

# Beobachtungsbasierte Erfassung von Müdigkeit im Kfz – die TUBS-Skala

**Elisabeth Dittrich, Stefan Brandenburg und Manfred Thüring**

*Schlüsselwörter: Müdigkeit im Kfz, Fremdbewertung, psychometrische Gütekriterien, Bewertungsverfahren*

## Zusammenfassung

Müdigkeit im Straßenverkehr ist eine der Hauptunfallursachen (Evers & Auerbach, 2005). In Zukunft soll eine automatische Erkennung von Müdigkeit im Kraftfahrzeug helfen, die Unfallhäufigkeit zu reduzieren. Die dafür entwickelten Assistenzsysteme werden u.a. durch Fremdeinschätzung der Fahrermüdigkeit evaluiert. Die zwei der hierfür am häufigsten verwendeten Skalen (Wierwille & Ellsworth, 1994; Kolrep, Rimini-Döring, Oehme, Jürgensohn & Altmüller, 2005) erfüllen die herkömmlichen psychometrischen Gütekriterien allerdings nur teilweise (Dittrich, Brandenburg & Thüring, 2009) und bringen unterschiedliche Vor- und Nachteile mit sich, die die Genauigkeit, mit der Müdigkeitsindikatoren den einzelnen Bewertungsstufen zugewiesen werden, und die Differenziertheit ihres Wertebereich betreffen. Eine optimierte Müdigkeitsbewertung würde demnach durch ein Verfahren ermöglicht, das die Vorteile beider Skalen vereint und dabei die psychometrischen Gütekriterien möglichst umfassend erfüllt. Ergebnis dieser Überlegungen ist die in diesem Beitrag vorgestellte TU Berlin Schläfrigkeitsskala (TUBSS).

## Abstract

Drivers' sleepiness is one of the most frequent causes of accidents (Evers & Auerbach, 2005). For the future, automatic sleepiness detection systems aim to reduce the frequency of accidents. These systems are often validated by comparing their output with the sleepiness estimation of trained raters. An analysis of the two most common scales used for this purpose (Wierwille & Ellsworth, 1994; Kolrep, Rimini-Döring, Oehme, Jürgensohn & Altmüller, 2005) revealed however a restricted fulfillment of central psychometric criteria (Dittrich, Brandenburg & Thüring, 2009). Further, both scales differ in basic characteristics, such as the detailed assignment of sleepiness indicators to ratings and the range of possible sleepiness values. An optimized estimation procedure for sleepiness would combine the advantages of both scales and fully comply with the psychometric criteria. A possible solution, the TUBSS scale, is presented in this paper.

## Einleitung

Im Jahr 2008 wurden 319.813 Verkehrsunfälle mit Personenschaden in Deutschland registriert (Statistisches Bundesamt, 2009). Müdigkeit wird in ca. 25 Prozent der Fälle als Ursache für ein Verunglücken im Straßenverkehr angenommen (GDV, 2003). Fahrerassistenzsysteme (FAS) bieten eine Möglichkeit, eine Ermüdung des Fahrers automatisch zu erkennen und ihn bei Fahruntauglichkeit zu warnen. Obwohl derartige Systeme bereits auf dem Markt eingeführt wurden, ist die Güte, mit der sie Müdigkeit erfassen, noch weit gehend ungeklärt.

Videobasierte Beobachtungsverfahren ermöglichen es, diese Güte zu bestimmen. Dabei dienen Videosequenzen geschulten Müdigkeitsbewertern als Urteilsgrundlage, um dem aktuellen Fahrerzustand eine Müdigkeitsstufe zuzuordnen. Eine möglichst objektive Einschätzung der Müdigkeit wird durch die Verwendung von standardisierten Skalen gewährleistet. Zwei der gebräuchlichsten Skalen zur Fremdeinschätzung stammen von Wierwille und Ellsworth (1994) sowie von der Human Factors Consult GmbH Berlin (Kolrep, Rimini-Döring, Oehme, Jürgen-

sohn & Altmüller, 2005). Die Skala von Wierwille und Ellsworth (1994) umfasst vier Müdigkeitsstadien (von 1 = nicht müde bis 4 = extrem müde / Sekundenschlaf), die durch Verhaltens- und Augenparameter beschrieben werden. In Abgrenzung dazu geht die Skala von Kolrep et al. (2005) von neun Stadien (1 = wach bis 9 = Schlaf) der Müdigkeit aus. Sowohl aus wissenschaftlicher Perspektive als auch für die praktische Anwendung stellt sich die Frage, welche der beiden Skalen besser geeignet ist Müdigkeit einzuschätzen.

