

- drivers. *Transportation Research Part F: Traffic Psychology & Behaviour*, 1F, 107–121.
- [9] Lajunen, T., Parker, D. & Summala, H. (2004). The Manchester Driver Behaviour Questionnaire: A cross-cultural study. *Accident Analysis and Prevention*, 36, 231–238.
- [10] Lajunen, T. & Summala, H. (2003). Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses. *Transportation Research Part F*, 6, 97–107.
- [11] Mesken, J., Lajunen, T. & Summala, H. (2002). Interpersonal violations, speeding violations and their relation to accident involvement in Finland. *Ergonomics*, 45, 469–483.
- [12] Parker, D., Reason, J. T., Manstead, A. S. R. & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics*, 38, 10361048.
- [13] Reason, J. T. (1993). Self-report questionnaires in cognitive psychology: Have they delivered the goods? In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control* (pp. 406–423). Oxford: Clarendon.
- [14] Reason, J. T., Manstead, A., Stradling, S., Baxter, J. & Campbell, K. (1990). Errors and violations on the roads: A real distinction? *Ergonomics*, 33, 1315–1332.
- [15] Schlag, B. (1994). Fahrverhalten älterer Autofahrer/innen. In U. Tränkle (Hrsg.), *Autofahren im Alter* (S. 161–172). Köln: Verlag TÜV Rheinland; Bonn: Deutscher Psychologen Verlag.
- [16] Seidler, C. (2005). *Konzentration und Autofahren*. Unveröffentlichte Diplomarbeit, Technische Universität Dresden.
- [17] Statistisches Bundesamt (2006). Unfallgeschehen im Straßenverkehr. Presseexemplar. [http://www.destatis.de/themen/d/thm\\_verkehr.php](http://www.destatis.de/themen/d/thm_verkehr.php) [29.01.2007]
- [18] Sullman, M. J. M., Meadows, M. L. & Pajo, K. B. (2002). Aberrant driving behaviours amongst New Zealand truck drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5, 217–232.
- [19] Underwood, G., Chapman, P., Wright, S. & Crundall, D. (1999). Anger while driving. *Transportation Research Part F*, 2, 55–68.



Dipl.-Psych. PD Dr. Carmen Hagemeister ist Wissenschaftliche Assistentin an der TU Dresden. Ihr Forschungsschwerpunkt ist Leistungsdiagnostik, vor allem Konzentration. Außerdem beschäftigt sie sich mit der Wahrnehmung von Verkehrssituationen durch Verkehrsteilnehmer, und zwar sowohl unter kognitiven und emotionalen als auch sozialen Aspekten.



Dipl.-Psych. Christina Enderlein arbeitet seit dem Abschluss ihres Psychologiestudiums an der TU Dresden in einem Bildungsträger. Tätigkeitsschwerpunkte sind die Erstellung von Eignungsgutachten und die individuelle Förderung benachteiligter Jugendlicher in der Berufsvorbereitung.

#### Anschrift:

Carmen Hagemeister · Technische Universität Dresden · 01062 Dresden

[hagemeis@psychologie.tu-dresden.de](mailto:hagemeis@psychologie.tu-dresden.de)

Die Fragebögen können Sie per Mail erhalten, bitte mailen Sie an [hagemeis@psychologie.tu-dresden.de](mailto:hagemeis@psychologie.tu-dresden.de).

# Zur summativen Evaluation von Maßnahmen der Verkehrserziehung und -aufklärung\*

I. Utzmann

**Zusammenfassung:** Verkehrserziehung und -aufklärung zielen darauf ab, das (Verkehrs-)verhalten zu beeinflussen und damit die Verkehrssicherheit zu erhöhen. Die Wirksamkeit von Verkehrserziehungsmaßnahmen wird jedoch kontrovers eingeschätzt. Im folgenden Artikel werden methodische Aspekte bei Wirkungsmessungen im Bereich der Verkehrserziehung und -aufklärung beschrieben. Dabei werden die Ergebnisse eines systematischen „Reviews“, der eine Analyse und Bewertung von insgesamt 36 internationalen Studien beinhaltet, berücksichtigt.

## About summative evaluation of road safety education and information measures

**Abstract:** Road safety education and information measures aim to improve road safety by influencing traffic behaviour. There is a controversial public discussion, whether traffic safety education measures are effective or not. This article describes methodological aspects for the process of effect-measuring in the field of road safety education and campaigns. Results of a „systematic review“ are taken into account. The systematic review contains an analysis and assessment of 36 international studies.

**Dokumentation:** Utzmann, I.: Zur summativen Evaluation von Maßnahmen der Verkehrserziehung und -aufklärung. Zeitschrift für Verkehrssicherheit 54 (2008) Nr. 1, S. 25

**Schlagwörter:** Verkehrsteilnehmer (1755), Sicherheit (1665), Erziehung (1682), Werbung (1692), Marketing (0162), Bewertung/Evaluation (9020), Wirksamkeitsuntersuchung (9112), Verfahren (9102), Messung (6136)

## 1 Einleitung und Hintergrund

Verkehrserziehung und -aufklärung wird allgemein als eine der drei Säulen der Verkehrssicherheit angesehen, die sog. drei „E“ (Education, Engineering, Enforcement). Allerdings liegen zur Wirksamkeit von Maßnahmen der Verkehrserziehung und -aufklärung widersprüchliche Ergebnisse vor. Dabei ist festzustellen,

dass gerade hinsichtlich der Frage, wie eine Wirkung im Einzelnen nachgewiesen werden kann, Unsicherheit besteht (vgl. [1]). Zudem ist festzustellen, dass durchgeführte Evaluationsstudien im Bereich der Verkehrserziehung häufig den methodischen Anforderungen nicht entsprechen (vgl. [2], [3], [4]). In einzelnen Bereichen der Verkehrserziehung und -aufklärung, z. B. bei Programmen für Fußgänger oder bei Aktionen und Seminaren, wurden zudem Wirkungsmessungen bisher äußerst selten durchgeführt (vgl. [2], [5]).

Bereits durchgeführte Metastudien in diesem Bereich beziehen sich oft auf bestimmte Verkehrserziehungsmaßnahmen, wie z. B. Aufklärungsmaßnahmen für Pkw-Fahrer (z. B. [14]), oder beziehen sich auf einen früheren Stand (z. B. [13], [6]). Andere Metastudien zielen auf die Bewertung einer Maßnahme ab, und gehen dabei nur am Rande auf methodische Aspekte ein (z. B. [3]).

Der folgende Artikel fokussiert die Fragestellung, welche methodischen Schritte bei einer Wirkungsmessung im Bereich Verkehrserziehung und -aufklärung im Einzelnen durchzuführen sind und auf welche Aspekte hierbei besonders zu achten ist. Dabei werden die Ergebnisse eines systematischen Reviews, der 36 internationale Studien, die eine

\* Peer-Reviewed Article  
Eingereicht August 2007; angenommen 25.11.2007

solche Wirkungsmessung durchgeführt haben, beinhaltet, mit einbezogen.

