

FE 82.326

Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung

## **Manual: Prinzipien zur Erstellung von multimedialen Testaufgaben für die Fahrerlaubnisprüfung**

19.12.2011

Dipl.-Psych. Sarah Malone

Dr. Antje Biermann

Univ.-Prof. Dr. Roland Brünken

Univ.-Prof. Dr. Susanne Buch

Fachrichtung Bildungswissenschaften

Universität des Saarlandes

Im Auftrag der Bundesanstalt für Straßenwesen,

Bergisch Gladbach

# Inhaltsverzeichnis

Inhaltsverzeichnis.....	2
1 Einleitung: Die theoretische Fahrerlaubnisprüfung als psychometrischer Test und Möglichkeiten der Optimierung .....	4
2 Gütekriterien Psychometrischer Tests.....	5
2.1 Hauptgütekriterien .....	6
2.1.1 Objektivität.....	6
2.1.2 Reliabilität .....	8
2.1.3 Validität.....	9
2.2 Nebengütekriterien .....	11
2.2.1 Ökonomie .....	12
2.2.2 Testfairness.....	13
2.3 Implikationen für die Erstellung von Aufgaben für die Fahrerlaubnisprüfung.....	15
3 Kriteriumsorientiertes Testen .....	16
3.1 Kriteriumsorientierte vs. normorientierte Testung .....	16
3.2 Vorgehensweise bei der Konstruktion kriteriumsorientierter Tests .....	17
3.3 Lehrziele .....	21
3.4 Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung .....	24
4 Typen von Testaufgaben .....	25
4.1 Gebundene Antwortformate .....	25
4.1.1 Richtig-Falsch-Aufgaben .....	26
4.1.2 Mehrfachwahlaufgaben.....	26
4.1.3 Zuordnungsaufgaben.....	29
4.1.4 Ordnungs- oder Umordnungs-Aufgaben.....	30
4.2 Freie Antwortformate .....	31
4.2.1 Ergänzungs-Aufgaben.....	31
4.2.2 Kurzaufsatz-Aufgaben .....	32
4.3 Handlungsnahe Antwortformate.....	33
4.3.1 Reaktionszeitaufgaben .....	34
4.3.2 Blickbewegungsaufgaben.....	35
4.3.3 Fahrsimulation.....	36

4.4	Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung .....	37
5	Gestaltung multimedialer Lern- und Prüfmaterialien .....	38
5.1	Gestaltungsdimensionen für multimediales Aufgabenmaterial.....	39
5.1.1	Kodalität .....	39
5.1.2	Modalität .....	39
5.1.3	Dynamik/Animation.....	40
5.1.4	Interaktivität/ Adaptivität .....	41
5.1.5	Immersion.....	43
5.2	Gestaltungsprinzipien .....	44
5.2.1	Generelle Gestaltungsprinzipien für Lernmaterial.....	44
5.2.2	Prinzipien zur Gestaltung dynamisch präsentierten Lernmaterials.....	51
5.3	Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung .....	55
6	Zusammenfassung.....	56
7	Checklisten zur Gestaltung von Testaufgaben.....	57
7.1	Checkliste Gestaltung von Texten.....	57
7.2	Checkliste Gestaltung von Abbildungen .....	57
7.3	Checkliste Gestaltung von Animationen .....	58
7.4	Checkliste Gestaltung von Multimediamaterial .....	58
8	Literaturverzeichnis.....	59

# 1 Einleitung: Die theoretische Fahrerlaubnisprüfung als psychometrischer Test und Möglichkeiten der Optimierung

Um dem überhöhten Unfallrisiko, dem Fahrer kurz nach Erhalt ihrer Fahrerlaubnis ausgesetzt sind, entgegen zu wirken, werden in Deutschland sowie international unterschiedliche Maßnahmen geprüft. Neben Ansätzen, die eine Verbesserung der *Fahrausbildung* verfolgen, besteht ein Ansatz darin, die Fahrerlaubnisprüfung zu optimieren und damit zu gewährleisten, dass Personen sicher identifiziert werden, die noch nicht über genügend Fahrkompetenzen verfügen, um sich sicher im Verkehr zu bewegen.

Bei der Erstellung von geeigneten Testaufgaben für die theoretische Fahrerlaubnisprüfung ist es von zentraler Bedeutung, dass durch die Aufgaben relevante Kompetenzbereiche standardisiert und zuverlässig gemessen werden können, also die Qualitätsstandards (psychometrische Gütekriterien) erfüllt werden. Weiterhin sollten die Aufgaben so gestaltet sein, dass sie bestimmte Personengruppen nicht systematisch vom Bestehen ausschließen (Testfairness). Darüber hinaus sollte ihre instruktionale Gestaltung an empirisch fundierten lehr-lerntheoretischen Kriterien ausgerichtet sein. Die Verwendung computerbasierter Testverfahren bietet dabei sowohl einige Vorteile hinsichtlich der Test-Gütekriterien als auch hinsichtlich handlungsnäherer Aufgabenformate.

Die theoretische Fahrerlaubnisprüfung wurde in Deutschland bereits von Papier-Bleistift-Testung auf Computertestung umgestellt. Allein diese Umstellung hat das Potential, die Güte der theoretischen Fahrerlaubnisprüfung als psychometrisches Testverfahren zu verbessern. Um jedoch die Vorteile neuer Medien zur Kompetenzerfassung optimal zu nutzen, bietet es sich an, auch die Möglichkeit des Einsatzes innovativer Präsentations- und Aufgabenformate zu erwägen, wodurch nicht nur eine Prüfung deklarativen sondern auch prozeduralen Wissens ermöglicht wird.

Zunächst werden in Kapitel 2 die Gütekriterien psychometrischer Testverfahren vorgestellt. Neben den Hauptgütekriterien Objektivität, Reliabilität und Validität werden die beiden Nebengütekriterien Ökonomie und Testfairness behandelt, die für die Fahrerlaubnisprüfung als besonders relevant angesehen werden.

Da es sich bei der Fahrerlaubnisprüfung um einen kriteriumsorientierten Test handelt, wird in Kapitel 3 zunächst die Vorgehensweise bei der Konstruktion kriteriumsorientierter Tests dargestellt und darauf aufbauend werden Implikationen für die Vorgehensweise bei der Erstellung von Aufgaben und Aufgabenformaten für die theoretische Fahrerlaubnisprüfung abgeleitet.

Kapitel 4 behandelt die konkrete Gestaltung von Testaufgaben. Es werden Typen von Testaufgaben beschrieben und es wird auf ihre Eignung für den Einsatz in der theoretischen Fahrerlaubnisprüfung eingegangen. Dabei werden zum einen die klassischen gebundenen und freien Antwortformate vorgestellt; zum anderen werden anschließend innovative handlungsnaher Antwortformate präsentiert und ihr potentieller Nutzen für die Verbesserung der theoretischen Fahrerlaubnisprüfung dargelegt.

In Kapitel 5 wird die konkrete Gestaltung von Lernmaterial behandelt und es werden Ableitungen für die Gestaltung von Testmaterial für die Fahrerlaubnisprüfung gezogen. Dabei wird die Präsentation multimedialer Materials fokussiert, wobei Erkenntnisse aus der Lernforschung auf das Prüfen mit Multimediamaterial übertragen werden. Von besonderem Interesse sind dabei die Gestaltungsprinzipien für dynamisch präsentiertes Bildmaterial.

Im Anschluss an jedes Kapitel werden Implikationen für die Konstruktion von Aufgaben für die theoretische Fahrerlaubnisprüfung formuliert. Daher kann das Manual als Orientierungshilfe bei der Erstellung neuer Aufgaben und neuartiger Aufgabentypen für die Fahrerlaubnisprüfung dienen.

## **2 Gütekriterien Psychometrischer Tests**

Bei der theoretischen Fahrerlaubnisprüfung handelt es sich um einen Leistungstest, dessen Bestehen für die Prüflinge in höchstem Maße bedeutsam ist. Eine Fahrerlaubnis zu besitzen hat einerseits in der Regel vor allem in strukturschwachen Gebieten eine besonders hohe Relevanz für berufstätige Personen. Andererseits stellen Personen, die noch nicht über ausreichende Fahrkompetenzen verfügen, eine potentielle Gefahr für ihr eigenes Leben sowie für das Leben anderer Verkehrsteilnehmer dar. Die diagnostische Entscheidung über Bestehen oder Nicht-Bestehen ist demnach schwerwiegend und diagnostische Fehlentscheidungen sollten weitestgehend vermieden werden. Dabei sind zwei Arten der Fehlklassifikation möglich, die Alpha- und Beta-Fehler bzw. Fehler 1. und 2. Art genannt werden. Bild 2-1 veranschaulicht die beiden Arten der diagnostischen Fehlklassifikation durch die Fahrerlaubnisprüfung. Unterläuft bei der Klassifikation der Prüflinge in Können und Nicht-Können durch die Fahrerlaubnisprüfung ein Beta-Fehler, so bedeutet dies, dass die Kompetenz des betreffenden Prüflings überschätzt wurde. Er hat demnach die Prüfung bestanden, obwohl er in Wirklichkeit noch nicht kompetent genug ist. Im Gegensatz dazu liegt ein Alpha-Fehler dann vor, wenn ein Prüfling durch die Prüfung unterschätzt wurde. Der betreffende Prüfling hat die Prüfung demnach nicht bestanden, obwohl er eigentlich kompetent ist.

		Prüfungsergebnis	
		bestanden	nicht bestanden
Prüfling ist tatsächlich ...	kompetent	korrekte Klassifikation	Alpha-Fehler
	nicht kompetent	Beta-Fehler	korrekte Klassifikation

**Bild 2-1 Alpha- und Beta-Fehler bei der diagnostischen Entscheidung durch die Fahrerlaubnisprüfung**

Um zu gewährleisten, dass die Fehlklassifikationsrate möglichst gering ausfällt, sollte die theoretische Fahrerlaubnisprüfung den Gütekriterien psychometrischer Tests Genüge leisten. Im Folgenden werden die Haupt- und Nebengütekriterien und die Möglichkeiten ihrer Gewährleistung und Überprüfung behandelt.

## 2.1 Hauptgütekriterien

### 2.1.1 Objektivität

Das erste Kriterium, anhand dessen die Güte eines psychometrischen Testverfahrens eingeschätzt werden kann, wird als Objektivität bezeichnet.

Definition: Objektivität bezeichnet das Ausmaß, in dem Testergebnisse unabhängig vom jeweiligen Untersuchungsleiter sind (Lienert & Raatz, 1998).

Objektive Testverfahren liefern demnach Ergebnisse, die in keinem Zusammenhang mit der Person stehen, die den Test durchgeführt und geleitet hat. Das bedeutet, dass man davon ausgehen kann, dass verschiedene Untersuchungsleiter bei derselben Messung zu einem identischen Ergebnis bezüglich des im Test gemessenen Merkmals gelangen.

Es werden drei Arten der Objektivität unterschieden: *Durchführungs-*, *Auswertungs-* und *Interpretationsobjektivität* (Amelang & Schmidt-Atzert, 2006).

Unter *Durchführungsobjektivität* versteht man das Ausmaß, in dem Unregelmäßigkeiten bei der Testdurchführung zu unterschiedlichem Verhalten und damit zu unterschiedlichen Testergebnissen der Teilnehmer führen. Eher gering ist die Durchführungsobjektivität einzuschätzen, wenn dem Versuchsleiter beispielsweise freigestellt würde, wie er den Testteilnehmern den Ablauf des Tests und die Aufgabenstellungen erklärt. Hohe Durchführungsobjektivität kann hingegen durch weitestgehende Standardisierung der Testsituation gewährleistet werden.

Wichtig ist, dass es feste Zeitvorgaben für die Bearbeitung gibt. Begünstigend ist zudem, die Interaktion zwischen Testleiter und Teilnehmer möglichst gering zu halten.

Unerlässlich für jedes Testverfahren ist gleichwertiges Testmaterial für alle Testteilnehmer. Hier wird die sogenannte *spezifische Objektivität* relevant, die besagt, dass das Testergebnis unabhängig von der Auswahl von Testaufgaben aus einem (hypothetischen) Aufgabenpool sein sollte (Rost, 2004).

Die *Auswertungsobjektivität* bezieht sich darauf, ob jegliches Testverhalten einer Person eindeutig einem Testwert zugeordnet werden kann. Für Leistungstests liegt meist ein Lösungsschlüssel zur Auswertung vor, dem zweifelsfrei entnommen werden kann, welche Antworten als richtige Lösung gewertet werden. Auswertungsobjektivität kann durch den Einsatz standardisierter Auswertungsroutinen, z.B. mithilfe von Schablonen oder Computerprogrammen gewährleistet werden. Als problematisch hinsichtlich ihrer Auswertungsobjektivität gelten z.B. so genannte projektive Verfahren. Freiantworten, die bei projektiven aber auch anderen Testverfahren gefordert werden, können in den meisten Fällen nur schwer eindeutig bewertet werden. Eine Unterform der Auswertungsobjektivität, die *Signierobjektivität*, bezieht sich auf diesen Aspekt der Objektivität, nämlich, ob die Kodierung von Freiantworten unabhängig von der Person des Kodierers ist (Rost, 2004).

Um die Auswertungsobjektivität quantitativ zu bestimmen, wird die Übereinstimmung der Ergebnisse eines Tests zwischen mindestens zwei Auswertern berechnet. Dazu wird mehreren Auswertern die identische Testbearbeitung einer Testperson vorgelegt, für die sie einen Punktwert bestimmen sollen. Hohe Übereinstimmungen dieses Punktwertes zwischen den Auswertern sprechen für eine hohe Auswertungsobjektivität.

*Interpretationsobjektivität* bezeichnet das Ausmaß, in dem das gleiche Testergebnis von unterschiedlichen Personen gleichermaßen interpretiert oder bewertet wird. Zur Wahrung der Interpretationsobjektivität liegen für die meisten standardisierten Testverfahren Kategoriensysteme vor, die bestimmen, wie die Testwerte zu interpretieren sind. Dies ermöglicht eine inhaltliche Interpretation der Rohwerte, die eine Person in einem Test erreicht hat. In der Regel liegen auch sogenannte Normtabellen vor, mit deren Hilfe z.B. die Leistung einer Person in die Leistungsverteilung der Personen ihrer Altersgruppe oder Gruppe der Personen mit dem gleichen Bildungsgrad oder ähnlichem eingeordnet wird. Zur Gewährleistung der Interpretationsobjektivität werden demnach stets aktuelle Normen benötigt. Im Falle der theoretischen Fahrerlaubnisprüfung stellt die Interpretationsobjektivität kein Problem dar, da Personen bis zu einer festgesetzten Fehlerzahl die Prüfung auf jeden Fall bestehen und bei Überschreitung der kritischen Fehlerzahl immer durch die Prüfung fallen.