Um diese Frage zu beantworten, wurden beide Verfahren einer Überprüfung ihrer psychometrischen Gütekriterien Objektivität und Reliabilität unterzogen (Dittrich et al., 2009). Als wichtigstes Ergebnis konnte festgestellt werden, dass beide Bewertungsverfahren die psychometrischen Kriterien nur bedingt erfüllen. Auf Basis der gewonnenen Erkenntnisse wurde die TU Berlin Schläfrigkeitsskala (TUBS-Skala) entworfen. Dieses Verfahren strebt an, die Vorteile der von Wierwille und Ellsworth (1994) entwickelten Skala mit denen der Skala von Kolrep et al. (2005) zu vereinen.

Im Folgenden werden die drei Skalen zunächst vorgestellt und anschließend auf Basis einer experimentellen Untersuchung miteinander verglichen.

### **Müdigkeitsbestimmung nach Wierwille und Ellsworth (1994)**

Die Skala von Wierwille und Ellsworth (kurz „Wierwille-Skala“) beinhaltet vier Stufen der Fahrermüdigkeit, die von „1 = nicht müde“ bis „4 = extrem müde“ reichen. Jede Stufe wird durch Augenparameter und Verhaltensindikatoren differenziert beschrieben. Besonders charakteristisch sind Veränderungen im Blickverhalten. Der wache Fahrer (Stufe 1) ist durch schnelle Blicke gekennzeichnet, die bis Stufe 3 stetig abnehmen und im starren Blick enden. Im Laufe der Ermüdung (ab Stufe 2) verringert sich die Lidschlaghäufigkeit zunehmend, während sich die Dauer des Lidschlags verlängert. Bei Sekundenschlaf (Stufe 4) treten schließlich überlange Lidschlüsse auf, und die Augen können sekundenlang geschlossen bleiben. Als Verhaltensindikatoren für steigende Müdigkeit gelten eine immer bequemer werdende Sitzposition und ein Erschlaffen der Gesichtsmuskulatur, was zu einer zunehmend ausdrucksloseren Mimik führt. Des Weiteren können selbstaktivierende Manierismen (wie z. B. Augenreiben) auftreten. Beim Aufwachen aus dem Sekundenschlaf sind ruckartige Bewegungen, ein plötzliches Wiederaufrichten des Körpers und häufig auch Schreckreaktionen zu beobachten, die sich vor allem in der Mimik des Fahrers widerspiegeln, aber auch zu überschießenden Lenkbewegungen führen können.

### **Die Müdigkeitseinschätzungsskala von HFC (Kolrep et al., 2005)**

Im Unterschied zur Wierwille-Skala postuliert der Ansatz von HFC (kurz „HFC-Skala“) neun Stadien der Müdigkeit, die von „1 = wach“ bis „9 = Schlaf“ reichen. Zur Einordnung eines Fahrerzustands in diese Stadien dienen neun Indikatoren. Anders als bei der Wierwille-Skala sind konkrete Ausprägungen dieser Indikatoren den einzelnen Müdigkeitsstadien allerdings nicht direkt zugeordnet. Vielmehr gibt die Indikatorliste lediglich Hinweise darauf, wie sich Müdigkeit generell äußern kann. Zufallende Augen, schwere Lider, Gähnen und Seufzen sowie Peripherhandlungen weisen auf Müdigkeit hin. Aktive Blickbewegungen, hohe Körperspannung und aufrechte Sitzhaltung, Körperbewegungen sowie Korrekturhandlungen beim Lenken charakterisieren den wachen Fahrerzustand. Ein Müdigkeitsbeurteiler bewertet den Fahrer im Videoausschnitt hinsichtlich des Vorhandenseins und der Ausprägtheit dieser Verhaltensaspekte und gibt anschließend sein Urteil über die Fahrermüdigkeit ab.