### Einordnung der summativen Evaluation

Die summative Evaluation hat die Aufgabe, mit größtmöglicher Zuverlässigkeit festzustellen, ob eine Maßnahme die beabsichtigte Wirkung hatte bzw. ob die Ziele der Maßnahme erreicht wurden. Entscheidend ist dabei, dass hierbei externe Störfaktoren ausgegrenzt werden und die tatsächliche Wirkung gemessen wird (vgl. [6], [7]). In Ergänzung hierzu befasst sich die formative Evaluation (oder auch Prozessevaluation) mit dem Ablauf, der Organisation und Durchführung der Maßnahme, mit dem Ziel, diese zu optimieren. Ob überhaupt, in welchem Ausmaß und mit welchem Ziel eine Evaluation durchgeführt werden soll, muss in jedem Einzelfall zunächst begründet und entschieden werden.

### Schritte der Wirkungsmessung

Obwohl jede Wirkungsmessung in Abhängigkeit von der genauen Ausgestaltung der Maßnahme (Dauer, räumlicher Geltungsbereich, Zielgruppe etc.) variiert, kann diese zunächst in zehn Schritte gegliedert werden, wobei nicht alle Schritte bei jeder Wirkungsmessung gleich relevant sind und sich hinsichtlich Komplexität und Umfang unterscheiden. Die einzelnen Schritte verlaufen teilweise (zeitlich) parallel zueinander oder sind inhaltlich miteinander verknüpft. Die einzelnen Schritte sind:

1. Darstellung der Ziele der Maßnahmen und des Wirkungsmodells
2. Darstellung der Hypothesen und Ziele der Wirkungsmessung
3. Darstellung der Erhebungseinheit
4. Auswahl und Entwicklung der/des Erhebungsinstrumente/-s
5. Bestimmung der Ergebnisindikatoren
6. Auswahl und Entwicklung des/der Untersuchungsdesigns
7. Stichprobenbildung
8. Datenerhebung
9. Datenanalyse/Anwendung statistischer Verfahren
10. Dateninterpretation

Es hängt vom vorgesehenen Untersuchungsdesign ab, wann mit der summativen Evaluation bzw. Konzeption und Vorbereitung der Wirkungsmessung begonnen wird. Im Idealfall werden zumindest Überlegungen hierzu bereits während der Maßnahmenentwicklung bzw. vor Maßnahmenumsetzung angestellt. Will man einen Vorher-Nachher-Vergleich vornehmen, so kann dies beispielsweise – in Abhängigkeit der Maß-

nahmenausgestaltung, aber auch der zu betrachtenden Ergebnisindikatoren – nur vor der Maßnahmenumsetzung sinnvoll erfolgen.

### Systematischer Review

Der systematische Review fasst die Ergebnisse einzelner Studien anhand vorgegebener Kriterien deskriptiv zusammen, wobei auch empirische und methodische Stärken und Schwächen der Einzelstudien diskutiert werden. Der systematische Review ist von der Metaanalyse im engeren Sinne abzugrenzen, bei der die quantitative Zusammenfassung von Effektgrößen im Vordergrund steht (vgl. [8]).

Im Rahmen einer Dissertation an der Universität Wuppertal [9] wurde ein systematischer Review von insgesamt 36 internationalen Studien durchgeführt. Ziel war es, methodische Vorgehensweisen bei praktischen Anwendungsbeispielen im Detail nachzuvollziehen, zu analysieren und zu bewerten. Betrachtet wurden Studien, die eine summative Evaluation einer Verkehrserziehungs- und Aufklärungsmaßnahme beinhalteten. In diesem Zusammenhang wurde umfassend nach dokumentierten Wirkungsmessungen von Maßnahmen der Verkehrserziehung und -aufklärung in deutschen wie internationalen Datenbanken recherchiert (Veröffentlichungen ab 1980).

Aus den gesammelten rund 180 Studien wurden zunächst ca. 80 ausgewählt, die hinsichtlich Form und Inhalt für die Untersuchung geeignet waren. Auswahlkriterien waren:

- Es wurde tatsächlich eine summative Evaluation durchgeführt (Die erste Auswahl anhand des Titels bzw. Kurzbeschreibung der Untersuchung konnte dies nicht sicherstellen).
- Eine Beschreibung der methodischen Vorgehensweise in ausreichender Detailschärfe war gegeben (Überprüft wurde, ob die oben erwähnten methodischen Schritte in einer gewissen Ausführlichkeit dargestellt wurden).<sup>1</sup>
- Sprache: deutsch oder englisch

Um die Anzahl der Wirksamkeitsstudien weiter einzuschränken, wurden nur Veröffentlichungen ab dem Jahr 2000 einbezogen. Auf diese Weise reduzierte sich die Anzahl der Studien auf 30. Danach wurden aus den Veröffentlichungen von 1980 bis 1999 weitere sechs gezielt ausgewählt, die sich mit Maßnahmen und Zielgruppen beschäftigen, die in der bis dahin gewonnenen Auswahl nur in geringem Umfang oder gar nicht vertreten waren.

Insgesamt gingen somit 36 Studien aus 13 Ländern in die Analyse ein (siehe Bild 1), in denen die Bandbreite der Verkehrs-

erziehung und -aufklärung in Bezug auf die wesentlichen Unterscheidungsmerkmale, wie Zielgruppe, räumlicher Geltungsbereich, Dauer etc. widergespiegelt wird. Neben allgemeinen Informationen, wie z. B. Name der Maßnah-

- Polizeiliche Maßnahmen zur Erhöhung der Verkehrssicherheit von Kindern, Deutschland, Aufklärung vor Ort, teils kombiniert mit Repression
- Police driver training, UK, Seminar/Training
- „Fairkehr“, Deutschland, Kampagne ohne Personeneinsatz
- Speak out!, Norwegen, Kampagne ohne Personeneinsatz
- 55Alive, Kanada, Seminar
- Fußgängersicherheitskampagne in der Innenstadt, Neuseeland, Kampagne mit Personeneinsatz/Aktion
- Festive Drink Driving, UK, Kampagne mit Personeneinsatz
- „DRUGS“, Deutschland, Seminar
- Begleitetes Fahren, Frankreich, Reguliäre Fahrausbildung, Seminar/Fahrpraxis
- Kurs für wiederholt alkoholauffällige Fahrzeuglenker, Schweiz, Seminar
- Dummies never die, Schweiz, Kampagne mit Personeneinsatz
- Training für Mopedfahrer, Niederlande, Training
- „Safety Central“, USA; Multiplikatoren/Seminar
- „Bicycle Helmets are IN“, Kanada/ Schweden; Kampagne mit Personeneinsatz
- Wiser Driver course, Australien, Seminar
- Physik und Straßenverkehr, Deutschland, Multiplikatoren/ Seminar
- Liebe dein Leben – B216, Deutschland, Kampagne ohne Personeneinsatz
- Darauf fahr ich ab..., Deutschland, Kampagne ohne Personeneinsatz
- DRIVE, UK, Multiplikatoren
- Theorietest für die Führerscheinprüfung, UK, Reguliäre Fahrausbildung, Theorietest
- Never saw the day, Irland, Theaterstück
- Foolspeed, UK Kampagne ohne Personeneinsatz
- TV-spots gegen Drinkdriving bei jungen Männern, Neuseeland, rein medial
- Fair fahren, sicher ankommen, Deutschland, Kampagne mit Personeneinsatz (mehrere Maßnahmen kombiniert)
- Computer-based pedestrian training, UK, Computer-basierte Simulation
- PC-based risk awareness-training, USA, Computer-basiertes Training
- Schulische Radfahrausbildung, Deutschland, Seminar/Training
- Travel Pack, UK, rein medial
- Vorfahrschulzerziehung mit praktischem Inhalt, Australien, Seminar/Training
- Einführung einer neuen Fahrausbildung, Finnland, Reguliäre Fahrausbildung, Seminar/Training
- Freundliche Zone, Schweiz, Kampagne ohne Personeneinsatz
- Helmpflicht und Fahrradsicherheitskampagne, USA, Kampagne
- CAN-BIKE, Kanada, Seminar/Training
- Road Safety in large companies, Schweden, Gruppendiskussionen, Training, Kampagne mit Personeneinsatz
- Modellversuch „Jugend fährt sicher“, Deutschland, Training/Seminar
- Ältere Menschen als Fußgänger im Straßenverkehr, Deutschland, Seminar