Weiterführende Klassifikationen werden durch diese Prüfung nicht vorgenommen. Auch ist für jede einzelne Aufgabe eindeutig festgelegt, wann diese als richtig bzw. falsch gelöst interpretiert wird. Es werden keine Teilpunkte vergeben.

Objektive Messungen gelten als Voraussetzung für die im Folgenden erläuterten Gütekriterien Reliabilität und Validität einer Messung.

### **2.1.2 Reliabilität**

Das zweite Gütekriterium Reliabilität wird auch Zuverlässigkeit einer Messung genannt.

Definition: Unter Reliabilität oder Zuverlässigkeit versteht man die Genauigkeit, mit der ein Test ein bestimmtes Merkmal misst (Lienert & Raatz, 1998).

Unter dem Gesichtspunkt der Reliabilität wird die Präzision einer Messung bewertet. Eine Reliabilitätsmessung kann auf verschiedene Arten erfolgen, wobei die Entscheidung für ein Verfahren von der Beschaffenheit des Merkmals, das erfasst werden soll, abhängt. Man kann zwischen *Retest-Reliabilität*, *Paralleltestreliabilität* sowie Verfahren zur Messung der *internen Konsistenz* unterscheiden (Lienert & Raatz, 1998).

Unter *Retest-Reliabilität* wird der Zusammenhang zwischen den Ergebnissen identischer Testpersonen bei zwei Messungen unter Verwendung des gleichen Messinstrumentes verstanden. Wenn Personen, die zum zweiten Mal mit dem gleichen Messverfahren getestet werden, vergleichbare Ergebnisse zur ersten Testung erzielen, kann die Retest-Reliabilität des Testverfahrens als gesichert gelten. Dabei muss allerdings beachtet werden, dass die Reliabilität nicht höher sein kann als die Stabilität des Merkmals selbst. Verändert sich die Ausprägung eines Merkmals bei einer Person über die Zeit, sollte der Zusammenhang des Ergebnisses der ersten Testung mit dem Ergebnis der Testwiederholung eher gering ausfallen. Bei Merkmalen, die stabil bleiben, wie z.B. dem Intelligenzquotienten, erscheint eine Überprüfung der Reliabilität mit der Retest-Methode geeignet. Eher ungeeignet ist sie hingegen bei Leistungstests, bei denen man besonders bei einer zeitnahen Wiederholung mit Lern- oder Übungseffekten rechnen kann.

Um die Reliabilität von Testverfahren zu bestimmen, die zeitlich wenig stabile Merkmale messen, eignet sich die *Paralleltestmethode*. Zur Bestimmung der *Paralleltestreliabilität* werden den Teilnehmern parallele Formen eines Tests vorgelegt und die Übereinstimmung der Ergebnisse berechnet. Zwei Testformen können dann als parallel betrachtet werden, wenn die Mittelwerte und Standardabweichungen der Ergebnisse identisch sind. Die Paralleltestreliabilität ist als gegeben einzuschätzen, wenn die Ergebnisse der Personen in den

unterschiedlichen Testformen weitestgehend übereinstimmen. Nachteilig bei dieser Methode der Reliabilitätsbestimmung ist, dass parallele Testformen erstellt werden müssen.

Die Nachteile der Bestimmung der Reliabilität mithilfe einer Testwiederholung oder dem Einsatz von Paralleltests kann man durch Verfahren der internen Konsistenz vermeiden. Der Test muss dabei den Teilnehmern nur ein einziges Mal vorgegeben werden und es sind keine parallelen Testversionen notwendig. Übungs- und Lerneffekte können vermieden werden und die Stabilität des Merkmals selbst hat keinen Einfluss auf die Reliabilitätsmessung.

Zur Bestimmung der Reliabilität mithilfe der *Split-Half-Methode* wird der Test in zwei gleiche Hälften geteilt. Die Halbierung kann auf unterschiedliche Arten erfolgen (z.B. per Zufall, *odd-even-Methode*, Gleichverteilung von Items mit ähnlichen Kennwerten, etc.). Ein hoher Korrelationskoeffizient der Ergebnisse der beiden Testhälften, also ein hoher Zusammenhang zwischen Leistungen in beiden Testhälften, lässt auf hohe Reliabilität des Testverfahrens schließen. Eine Verallgemeinerung der Split-Half-Methode stellt die *Konsistenzanalyse* dar. Dabei wird jede Aufgabe des Tests als eigenständiger Testteil betrachtet. Die sogenannte innere oder interne Konsistenz ist umso höher, je stärker die Testteile untereinander und damit mit der Gesamtskala positiv korrelieren. Ein bekanntes Maß zur Bestimmung der internen Konsistenz ist Cronbachs  $\alpha$ -Koeffizient (Cronbach, 1951). Für den Einsatz der Split-Half-Methode und der Konsistenzanalyse zur Bestimmung der Reliabilität wird allerdings vorausgesetzt, dass das zu messende Konstrukt homogen ist, d.h. dass das Konstrukt nicht aus mehreren unterscheidbaren Teilkonstrukten besteht.

### 2.1.3 Validität

Das dritte der Hauptgütekriterien – die Validität – stellt das wichtigste Gütekriterium dar.

Definition: Die Validität gibt den Grad der Genauigkeit an, mit dem ein Test dasjenige (Persönlichkeits-) Merkmal misst, das er messen soll bzw. vorgibt zu messen (Lienert & Raatz, 1998).

Es müssen demnach bei der Validierung eines Testverfahrens Belege dafür gesammelt werden, dass das Testergebnis in einem hohen positiven Zusammenhang mit der tatsächlichen Ausprägung des interessierenden Merkmals steht. Weist ein Test eine hohe Validität auf, erlaubt dies den Rückschluss vom Testergebnis auf Verhaltensmerkmale außerhalb der Testsituation, also auf die Generalisierbarkeit der Testergebnisse (*ökologische Validität*). Bei der Beurteilung der Validität eines Testverfahrens werden in der Regel verschiedene Aspekte

oder Arten der Validität berücksichtigt. Dazu gehören Inhaltsvalidität, Konstruktvalidität und kriteriumsbezogene Validität (Amelang & Schmidt-Atzert, 2006).

Zur Bestimmung der *Inhaltsvalidität* wird überprüft, ob das Testverfahren das interessierende Merkmal repräsentativ erfasst. Für diese Art der Validität, bei der eingeschätzt werden soll, ob es sich bei den zu einem Test zusammengefassten Testaufgaben um eine repräsentative Auswahl aus dem Itempool handelt, gibt es keinen Kennwert. Die Einschätzung der inhaltlichen Validität erfolgt in der Regel durch Expertenurteile.

Inhaltlich valide Tests sind häufig auch *augenscheinvalide*. Die Bestandteile augenscheinvalider Tests erscheinen selbst dem Laien gerechtfertigt. Auch wenn der Augenscheinvalidität eine gewisse Bedeutung hinsichtlich der allgemeinen Akzeptanz eines Testverfahrens zukommt, ist sie aus der wissenschaftlichen Perspektive nicht als ausreichend zu betrachten. Das Gütekriterium Validität gilt für ein Testverfahren erst als erwiesen, wenn auch empirische Belege für seine Gültigkeit vorliegen.

Die *Konstruktvalidität* bezeichnet das Ausmaß, in dem die Operationalisierung eines Konstruktes in einem Test Rückschlüsse auf dieses nicht direkt messbare Konstrukt zulässt. Man kann diese Art der Validität mit Hilfe des Zusammenhangs zwischen Ergebnissen im zu beurteilenden Test und anderen Testverfahren bestimmen. Dabei kann man zwischen *konvergenter* und *diskriminanter Validität* unterscheiden. Um die konvergente Validität zu ermitteln, wird die Übereinstimmung zwischen den Ergebnissen in dem zu validierenden Test mit Testergebnissen anderer Tests, die das gleiche oder ein ähnliches Konstrukt messen und bereits etabliert sind, bestimmt. Hohe Korrelationen deuten darauf hin, dass beide Tests etwas Ähnliches erfassen. Umgekehrt ist die fehlende Übereinstimmung zwischen dem zu validierenden Test und einem Test, der ein anderes Konstrukt messen soll, als Nachweis für die diskriminante Validität zu werten.

Zur Bestimmung der *kriteriumsbezogenen Validität* eines Testverfahrens werden die Zusammenhänge der Testergebnisse mit einem so genannten *Außenkriterium* ermittelt. Bei Außenkriterien handelt es sich um Verhaltensmaße, die außerhalb der Testsituation erfasst werden und Rückschlüsse auf die tatsächliche Ausprägung des zu messenden Merkmals zulassen. Ein Kriterium für das Konstrukt Fahrkompetenz, das durch die Fahrerlaubnisprüfung gemessen werden soll, kann z.B. die Fahrerfahrung, gemessen in Kilometern, sein. Durch ihre praktische Übung und Erfahrung sollten Fahrer, die bereits größere Strecken mit einem Verkehrsmittel zurückgelegt haben, über mehr Fahrkompetenz verfügen als unerfahrene Fahrer. Die Fahrerlaubnisprüfung wäre demnach dann als kriteriumsvalide zu bezeichnen, wenn es hohe positive Zusammenhänge zwischen dem

Testergebnis und der Anzahl der gefahrenen Kilometer gäbe. Weitere sinnvolle Kriterien anhand derer man die Kriteriumsvalidität der Fahrerlaubnisprüfung bestimmen könnte, wären z.B. die Unfallhäufigkeit oder die Anzahl der Verwarnungsgelder in einem bestimmten Zeitintervall.

Man kann zwei Arten der kriteriumsbezogenen Validität unterscheiden. Zum einen lässt sich die *Übereinstimmungsvalidität* (auch *konkurrente* Validität) durch die Höhe des Zusammenhangs zu gleichzeitig mit dem Test erhobenen Kriterien bestimmen. Z.B. könnte ein Kriterium zur Bestimmung der konkurrenten Validität das Ergebnis einer standardisierten Befragung des Fahrlehrers zu den Fähigkeiten des Prüflings sein, die zur gleichen Zeit wie die theoretische Fahrerlaubnisprüfung stattfindet. Die *Vorhersagevalidität* (auch *prognostische* Validität) hingegen lässt sich über den Zusammenhang von Testergebnis und einem zukünftigen Kriterium ermitteln. So stützen z.B. negative Korrelationen zwischen den Ergebnissen in der theoretischen Fahrerlaubnisprüfung und der Anzahl selbst verschuldeter Unfälle im ersten Führerscheinjahr die prognostische Validität der Prüfung. Tabelle 2-1 veranschaulicht hypothetische Beziehungen zwischen den Testergebnissen in der theoretischen Fahrerlaubnisprüfung und verschiedenen Kriterien, die für ihre *hohe* konkurrente und prognostische Validität sprächen.

	<b>Testergebnis in der theoretischen Fahrerlaubnisprüfung</b>	
	<b>Hohe Punktzahl</b>	<b>Niedrige Punktzahl</b>
<b>Konkurrente Validität</b>	<ul style="list-style-type: none"> <li>• hohe Punktzahl in der Befragung des Fahrlehrers</li> <li>• gute Leistungen bei einer Simulatorfahrt</li> </ul>	<ul style="list-style-type: none"> <li>• niedrige Punktzahl in der Befragung des Fahrlehrers</li> <li>• schlechte Leistungen bei einer Simulatorfahrt</li> </ul>
<b>Prognostische Validität</b>	<ul style="list-style-type: none"> <li>• keine selbstverschuldeten Unfälle im ersten Führerscheinjahr</li> <li>• keine Verwarnungsgelder</li> <li>• positive Einschätzung der Entwicklung der Fahrkompetenzen durch die Begleitperson</li> </ul>	<ul style="list-style-type: none"> <li>• mehrere selbstverschuldete Unfälle im ersten Führerscheinjahr</li> <li>• mehrere Verwarnungsgelder</li> <li>• negative Einschätzung der Entwicklung der Fahrkompetenzen durch die Begleitperson</li> </ul>

**Tabelle 2-1** Mögliche Hinweise auf eine hohe konkurrente und prognostische Validität der theoretischen Fahrerlaubnisprüfung

## **2.2 Nebengütekriterien**

Neben den drei Hauptgütekriterien psychometrischer Tests gibt es eine Vielzahl von Nebengütekriterien. Ihre Relevanz hängt von der Fragestellung ab, auf die ein Test abzielt,

von der Wichtigkeit einer diagnostischen Entscheidung, die auf der Grundlage des Testergebnisses getroffen werden soll und von der Anzahl von Personen, über deren Leistung oder Merkmalsausprägung mit Hilfe des Testverfahrens entschieden werden soll.

Für Tests, wie beispielsweise die Fahrerlaubnisprüfung, deren Ergebnis hohe persönliche und gesellschaftliche Wichtigkeit besitzt und die besonders viele Personen absolvieren, sind besonders die Nebengütekriterien *Ökonomie* und *Testfairness* relevant.

### **2.2.1 Ökonomie**

Mit dem Begriff Ökonomie wird die Wirtschaftlichkeit eines Testverfahrens bezeichnet (Moosbrugger & Kelava, 2008). Bei einer Testung entstehen in der Regel Kosten, die vom Testteilnehmer, verschiedenen Institutionen oder beiden zu gewissen Teilen getragen werden müssen. Es wird in der Regel angestrebt, die Kosten einer Testung so gering zu halten, wie es unter Wahrung der Hauptgütekriterien möglich ist. Ökonomiebestrebungen dürfen demnach nicht zu Lasten der Güte des Testverfahrens gehen.

Testverfahren können auf den Ebenen Anschaffung, Durchführung und Auswertung eher ressourcensparend oder zeit- und kostenintensiv sein. Im Folgenden wird diskutiert, inwiefern die theoretische Fahrerlaubnisprüfung ökonomischen Kriterien genügt und wie sich möglicherweise die Umstellung von einer Papier-Bleistift-basierten Version auf computerbasiertes Testen auf die Ökonomie ausgewirkt hat.

Die Anschaffung des Testmaterials war bei der ursprünglichen theoretischen Fahrerlaubnisprüfung hoch, denn das Testmaterial – vor der Umstellung noch in Form von Testbogen aus Papier – wurde bei der Testung verbraucht. Die Bereitstellung der aktuell notwendigen Computerhardware und Testsoftware kann allerdings ebenfalls beachtliche ökonomische Belastungen darstellen. Allerdings handelt es sich dabei um einmalige Ausgaben.