### **Die TU-Berlin-Schläfrigkeitsskala (TUBSS)**

Mit der Entwicklung der TUBS-Skala wird das Ziel verfolgt, die Vorteile der Wierwille-Skala mit denen der HFC-Skala zu vereinen. Zum einen wird eine exakte Zuordnung von Müdig-

keitsstufen zu einzelnen Müdigkeitsindikatoren vorgenommen, die der von Wierwille und Ellsworth (1994) entspricht. Zum anderen wird eine mehrstufige Skala ähnlich der HFC-Skala angestrebt, um einen differenzierteren Wertebereich für die Müdigkeitseinschätzung zur Verfügung zu stellen. Um dies zu erreichen, wird ein zweistufiger Müdigkeitsbewertungsprozess eingeführt. Nach einer ersten Einordnung durch den Rater des Videoausschnitts zu einer der vier Müdigkeitsstufen anhand der Wierwille-Kriterien erfolgt ein zweiter Bewertungsschritt. In diesem werden die Stufen 2 (schläfrig/ müde) und 3 (sehr müde) hinsichtlich ihrer Tendenz zu den jeweils angrenzenden Stadien beurteilt. Wenn beispielsweise im ersten Einschätzungsschritt die Müdigkeitsstufe 2 vergeben wurde, kann im zweiten Schritt angegeben werden, ob eine Tendenz in Richtung wach (hin zu Müdigkeitsstufe 1) oder sehr müde (hin zu Müdigkeitsstufe 3) besteht. Wird keine derartige Tendenz angegeben, wird der Wert der ersten Einschätzung beibehalten. Aus diesem Vorgehen ergibt sich eine 8-stufige Skala von 1 (nicht müde) bis 8 (Sekundenschlaf), wobei die Stufen 2 bis 7 eine differenzierte Einschätzung der Fahrermüdigkeit im sicherheitskritischen Bereich ermöglichen.

Zur Gewährleistung der Objektivität wurden der Ablauf des Verfahrens (Durchführungsobjektivität), die Vergabe der Ratingwerte (Auswertungsobjektivität) sowie die Interpretation im Sinne von Müdigkeitsstufen (Interpretationsobjektivität) standardisiert (vgl. Bortz und Döring, 2003). Hierauf aufbauend wurde ein Training entwickelt, das jeder Rater vor seinem Einsatz zu durchlaufen hat.

## **Fragestellung**

Die drei vorgestellten Verfahren zur Fremdeinschätzung von Müdigkeit im Kraftfahrzeug unterscheiden sich grundlegend in zwei Aspekten. Zum einen skalieren sie die Müdigkeit mit unterschiedlichen Auflösungsgraden. Zum anderen gibt die Wierwille-Skala konkrete Ausprägungen von Indikatoren vor, bei denen eine bestimmte Müdigkeitsbewertung vergeben werden kann. Die HFC-Skala weist dagegen keine spezifische Zuordnung von Verhaltensaspekten und deren Ausprägung zu Müdigkeitsstufen auf, erlaubt allerdings eine differenziertere Einschätzung der Fahrermüdigkeit. Die TUBS-Skala zielt darauf ab, die Vorteile beider Verfahren zu vereinen. Da ein Verfahren zur Einschätzung von Müdigkeit ein objektives, reliables und möglichst differenziertes Müdigkeitsurteil erlauben soll, ist zu klären, welcher der drei Ansätze diesem Anspruch am ehesten genügt. Drei zentrale Fragen stellen sich in diesem Zusammenhang:

12. Unterscheiden sich die Müdigkeitseinschätzungen, die mit den drei Skalen erhoben werden voneinander? Ein statistischer Vergleich, der diese Frage beantwortet, muss natürlich standardisierte Werte benutzen, damit die unterschiedlichen Skalierungen nicht zu einer Verzerrung führen.
13. Wie objektiv misst jede der drei Skalen Müdigkeit? „Die Objektivität eines Tests gibt an, in welchem Ausmaß die Testergebnisse vom Testanwender unabhängig sind.“ (Bortz & Döring, 2003; S. 194). Eine solche Unabhängigkeit sollte sich nicht nur darin äußern, dass die Bewertungen verschiedener Urteiler, die dasselbe Verfahren benutzen, hoch miteinander korrelieren (Interrater-Reliabilität), sondern auch darin, dass die Unterschiede zwischen ihnen möglichst gering ausfallen.
14. Wie reliabel sind die drei Skalen jeweils? „Die Reliabilität eines Tests kennzeichnet den Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird.“ (Bortz & Döring, 2003; S. 195). Ein gängiges Maß hierfür ist die sog. Retest-Reliabilität, bei der ein Urteiler einen Sachverhalt zweimal bewertet. Je höher die Werte der beiden so gewonnenen Messreihen miteinander korrelieren (Intrarater-Reliabilität), und je geringer die Differenz zwischen ihnen ausfällt, desto reliabler ist das Verfahren.

Um die drei Fragen zu beantworten, wurden alle drei Verfahren hinsichtlich der psychometrischen Gütekriterien Objektivität und Reliabilität in einem Experiment überprüft.