Bild 1: Übersicht der Studien des systematischen Reviews: Name, Land und Umsetzungsform

<sup>1</sup> Wobei es nicht Voraussetzung war, dass alle methodischen Schritte im Detail beschrieben wurden. Dies war nur äußerst selten der Fall.

me, Erscheinungsjahr, Land etc., wurden im systematischen Review insbesondere Aspekte zur Maßnahme und zu den zehn Schritten der Wirkungsmessung erhoben.

## 2 Schritte der Wirkungsmessung und Ergebnisse des systematischen Reviews

Im Folgenden werden anhand der einzelnen Schritte methodische Aspekte verdeutlicht und es wird jeweils auf relevante Ergebnisse des systematischen Reviews hingewiesen.

### Schritt 1: Darstellung der Ziele der Maßnahmen und des Wirkungsmodells

Die Darstellung der Ziele (und Zielgruppen) der Maßnahmen ist eine Grundvoraussetzung für die gesamte Evaluation, da der Erfolg bzw. die Wirkung einer Maßnahme letztlich nur an ihren Zielen gemessen werden kann. So weit es geht, sollten hier quantitative Angaben gemacht werden. Eine quantifizierte Zielangabe könnte z. B. lauten: „Die Helmtragequote soll sich bei Kindern zwischen fünf und zwölf Jahren von 1,3% auf 20% erhöhen.“<sup>2</sup>

Das wesentliche Ziel der Verkehrserziehung und -aufklärung ist die Beeinflussung von menschlichem Verhalten. Bei der Konzeption, aber auch bei der Wirkungsmessung von Verkehrserziehungsmaßnahmen ist es daher von Bedeutung, generelle Kenntnisse über Beeinflussungsmöglichkeiten menschlichen Verhaltens mit einzubeziehen. Es sollte klar sein, in welcher Weise die geplante bzw. durchgeführte Maßnahme auf das menschliche Verhalten einwirken soll. Hierfür ist es sinnvoll, sich auf Theorien und Modelle zur Verhaltensbeeinflussung zu beziehen. Ein Beispiel für ein solches Wirkungsmodell ist die angewandte Verhaltensanalyse, die ihre Wurzeln und ihre weiteste Verbreitung in den USA hat. Weitere Beispiele sind die Theorie des geplanten Verhaltens (vgl. in [10]) oder das Stufenmodell (vgl. in [7]). Verschiedene bereits durchgeführte Metaanalysen im Bereich Verkehrserziehung kamen zu dem Ergebnis, dass Studien, die auf theoretischen Modellen basieren, effektiver sind als Studien, die sich nicht auf solche beziehen (z. B. [11]).

Auch wenn die Zieldefinition und die Angabe des Wirkungsmodells Bestandteile der Maßnahmenentwicklung sind und als Voraussetzung für die Wirkungsmessung gewertet werden können, sollten diese im Rahmen der summativen Evaluation zu Beginn explizit genannt werden.

Im Rahmen des systematischen Reviews wurden 11 Studien gefunden, bei

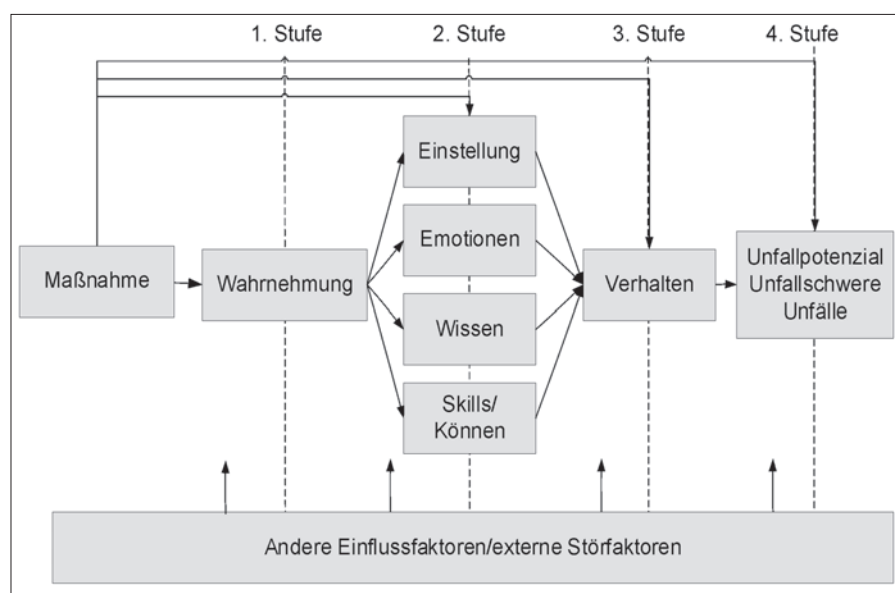


Bild 2: Stufen der Wirkungsmessung

denen das Ziel der Maßnahme gar nicht oder nur sehr grob beschrieben ist (z. B. „Erhöhung der Verkehrssicherheit“), nur in zwei der 36 Studien werden quantitative Zielvorgaben gemacht. Ein Wirkungsmodell wurde lediglich in acht Studien explizit angegeben.