Die Aktualisierung des Testmaterials muss im Falle der Fahrerlaubnisprüfung relativ häufig erfolgen, nämlich immer dann, wenn neue Regeln eingeführt oder bestehende geändert werden, die auch abgeprüft werden sollen. Bei der ursprünglichen Testversion mussten aktualisierte Fragebögen erstellt sowie gedruckt und verteilt werden. In der Computerversion muss lediglich das Testprogramm, wenn möglich serverbasiert, aktualisiert werden.

Während die Erstellung paralleler Aufgaben und Testversionen bei der Fahrerlaubnisprüfung in ihrer Papier-Bleistift-Version noch aufwendig und unflexibel war, kann dies in einer computerbasierten Testversion relativ unproblematisch erfolgen. Parallele Testversionen können durch eine automatische Zusammenstellung von Testaufgaben, die in zufälliger Reihenfolge präsentiert werden können, aus dem Itempool generiert werden. Auch die

Reihenfolge von Antwortalternativen bei Multiple-Choice-Aufgaben kann automatisch variiert werden. Die einfache Einbindung neuer Aufgaben und Aufgabenformate geht schnell und ist damit ökonomisch.

Die Durchführung der theoretischen Fahrerlaubnisprüfung ist durch die Möglichkeit der Gruppentestung in der ursprünglichen, sowie in der neuen Darbietungsform als ökonomisch zu betrachten. Bei der computerbasierten Testung ist jedoch die Bereithaltung und Verteilung unterschiedlicher Paralleltestformen nicht mehr notwendig.

In der Papier-Bleistift-Version der theoretischen Fahrerlaubnisprüfung musste die Auswertung einzeln mithilfe von Auswertungsschablonen erfolgen. Deutliche Zeitersparnis bietet bei der computerbasierten Version die Möglichkeit der direkten automatischen Auswertung der Testergebnisse.

Noch ökonomischer wäre das adaptive Testen, das bei einer Computertestung und geeigneten Aufgaben gut durchführbar ist (*Computerisiertes Adaptives Testen, CAT*). Dabei ist eine Reduktion der Testzeit bei gleicher Güte der Testung zu erwarten, da jedem Testteilnehmer individuell Aufgaben zugeteilt werden, die genau seinem Leistungsniveau entsprechen. Zu leichte und zu schwierige Aufgaben, die keinerlei zusätzliche Informationen über die Kompetenz der Person liefern, fallen daher weg und die Testlänge verkürzt sich, ohne an Informationskraft einzubüßen.

### **2.2.2 Testfairness**

Unter fairem Testen wird verstanden, dass Personen nicht aufgrund ihrer Zugehörigkeit zu einer besonderen Personengruppe (sozioökonomischer Hintergrund, Geschlecht, Zugehörigkeit zu religiösen Gruppen etc.) systematisch bessere oder schlechtere Leistungen erbringen als Personen, die dieser Gruppe nicht zugehören (Moosbrugger & Kelava, 2008). Dieses Gütekriterium erhält besondere Wichtigkeit, wenn wichtige diagnostische Entscheidungen für eine große Anzahl an Personen auf der Grundlage ihres Ergebnisses in diesem Test getroffen werden.

Man kann die Fairness eines Testverfahrens schon bei der Konstruktion neuer Aufgaben oder Aufgabenformate positiv beeinflussen. So sollte man Fragestellungen, Beispiele oder Formulierungen vermeiden, die von unterschiedlichen Personengruppen unterschiedlich gut verstanden werden. Die verwendeten Begriffe könnten in Abhängigkeit von der Gruppenzugehörigkeit entweder weniger familiär sein, andere Bedeutungen könnten salienter sein oder sie könnten manchen Personen gänzlich unbekannt sein. Wichtig erscheint zudem, dass neue Aufgaben vor ihrem Einsatz auf ihre Testfairness geprüft werden, indem sie Vertretern aller Personengruppen zur Bearbeitung vorgelegt werden. Die dabei ermittelten,

nach Personengruppen getrennten Reliabilitäts-, Validitäts- und Objektivitätsindikatoren sollten weitestgehend übereinstimmen.

Bei einem computerbasierten Testverfahren, wie es die theoretische Fahrerlaubnisprüfung in Deutschland ist, sollte geprüft werden, ob eine Abhängigkeit von Testergebnissen mit Computererfahrung oder Computerängstlichkeit vorliegt, die zu kulturell, ethnisch oder geschlechtlich bedingter Benachteiligung beim computerbasierten Testen führen kann (Jurecka & Hartig, 2007). Es besteht zudem die Möglichkeit, den Umgang mit dem Testmaterial zu üben und somit allen Teilnehmern gleiche Chancen zu bieten. Den Fahrschülern sollte demnach ebenfalls die Möglichkeit gegeben werden, vor der Prüfung die Aufgaben in der Form zu üben, wie sie auch in der Prüfung selbst dargeboten werden.

Eine Besonderheit stellt der Einfluss intelligenzkorrelierter Merkmale der Person auf die Testleistung dar. Wenn davon auszugehen ist, dass das zu messende Merkmal ebenfalls in positivem Zusammenhang mit der Intelligenz steht, ist ein positiver Zusammenhang zwischen der Testleistung und Intelligenz bzw. der Ausprägung intelligenzkorrelierter Merkmale zu erwarten. Als problematisch ist zu werten, dass ein so genannter *Testbias* entsteht, wenn davon auszugehen ist, dass der Hauptanteil der Varianz in den Testleistungen durch Intelligenzunterschiede erklärt wird. Ein Testbias kann durch die Gestaltung des Testmaterials begünstigt werden. Es besteht nämlich die Möglichkeit, dass inhaltlich gleiche, aber anders gestaltete Testaufgaben in unterschiedlichem Ausmaß mit intelligenzkorrelierten Merkmalen zusammenhängen. In diesem Fall ist diejenige Präsentationsform zu wählen, bei der die geringsten Zusammenhänge zwischen Testleistung und Intelligenz zu erwarten sind.

Den Einfluss von sprachlichen Fähigkeiten, die selbst keinen Einfluss auf die Fahrkompetenz haben sollten, kann man durch den weitestgehenden Verzicht von Textpassagen in den Prüfungsaufgaben reduzieren. In der Fahrerlaubnisprüfung besteht die Möglichkeit, textuelle Beschreibungen von Verkehrssituationen durch Bilder, Bildreihen oder Videos zu ersetzen oder optional auditiv darzubieten. Komplizierte Satzgefüge sollten, ebenso wie Fremdwörter und wenig gebräuchliche Formulierungen in einer schriftlichen wie auch in einer auditiven Präsentation gänzlich vermieden werden.

Es hat sich zudem gezeigt, dass reine Textaufgaben der derzeitigen Fahrerlaubnisprüfung in einem deutlichen Zusammenhang mit räumlichen Fähigkeiten stehen, während dies bei Aufgaben mit Bildmaterial in geringerem Maße der Fall ist (siehe Bericht „*Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung*“). Auch dieser Befund spricht dafür, auf reine Textaufgaben zu verzichten.

## **2.3 Implikationen für die Erstellung von Aufgaben für die Fahrerlaubnisprüfung**

Bei der Fahrerlaubnisprüfung handelt es sich um ein psychometrisches Testverfahren, dessen Ergebnis die Grundlage einer für den Prüfling in höchstem Maße bedeutsamen Klassifikationsentscheidung darstellt. Um Fehlklassifikationen zu vermeiden, sollte bei der Erstellung neuer Aufgaben und Aufgabenformate gewährleistet werden, dass die wichtigsten Gütekriterien erfüllt sind.

Die Wahrung der Objektivität kann durch die möglichst hohe Standardisierung der Durchführung, Auswertung und Interpretation der Testergebnisse gewährleistet werden. Diese Forderung wird durch das computerbasierte Testen optimal erfüllt. Die Reliabilität von Testverfahren kann durch verschiedene Verfahren bestimmt werden. Die Reliabilität neuer Aufgaben und Aufgabenformate sollte vor ihrem tatsächlichen Einsatz in der Prüfung in angemessener Form überprüft werden. Das gleiche gilt für die Validität, wobei die kriteriumsbezogene Validität, also der Zusammenhang zwischen Testergebnis und Außenkriterien, als besonders relevant erachtet wird.

Ergänzend zu den drei Hauptgütekriterien sollten auch für die beiden Nebengütekriterien Ökonomie und Testfairness, die bei besonderer Tragweite diagnostischer Entscheidungen besonders relevant sind, Sorge getragen werden.

Das computerbasierte Testen hat sich als besonders ökonomisch erwiesen, weshalb neue Aufgaben und Aufgabenformate für die theoretische Fahrerlaubnisprüfung ebenfalls computerbasiert dargeboten werden sollten. Zudem sollte die Möglichkeit eines Einsatzes adaptiver Testverfahren geprüft werden.

Zur Wahrung der Testfairness in der Fahrerlaubnisprüfung sollte der Einfluss verschiedener Personenmerkmale auf die Testleistung in neuen Aufgaben und Aufgabenformaten immer vor einem Einsatz überprüft werden. Um den Einfluss von Computererfahrung und -ängstlichkeit auf die Leistung in der theoretischen Fahrerlaubnisprüfung zu minimieren, sollte den Fahrschülern die Möglichkeit zur Übung mit den computerbasierten Aufgaben gewährt werden.

Testmaterial sollte darüber hinaus so gestaltet werden, dass der Zusammenhang zwischen Testleistung und Intelligenz bzw. intelligenzkorrelierten Fähigkeiten sowie den sprachlichen Fähigkeiten möglichst gering ist.

### **3 Kriteriumsorientiertes Testen**

Die im vorangegangenen Kapitel beschriebenen Hauptgütekriterien gelten für sämtliche Test- und Fragebogenverfahren. Für die Messung von alltagsrelevanten Kompetenzen, wie die Fahrkompetenz eine ist, spielt insbesondere die Übertragung der gemessenen Werte auf tatsächliches Verhalten eine Rolle. Bei der Bewertung der Testergebnisse ist besonders relevant, inwiefern zuvor festgelegte Leistungskriterien erfüllt werden, die Rückschlüsse auf tatsächliches kompetentes Verhalten ermöglichen.

#### **3.1 Kriteriumsorientierte vs. normorientierte Testung**

Je nach diagnostischer Zielsetzung können zwei unterschiedliche Zugänge zur Messung psychologischer Merkmale unterschieden werden. *Normorientierte* Tests dienen dazu, interindividuelle Unterschiede von Personen festzustellen, wobei die Testleistung relativ zur Leistung einer Referenzpopulation interpretiert wird. Das *kriteriumsorientierte* (auch *lehrzielorientierte*) Testen ermöglicht die Beurteilung, ob bzw. wie gut eine Person ein zuvor festgelegtes Lehrziel erreicht hat (Lehrzielorientierte Klassifikation; Klauer & Leutner, 2007). Als Lehrziel kann nach Klauer (1987, S. 14) „die geforderte Ausprägung einer Kompetenz“ betrachtet werden. Die Interpretation der Testleistung wird beim kriteriumsorientierten Testen also anders als beim normorientierten Vorgehen unabhängig von einer Referenzpopulation vorgenommen. Bei der Fahrerlaubnisprüfung handelt es sich um einen kriteriumsorientierten Test, da die Testleistung einer Person nicht an der Leistung anderer gemessen wird. Ausschlaggebend für die Bewertung eines Prüflings ist, ob er mit seiner Leistung ein zuvor festgelegtes Leistungskriterium, auch Lehr- bzw. Lernziel genannt, erreicht hat.

Testverfahren sollten von Beginn an so konstruiert werden, dass sie entweder eine kriteriumsorientierte oder eine normorientierte Interpretation der Ergebnisse unterstützen (Shrock & Coscarelli, 2000). Zentral bei der Konstruktion kriteriumsorientierter Tests ist das Konzept der Kontenvalidität. Diesem Konzept zufolge muss der Test die Gesamtheit einer wohl definierten Menge der Aufgaben enthalten oder repräsentieren (Klauer, 1987). Da ein Test aus ökonomischen Gründen selten alle möglichen Aufgaben enthält, muss gewährleistet sein, dass die Leistung einer Person bei einer zufälligen Auswahl von Items derjenigen Leistung entspricht, welche die Person erreichen würde, wenn sie sämtliche Aufgaben bearbeiten würde. Die ausgewählten Aufgaben wären demnach repräsentativ für die (hypothetische) Gesamtmenge aller Aufgaben. Nach Nussbaum (1987) kann beim kriteriumsorientierten Testen eine Auswahl von Aufgaben für Paralleltests nach formalen Kriterien wie Trennschärfe oder Schwierigkeit nicht erfolgen. Stattdessen soll eine

