht passieren, dass ein oder mehrere Kriterien vernachlässigt werden. In der Praxis ist dies jedoch nicht leicht zu realisieren, weil die Optimierung eines Kriteriums häufig zu einer **Verschlechterung bei einem anderen Kriterium führt**. Beispielsweise führt die Maximierung der *internen* Validität häufig zu einer Reduzierung der *externen* Validität und umgekehrt.

Die verschiedenen Methoden der Fahrverhaltensbeobachtung – von *Naturalistic Driving Studies (NDS)* über experimentelle Realfahrzeugstudien bis hin zu Experimenten im Fahrsimulator und einfachen Formen simulierten Fahrens im Labor – haben hinsichtlich dieser Qualitätskriterien alle ihre typischen Stärken und Schwächen. Im **Teil 2** dieses Whitepapers werden diese Methoden näher beschrieben und ihre Vor- und Nachteile hinsichtlich der genannten wissenschaftlichen Qualitätsstandards anhand zahlreicher Beispiele aus der Praxis erläutert. Ziel ist es, dem Leser eine nützliche Entscheidungshilfe an die Hand zu geben, um für den jeweiligen Untersuchungszweck die jeweils geeignetste Methode auszuwählen.

Teil 2: Vor- und Nachteile verschiedener Methoden der Fahrverhaltensbeobachtung

9. LITERATUR

- Alliance of Automobile Manufacturers (2006). Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems, Washington, D.C.: Alliance of Automobile Manufacturers.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Campbell, N. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation*. Boston: Houghton Mifflin.
- Cook, T. D. & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, 45, 545-580.
- Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, 7, 585-594.
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, 7, 562-571.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimente. In J. Bredenkamp & H. Feger (Hrsg.), *Hypothesenprüfung*. Göttingen: Hogrefe.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645-654.
- Koole, S. & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7, 608-614.
- Landis, J. R. & Koch G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lietz, H., Petzoldt, T., Henning, M., Haupt, J., Wanielik, G., Krems, J., Mosebach, H., Schomerus, J., Baumann, M. & Noyer, U. (2011). *Methodische und technische Aspekte einer Naturalistic Driving Study*. FAT-Schriftenreihe 229, Forschungsvereinigung Automobiltechnik (FAT).
- Makel, M. C., Plucker, J. A. & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537-542.
- Pashler, H. & Harris, Ch. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531-536.
- Rosnow, R. L & Rosenthal, R. (2007). *Beginning behavioral research. A conceptual primer*. Upper Saddle River, NJ: Pearson Education.
- Roethlisberger, F. J. & Dickson, W. J. (1964). *Management and the worker*. Cambridge, Mass.: Harvard University Press.

Die Ergoneers GmbH wurde 2005 als Spin-off des Lehrstuhls für Ergonomie der Technischen Universität München gegründet. Heute ist das Unternehmen mit weltweiten Standorten in Deutschland (Hauptsitz bei München) und USA sowie zahlreichen Vertriebspartnern ein international wichtiger Partner für die Branchen Transport und Automotive, Marktforschung und Nutzerfreundlichkeit (Usability), Wissenschaft und Forschung sowie Sport und Biomechanik.

Neben der Entwicklung, Herstellung und dem Vertrieb von Mess- und Analysesystemen zur Erforschung von Verhalten und zur Optimierung der Interaktion von Mensch und Maschine bietet Ergoneers umfassende Kompetenz in allen Phasen des Studienablaufs. Zur Ergoneers-Produktpalette zählt vor allem die 360-Grad-Lösung D-LAB, eine umfassende Erfassungs- und Auswertungsplattform für Nutzer- und Verhaltensstudien, mit deren Software-Modulen sich Daten in den Bereichen Eye-Tracking, Datastream, Video, Audio, Physiologie und CAN-Bus messen und analysieren lassen. Mit dem Ergoneers-Blickerfassungssystem Dikablis liefert Ergoneers zudem die passende Hardware, um professionelles Eye-Tracking im realen oder virtuellen Umfeld zu betreiben.

Ergoneers Group
Gewerbering 16
82544 Egling
Germany

T +49.8176.99894-0
F +49.8176.99894-15

Ergoneers of North America, Inc.
111 SW 5th Ave
Suite 3150
Portland, OR 97204
USA

T +1.503.444.3430

info@ergoneers.com
www.ergoneers.com

ERGONEERS
FROM SCIENCE TO INNOVATION