### Schritt 2: Darstellung der Ziele der Wirkungsmessung und der Hypothesen

Ziel einer Evaluationsstudie ist es, Erkenntnisse über einen vermuteten Wirkungszusammenhang zu gewinnen. Zu Beginn der Wirkungsmessung sind die Ziele der Evaluation und einzelne Wirkungsbereiche zu nennen. Diese bauen auf den Zielen der Maßnahmen bzw. den Wirkungsmodellen auf. Im Einzelnen können die Erkenntnisse aus Unfallkennzahlen, Verhaltensmerkmalen, Wissens- und Einstellungsänderungen, der Veränderung motorischer Fähigkeiten oder schlicht der Wahrnehmung einer Maßnahme abgeleitet werden. Wirkungsmodelle zu Maßnahmen der Verkehrserziehung sind oft nicht nur auf einen einzelnen Wirkungsbereich ausgerichtet. Daher ist es oft sinnvoll, die Wirkungen auf mehreren Stufen zu messen. Hinzu kommt, dass es oft schwierig ist, den tatsächlichen Ursache-Wirkungs-Zusammenhang nachzuweisen. Eine Betrachtung mehrerer Wirkungsstufen kann das Gesamtergebnis der Studie auf eine fundiertere Basis stellen. Letztlich hängt es von der genauen Ausgestaltung der Maßnahme ab, welche Wirkungsstufe(n) relevant ist/sind. Bild 2 verdeutlicht die unterschiedlichen Wirkungsstufen.

Ein Wirkungsmodell bzw. eine Theorie ist idealerweise auch die Basis für die Formulierung von Hypothesen (vgl. [8]). Mit der Hypothesenbildung wird das Ziel der Wirkungsmessung festgeschrieben.

Wissenschaftliche oder auch inhaltliche Hypothesen stellen verbale Behauptungen über kausale oder nicht kausale Beziehungen zwischen Variablen dar. Im ersten Ableitungsschritt sind solche inhaltlichen Hypothesen in der Regel recht allgemein gehalten. Um bei der Datenanalyse statistische Verfahren anwenden zu können, sind diese wissenschaftlichen Hypothesen in statistische Hypothesen umzuformulieren. Für die Überprüfung von Mittelwertsunterschieden aus Stichproben zweier Populationen könnte die Hypothese lauten: Zwischen den Mittelwertsparametern  $\mu_1$  und  $\mu_2$  der Populationen besteht ein Unterschied. Die dazugehörige Nullhypothese würde keinen Unterschied zwischen den Gruppen konstatieren. In quantitativer Form könnte das sog. Hypothesenpaar folgendermaßen dargestellt werden:

$H_1: \mu_1 \neq \mu_2; H_0: \mu_1 = \mu_2$  (vgl. [8]).

Bei den untersuchten 36 Studien wurde nur in acht Fällen eine Hypothese angegeben. Dies ist besonders problematisch, da im weiteren Verlauf der Studien statistische Hypothesentests Verwendung fanden. (Siehe Schritt 9)

### Schritt 3: Darstellung der Erhebungseinheit

Im Optimalfall entspricht die Erhebungseinheit der Zielgruppe der Maßnahme. Da dies nicht immer möglich ist, ist es insbesondere zur Bewertung der methodischen Qualität notwendig, die Erhebungseinheit anzugeben. Die Zielgruppe eines Programms könnten z. B. allgemein „Schulkinder“ sein. Für die Evaluation könnten z. B. die Schulkinder einer bestimmten Stadt die Erhebungseinheit sein. Die Erhebungseinheit bildet die Grundlage für die Stichprobenbildung. (Siehe Schritt 7)

<sup>2</sup> Beispiel aus dem systematischen Review vgl. [9], S. 260

#### Schritt 4: Auswahl und Entwicklung des/der Erhebungsinstrumente(s)

Nicht immer müssen die Daten, anhand derer die Wirksamkeit einer Maßnahme festgestellt werden soll, selbst erhoben werden. Bei Evaluationsstudien aber, in denen eigene Erhebungen durchgeführt werden, spielt die Entwicklung des Erhebungsinstrumente eine besondere Rolle für die Reliabilität und Validität der Wirkungsmessung. Die klassischen Erhebungsinstrumente der Wirkungsmessung sind die Befragung, der Test und die Beobachtung (vgl. [6]). Das Erhebungsinstrument hat Einfluss auf die Auswahlmöglichkeiten der Ergebnisindikatoren. Tabelle 1 zeigt, welche Erhebungsverfahren zur Erfassung welcher Ergebnisindikatoren (und Wirkungsbereiche) geeignet sind.

Im Rahmen des systematischen Reviews wurden in den untersuchten 36 Studien 63 Erhebungsinstrumente angewandt.<sup>3</sup> Bild 3 verdeutlicht, dass dabei die schriftliche Befragung deutlich am häufigsten verwendet wurde.

Die schriftliche Befragung hat gegenüber den anderen Erhebungsinstrumenten den Vorteil, dass sie hinsichtlich der Ergebnisindikatoren sehr flexibel ist. Problematisch für die Ergebnisbewertung können allerdings eine niedrige Rücklaufquote sein (da der Befragungsrücklauf nicht als Zufallsstichprobe gewertet werden kann) bzw. – bei in mehreren Zeitintervallen wiederholten Befragungen – ein möglicher Datenschwind.

#### Schritt 5: Bestimmung der Ergebnisindikatoren

Ein guter Ergebnisindikator sollte sich möglichst direkt auf die Ziele der Maßnahmen beziehen. Daher sind Ergebnisindikatoren spezifisch je nach Anwendungsfall und Wirkungsbereich. Tabelle 1 verdeutlichte bereits den Zusammenhang von Erhebungsinstrument und Ergebnisindikator. Im Rahmen der statistischen Auswertung entsprechen die Ergebnisindikatoren den abhängigen Variablen. Tabelle 2 verdeutlicht beispielhaft die Skalenniveaus verschiedener Ergebnisindikatoren, die Auswirkungen auf die Auswahl des passenden statistischen Verfahrens haben.

Im systematischen Review wurden insgesamt 107 Ergebnisindikatoren gefunden, durchschnittlich also rund drei pro Studie. Die Häufigkeit der einzelnen Ergebnisindikatoren auf den verschiedenen Wirkungsstufen zeigt Bild 4.