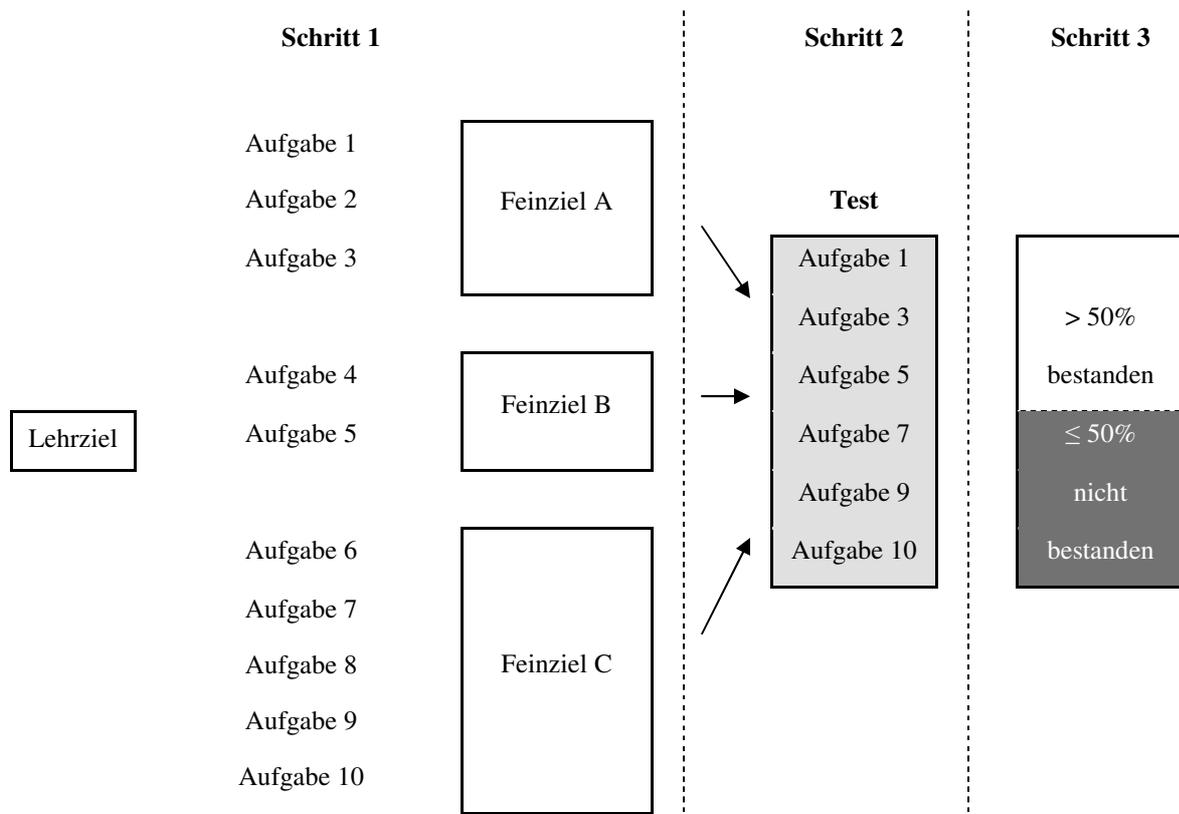
Zufallstichprobe von Items aus dem Itemuniversum gezogen werden, weshalb auch von zufallsparellen Tests gesprochen wird. Gleiche Mittelwerte, Varianzen und Kovarianzen, also äquivalente Bedingungen, sind dabei – im Gegensatz zu den klassischen Verfahren – nicht notwendig, wobei dies bei steigender Itemzahl zunehmend dennoch gewährleistet wäre. Um dem Anspruch der Kontenvalidität zu genügen, erfolgt die Konstruktion kriteriumsorientierter Tests auf eine andere Art als es bei normorientierten Verfahren üblicherweise der Fall ist (siehe 3.2). Bei der Konstruktion normorientierter Testverfahren ermittelt man für gewöhnlich mit Hilfe eines Pretests – der alle formulierten Items enthält – die Schwierigkeiten und Trennschärfen der einzelnen Aufgaben. Bei der Auswahl der Aufgaben für die Testendform werden üblicherweise die Testaufgaben, die eine hohe positive Trennschärfe aufweisen, beibehalten, diejenigen mit niedriger Trennschärfe in der Regel eliminiert. Ein weiteres Selektionskriterium ist die Schwierigkeit von Testaufgaben, wobei innerhalb des Tests die Schwierigkeiten der Aufgaben üblicherweise möglichst weit streuen, aber überwiegend im Mittelbereich liegen sollten (Amelang & Schmidt-Atzert, 2006). Aufgaben mit hoher Trennschärfe werden bei diesem Vorgehen beibehalten, selbst wenn sie sehr schwierig oder sehr leicht sind.

Eine normorientierte Testentwicklung für Aufgaben zur Messung von Teilaspekten der Fahrkompetenz erscheint nicht sinnvoll, da kein Leistungsvergleich zwischen den Teilnehmern der Fahrerlaubnisprüfung angestrebt wird, sondern eine Klassifikation in Könnern und Nichtkönnern, die eine Entscheidung zwischen bestanden und nicht bestanden ermöglicht. Im Folgenden wird daher das Vorgehen für die Konstruktion kriteriumsorientierter Tests ausführlich beschrieben.

### ***3.2 Vorgehensweise bei der Konstruktion kriteriumsorientierter Tests***

Nach Klauer und Leutner (2007) kann die Entwicklung eines kriteriumsorientierten Tests in drei Schritten erfolgen (vgl. Bild 3-1). Zunächst wird in einem ersten Schritt das Lehrziel als Aufgabenmenge definiert. Wenn das Lehrziel in wichtige Teilziele untergliedert werden kann, ist dies für das weitere Vorgehen von Vorteil. Für jedes Feinziel können spezielle Aufgaben zu diesem eng umschriebenen Inhaltsbereich konstruiert werden. In einem zweiten Schritt wird aus der Aufgabenmenge eine repräsentative oder zufällige Stichprobe von Aufgaben zusammengestellt. Dabei sollte festgelegt werden, in welchem Verhältnis Aufgaben der Teilbereiche in die Entscheidung eingehen sollen. Zum Schluss sollte man das Lehrziel quantitativ festlegen, indem man den angestrebten Kompetenzgrad bestimmt. Ein

quantitatives Kriterium könnte z.B. die Prozentzahl richtiger Lösungen sein, die eine Person erreichen muss, um als Köhner eingestuft zu werden.



**Bild 3-1 Die Entwicklung eines kriteriumsorientierten Tests in drei Schritten (Klauer & Leutner, 2007)**

Shrock und Coscarelli (2000) beschreiben in elf Schritten, wie bei der Konstruktion kriteriumsorientierter Tests vorgegangen werden sollte:

1. Anforderungsanalyse
2. Herstellung der Inhaltsvalidität der Ziele
3. Itemgenerierung
4. Erstellung des Bewertungsinstrumentes
5. Bestimmung der Inhaltsvalidität der Items und Beurteilungsskalen
6. Pilotierung einer vorläufigen Testversion
7. Itemanalyse
8. Erstellung paralleler Testversionen
9. Bestimmung von Cut-Off-Werten
10. Reliabilitätsbestimmung
11. Bekanntgabe der Testwerte

Im Folgenden wird dargestellt, wie innerhalb der aufgeführten Schritte vorzugehen ist.

## 1. Anforderungsanalyse

Den ersten Schritt der Konstruktion eines lehrzielorientierten Tests stellt die Aufstellung eines oder mehrerer Lehrziele dar. Dazu müssen die Aufgaben identifiziert werden, die geeignet erscheinen, eine Person als Könnler oder Nichtkönnler zu klassifizieren. An dieser Stelle soll auch, sofern möglich, eine Aufteilung in Feinziele erfolgen, welche in eine Hierarchie gebracht werden.

## 2. Inhaltsvalidität der Lehrziele

Die aufgestellten Lehrziele sollen anschließend auf ihre Inhaltsvalidität hin geprüft werden. Lehrziele sind dann inhaltsvalide, wenn ihre Relevanz durch Experten bestätigt wurde.

## 3. Itemgenerierung

Die Aufgabenkonstruktion erfolgt in Anlehnung an die Lehrziele. Wenn Lehrziele auf unterschiedlichen kognitiven Ebenen vorliegen, kann es von Vorteil sein, entsprechend auch unterschiedliche Aufgaben- und Antwortformate einzusetzen.

## 4. Bewertungsinstrument

Abhängig von den gewählten Aufgaben- und Antwortformaten sollte nun festgelegt werden, wie die Leistung einer Person darin bewertet wird.

## 5. Inhaltsvalidität der Testaufgaben und Beurteilungsskalen

Die Inhaltsvalidität kann durch Expertenurteile überprüft werden. Ausgewiesene Experten werden dabei mit den Lehrzielen vertraut gemacht und sollen anschließend einschätzen, ob die entwickelten Aufgaben die Erreichung der Lehrziele abprüfen und die Beurteilungsskala für die Bewertung der Testleistung geeignet ist.

## 6. Pilotierung der vorläufigen Testversion

Eine erste Pilotuntersuchung mit den erstellten Aufgaben sollte durchgeführt werden, um Rückmeldung bezüglich Verbesserungsmöglichkeiten des Tests zu erhalten. Die Stichprobe sollte in ihrer Zusammensetzung der Stichprobe entsprechen, für die der eigentliche Test vorgesehen ist.

## 7. Itemanalyse

Im Anschluss an den ersten Einsatz der Testaufgaben sollten diese bezüglich ihrer statistischen Kennwerte analysiert werden. Dabei werden Schwierigkeiten und Trennschärfen der einzelnen Aufgaben berechnet. Zudem wird überprüft, inwieweit Inter-Item-Korrelationen vorliegen. Bei Mehrfachwahlaufgaben, bei denen den Teilnehmern neben einer oder mehreren richtigen Antworten auch falsche Antworten (Distraktoren) zur Auswahl stehen, werden die Wahrscheinlichkeiten für die Wahl jeder einzelnen Falschantwort ermittelt.

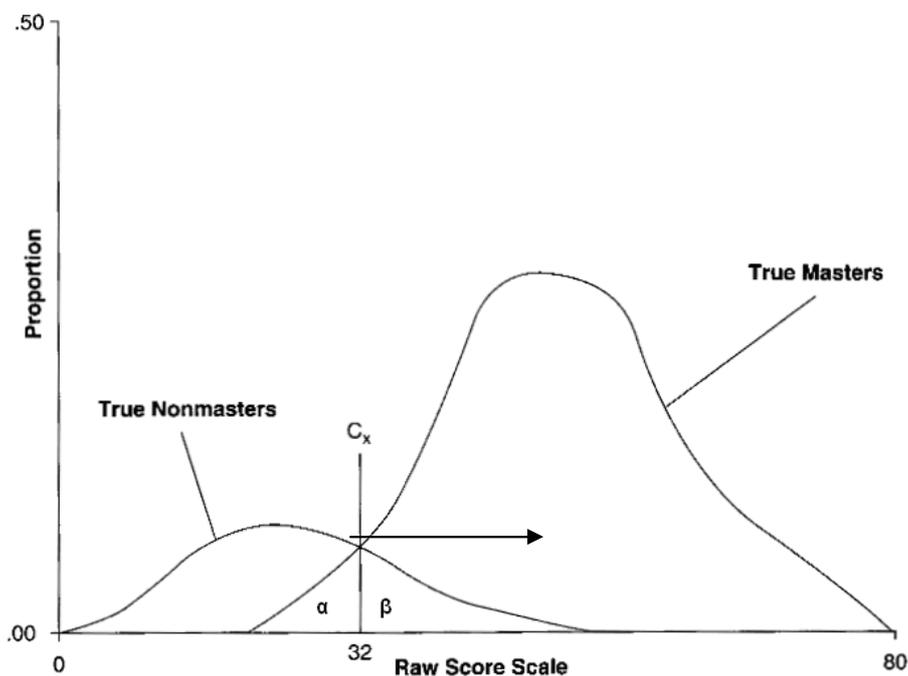
Die Itemstatistiken können Aufschluss darüber geben, ob eine Verbesserung des Tests notwendig ist.

#### 8. Erstellung paralleler Testversionen

In parallelen Testformen sollte durch sorgfältige Zusammenstellung der Items gewährleistet werden, dass alle Inhaltsbereiche abgedeckt werden. Aus diesen Inhaltsbereichen können zufällig Aufgaben ausgewählt werden.

#### 9. Bestimmung der Cut-Off-Werte

Im nächsten Schritt müssen Cut-Off-Werte bestimmt werden. Erreicht eine Person diesen oder einen höheren Punktwert, hat sie den Test bestanden und wird als Könnler klassifiziert. Alle Personen, die unter diesem kritischen Leistungswert bleiben, werden als Nicht-Könnler eingeteilt und haben die Prüfung nicht bestanden. Nach Klauer (1987) kann davon ausgegangen werden, dass Könnler praktisch keine Fehler machen. Zufällige Irrtümer können allerdings auch bei Könnern nicht ausgeschlossen werden, weshalb eine 100%ige Trefferquote ein zu hartes Kriterium darstellen würde. Es gibt unterschiedliche Methoden, diese Cut-Off-Werte sinnvoll festzulegen (z.B. Klauer, 1987).



**Bild 3-2. Hypothetische Verteilungen zur Veranschaulichung der Methode der Kontrastierenden Gruppen (aus Cizek & Bunch, 2007)**

Zur Anwendung der Methode der Kontrastierenden Gruppen (*Contrasting Groups Method*) muss zunächst ein Pool definitiver Könnler und definitiver Nicht-Könnler identifiziert werden (Shrock & Coscarelli, 2000). Dafür wird beiden Gruppen der Test zur Bearbeitung

dargeboten. Anschließend werden die Ergebnisse der Teilnehmer beider Gruppen in Häufigkeitsverteilungen (oder in Verteilungen der prozentualen Häufigkeit) getrennt graphisch dargestellt (vgl. Beispielverteilungen in Bild 3-2). An dem Punkt, in dem sich die beiden Kurven schneiden, sollte ein erster Cut-Off-Wert gesetzt werden (im Beispiel bei 32 Punkten). An dieser Stelle ist die Rate der Fehlklassifikationen am geringsten. Dieser Punkt kann soweit verschoben werden, bis alle Nicht-Könnler unterhalb des Cut-Off-Wertes liegen (siehe Pfeil). Dies führt allerdings auch zu einer Erhöhung der Wahrscheinlichkeit, Könnler fälschlicherweise als Nichtkönnler zu klassifizieren. Die Festlegung eines Cut-off-Wertes kann daher in der Regel nicht rein rechnerisch erfolgen, sondern setzt immer auch eine normative Entscheidung darüber voraus, welches Gewicht man welchem Fehlertyp (Alpha- oder Beta-Fehler, vgl. Kapitel 2) beimisst.

#### 10. Reliabilitätsbestimmung

Die Reliabilität ist neben Objektivität und Validität eines der Hauptgütekriterien, mit dem man ein Testverfahren bewerten kann. Die Reliabilität bezeichnet die Messgenauigkeit eines Tests, also die Höhe des Anteils der Varianz der wahren Werte an der Varianz der beobachteten Testwerte (Bühner, 2004). Zur Bestimmung der Reliabilität kriteriumsorientierter Tests gibt es verschiedene Methoden, die an dieser Stelle jedoch nicht eingehender dargestellt werden sollen (vgl. Fricke, 1974).

#### 11. Bekanntgabe der Testwerte

Im Falle einer kriteriumsorientierten Klassifikation wird dem Testteilnehmer als Testergebnis lediglich mitgeteilt, ob er den Test bestanden hat oder nicht.

### **3.3 Lehrziele**

Der erste Schritt bei der Konstruktion eines kriteriumsorientierten Tests entspricht der Festlegung eines Lehrziels. Sollen Lernende am Ende ihrer Ausbildung über eine bestimmte Qualifikation verfügen, ist es sinnvoll, im Vorfeld Lehrziele festzulegen (Klauer & Leutner, 2007). An diesen Lehrzielen orientiert sich inhaltlich die Aufgabenerstellung.

Lehrziele können zunächst grob nach allgemein-psychologischen Funktionsbereichen kategorisiert werden (Kerres, 2001). Üblich ist eine Differenzierung in kognitive, affektive und psychomotorische Lehrziele, die sich durch die Arbeiten der Gruppe um Bloom (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) durchgesetzt hat. Zur Einordnung von Lehrzielen für den kognitiven Bereich hat Bloom (1974) ein als Lehrzieltaxonomie bezeichnetes Klassifikationssystem entwickelt, das aus sechs Hauptkategorien ansteigender Komplexität besteht (vgl. Tabelle 3-1).