## **Methode**

### **Stichprobe**

An der vorliegenden Untersuchung nahmen insgesamt 16 Personen teil. Davon generierten vier weibliche Versuchspersonen im Alter von 22 bis 28 Jahren ( $MW = 24.75$ ,  $SD = 2.5$ ) das Videomaterial und 12 Probanden beurteilten die Müdigkeit der vier Fahrer. Die vier weiblichen und acht männlichen Rater waren zwischen 20 und 29 Jahre alt ( $MW = 25.75$ ,  $SD = 3.37$ ).

### **Versuchsmaterial**

Zur Erstellung des Versuchsmaterials dienten vier ca. zweistündige Fahrten, die im Fahrsimulator der Technischen Universität Berlin und der Volkswagen AG absolviert wurden. Die Probanden hatten die Aufgabe, einem vorausfahrenden Fahrzeug mit 100 km/h in konstantem Abstand zu folgen. Dabei wurde das Gesicht der Personen von einer Kamera aufgezeichnet. Aus jedem der vier Videos wurde alle zehn Minuten ein einminütiger Videoausschnitt extrahiert. Zusätzlich dazu wurde je ein manuell gesichteter Sekundenschlaf pro Person als weiterer einminütiger Ausschnitt dem Versuchsmaterial hinzugefügt. Insgesamt wurden so 59 einminütige Videosequenzen gewonnen.

### **Versuchsablauf**

Die 12 Probanden des Versuchs wurden zunächst mit einer der drei Skalen trainiert. Nach der Trainingsphase absolvierten sie im Abstand von einer Woche zwei Versuchsdurchläufe. Dabei wurden ihnen jeweils die 59 Videosequenzen in randomisierter Reihenfolge dargeboten, die sie sukzessiv anhand der jeweiligen Skala bewerteten.

### **Versuchsdesign**

In der vorliegenden Untersuchung wurden zwei unabhängige Variablen verwendet. Zum einen war dies die verwendete Müdigkeitseinschätzungsskala als Zwischensubjektfaktor mit den drei Ausprägungen „HFC-Skala“, „Wierwille-Skala“ und „TUBS-Skala“. Zum anderen wurde der Messzeitpunkt der Müdigkeitsbewertung als Innersubjektfaktor mit den beiden Stufen „erster Zeitpunkt“ ( $t_1$ ) und „zweiter Zeitpunkt“ ( $t_2$ ) variiert. Als abhängige Variable diente die Müdigkeitsbewertung der zwölf Rater.

### **Auswertung und Ergebnisse**

Um die drei oben genannten Fragestellungen zu bearbeiten, wurden sowohl Zusammenhangsmaße berechnet, als auch Mittelwertunterschiede mit Hilfe inferenzstatistischer Verfahren überprüft. Bei allen Mittelwertdarstellungen und -vergleichen wurden die Müdigkeitsbewertungen z-standardisiert, um eine Vergleichbarkeit der Bewertungen zwischen den drei Skalen zu gewährleisten. Weiterhin wurden Effektgrößen für jeden der durchgeführten Signifikanztests berechnet, um die Bedeutsamkeit der gefundenen Unterschiede zu beurteilen. Dabei steht ein partielles eta-square ( $\eta^2_{\text{part}}$ ) von  $\eta^2_{\text{part}} = 0.01$  für einen kleinen,  $\eta^2_{\text{part}} = 0.06$  für einen mittleren und  $\eta^2_{\text{part}} = 0.14$  für einen großen Effekt (Sedlmeier & Renkewitz, 2008).

Zur Beantwortung der Frage nach möglichen Bewertungsunterschieden aufgrund der drei Skalen (Fragestellung 1) wurden zunächst die z-Werte jedes Videoausschnitts pro Skala über die Rater gemittelt. Ein Vergleich per ANOVA mit Messwiederholung, bei dem die Videos als Zufallsvariable fungieren, zeigt einen signifikanten und großen Mittelwertunterschied zwischen den Skalen ( $F(2,348) = 12.53$ ,  $p < 0.001$ ,  $\eta^2_{\text{part}} = 0.17$ ). Dabei werden die Fahrer mit der HFC- und der TUBS-Skala als signifikant müder eingeschätzt als mit der Wierwille-Skala.