Tabelle 1: Wirkungsbereiche, Indikatoren und Erhebungsinstrumente

| Wirkungsbereich                       | Ergebnisindikatoren                                                                                                                                            | Erhebungsinstrument                                                                        |
|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| <b>Wahrnehmung</b>                    | Wahrnehmung der Maßnahme, Erinnerung an bestimmte Inhalte                                                                                                      | Befragung                                                                                  |
| <b>Einstellung, Emotionen, Wissen</b> | Einstellungsänderung<br>Hervorgerufene Emotionen<br>Wissensänderung                                                                                            | Befragung                                                                                  |
| <b>Skills</b>                         | Änderung motorischer Fertigkeiten/Skills                                                                                                                       | Test Befragung (Selbsteinschätzung)                                                        |
| <b>Verhalten</b>                      | Reduzierung der Verkehrsverstöße (spezielle Delikte, z. B. Alkoholdelikte)<br>Einhaltung von Regeln (z. B. Gurtbenutzung)<br>Reduzierung der Geschwindigkeiten | Befragung (Selbsteinschätzung)<br>Auswertung vorhandener Datensätze<br>Beobachtung/Messung |
|                                       | Zunahme von Rücksichtnahme (z. B. Anhalten an Zebrastreifen)                                                                                                   | Befragung (Selbsteinschätzung)<br>Beobachtung                                              |
|                                       | Veränderung eines bestimmten Verhaltens (z. B. „Rotlichtgeher“)                                                                                                | Befragung (Selbsteinschätzung)<br>Beobachtung                                              |
|                                       | Reduzierung der „Fast-Unfälle“                                                                                                                                 | Befragung (Selbsteinschätzung)<br>Beobachtung                                              |
|                                       | Weitere spezielle Messkriterien (z. B. Rückfallquote nach Nachschulungskursen)                                                                                 | Befragung (Selbsteinschätzung)<br>Auswertung vorhandener Datensätze                        |
| <b>Unfälle, Unfallschwere</b>         | Reduzierung der Unfallzahlen                                                                                                                                   | Befragung (Selbsteinschätzung)<br>Auswertung vorhandener Datensätze                        |
|                                       | Reduzierung der Unfallschwere<br>Verletztenzahlen                                                                                                              | Befragung (Selbsteinschätzung)<br>Auswertung vorhandener Datensätze                        |
|                                       | Befragung (Selbsteinschätzung)<br>Versicherungsakten                                                                                                           | Reduzierung der Schadensdaten aus den<br>Auswertung vorhandener Datensätze                 |

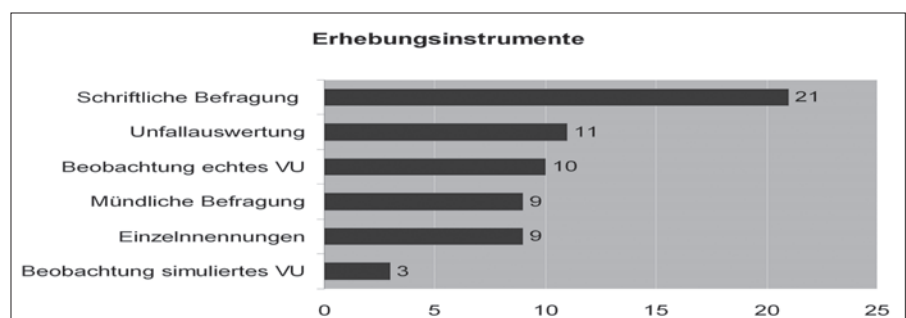


Bild 3: Erhebungsinstrumente im Rahmen des Reviews, n = 63 (VU = Verkehrsumfeld)

Tabelle 2: Skalenniveau verschiedener Ergebnisindikatoren

| Skalenniveau              | Ergebnisindikatoren (= Zielgrößen) als Beispiel                                                                                       |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| <b>Binär</b>              | Helmträger/Kein HelmträgerRotlichtgeher/GrünlichtgeherRichtig-/Falsch-Antworten                                                       |
| <b>Nominal</b>            | z. B. Gruppen bei einer Beobachtung. „trägt keinen Helm“, „trägt Helm, aber falsch benutzt“, „trägt keinen Helm“                      |
| <b>Ordinal</b>            | Zustimmungsskalen (stimme gar nicht, etwas, sehr zu)Einstufung im Verhaltenstest (sehr gut gebremst, gut gebremst, schlecht gebremst) |
| <b>Intervall/Rational</b> | Punkte in einem Wissenstest Punkte bei einem Leistungstest Schadensdaten (Schadenshöhe)Geschwindigkeiten                              |

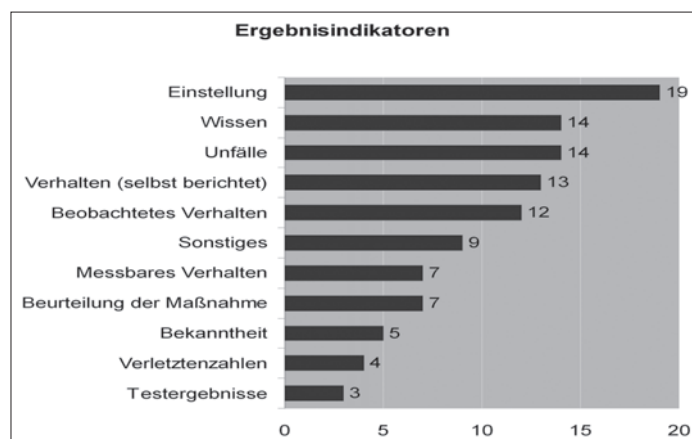


Bild 4: Ergebnisindikatoren im systematischen Review, n = 107

<sup>3</sup> Im Folgenden werden die Ergebnisse des systematischen Reviews teilweise auf die 36 Studien bezogen, teilweise auf die 63 Erhebungsmethoden.

### Schritt 6: Auswahl und Entwicklung des Untersuchungsdesigns

Um den Effekt einer Maßnahme zu messen, wird bei der Wirkungsmessung oft eine Gruppe oder ein Untersuchungsraum, bei der bzw. in dem die Maßnahmen durchgeführt wurden, mit einer Gruppe oder einem Raum ohne entsprechende Maßnahme (Kontrollraum, -gruppe) verglichen. Die Versuchsbedingungen sollten sich im Idealfall nur im Hinblick auf die Intervention unterscheiden. Als allgemeines Prinzip gilt: Je strenger das bei der Wirkungsanalyse angewendete Untersuchungsdesign ist, desto verlässlicher sind die geschätzten Programmeffekte (vgl. [12]). Untersuchungsdesigns lassen sich in der Hauptsache nach folgenden Aspekten unterscheiden:

- Art und Anzahl der Untersuchungs- bzw. Kontrollgruppen
- Art der (echten) Kontrollgruppe (z. B. randomisiert, konstruiert)
- Anzahl der Erhebungszeitpunkte
- Nachher-Erhebungen an derselben oder einer anderen Stichprobe
- Eigene Erhebung oder Auswertung vorhandener Daten

Als Grundformen für Untersuchungsdesigns können gelten:

- 1) Nur-Nachher-Untersuchung (ohne Kontrollgruppe)
- 2) Retrospektive Vorher-Nachher-Untersuchung (ohne Kontrollgruppe)
- 3) Vorher-Nachher-Untersuchung (ohne Kontrollgruppe)
- 4) Mehrfach-Untersuchung (ohne Kontrollgruppe)
- 5) Nur-Nachher-Untersuchung mit Kontrollgruppe (Mit-Ohne-Vergleich)
- 6) Retrospektive Vorher-Nachher-Untersuchung mit Kontrollgruppe
- 7) Vorher-Nachher-Untersuchung mit Kontrollgruppe
- 8) Vorher-Nachher-/Mit-Ohne-Untersuchung (mit Kontrollgruppe)
- 9) Mehrfach-Untersuchung mit Kontrollgruppe

Diese neun Designs unterscheiden sich deutlich hinsichtlich ihrer methodischen Qualität und damit hinsichtlich der Verlässlichkeit der auf die eine oder andere Art gewonnenen Ergebnisse. Eine grobe Übersicht gibt die folgende Tabelle 3. Für eine endgültige Bewertung der methodischen Qualität eines Untersuchungsdesigns ist jedoch die Betrachtung weiterer Aspekte notwendig. Relevant ist z. B., dass Untersuchungsgruppe bzw. -raum und Kontrollgruppe bzw. -raum möglichst gleich groß bzw. möglichst ähnlich gewählt sind. Entscheidend ist ferner, auf welche Art Untersuchungs- und Kontrollgruppe gebildet wurden.