Hauptkategorien	Beschreibung
1. Wissen, Kenntnis	Sachverhalte wiedergeben können
2. Verständnis	Mit eigenen Worten wiedergeben und interpretieren können
3. Anwendung	Regeln, Gesetzmäßigkeiten auf Sonderfälle übertragen und so Probleme lösen
4. Analyse	Sachverhalte in ihre Struktur zerlegen
5. Synthese	Elemente zu einem neuen Komplex zusammenführen
6. Bewertung	Sachverhalte nach Kriterien beurteilen

**Tabelle 3-1 Einteilung kognitiver Lehrziele nach Bloom (1974)**

Oftmals werden nicht ausschließlich kognitive Lehrziele, sondern auch affektive oder psychomotorische Lehrziele angestrebt. Eine Klassifikation affektiver Lehrziele findet sich bei Krathwohl, Bloom und Masia (1975). Affektive Lehrziele können in fünf Kategorien eingeteilt werden (vgl. Tabelle 3-2).

Hauptkategorien	Beschreibung
1. Aufnehmen	Aufmerksam werden
2. Reagieren	Sich einer Tätigkeit oder einem Phänomen widmen
3. Werten	Glauben oder Überzeugung
4. Wertordnung	Organisation von Werten
5. Bestimmtsein durch Werte	Lebensphilosophie, Weltanschauung

**Tabelle 3-2 Einteilung affektive Lehrziele nach Krathwohl, Bloom und Masia (1975)**

Demnach beziehen sich affektive Lernziele auf Interessen, Werte oder Einstellungen sowie die Fähigkeit moralische Werturteile zu bilden und sein eigenes Handeln danach auszurichten. Auch für den psychomotorischen Bereich wurden Einteilungsmöglichkeiten von Lehrzielen erarbeitet (z.B. Dave, 1968; Harrow, 1972). Unter psychomotorischen Lehrzielen wird die Beherrschung von Bewegungsabläufen und komplexen Verhaltensweisen verstanden. Diese erfordern unterschiedlich starke Regulation und Koordination, die durch Automatisierung reduziert werden kann. In Tabelle 3-3 wird die Einteilung von Lehrzielen im psychomotorischen Bereich von Dave (1968) dargestellt.

Hauptkategorien	Beschreibung
1. Imitation	Verhalten nachahmen
2. Manipulation	Durch Anwendung Handlungsablauf erlernen
3. Präzision	Handlungsablauf verfeinern
4. Integration	Koordination von Bewegungsabläufen
5. Naturalisierung	Automatisiertes Verhalten

**Tabelle 3-3 Einteilung psychomotorischer Lehrziele nach Dave (1968)**

Aufbauend auf diesen Einteilungen kann zur Organisation der Feinziele eine Lehrzielmatrix (Tylermatrix) erstellt werden (vgl. Tabelle 3-4). Sie enthält eine Inhaltsdimension (Lerninhalte) sowie eine Prozessdimension (Leistungsaspekte), in der die oben dargestellten Kategorien zu finden sein können. Zu jeder Zelle dieser Matrix können Lern- und Prüfungsaufgaben erstellt werden. Eine solche Matrix eignet sich somit zur Unterrichtsplanung, zur Entwicklung lehrzielorientierter Tests, aber auch zur Beurteilung eines bestehenden Lehrangebots oder einer Prüfung.

		Lerninhalte		
		Lerninhalt 1	Lerninhalt 2	Lerninhalt 3
Leistungsaspekte	Wissen	Aufgabenkategorie 1.1	Aufgabenkategorie 1.2	Aufgabenkategorie 1.3
	Verstehen	Aufgabenkategorie 2.1	Aufgabenkategorie 2.2	Aufgabenkategorie 2.3
	Anwenden	Aufgabenkategorie 3.1	Aufgabenkategorie 3.2	Aufgabenkategorie 3.3

**Tabelle 3-4 Schematische Darstellung einer Tylermatrix für kognitive Lehrziele**

Die obersten Ziele der Fahrschulerausbildung sind laut Fahrschüler-Ausbildungsordnung (§ 1 Fahrschüler-Ausbildungsordnung (FahrschAusbO)<sup>1</sup> die Erziehung des Fahrschülers zu einem sicheren, verantwortungsvollen und umweltbewussten Verkehrsteilnehmer sowie die Vorbereitung des Fahrschülers auf die Fahrerlaubnisprüfung. Zur Definition von Feinzielen muss konkretisiert werden, wodurch bestimmt werden kann, wann ein Fahrer als sicher, verantwortungsvoll und umweltbewusst gelten kann. Aufbauend auf vier unterschiedlichen

<sup>1</sup> Fahrschüler-Ausbildungsordnung vom 18. August 1998 (BGBl. I S. 2307, 2335), zuletzt geändert durch Artikel 3 der Verordnung vom 18. Juli 2008 (BGBl. I S. 1338)

Ebenen wurden innerhalb des EU-Projektes GADGET (*Guarding Automobile Drivers through Guidance, Education and Technology*) Richtlinien für die Ausbildung der Fahrschüler erarbeitet, die in der so genannten GDE (*Goals for Driver Education*)-Matrix strukturiert wurden (Siegrist, 1999). Dabei wurden zahlreiche Feinziele für die Fahrausbildung formuliert. Der in Tabelle 3-5 dargestellte Auszug aus der GDE-Matrix veranschaulicht die Systematik mit einigen Beispielen.

		Wesentliche Ausbildungsinhalte		
		Wissen und Können	Risikoerhöhende Faktoren	Selbstbeurteilung
Hierarchische Ebenen des Fahrverhaltens	<b>Lebensziele und Fähigkeiten für das Leben (generell)</b>	<ul style="list-style-type: none"> <li>• Lebensstil und -umstände</li> <li>• Motive</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Risikoakzeptanz</li> <li>• Selbstwertgefühl stärken</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Riskante Tendenzen</li> <li>• ...</li> </ul>
	<b>Absichten und sozialer Kontext (bezogen auf Fahrt)</b>	<ul style="list-style-type: none"> <li>• Notwendigkeit der Fahrt</li> <li>• Fahrzweck</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Zustand des Lenkers</li> <li>• Extra-Motive</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Persönliche Fähigkeit zur Planung</li> <li>• ...</li> </ul>
	<b>Beherrschen von Verkehrssituationen</b>	<ul style="list-style-type: none"> <li>• Verkehrsregeln</li> <li>• Geschwindigkeitsanpassung</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Falsche Erwartungen</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Persönlicher Fahrstil</li> <li>• ...</li> </ul>
	<b>Fahrzeugbedienung</b>	<ul style="list-style-type: none"> <li>• Kontrolle der Richtung und Position</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Ungenügende Automatismen/Fähigkeiten</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Stärken und Schwächen beim elementaren Fahrenkönnen</li> <li>• ...</li> </ul>

Tabelle 3-5 Auszug aus der GDE-Matrix nach (Siegrist, 1999)

Um für die theoretische Prüfung Lehrziele zu bestimmen, muss außerdem in Betracht gezogen werden, ob die Erreichung dieser Lehrziele überhaupt mit Hilfe der theoretischen Fahrerlaubnisprüfung kontrolliert werden kann, bzw. ob sich ihre Vermittlung überhaupt als Teil des theoretischen Unterrichts eignet.

### **3.4 Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung**

Bei der Fahrerlaubnisprüfung handelt es sich um einen kriteriumsorientierten Test, da die Testleistung einer Person nicht an der Leistung anderer gemessen wird. Ausschlaggebend für die Bewertung eines Prüflings ist, ob er mit seiner Leistung ein zuvor festgelegtes Leistungskriterium, auch Lehrziel genannt, erreicht hat. Bei der Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung sollte man sich an den Prinzipien der mehrschrittigen Konstruktion kriteriumsorientierter Tests orientieren. Ein erster Schritt besteht dabei darin, angemessene Lehrziele auf unterschiedlichen Ebenen zu erstellen.

Für die Ausbildung von Fahrschülern sind Lehrziele in allen drei Funktionsbereichen (kognitiv, affektiv und psychomotorisch) relevant. Für jedes Feinziel muss individuell entschieden werden, ob und wie die Erfassung der Erreichung dieses Lehrzieles innerhalb der theoretischen Fahrerlaubnisprüfung erfolgen kann.

## **4 Typen von Testaufgaben**

Bei der Erstellung jeglicher Testverfahren ist es notwendig, sich für einen oder mehrere Aufgabentypen zu entscheiden. Aufgabentypen legen fest, wie der Inhalt der Aufgabe präsentiert wird und in welcher Form die Testteilnehmer darauf antworten müssen. Der Wahl des Aufgabentyps sollte besondere Bedeutung bei der Erstellung eines Testverfahrens gewidmet werden. Die Wahl des angemessenen Aufgabentyps ist allerdings nicht vom thematischen Inhalt abhängig. Viel eher ist es möglich, zu jedem Inhaltsbereich Aufgaben jeglicher Art zu erstellen und einzusetzen (Lienert & Raatz, 1998). Relevant ist jedoch für die Auswahl der Aufgabenformate, inwieweit Lehrziele auf den verschiedenen Ebenen dadurch abgeprüft werden können. Für alle Lehrziele müssen, angepasst an die entsprechende Ebene des Lehrziels, Aufgaben im jeweils passenden Antwortformat erstellt werden. Beispielsweise ist es nicht möglich, die Erreichung eines Lehrziels auf der Ebene Transfer mit einer Aufgabe zu überprüfen, in der durch den Aufgabentyp bedingt lediglich Faktenwissen abgefragt wird. Bei den Aufgabentypen kann man grob zwischen Aufgaben mit gebundenem und freiem Antwortformat unterscheiden, seltener gibt es auch Aufgaben mit atypischem Antwortformat. Im Folgenden werden einige Antwortformate mit ihren Vor- und Nachteilen vorgestellt.

### **4.1 Gebundene Antwortformate**

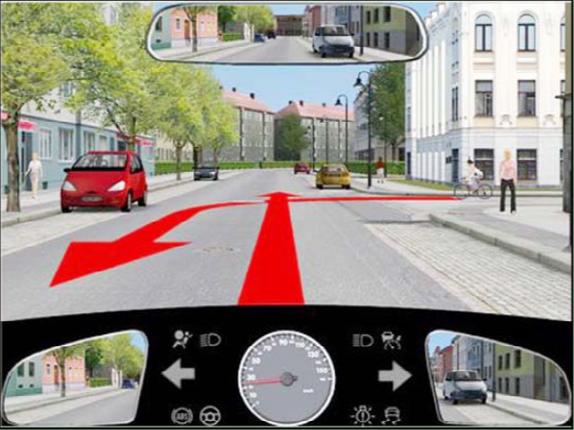
Gebundenen Antwortformaten ist gemeinsam, dass dem Teilnehmer zur Lösung eine oder mehrere Alternativen vorgegeben werden, unter welchen er die richtige bzw. zutreffende identifizieren soll. Besonders vorteilhaft ist die Möglichkeit der ökonomischen Auswertung gebundener Antwortformate zu werten, da sie meist bequem und schnell mithilfe von Schablonen oder Scannern bzw. wenn sie am Computer präsentiert werden, mithilfe eines Auswertungsprogramms ausgewertet werden können.

Aufgrund ihrer eindeutigen Auswertungsroutinen sind Aufgaben mit gebundenem Antwortformat auch als besonders objektiv einzuschätzen.

### 4.1.1 Richtig-Falsch-Aufgaben

Bei Richtig-Falsch-Aufgaben handelt es sich um ein so genanntes dichotomes Antwortformat, da nur zwei mögliche Lösungen vorliegen (vgl. Beispielaufgabe 4-1). Als Aufgabenstamm wird dabei in der Regel eine Aussage formuliert, die mit richtig oder falsch bzw. ja oder nein beantwortet werden kann.

Problematisch ist bei diesem Antwortformat die hohe Wahrscheinlichkeit, die Aufgabe durch Raten zu lösen, ohne die richtige Antwort wirklich zu kennen. Die so genannte Ratewahrscheinlich beträgt bei diesem Antwortformat 50%.



**Der Radfahrer darf vor mir fahren.**

Richtig

Falsch

Beispielaufgabe 4-1 Richtig-Falsch-Aufgabe aus dem Verkehrskontext<sup>2</sup>

### 4.1.2 Mehrfachwahlaufgaben

Mehrfach-Wahlaufgaben, häufig auch als Multiple-Choice-Aufgaben bezeichnet, stellen den Teilnehmer vor die Aufgabe, aus mehreren (mindestens drei) Antwortalternativen die richtige bzw. zutreffende zu wählen. Die falschen Antwortalternativen werden dabei als *Distraktoren* bezeichnet.

---

<sup>2</sup> Die mit <sup>2</sup> gekennzeichneten Beispielaufgaben in diesem Kapitel stammen aus einer Übersetzung des CIECA Theory Test Project WP300-Report (2009) und wurden von der Arbeitsgemeinschaft Technischen Prüfstellen für den Kraftfahrzeugverkehr (TÜV/DEKRA arge tp 21) für dieses Manual zur Verfügung gestellt.

**Sie fahren mindestens 20 km/h schneller als der LKW. Was ist richtig?**



- X Den LKW auf der rechten Seite überholen und auf die Autobahn auffahren.
- O Geschwindigkeit senken und hinter dem LKW auf die Autobahn fahren.
- O Am Ende des Beschleunigungsstreifens anhalten und dann auf die Autobahn auffahren.

**Beispielaufgabe 4-2 Mehrfachwahlaufgabe aus dem Verkehrskontext mit einer richtigen und zwei falschen Antwortmöglichkeiten<sup>2</sup>**

Theoretisch sollte sich mit jeder Hinzunahme eines Distraktors die Wahrscheinlichkeit, die richtige Lösung zu erraten, verringern. Allerdings setzt dies voraus, dass es sich um Distraktoren handelt, die für Personen, die die richtige Antwort nicht kennen, relativ attraktiv und sinnvoll erscheinen. Die Wahl geeigneter Distraktoren stellt das Hauptproblem bei der Konstruktion von Aufgaben mit Mehrfach-Wahl-Antwortformat dar. Geeignete Distraktoren sollten plausibel sein und der richtigen Lösung ähneln (Jankisz & Moosbrugger, 2008). Bei den klassischen Mehrfachwahlaufgaben findet sich nur eine richtige Lösung unter den Antwortalternativen (vgl. Beispielaufgabe 4-2).