Zur Einschätzung der Objektivität (Fragestellung 2) wurde zunächst die Interrater-Reliabilität berechnet (Tabelle 1). Dabei ergab sich für alle drei Skalen eine sehr hohe durchschnittliche Übereinstimmung der Bewerter mit den folgenden Werten: HFC-Skala  $r = 0.80$ ,  $p < 0.001$ , Wierwille-Skala  $r = 0.76$ ,  $p < 0.001$  und TUBS-Skala  $r = 0.84$ ,  $p < 0.001$ .

Diese Analyse wird ergänzt durch einen Vergleich der mittleren Ratings zwischen den Bewertern anhand von Varianzanalysen mit Messwiederholung über einen Faktor. Dabei zeigt sich, dass sich die Bewerter in ihren Müdigkeitseinschätzungen bei Verwendung der HFC-Skala signifikant voneinander unterscheiden ( $F(3,174) = 92.09$ ,  $p < 0.001$ ,  $\eta^2_{\text{part}} = 0.61$ ). Aufgrund des partiellen  $\eta^2$  ( $\eta^2_{\text{part}}$ ) ist der zugrunde liegende Effekt als extrem groß zu bewerten. Auch für die Wierwille-Skala zeigen sich signifikante Unterschiede zwischen den Ratern, die aber lediglich als mittelgroß einzustufen sind ( $F(3,174) = 7.61$ ,  $p < 0.001$ ,  $\eta^2_{\text{part}} = 0.11$ ) und sehr viel geringer ausfallen als bei der HFC-Skala. Vergleichbar zu den Ergebnissen der Wierwille-Skala existieren auch bei der TUBS-Skala signifikante Unterschiede zwischen den Bewertern, die ebenfalls um ein vielfaches geringer sind als bei der HFC-Skala ( $F(3,174) = 22.16$ ,  $p < 0.001$ ,  $\eta^2_{\text{part}} = 0.27$ ).

Die Untersuchung der *Reliabilität* (Fragestellung 3) zeigt, dass sich alle drei Skalen durch eine hohe Retest-Reliabilität auszeichnen. Als Werte ergaben sich für die HFC-Skala  $r = 0.93$  ( $p < 0.001$ ), für die Wierwille-Skala  $r = 0.90$  ( $p < 0.001$ ) und die TUBS-Skala  $r = 0.93$  ( $p < 0.001$ ). Allerdings weisen auch hier Mittelwertdifferenzen auf Unterschiede zwischen den Skalen hin. Vergleicht man die Urteile zu den beiden Messzeitpunkten miteinander, so zeigt sich, dass die Müdigkeit im Durchschnitt im zweiten Durchgang mit allen drei Skalen als geringer eingestuft wird als im ersten. Dieser Effekt ist bei der HFC-Skala signifikant und extrem groß ( $F(1,174) = 89.91$ ,  $p < 0.001$ ,  $\eta^2_{\text{part}} = 0.60$ ), bei der Wierwille-Skala zwar signifikant, aber lediglich mittelgroß, ( $F(1,174) = 5.89$ ,  $p = 0.02$ ,  $\eta^2_{\text{part}} = 0.09$ ) und bei der TUBS-Skala nicht signifikant und vernachlässigbar klein ( $F(1,174) = 1.09$ ,  $p = 0.30$ ,  $\eta^2_{\text{part}} = 0.01$ ).

Tab. 1. Interrater-Korrelationen

		Müdigkeitsbewerter				
			VP2	VP3	VP4	
Skala	HFC	VP1	0.78	0.82	0.80	
		VP2		0.86	0.78	
		VP3			0.79	
	Wierwille		VP6		VP7	VP8
		VP5	0.82	0.75	0.88	
		VP6		0.66	0.86	
		VP7			0.73	
	TUBSS		VP10	VP11	VP12	
		VP9	0.73	0.75	0.72	
		VP10		0.87	0.90	
			VP11		0.93	

## Diskussion

Um die Güte von FAS zur automatischen Erkennung der Fahrermüdigkeit ermitteln zu können, werden objektive und reliable Beobachtungsverfahren benötigt. In der vorliegenden Arbeit wurden drei Müdigkeitseinschätzungsskalen hinsichtlich ihrer psychometrischen Gütekriterien überprüft. Dabei interessierte, (1) ob sich die Urteile von geschulten Bewertern in Abhängigkeit von der verwendeten Skala unterscheiden, (2) wie objektiv jede der drei Skalen Müdigkeit misst und (3) wie reliabel die bei einer Messung erzielten Ergebnisse sind.