Tabelle 3: Methodisches Level von Untersuchungsdesigns (grobe Einteilung)

| Qualität der Untersuchungsdesigns (Level)                                                                                                                         |                                                                                                                                                                                                                           |                                                                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| Eher hoch                                                                                                                                                         | Mittel                                                                                                                                                                                                                    | Eher niedrig                                                                                                            |
| 7) Vorher-Nachher-Untersuchung mit Kontrollgruppe<br>8) Vorher-Nachher-/Mit-Ohne-Untersuchung (mit Kontrollgruppe)<br>9) Mehrfach-Untersuchung mit Kontrollgruppe | 5) Nur-Nachher-Untersuchung mit Kontrollgruppe<br>6) Retrospektive Vorher-Nachher-Untersuchung mit Kontrollgruppe<br>3) Vorher-Nachher-Untersuchung (ohne Kontrollgruppe)<br>1) Nur-Nachher-Untersuchung (Kontrollgruppe) | 1) Nur-Nachher-Untersuchung (ohne Kontrollgruppe)<br>2) Retrospektive Vorher-Nachher-Untersuchung (ohne Kontrollgruppe) |

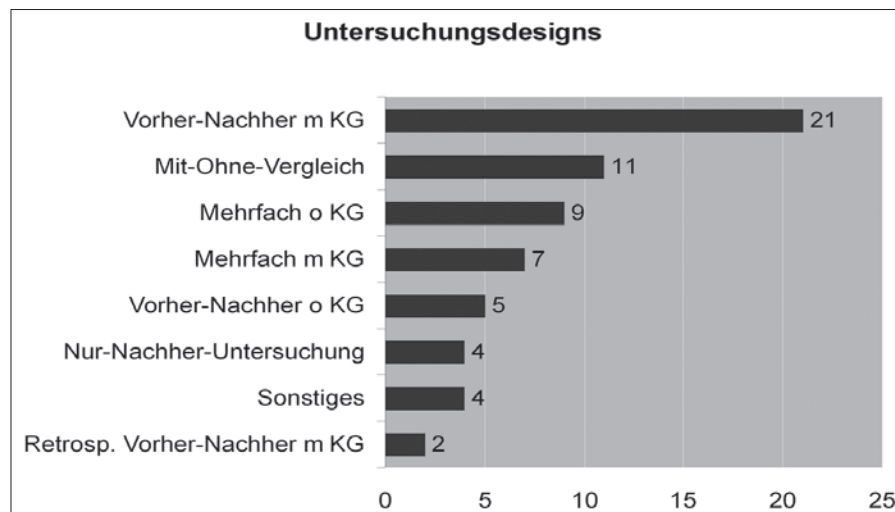


Bild 5: Untersuchungsdesigns im systematischen Review,  $n = 63$  KG = Kontrollgruppe, o = ohne, m = mit

Tabelle 4: Stichprobenarten nach BORTZ, DÖRING [8]

| Probabilistische Stichproben                                                                         | Nicht probabilistische Stichproben                                                                |
|------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Einfache Zufallsstichprobe<br>Geschichtete Stichprobe<br>Klumpenstichprobe<br>Mehrstufige Stichprobe | Ad-hoc-Stichprobe<br>Theoretische Stichprobe<br>Quotenstichprobe<br>Stichprobe nach Verfügbarkeit |

Die im systematischen Review betrachteten Studien haben rund zur Hälfte Designs „mittlerer“ und „eher niedriger“ Qualität verwendet, wie Bild 5 zeigt. Wie bereits bemerkt, ist zur Beurteilung der Qualität des Untersuchungsdesigns ein detaillierter Blick erforderlich. In den untersuchten Studien wurden die Kontrollgruppen beispielsweise sehr häufig nach Verfügbarkeit (12 Untersuchungsdesigns) gebildet oder konstruiert (8 Untersuchungsdesigns).

### Schritt 7: Stichprobenbildung

Die Einhaltung bestimmter Grundsätze zur Stichprobenbildung ist Voraussetzung für die differenzierte Datenanalyse mittels statistischer Verfahren (insbesondere der schließenden Statistik) und letztlich auch Voraussetzung für die Interpretation und Schlussfolgerungen aus den empirisch gewonnenen Ergebnissen. Ziel ist die repräsentative bzw. unverzerrte Selektion eines Teils einer relevanten und sinnvoll abgegrenzten Grundgesamtheit (Erhebungseinheit).

Die beste Gewähr für größtmögliche Repräsentativität bietet die Zufallsstichprobe. Eine Zufallsauswahl der Einheiten ist sichergestellt, wenn jedes Element der Grundgesamtheit die gleiche Chance hat, ausgewählt zu werden. Bei Stichproben, die diese Voraussetzung erfüllen, spricht man von probabilistischen Stichproben. Die Verfahren zur Stichprobenbildung können somit grob in probabilistische und nicht probabilistische Stichproben eingeteilt werden. Bei der theoretischen Stichprobe werden aufgrund theoretischer Überlegungen typische oder untypische Fälle bewusst ausgesucht.

Neben der Methode der Stichprobengewinnung ist auch die Stichprobengröße von besonderer Bedeutung für die methodische Qualität bzw. die Repräsentativität der Stichprobe (Bild 6). Der systematische Review zeigt, dass in der Praxis sehr oft theoretische Stichproben bzw. Stichproben nach Verfügbarkeit gebildet werden. Die theoretischen Stichproben sind dabei hinsicht-

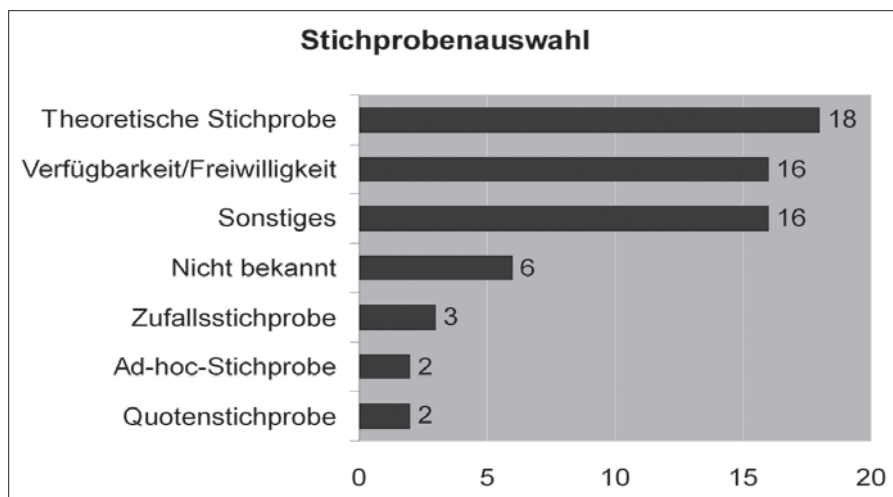


Bild 6: Gefundene Stichprobenarten im systematischen Review,  $n = 63^4$

lich der spezifischen Repräsentativität teilweise als sehr genau einzustufen. Dennoch kann festgehalten werden, dass, von einem strengen Standpunkt aus betrachtet, die Verfahren zur Stichprobenbildung den Einsatz von Hypothesentests nur sehr eingeschränkt ermöglichen.