Auch wenn die Wahrscheinlichkeit, die richtige Lösung zu erraten, durch geeignete Distraktoren verringert werden kann, besteht dennoch die Möglichkeit, trotz fehlender Kenntnisse die richtige Antwort per Zufall zu wählen. Wenn sie eine Aufgabe nicht mithilfe ihres Wissens lösen können, haben die Testteilnehmer prinzipiell zwei Möglichkeiten sich zu verhalten. Sie können per Zufall eine der Antwortalternativen wählen, oder sie können die Aufgabe unbeantwortet lassen. Personen, die die richtige Lösung per Zufall wählen, würden bei gleichem Kenntnisstand im Test möglicherweise bessere Leistungen erzielen als Personen, die die Aufgabe gar nicht beantworten. Um diesem Problem entgegen zu treten, kann man Falschantworten mit Punktabzügen sanktionieren.

Es ist auch möglich, mehrere richtige Lösungen vorzugeben (vgl. Beispielaufgabe 4-3). Dabei kann man zusätzlich entscheiden, ob man dem Testteilnehmer in der Instruktion bereits mitteilt, wie viele richtige Antworten ausgewählt werden sollen, oder ihnen auch diese Entscheidung überlässt (was in der Regel die Schwierigkeit der Aufgabe beeinflusst).

### Auf was müssen Sie in dieser Situation achten?



- X Ein langsames Fahrzeug hinter der Kurve, welches in die selbe Richtung fährt wie Sie
- X Schnell entgegenkommende Fahrzeuge, welche gegebenenfalls die Kurve schneiden
- X Wildwechsel hinter der Kurve

#### Beispielaufgabe 4-3 Mehrfachwahlaufgabe aus dem Verkehrskontext mit drei richtigen Antwortmöglichkeiten<sup>2</sup>

Wenn geeignete Distraktoren ausgewählt wurden, reduziert dies die Ratewahrscheinlichkeit für Mehrfachwahlaufgaben erheblich.

Die Art der Bewertung von Mehrfachwahlaufgaben kann sich zudem auf das Antwortverhalten des Prüflings auswirken. Bei jeder gegebenen Antwortalternative hat der Testteilnehmer die Möglichkeit, sie entweder als richtige Antwort zu markieren oder nicht. Tabelle 4-1 veranschaulicht, dass er die Aufgabe richtig lösen kann, indem er die Antwortalternative ankreuzt, wenn es sich dabei wirklich um die richtige Lösung handelt, oder indem er sie nicht ankreuzt, wenn es sich tatsächlich um eine falsche Antwortalternative handelt. Der Teilnehmer hat auch zwei Möglichkeiten, sich bei einer Antwortalternative falsch zu verhalten: er kreuzt die Alternative an, obwohl sie falsch ist oder er kreuzt sie nicht an, obwohl sie richtig ist.

		Antwortalternaive ist...	
		richtig	falsch
Antwort des Prüflings lautet...	r	richtige Antwort ✓	falsche Antwort ✗
	f	falsche Antwort ✗	richtige Antwort ✓

**Tabelle 4-1 Bearbeitung einer Antwortalternative einer Mehrfachwahlaufgabe**

Bei der Bewertung einer Testleistung gibt es unterschiedliche Möglichkeiten (vgl. Tabelle 4-2).

		Alternative	
		r	f
Antw.	r	1	0
	f	0	0

**Bewertungsmöglichkeit  
1**

		Alternative	
		r	f
Antw.	r	1	0
	f	0	1

**Bewertungsmöglichkeit  
2**

		Alternative	
		r	f
Antw.	r	1	-1
	f	-1	1

**Bewertungsmöglichkeit  
3**

**Tabelle 4-2 Bewertungsmöglichkeiten einer Antwortalternative einer Mehrfachwahlaufgabe**

Bei der Festlegung von Bewertungsvorgaben für Mehrfachwahlaufgaben muss entschieden werden, ob nur die Identifikation einer richtigen Antwortalternative belohnt wird (Bewertungsmöglichkeit 1) oder auch die Identifikation einer falschen Antwortalternative (Bewertungsmöglichkeit 2). Zudem muss festgelegt werden, ob falsches Antwortverhalten mit Punktabzug sanktioniert wird (Bewertungsmöglichkeit 3). Die Sanktionierung von Falschantworten erhöht die Schwierigkeit einer Aufgabe, da sie die Wahrscheinlichkeit, dass die Teilnehmer versuchen, die richtige Lösung zu erraten, reduziert. Die Testteilnehmer sollten immer über die Art der Bewertung der Mehrfachwahlaufgaben informiert werden.

### 4.1.3 Zuordnungsaufgaben

Bei Zuordnungsaufgaben besteht die Aufgabe der Testteilnehmer darin, sinnvolle Paare aus vorgegebenen Begriffen oder Bildmaterial zu bilden. Im Aufgabenstamm wird in der Regel eine Zuordnungsvorschrift formuliert (vgl. Beispielaufgabe 4-4 und Beispielaufgabe 4-5).

**Ordnen Sie bitte jedem Bundesland die richtige Hauptstadt zu. Verbinden Sie dazu die zusammengehörigen Namen.**

**Sachsen**

**Bayern**

**Baden-Württemberg**

**Rheinland-Pfalz**

**München**

**Mainz**

**Dresden**

**Stuttgart**

**Beispielaufgabe 4-4 Typische Zuordnungsaufgabe**

Bitte ordnen Sie die Gefahren den Situationen zu (drag & drop).

The image shows three driving scenarios from a driver's perspective, each with a set of hazard icons below it. The first scenario shows a road with a yellow arrow pointing left on the dashboard, indicating a lane change or merging. The second scenario shows a car skidding on a wet road, indicated by a yellow arrow pointing right on the dashboard. The third scenario shows a pedestrian crossing with a yellow arrow pointing left on the dashboard, indicating a pedestrian crossing.

Scenario 1: Road with yellow arrow pointing left. Icons: Down arrow, Up arrow, Car.

Scenario 2: Car skidding on wet road. Icons: Skidding car, Spill.

Scenario 3: Pedestrian crossing with yellow arrow pointing left. Icons: Pedestrian, Child, Adult.

#### Beispielaufgabe 4-5 Zuordnungsaufgabe aus dem Verkehrskontext<sup>2</sup>

Dieser Aufgabentyp eignet sich gut zur Wissensüberprüfung, da die Ratewahrscheinlichkeit aufgrund der in der Regel sehr großen Zahl an Lösungsvarianten äußerst gering ausfällt. Allerdings ist zu beachten, dass die Antwortalternativen nicht unabhängig voneinander sind. Die Kombinationsmöglichkeiten verringern sich, wenn Paare gebildet werden. Auch mit diesem Aufgabenformat können lediglich Leistungen auf der Ebene der Wiedererkennung abgeprüft werden.

#### 4.1.4 Ordnungs- oder Umordnungs-Aufgaben

Beim Aufgabentyp der Ordnungsaufgaben sollen Begriffe oder Bildmaterial in eine sinnvolle Reihenfolge gebracht werden (vgl. Beispielaufgabe 4-6). Im Aufgabenstamm wird nicht immer angegeben, nach welcher Vorschrift das Material geordnet werden soll. Oft muss die sinnvolle Reihenfolge von den Teilnehmern erschlossen werden.

**Bitte ordnen sie die Gefahren nach ihrer Wahrscheinlichkeit.**



1	Längsverkehr		
2	kreuzender Verkehr		
3	Hindernisse		
4	Straßenverhältnisse		
5	Fußgänger	▼	▲
6	Tiere		

**Beispielaufgabe 4-6 Umordnungsaufgabe aus dem Verkehrskontext<sup>2</sup>**

## **4.2 Freie Antwortformate**

Auf der Ebene der zu erbringenden Leistung von Seiten der Testteilnehmer ist bei den beschriebenen gebundenen Aufgabenformaten zu beachten, dass lediglich Wiedererkennungslösungen abgeprüft werden, was kognitiv weniger anspruchsvoll ist als das eigenständige Formulieren der richtigen Lösung.

Im Gegensatz zu gebundenen Antwortformaten stehen den Testteilnehmern bei der Bearbeitung der Aufgaben mit freiem Antwortformat keine Lösungsalternativen zur Verfügung. Dies kann zwar Nachteile bei der objektiven und ökonomischen Auswertung der Antworten mit sich bringen, reduziert aber die Wahrscheinlichkeit, dass die Lösung erraten wird, auf ein Minimum.

### **4.2.1 Ergänzungs-Aufgaben**

Bei den so genannten Ergänzungsaufgaben besteht die Aufgabe zumeist darin, einzelne fehlende Begriffe in einem Lückentext oder am Ende einer Aussage zu ergänzen. Objektivität wird in der Regel durch eine standardisierte Instruktion sowie Auswertung erreicht.

In der derzeitigen theoretischen Fahrerlaubnisprüfung wird dieses Antwortformat eingesetzt. Allerdings kommt diese Aufgabenart nur zum Einsatz, wenn es sich bei der erfragten Antwort, wie in Beispielaufgabe 4-7 um eine Zahl handelt. Aufgaben, in denen die Anforderung darin besteht, Wörter oder Begriffe einzusetzen, werden in der derzeitigen theoretischen Fahrerlaubnisprüfung nicht eingesetzt. Dies erleichtert die Auswertung insofern, als dass nur eine richtige Antwort existiert, nämlich die Lösungszahl. Zahlen die von dieser Zahl abweichen, werden als falsch gewertet.

**Welche zulässige Höchstgeschwindigkeit gilt allgemein innerhalb geschlossener Ortschaften?**

**Antwort:**  km/h

**Beispielaufgabe 4-7 Ergänzungsaufgaben aus der derzeitigen theoretischen Fahrerlaubnisprüfung**

### **4.2.2 Kurzaufsatz-Aufgaben**

Die Lösung einer Kurzaufsatz-Aufgabe kann aus wenigen Worten bis zu einem Essay bestehen, wobei die Antwort auf eine Aufgabenstellung in eigenen Worten formuliert wird.

**Auf was müssen Sie in dieser Situation achten?**



**Die Kurve ist steil.**

**Beispielaufgabe 4-8 Kurzaufsatzaufgabe aus dem Verkehrskontext**

Beispielaufgabe 4-8 mit der beispielhaften Lösung der Aufgabe verdeutlicht die Schwierigkeit der Bewertung solcher Aufgaben. Die Beispiellösung stellt zwar einen wichtigen Aspekt der dargestellten Verkehrssituation dar, jedoch wird nicht dargestellt, inwieweit dies zu einer Gefahr führen kann (z.B. Bezug zur Geschwindigkeit des Fahrzeuges). Außerdem fehlen andere Aspekte der Verkehrssituation, die ebenfalls relevant sind (z.B. möglicherweise liegengeliebene Fahrzeug hinter der Kurve).

Die Bewertung der Leistung in einer solchen Aufgabe birgt unterschiedliche Probleme. Insbesondere die Standardisierung der Auswertung erweist sich in der Regel als schwierig.

Eine Schwierigkeit stellt dabei die Vorgabe von Standards dar. Es muss vorab festgelegt werden, welche Anforderungen erfüllt sein müssen, damit die Aufgabe als gelöst gewertet wird, bzw. damit eine bestimmte Anzahl von Teilpunkten vergeben wird. Häufig werden Schlüsselwörter im Vorhinein festgelegt, für deren Verwendung im Antworttext Punkte vergeben werden. Werden diese Schlüsselwörter lediglich umschrieben oder mit einem Synonym benannt, entsteht ein Ermessensspielraum für den Beurteiler, wodurch die Objektivität der Messung eingeschränkt wird. Zur Begrenzung dieser Problematik sollte vor der Bewertung festgelegt werden, inwieweit die Qualität der Formulierung bewertet werden soll sowie ob und für welche Leistungen Teilpunkte vergeben werden. Wenn ein bestimmter Bewertungsspielraum gegeben ist, empfiehlt es sich, die Aufgaben von mindestens zwei Beurteilern bewerten zu lassen. Die Höhe der Übereinstimmung der Beurteiler gibt dabei Auskunft über die Interrater-Reliabilität der Aufgabe und ist zudem ein Maß für die Auswertungsobjektivität der Messung (Moosbrugger & Kelava, 2008).

Um ein gewisses Maß an Standardisierung zu schaffen, die Auswertung zu erleichtern und den Prüflingen Sicherheit zu gewähren, können auch formale Vorgaben für die Bearbeitung der Aufgaben gemacht werden. Z.B. kann man in der Instruktion festlegen, ob nur mit einem Wort oder einer Nennung von mehreren Wörtern (Festlegung einer genauen oder ungefähren Anzahl), in einem Satz oder mehreren (z.B. „in maximal drei Sätzen“) etc..

Offene Antwortformate, die die Produktion eines Textes erfordern, können zudem einen Validitätseinschränkenden Effekt haben. Sie können in stärkerem Maße als die anderen Antwortformate einem Testbias unterliegen. Das bedeutet, dass die Qualität der gegebenen Antwort in besonders hohem Maße mit bestimmten Personenmerkmalen, wie z.B. sprachlichen Fähigkeiten, zusammenhängt, deren Erfassung zumeist in den Testverfahren nicht intendiert ist.

Auch unter ökonomischen Gesichtspunkten ist das offene Antwortformat kritisch zu betrachten. Bearbeitung und Auswertung erfordern in der Regel einen erhöhten Zeitaufwand. Für freie Antwortformate und insbesondere Kurzaufsatz-Aufgaben spricht jedoch die Möglichkeit, Lehrziele auf höheren Ebenen (z.B. Anwendung, Transfer, Werten) und auch besondere Merkmale (Kreativität, persönliche Wichtigkeiten) abzufragen (Jankisz & Moosbrugger, 2008; Rost, 2004).

### **4.3 Handlungsnahe Antwortformate**

In Kapitel 3.3 wurden kognitive, affektive und psychomotorische Lehrziele unterschieden. Mit den in den vorangegangenen Abschnitten vorgestellten gebundenen und freien

Antwortformaten kann die Erreichung von kognitiven und affektiven Lehrzielen auf der Wissens- und Verständnisebene angemessen überprüft werden. Es bleibt dabei allerdings fraglich, ob das erworbene Wissen oder erworbene Werte sich auch in tatsächlichem Verhalten zeigen und ob die Kenntnisse und Einstellungen auf andere Situationen transferiert werden können. Psychomotorische Lehrziele können jedoch mithilfe herkömmlicher Aufgabentypen kaum überprüft werden.