Ein Vergleich der mittleren Urteile unter allen drei Skalen verdeutlicht, dass die erste Frage zu bejahen ist. Es macht einen Unterschied, welche Skala verwendet wird, wobei auf Basis der

HFC- und der TUBS-Skala im Durchschnitt die Müdigkeit der Fahrer höher eingeschätzt wird als auf Basis der Wierwille-Skala.

Die beiden Fragen nach der Objektivität und Reliabilität müssen differenziert beantwortet werden. Betrachtet man die Interrater- und Intrarater-Korrelationen, so beeindrucken alle Verfahren durch ausgesprochen hohe Werte. Analysiert man allerdings zusätzlich die entsprechenden inter- und intraindividuellen Mittelwertunterschiede, so zeigen sich Vorteile für die Wierwille- und die TUBS-Skala. Mit ihnen werden sowohl innerhalb der Bewertergruppen homogenere Ergebnisse erzielt, als auch konsistentere Urteile eines Bewerbers zu verschiedenen Messzeitpunkten gewonnen. Zwar zeigen die Analysen signifikante intra- und interindividuelle Unterschiede auf, die die Objektivität und Reliabilität beider Verfahren schmälern, doch sind die zugrunde liegenden Effekte gering und daher nicht von praktischer Relevanz.

Als Fazit kann daher festgehalten werden, dass es mit der TUBS-Skala gelungen ist, die Vorteile der Wierwille- und der HFC-Skala in einem einheitlichen Verfahren zu integrieren. Das zweistufige Vorgehen bei der Beurteilung von Müdigkeit gewährleistet nicht nur eine genaue Zuordnung von Indikatoren zu vier zentralen Müdigkeitswerten, sondern erlaubt darüber hinaus durch die Bereitstellung von Übergangswerten die Spezifizierung von Tendenzen und damit eine weitergehende Differenzierung der Urteile.

Hinsichtlich der Gütekriterien Objektivität und Reliabilität kann die TUBSS als gut abgesichert angesehen werden; in Bezug auf das dritte zentrale psychometrische Gütekriterium - die Validität - besteht allerdings noch Forschungsbedarf. Da es bisher kein allgemein akzeptiertes Außenkriterium für Müdigkeit gibt, kann generell für beobachtungsbasierte Verfahren, wie die TUBSS, keine Kriteriumsvalidität ermittelt werden. In Sachen Inhaltsvalidität spricht allerdings einiges für die neue Skala, da durch die Einbeziehung der Wierwille-Indikatoren auf Beobachtungsebene „das zu messende Konstrukt [Müdigkeit] in seinen wichtigsten Aspekten erschöpfend erfasst“ wird (Bortz und Döring, 2003, S. 199). Unsere zukünftige Forschung wird sich um die Validierung der TUBS-Skala anhand verlässlicher Außenkriterien bemühen und außerdem durch die experimentelle Überprüfung weiterer differenzierter Hypothesen zur Fahrerermüdung eine möglichst hohe Konstruktvalidität der Skala anstreben.

## Literatur

- Bortz, J. & Döring, N. (2003). *Forschungsmethoden und Evaluation*. Berlin, Heidelberg, New York: Springer.
- Dittrich, Brandenburg & Thüring (2009). *Verfahren zur Erfassung von Müdigkeit im Kfz. Blutalkohol*, 46 (2), S.148.
- Evers, C., & Auerbach, K. (2005). *Verhaltensbezogene Ursachen schwerer Lkw-Unfälle*. In *Berichte der Bundesanstalt für Straßenwesen. Mensch und Sicherheit*. Heft M174.
- Kolrep, H., Rimini-Döring, M., Oehme, A., Jürgensohn, T., & Altmüller, T. (2005). *Wie sieht „müde“ aus? – Entwicklung und Validierung einer Skala zur Müdigkeitsbewertung von Kraftfahrern*. In L. Urbas, & C. Steffens (Hrsg.), 6. Berliner Werkstatt Mensch-Maschine-Systeme, Zustandserkennung und Systemgestaltung, 13. – 15. Okt. 2005, Berlin, S. 65-70. (Fortschrittsberichte VDI, Reihe 22 MMS, Nr. 22). Düsseldorf: VDI.
- Sedlmeier, P. & Renkewitz, F. (2008). *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson.
- Statistisches Bundesamt (2009). Abruf 14.Mai.2009, <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Verkehr/Verkehrsunfaelle/Tabellen/Content75/UnfaelleVerunglueckte,templateId=renderPrint.psml>
- Wierwille, W. W., & Ellsworth, L. A. (1994). *Evaluation of driver drowsiness by trained raters*. *Accident Analysis and Prevention*, 26(5), S. 571-581.