#### Schritt 8: Datenerhebung

Die Konzeption des Erhebungsinstruments (siehe Schritt 4) legt den Grundstein für die Qualität der Datenerhebung. Besonders bei der mündlichen Befragung, der Beobachtung und bei bestimmten Verhaltenstests ist bei dem Schritt der Datenerhebung große Sorgfalt anzuwenden. Relevant ist hier der sog. „Hawthorne-Effekt“ (auch Placebo-Effekt), der das Phänomen bezeichnet, dass Menschen allein auf Grund der Tatsache, dass sie an einer Untersuchung teilnehmen, das erwartete Verhalten zeigen. Generell sind daher, so oft es geht, Doppelblindversuche durchzuführen, bei denen weder der Proband noch der Befragter/Beobachter/Tester über das Untersuchungsdesign (insb. Zugehörigkeit des einzelnen Probanden zur Kontroll- oder Untersuchungsgruppe) Bescheid wissen.

Um die Verhaltensvarianz von Befragern/ Beobachtern/Testern möglichst zu minimieren, können gründliche Schulungen und zu deren Kontrolle Pre-Tests durchgeführt werden. Zudem sollte bei Beobachtungen der Grad der Beobachterübereinstimmung geprüft werden. Alternativ zur Datenerhebung kann ggf. auch vorhandenes Datenmaterial (z. B. amtliche Statistiken) ausgewertet werden. Im systematischen Review wurde nur bei 14 von 63 Erhebungsverfahren vorhandenes Datenmaterial verwendet.

Es ist bemerkenswert, dass bei diesen Fällen sehr häufig Untersuchungsdesigns mit methodisch „eher hohem“ Niveau verwendet wurden.

#### Schritt 9: Datenanalyse/Anwendung statistischer Verfahren

Eine Vielzahl von statistischen Verfahren kann bei einer Wirkungsmessung Anwendung finden. Die Verfahren der schließenden Statistik geben Auskunft darüber, inwieweit die empirisch gewonnenen Ergebnisse noch durch Zufall zu erklären sind oder ein systematischer Einfluss durch die untersuchte Maßnahme angenommen werden kann. Dazu werden sog. statistische Hypothesentests (Signifikanztests) verwendet. Es hängt von der Art der Daten und dem Untersuchungsdesign ab, welches statistische Verfahren zur Signifikanzprüfung geeignet ist.

Relevant für die Auswahl des Verfahrens (meist Signifikanztest) sind:

- Anzahl und Art der Einflussgrößen (unabhängige Variable): Wie viele Einflussgrößen mit wie vielen Ausprägungen und welcher Art werden betrachtet?

- Art der Zielgröße/Ergebnisindikator (abhängige Variable): Welche Verteilung der Zielgröße liegt in der Grundgesamtheit vor? Liegt eine Normalverteilung vor? Bei nicht verteilungsabhängigen Verfahren: Ist die Zielgröße binär (B), nominal (N) oder ordinal (O) skaliert?

Da bei Wirkungsmessungen besonders häufig Stichprobenvergleiche durchgeführt werden, wird im Folgenden auf diese eingegangen. Zu unterscheiden sind verbundene Gruppen (V) (vorher/nachher) von nicht verbundenen Gruppen (N) (Untersuchungs-/Kontrollgruppe). Bild 7 zeigt, welche statistischen Verfahren bei Stichprobenvergleichen für die verschiedenen Skalenniveaus von Ergebnisindikatoren relevant sind. In Tabelle 2 wurden beispielhaft Einflussgrößen und Ergebnisindikatoren den Skalenniveaus zugeordnet.

Insgesamt fanden 80 statistische Verfahren in 31 Studien des systematischen Reviews Anwendung. Dabei handelt es sich zu einem großen Teil um die klassische Verfahren t-Test, Chi-Quadrat-Test und Varianzanalyse. Diese machten mehr als die Hälfte der verwendeten Verfahren aus.

#### Schritt 10: Dateninterpretation

Bei der Dateninterpretation sind methodische Schwachstellen der Wirkungsmessung (die oft nicht zu verhindern sind) herauszustellen und bei der Ergebnismessung zu berücksichtigen. Außerdem soll überprüft werden, ob eine grundsätzliche Logik der Zusammenhänge vorliegt.

In 13 Studien des systematischen Reviews wird eine Wirkung nur in einigen untersuchten Aspekten festgestellt, eine positive Wirkung in allen untersuchten Bereichen (oder in dem einen relevanten Bereich) wird nur in fünf Studien abgeleitet. In vier Studien wird explizit keine Wirkung festgestellt. Positiv ist zu bemerken, dass in vielen Studien die

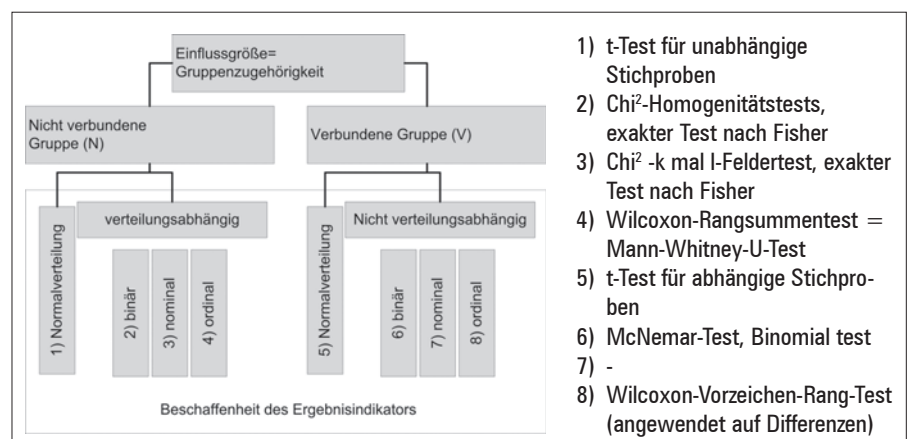


Bild 7: Statistische Verfahren im Rahmen der Wirkungsmessung

4 Bei den sonstigen Stichproben wurden überwiegend Kombinationen verschiedener Stichprobenverfahren verwendet, oder es lag keine Stichprobe im eigentlichen Sinne vor (z. B. Vollerhebung).