Für die Fahrausbildung sind allerdings kognitive und affektive Lehrziele auf der Anwendungsebene sowie psychomotorische Lehrziele besonders relevant. Es bedarf für die Fahrerlaubnisprüfung also auch Aufgaben- und Antwortformate, die näher an der tatsächlichen Handlung sind, deren positive Beeinflussung durch das Lernen angestrebt wird. Handlungsnahe Antwortformate wie Reaktionszeitaufgaben, Blickbewegungsaufgaben und Fahrsimulationen bieten die Möglichkeit, in Ergänzung zu den herkömmlichen Formaten, die durch den Fahrschulunterricht verfolgten kognitiven, affektiven und psychomotorischen Lehrziele der höheren Ebenen abzutesten.

#### **4.3.1 Reaktionszeitaufgaben**

Im Realverkehr ist schnelles Reagieren (z.B. bei Gefahr bremsen) sehr häufig notwendig. Eine Möglichkeit, diese Fähigkeit zur schnellen Reaktion in Verkehrssituationen zu erfassen, stellen Reaktionszeitaufgabenformate dar.

Populär sind die so genannten Gefahrenwahrnehmungs- oder Hazard Perception-Tests, die bereits in Großbritannien und dem australischen Bundesstaat Victoria seit einiger Zeit einen Teil der Fahrerlaubnisprüfung darstellen. Dabei wird zumeist von Testteilnehmern verlangt, auf die Entdeckung eines potentiellen Gefahrenhinweises (auf Fotos, in Videos, Computeranimationen oder in Bilderfolgen) zu reagieren (vgl. Beispielaufgabe 4-9). Als Reaktionszeit wird die Zeit gemessen, die vom ersten Auftauchen des kritischen Reizes bis zur Reaktion durch den Testteilnehmer vergeht. Dabei hat sich gezeigt, dass es sich bei diesen Reizen nicht unbedingt im Sinne der klassischen Hazard Perception-Aufgaben um Gefahrenreize handeln muss. Kritische Reize können auch Hinweise auf geltende Verkehrsregeln (z.B. Kreuzung, an der man die Vorfahrt gewähren muss), die eine Reaktion verlangen, sein (vgl. Bericht „*Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung*“).



**Beispielaufgabe 4-9 Reaktionszeitaufgabe Nr. 28 (aus Bericht „Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung“)**

Die Durchführung erfolgt in der Regel computerbasiert und daher in hohem Maße standardisiert und objektiv. Die Auswertung erfolgt ebenfalls computerisiert durch die Aufzeichnung der Reaktionszeiten und ist damit objektiv.

Auch das Setzen von Standards kann objektiv erfolgen. Für eine objektive Interpretation der Reaktionszeiten, die ein Testteilnehmer in einer Aufgabe zeigt, muss im Vorhinein für jede Aufgabe festgelegt werden, bis zu welchem Zeitpunkt noch spätestens reagiert werden darf, damit die Aufgabe noch als gelöst gewertet wird.

Für diese Aufgaben spricht neben ihrer Objektivität auch die erwiesene Validität. Erfahrene Fahrer sind Novizen in diesen Aufgaben überlegen (z.B. Deery, 1999; McKenna & Crick, 1991, 1994; Mills, Hall, McDonald, & Rolls, 1998) und außerdem erbringen Personen mit wenigen oder keinen Unfällen bessere Leistungen als Personen mit mehr Unfällen (Hull & Christie, 1993; Pelz & Krupat, 1974; Quimby, Maycock, Carter, Dixon, & Wall, 1986; Quimby & Watts, 1981).

Reaktionszeitaufgaben erfordern eine computerbasierte Testvorgabe, was bei der in Deutschland ohnehin PC-basierten theoretischen Fahrerlaubnisprüfung keine zusätzlichen ökonomischen Belastungen hervorruft.

### **4.3.2 Blickbewegungsaufgaben**

Das Absuchen der Umwelt (*Scanning*) dient beim Fahren dazu, potentielle Gefahren sowie Hinweise auf bestimmte Verhaltensvorschriften (z.B. Verkehrsschilder) so früh wie möglich zu entdecken und die Verkehrssituation, in der man sich befindet, zu überblicken. Das Scanning-Verhalten einer Person kann z.B. mithilfe der Aufzeichnung von Blickbewegungen

durch einen *Eye Tracker* erfasst werden. Mit der Eye Tracking-Methode kann zuverlässig beobachtet werden, ob der Proband relevante Bereiche des Szenarios, das er gerade beobachtet, fixiert (Chapman & Underwood, 1998).

Die Durchführung solcher Blickbewegungsmessungen kann standardisiert und objektiv erfolgen. Eine Standardisierung der Auswertung ist erschwert, wenn man anstrebt, sämtliche aufgenommene Blickbewegungsdaten der Personen auszuwerten. Man müsste dazu definieren, welche Blickabfolgen als richtig zu bewerten sind. Studien, innerhalb derer die Eye Tracking-Methode eingesetzt wurde, zeigten Unterschiede zwischen erfahrenen und unerfahrenen Fahrern im Absuchen der Umwelt (z.B. Deery, 1999; Mayhew & Simpson, 1995). Dennoch kann man daraus nicht schließen, dass das Blickverhalten von Erfahrenen als Muster für richtiges Blickverhalten gelten kann, da ungewiss ist, ob das Blickverhalten der Fahrexperten für Fahranfänger überhaupt angemessen ist.

Es besteht aber die Möglichkeit, nicht das gesamte Blickverhalten einer Person auszuwerten, sondern im Sinne einer Reaktionszeiterfassung nur zu bewerten, ob und wie schnell relevante Bereiche des beobachteten Verkehrsszenarios fixiert wurden. Außerdem kann erfasst werden, ob erlernte Blickabfolgen tatsächlich in den passenden Situationen richtig ausgeführt werden (z.B. richtige Kombination von Rück- und Seitenspiegeleinsatz und Schulterblick beim Auffahren auf die Autobahn). Eine standardisierte Auswertung und Interpretation der Daten wird so möglich.

Die Blickbewegungserfassung schließt allerdings nicht die Erfassung der Reaktion ein. Man kann allein daraus, dass eine Person einen Reiz angeschaut hat noch nicht darauf schließen, ob sie ihn als relevant bewertet hat und darauf reagieren würde. Diese Problematik schränkt bestimmte Validitätsaspekte ein.

Kritisch ist außerdem zu betrachten, dass der Einsatz der Methode des Eye-Tracking zurzeit noch als äußerst unökonomisch zu werten ist, was Durchführung, Auswertung und Anschaffungskosten betrifft.

### **4.3.3 Fahrsimulation**

Die Erreichung psychomotorischer Lehrziele der Fahrausbildung wird in der praktischen Fahrerlaubnisprüfung überprüft. Bestimmte Situationen, besonders Gefahren- oder Belastungssituationen, können oder sollen (zur Wahrung der Sicherheit) innerhalb dieser Prüfung allerdings nicht hervorgerufen werden. Eine Überprüfung, ob die Fähigkeiten ausgeprägt sind, die benötigt werden, um z.B. nach Erkennung einer Gefahr das richtige Verhalten angemessen auszuführen, kann daher nicht innerhalb der praktischen Fahrerlaubnisprüfung erfolgen. Die Prüfungsgestaltung und damit die Überprüfung der

Lehrziele ist weiterhin abhängig von den Gegebenheiten der jeweiligen Umwelt, in der die Prüfung durchgeführt wird. Beispielsweise kann das richtige Verkehrsverhalten im Zusammenhang mit Straßenbahnen in einer Stadt, die über kein Straßenbahnnetz verfügt, nicht überprüft werden.

Abhilfe für dieses Problem könnte der Einsatz von Fahrsimulatoren schaffen. In interaktiven Fahrsimulatoren kann der Proband sein Fahrzeug selbst durch einen virtuellen Parcours steuern. Wie realitätsnah die Visualisierungen und das Fahrgefühl, das unter anderem durch Force Feedback erzeugt wird, empfunden werden, ist von der Art des Simulators abhängig.

Um eine standardisierte Durchführung und Auswertung zu ermöglichen, müssen relevante Fahraufgaben und Verkehrskonstellationen definiert werden, die bei der Simulatorfahrt eintreffen müssen. Dem Testteilnehmer kann daher die Wahl der Route nicht überlassen werden, sondern er muss über eine standardisierte Strecke navigiert werden.

Die Auswertung und Interpretation kann für festgelegt Fahraufgaben computerbasiert und somit objektiv erfolgen, indem das aufgezeichnete Verhalten mit einer Musterlösung verglichen wird. Diese Musterlösungen festzulegen, erscheint jedoch sehr schwierig, da sich die Fahrten aufgrund der interaktiven Mitgestaltung jedes Testteilnehmers unterscheiden. Sinnvoller erscheint es, einzelne, objektiv feststellbare Parameter auszuwerten. Z.B. könnte automatisch ausgewertet werden, inwieweit sich der Testteilnehmer während der Fahrt an Geschwindigkeitsvorgaben hält oder wie genau der Sicherheitsabstand zu vorausfahrenden Fahrzeugen eingehalten wird. Es besteht zudem die Möglichkeit, die Simulatorfahrt oder einzelne Fahraufgaben und Situationen während der Simulatorfahrt ähnlich wie bei einer praktischen Prüfung, durch einen oder mehrere Prüfer bewerten zu lassen.

Der Aufwand, der mit dem Einsatz von Simulatoren in der Fahrerlaubnisprüfung einhergeht, erscheint derzeit noch relativ groß. In Abhängigkeit vom jeweiligen Simulatormodell können zum einen finanzielle und zum anderen auch räumliche Belastungen entstehen.

#### ***4.4 Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung***

In Kapitel 4 wurden Typen von gebundenen, offenen und handlungsnahen Antwortformaten vorgestellt. Gebundene Antwortformate weisen Vorteile bei der objektiven Auswertung auf. Besonders das Multiple-Choice-Antwortformat erscheint, unter der Voraussetzung des Einsatzes geeigneter Antwortalternativen, besonders geeignet. Andere Typen von gebundenen Antwortformaten bringen zusätzlich zum Multiple-Choice-Antwortformat keine weiteren Vorteile. In Anbetracht der Tatsache, dass die Anzahl unterschiedlicher Antwortformate, die

innerhalb eines relativ kurzen Tests wie der Fahrerlaubnisprüfung eingesetzt werden sollten, gering gehalten werden sollte (Lienert & Raatz, 1998), wird empfohlen, auf den Einsatz weiterer gebundener Antwortformate zu verzichten.

Ein Nachteil von Multiple-Choice-Aufgaben besteht darin, dass Prüflinge durch Raten oder Ausschluss von Antwortalternativen die Aufgaben lösen könnten, obwohl diejenigen Kenntnisse, die abgeprüft werden sollen, nicht vorhanden sind. Abhilfe könnten hier freie Antwortformate schaffen, da die Testteilnehmer ihre Antworten selbstständig formulieren müssen. Freie Antwortformate weisen allerdings den Nachteil der deutlichen Abhängigkeit von sprachlichen Fähigkeiten auf, insbesondere, wenn längere Texte verfasst werden sollen. Zudem ist die objektive Auswertung und Interpretation der Leistung schwieriger als bei gebundenen Antworttypen. Der Einsatz freier Antwortformate in der theoretischen Fahrerlaubnisprüfung sollte sich daher auf die Verwendung von Ergänzungsaufgaben beschränken, die nur das Einsetzen eines Wortes oder einer Zahl erfordern.

Die klassischen gebundenen und offenen Aufgaben weisen zudem insgesamt den Nachteil auf, dass ihre Bearbeitung sich von der tatsächlichen Handlung, die erlernt und geprüft werden soll, sehr stark unterscheidet. Rückschlüsse auf tatsächliches Verhalten, das sich in der Anwendung des Erlernten zeigen soll, können kaum gezogen werden.

Um diesem Problem entgegenzutreten, sollten ergänzend handlungsnaher Aufgabenformate eingesetzt werden, die die Anwendung des Erlernten überprüfen können. Besonders das Reaktionszeitaufgabenformat hat sich hinsichtlich seiner Güte bewährt und wäre mit relativ geringem Aufwand in die Fahrerlaubnisprüfung integrierbar.

Es kann daher empfohlen werden, neben Multiple Choice-Aufgaben auch handlungsnaher Antwortformate in die theoretische Fahrerlaubnisprüfung einzubinden.

## **5 Gestaltung multimedialer Lern- und Prüfmaterialien**

Neue computerbasierte Medien bieten prinzipiell ein hohes Potenzial zur Förderung des Kompetenzerwerbs und dessen Überprüfung (Jude & Wirth, 2007). Sie bieten zum einen die Möglichkeit, innovative Präsentationsformate zur Veranschaulichung der Lerninhalte zu nutzen (z.B. Videos, Animationen, Vertonung, Forced Feedback), zum anderen ermöglichen interaktive und adaptive Programme eine Anpassung des Lern- und Prüfungsmaterials an Voraussetzungen und Präferenzen des Anwenders. Verschiedene Merkmale der Gestaltung von Lern- und Prüfungsmaterial können sich auf die (Lern-) Leistung von Individuen in unterschiedlicher Weise auswirken. Daher werden im Folgenden verschiedene Gestaltungsdimensionen für multimediales Lern- und Prüfungsmaterial vorgestellt.

Anschließend werden Prinzipien aufgeführt, die bei der Erstellung von Lern- und Prüfungsmaterial generell beachtet werden sollten.

## **5.1 Gestaltungsdimensionen für multimediales Aufgabenmaterial**

Multimediales Aufgabenmaterial kann durch die Verortung auf unterschiedlichen Dimensionen klassifiziert werden. Die wichtigsten Dimensionen werden im Folgenden dargestellt.

### **5.1.1 Kodalität**

Informationen können in unterschiedlichen Zeichenformen präsentiert werden. Häufig werden Informationen in Form von Texten dargeboten. Lehrbücher bestehen zum größten Teil aus Text und auch das Lernen mit elektronischen Texten ist verbreitet. Neben dem Text ist auch die bildliche Darbietungsform verbreitet. Informationen werden in Büchern, Lernprogrammen oder im Internet durch Zeichnungen, Fotos oder Graphiken dargeboten. Bildliche Informationen können auch als Animationen, also computeranimierten Bilderfolgen (z.B. Videos), präsentiert werden.

Paivios Theorie der dualen Kodierung (1986) ist eine der ersten, die sich mit der Verarbeitung multipler Kodierungen von Lerninhalten beschäftigt. Es wird angenommen, dass Sprach- und Bildinformationen in unterschiedlichen kognitiven Subsystemen verarbeitet werden. Das imaginale und das verbale Subsystem sind dabei durch referentielle Verknüpfungen miteinander verbunden. Wird dem Lerner ein illustrierter Text dargeboten, entstehen zwei unterschiedliche mentale Repräsentationen, die inhaltlich aufeinander bezogen sind.