Frage der Wirksamkeit kritisch und mit Vorsicht beantwortet wird; zum Teil wird auf methodische Schwachstellen hingewiesen.

### 3. Ausblick und Fazit

Die Wirkungsmessung im Bereich der Verkehrserziehung und -aufklärung ist ein komplexes Feld, und sie bleibt auch immer einzelfallbezogen. So wird die hier in Ausschnitten wiedergegebene Forschungsarbeit es nicht ersetzen können, dass in jedem einzelnen Fall einer Wirkungsmessung individuelle Überlegungen angestellt werden müssen und ggf. fachliche Expertise hinzugezogen werden muss. Unter dem Effizienzgedanken ist es wünschenswert, dass Wirkungsmessungen dokumentiert und veröffentlicht werden – und zwar auch solche, bei denen keine erwarteten Effekte gefunden wurden. Darüber hinaus wäre es wünschenswert, dass diese auch gut zugänglich sind und ggf. vielleicht sogar an einer zentralen Stelle gesammelt werden. Aus eigener Erfahrung aus dieser Forschungsarbeit, aber auch aus den Erfahrungen anderer Autoren heraus, kann bestätigt werden, dass es sehr mühsam ist, entsprechende Studien zu finden (Dem Anhang zur Dissertation kann eine Kurzfassung aller betrachteten Wirkungsmessungen entnommen werden, vgl. [9]). Regelmäßige Metastudien zu verschiedenen typischen Verkehrserziehungsmaßnahmen könnten hier zukünftig die Ausgangslage verbessern. Dieser Artikel enthält kein Plädoyer, zukünftig Wirkungsmessungen bei sämtlichen Verkehrserziehungsmaßnahmen durchzuführen. Vielmehr sollten dann,

wenn eine Wirkungsmessung durchgeführt wird, methodische Standards bei den unterschiedlichen Vorgehensweisen eingehalten werden. Bezüglich „kleinerer“ Maßnahmen wird empfohlen, lieber beispielhaft methodisch anspruchsvolle Wirkungsmessungen durchzuführen als häufigere Wirkungsmessungen auf niedrigem Niveau.

Im Rahmen der Dissertation zur Wirkungsmessung ([9]) wurden insgesamt 22 Standards für eine Wirkungsmessung im Bereich Verkehrserziehung und -aufklärung definiert. Die Anwendbarkeit dieser Standards – auch bei kleineren Maßnahmen – konnte in dieser Arbeit anhand von Praxisbeispielen belegt werden. Im Einzelfall können auch Gründe dafür sprechen, einen Standard nicht zu berücksichtigen. Im Sinne einer Transparenz bzw. Qualitätssicherung von summarischen Evaluationen in diesem Bereich wäre es jedoch ratsam, diese Standards zu berücksichtigen und in Einzelfällen zu begründen, warum ein Standard nicht eingehalten werden konnte.

#### Literatur

- [1] Gerlach, J.; Kesting, T.; Limbourg, M.; Utmann, I. (2005). *Qualitätssicherung bei Programmen der Verkehrserziehung, -aufklärung und -weiterbildung*. Abschlussbericht 2005, unveröffentlichter Schlussbericht zum FE 82.211/2001 der Bundesanstalt für Straßenwesen. Wuppertal.
- [2] Duperrex, O.; Roberts, I.; Bunn, F. (2003). *Safety education of pedestrians for injury prevention*. In: The Cochrane Library, Issue 1. Oxford.
- [3] Elliott, B. (1993). *Road Safety Mass Media Campaigns: A Meta Analysis*. Sydney.
- [4] Elvik, R.; Vaa, T. (Hrsg.) (2005). *The Handbook of Road Safety Measures*. Amsterdam.
- [5] Harré, N.; Wrapson, W. (2004). The evaluation of a central-city pedestrian safety campaign. In: Rothengatter, J.A.; Groeger, J.A. (Hrsg.). *Transportation Research Part F*. 2004/05. 7(3) Pp. 167–79. Pergamon.
- [6] Pfaff, M. et al. (1982). *Methoden und Kriterien zur Überprüfung des Erfolgs von Aufklärungsmaßnahmen – Handbuch zur Durchführung von Wirksamkeitsanalysen*.

- Bericht zum Forschungsprojekt 7824 der Bundesanstalt für Straßenwesen – Bereich Unfallforschung, Köln 1982.
- [7] Becker, W.; Kistler, E. (1992). *Optimierung von Verkehrsaufklärungsmaßnahmen*. Forschungsberichte der Bundesanstalt für Straßenwesen, Heft 252, Bergisch Gladbach.
  - [8] Bortz, J.; Döring, N. (2003). *Forschungsmethoden und Evaluation*. Berlin, Heidelberg. 3. Auflage.
  - [9] Utmann, I. (2008). *Summative Evaluation der Verkehrserziehung und -aufklärung*. Methodische Vorgehensweisen bei der Wirkungsmessung von Maßnahmen der Verkehrserziehung und -aufklärung. Hamburg. Verlag Dr. Kovac.
  - [10] Flade, A.; Borchering, K. (2001). *Einflussgrößen und Motive der Fahrradnutzung im Alltagsverkehr*. In: Heinicke, G.; Dienel, H.-L. (Hrsg.). *Mobilitäts- und Verkehrsforschung*. Neue empirische Erhebungsinstrumente im Vergleich. Dokumentation eines Workshops des Bundesministeriums für Bildung und Forschung. S. 53–68. Berlin.
  - [11] Organisation for Economic Cooperation and Development (OECD) (Hrsg.) (1986). *Effectiveness of road safety education programmes*. Paris.
  - [12] Beckmann, K. J.; Bamberg, S.; Wehmeier, T. (2004b). *Leitlinien zur Evaluation verkehrlicher Maßnahmen und Programme*. Aachen, Gießen.
  - [13] OECD, Road Transport Research (1993). *Marketing für Verkehrssicherheit in der Praxis*. Berichte der Bundesanstalt für Straßenwesen, Reihe Mensch und Sicherheit, Heft M 11. Bremerhaven: Wirtschaftsverlag NW.
  - [14] Christ, R., Delhomme, P., Kaba, A., Mäkinen, T., Sagberg, F., Schulze, H. & Siegrist, S. (1999). *GADGET - Guarding Automobile Drivers through Guidance Education and Technology*. Final Report. Investigations on Influences upon Driver Behaviour - Safety Approaches in Comparison and Combination. Vienna: Austrian Road Safety Board (KfV).



Dr.-Ing. Iris Utmann hat am Fachbereich Bauingenieurwesen der Universität Wuppertal promoviert. Ihre Forschungsthemen sind Umwelt und Verkehr sowie menschliches Verhalten und Verkehr.

Anschrift:  
Dr.-Ing. Iris Utmann  
Forstmannstr. 8  
45239 Essen  
irisutmann@gmail.com

# Füller