In der derzeitigen Fahrerlaubnisprüfung werden Informationen zu den Aufgaben oft multikodal präsentiert. Die textuell dargebotenen Instruktionen und Antwortmöglichkeiten beziehen sich dabei auf ein Bild. Zukünftig wird angestrebt, auch kurze Bildsequenzen von Verkehrsszenarien in die Prüfung zu implementieren.

### **5.1.2 Modalität**

Lernumgebungen können danach charakterisiert werden, über welche Modalität die Lerner die Lerninhalte aufnehmen. In multimodalen Lernumgebungen wird der Lernstoff also mehreren Sinnesmodalitäten oft gleichzeitig dargeboten. Darin wird ein entscheidender Vorteil multimedialer Lerntechnologien gesehen. Die Präsentation von Informationen kann unterschiedliche Sinne ansprechen, über welche die Informationen aufgenommen werden. Sprachliche Informationen im Lern- oder Prüfmaterial können demnach visuell, in Form von Schrift, oder auditiv, in Form von gesprochenem Text, dargeboten werden. Die

Instruktionspsychologie widmet sich unter anderem der Frage, wann sprachliche Informationen dem Lernenden in welcher Modalität präsentiert werden sollen (z.B. Brünken, Plass, & Leutner, 2004; Mayer, 2005a).

Die derzeitige Fahrerlaubnisprüfung bezieht keine multimodalen Darbietungen ein. Text wird normalerweise ausschließlich schriftlich dargeboten. Lediglich Personen mit Leseschwierigkeiten können die Textpassagen über Kopfhörer hören.

Wenn in Zukunft auch Bildsequenzen in der Prüfung gezeigt werden, erscheint eine Anreicherung der visuellen Darstellung mit akustischen Informationen (z.B. Motorgeräuschen, Martinshorn), und damit die Einführung von multimodalen Präsentationen, sinnvoll.

### **5.1.3 Dynamik/Animation**

Die Dynamik als Gestaltungsmerkmal von Lernumgebungen oder Prüfungsaufgaben bestimmt, ob (hauptsächlich bildlich kodierte) Informationen statisch oder bewegt dargeboten werden. Als dynamisch werden z.B. Darstellungsformen wie Videomaterial oder Computeranimationen bezeichnet.

Es wird davon ausgegangen, dass es das Verständnis dynamischer Prozesse unterstützt, wenn man beim Lernen über bildliche dynamische Repräsentationen des Lerngegenstandes verfügt (Betrancourt, 2005). Allgemeine Vergleiche zwischen den Lerneffekten mit dynamischem gegenüber statischem Bildmaterial führten jedoch nicht zu konsistenten Ergebnissen. In einigen Studien konnte kein Vorteil dynamisch präsentierten Materials gegenüber statischen Bildern gefunden werden (vgl. Betrancourt, 2005). Tversky et al. (2002) zeigten, dass eine Überlegenheit dynamischen Lernmaterials in den meisten Fällen darauf zurückzuführen war, dass den Lernern durch die dynamische Präsentation mehr Informationen zu Verfügung gestellt wurden. In einer Meta-Analyse über 26 Studien zeigte sich jedoch, dass die dynamische Präsentation von Lernmaterial insgesamt einen moderaten Vorteil gegenüber einer statischen Präsentation aufweisen (Höffler & Leutner, 2007). Höhere Effekte wurden gefunden, wenn es sich um eher repräsentative als dekorative Animationen handelte, je realistischer das dargebotene Material war und wenn es darum ging, prozedurales Wissen im Gegensatz zu deklarativem aufzunehmen.

Ob dies nicht nur für das Lernen mit dynamischen Präsentationen, sondern auch für die Gestaltung von Testaufgaben gilt, ist bisher kaum empirisch überprüft.

Dynamische Präsentationen sollen Teil einer neuen theoretischen Fahrerlaubnisprüfung werden. Der Vergleich von dynamischen und statischen Präsentationen bei Aufgaben im Verkehrskontext zeigte in eigenen Studien, dass sich dynamische Darstellungen positiv auf

das Instruktionsverständnis von Fahrschülern auswirken können (Wissensaufgaben im Multiple-Choice-Format) aber auch die Schwierigkeit von Aufgaben erhöhen können (Reaktionszeitaufgaben) (vgl. Bericht „*Neue Aufgabenformate in der Theoretischen Fahrerlaubnisprüfung*“).

#### **5.1.4 Interaktivität/ Adaptivität**

Eine Lern- oder Prüfumgebung ist als interaktiv zu bezeichnen, wenn Nutzer und System jeweils auf eine Aktion des anderen reagieren (Renkl & Atkinson, 2007). Interaktivität ist somit eine Art von Feedback. Eine interaktive multimediale Lernumgebung ist dadurch gekennzeichnet, dass ihre Gestalt von der Aktivität des Lerners abhängt (Moreno & Mayer, 2007).

Beispiele für einfache Interaktionen sind Entscheidungen des Lernenden über die Menge des Lerninhalts, das Abrufen zusätzlicher Informationen (z.B. Videofilme, Bilder usw.) oder die Dauer bestimmter Sequenzen. Auch zur Gestaltung dieser Dimension liegen hauptsächlich Befunde vor, die sich auf das *Lernen* mit interaktiven Systemen beziehen. Meist wird herausgestellt, dass die Möglichkeit zur interaktiven Kontrolle, besonders bei dynamisch dargebotenem Lernmaterial, zu einer schnelleren Aufnahme des Lernstoffes beim Lerner führt (Schwan & Riempp, 2004). Mayer und Chandler (2001) konnten aus ihren Untersuchungen schließen, dass Interaktivität jedoch nur dann vorteilhaft für das Lernen ist, wenn dadurch der Lerner nicht kognitiv überlastet wird.

Einfache Interaktionen, die zu einem besseren Instruktionsverständnis führen können, sind auch für die theoretische Fahrerlaubnisprüfung sinnvoll. So können z.B. ansonsten benachteiligte Personen unterstützt werden, wobei jedoch die Durchführungsobjektivität nicht beeinflusst werden darf. Ein Beispiel für eine in dieser Hinsicht sinnvolle Interaktionsmöglichkeit stellt die optionale Übersetzung des Instruktions- oder Aufgabentextes für Nichtmuttersprachler dar bzw. die optionale auditive Darbietung von Text für Personen mit einer Leseschwäche. Auch eine selbstständige Auswahl der Reihenfolge der dargebotenen Inhaltsbereiche oder einzelner Aufgaben, sollte ermöglicht werden.

Problematisch in Hinsicht auf die Objektivität der Testung stellen sich Interaktionen dar, bei denen der Testteilnehmer aktiv darauf eingreifen kann, welche Inhalte ihm dargeboten werden. Daher ist bei hochgradig interaktiven Testsituationen, wie z.B. im Fahrsimulator zu gewährleisten, dass sich die einzelnen Testungen in ihrem Anspruch an die Testteilnehmer entsprechen. Der Einsatz zuvor definierter Fahraufgaben und fester Streckenvorgaben kann hierbei hilfreich sein.

Eine andere Form der Interaktion zwischen Testteilnehmer und Testprogramm stellt das computerisierte adaptive Testen (CAT) dar. Dabei erfolgt die Auswahl von Aufgaben geeigneter Schwierigkeit in Abhängigkeit der Leistung in den zuvor bearbeiteten Aufgaben. Der Vorteil des adaptiven Testens ist von ökonomischer Natur. Die Vorgabe von Aufgaben, die entweder zu leicht oder zu schwierig für eine Person sind, liefert keinerlei Informationen über die Ausprägung des interessierenden Merkmals. Daher sollen die Testteilnehmer beim adaptiven Testen nur Aufgaben bearbeiten, die ihrem Leistungsniveau entsprechen. Diese Vorgabe wird dadurch erreicht, dass die Person zunächst eine zufällige, relativ leichte Aufgabe aus dem Aufgabenpool lösen muss. Wenn diese Aufgabe richtig gelöst wurde, folgt eine schwierigere Aufgabe, wenn sie nicht gelöst wurde, eine leichtere und so weiter. Immer wieder wird die Fähigkeit der Person durch einen Rechenalgorithmus geschätzt. Sobald die Fähigkeitsschätzungen sich von Aufgabe zu Aufgabe nicht mehr deutlich unterscheiden, wird die Testvorgabe gestoppt. Durch diese Vorgehensweise wird Testzeit gespart, da weniger Aufgaben bearbeitet werden müssen.

Die Voraussetzung für adaptives Testen ist das Vorhandensein eines Itempools von Aufgaben, die nach einem geeigneten Messmodell skaliert sind. Da Personen mit unterschiedlichen Fähigkeiten Aufgaben unterschiedlicher Schwierigkeit vorgelegt werden, kann nicht im Sinne der klassischen Testtheorie davon ausgegangen werden, dass gleich viele richtig gelöste Aufgaben für identische Fähigkeiten der Personen sprechen. Sinnvolle Schätzungen der Fähigkeit setzen daher Aufgaben voraus, die konform mit den Annahmen der Item Response-Theorie sind. Bei den Modellen der Item Response-Theorie (z.B. Rasch-Modell) werden stochastische Beziehungen zwischen dem Antwortverhalten einer Person, der Fähigkeit der Person und der Aufgabenschwierigkeit angenommen. Dies ermöglicht es, Fähigkeitsschätzungen von Personen zu vergleichen, auch wenn die Personen eine unterschiedliche Anzahl von Aufgaben und Aufgaben unterschiedlicher Schwierigkeit bearbeitet haben.

Eine weitergehende Anpassung der Vorgabe der Testaufgaben an die individuelle Fähigkeit der Personen gegenüber dem einfachen computerbasierten adaptiven Testen bietet das so genannte Multidimensionale Adaptive Testen (MAT) (Frey & Seitz, 2009). Dieser Ansatz erscheint insbesondere für die Kompetenzmessung als vorteilhaft, da er im Gegensatz zu einer unidimensionalen Anpassung eine bessere Passung zwischen theoretischem Modell und dem statistischen Messmodell ermöglicht und zudem die Anzahl der vorgegebenen Testaufgaben deutlich verringert.

Im Fall der theoretischen Fahrerlaubnisprüfung ist der Einsatz adaptiven Testens zu erwägen, da eine mit dieser Art der Testvorgabe einhergehende Effizienzsteigerung bei einem derart häufig durchgeführten Verfahren enorm wäre.

### 5.1.5 Immersion

Immersion ist charakterisiert durch die sich verringende kritische Distanz zu dem was multimedial gezeigt wird und durch die steigende emotionale Involviertheit in das Geschehen (Grau, 2003). Der Zustand der Immersion kann in Virtual-Reality-Umgebungen (Computerspiele, Fahr- oder Flugsimulator) eintreten, die – vor allem visuelle – Eindrücke liefern, die der Wahrnehmung in der Realität ähneln (z.B. 360 Grad Sicht). Immersion kann in unterschiedlichen Abstufungen auftreten, wobei das *Flow*-Gefühl, das auch außerhalb multimedialer Umgebungen (z.B. beim Lesen, Lernen, Schachspielen) erlebt werden kann, als ein optimaler Extremzustand absoluter Versunkenheit in die aktuelle Aktivität bezeichnet wird (Jennett et al., 2008). Nach Grau (2003) kann Immersion intellektuell stimulierend wirken, ist aber in der Regel ein Prozess, der in seiner Übergangsphase mentale Ressourcen in Anspruch nimmt. Der Zustand der Immersion kann durch den Einbezug so vieler Sinne wie möglich verstärkt werden.

Ein Beispiel für das Einbeziehen mehrere Sinne und Hervorrufen von Immersion sind Motion-Base-Fahrsimulatoren. Hier sorgt ein perfektes Zusammenspiel von Hardware- (Force- Feedback-Lenkräder, bewegter Untergrund) und Softwarekomponenten (wirklichkeitsgetreue Visualisierungen und Geräusche) für ein realitätsnahes Fahrerlebnis.

Perfektioniert wird Virtual Reality und damit die Immersion im CAVE (Cave Automatic Virtual Environment): ein kubischer Raum, dessen sechs Innenflächen Projektionsflächen darstellen.

Ein Vorteil von Lern- und Prüfungsumgebungen, die Immersion erzeugen, ist, dass möglicherweise durch das wirklichkeitsgetreue Empfinden auch Emotionen erzeugt werden, die denen in einer realen Situation stark ähneln. So könnte man beispielsweise valide überprüfen, wie sich eine Person unter Stress oder bei Angst verhält. Die empirische Überprüfung des Einflusses von Immersion im Fahrsimulator auf die Validität einer Testung steht allerdings derzeit noch aus.

Je stärker Immersion erzeugt werden soll, desto wirklichkeitsgetreuer muss die Lern- und Testumgebung sein. Die Bereitstellung solcher immersiver Simulatoren ist allerdings derzeit sehr kostenintensiv und daher als unökonomisch zu betrachten. Als vorteilhaft zu bewerten wäre es, wenn bereits mit weniger aufwendigen Maßnahmen, wie z.B. der Einführung dynamischer Darstellungen, ein gewisses Maß an Immersion hervorgerufen werden könnte.

## **5.2 Gestaltungsprinzipien**

Die Gestaltung von Lern- und Testmaterial kann einen Einfluss auf die Informationsverarbeitung und somit auf die (Lern-) Leistung ausüben. Bei der Erstellung von Texten, Bildmaterial und Multimediamaterial zum Lernen und Prüfen sollten sogenannte Gestaltungsprinzipien berücksichtigt werden, um Wissensinhalte optimal und angepasst an ihre Funktion darzustellen.

### **5.2.1 Generelle Gestaltungsprinzipien für Lernmaterial**

Ballstaedt (1997) formuliert die folgenden drei Grundprinzipien der Gestaltung von Lernmaterial: Funktionalität, Einfachheit und Konsistenz.

Funktionalität bedeutet dabei, dass der Wissenserwerb das Hauptziel sein sollte, wobei geprüft werden muss, welche Aspekte des Lernmaterials lernförderlich oder -hinderlich sind.

Unter Einfachheit wird verstanden, dass die Informationsvermittlung auf das Wesentliche reduziert werden soll. Dabei muss geprüft werden, ob eine Information direkt zum Verständnis beiträgt oder nicht.

Weiterhin soll stets ein konsistenter Einsatz didaktischer Mittel erfolgen. Eingeführte Standards bezüglich Schreibstil, Farbverwendung, Anordnung von Texten und Bildern usw. sollen durchgängig gewahrt werden.

Diese drei Grundprinzipien gelten auch für die Gestaltung von Testmaterial, da sie zum Instruktionsverständnis und damit zur Reliabilität beitragen.

Zunächst werden im Folgenden Prinzipien zur Gestaltung von Text vorgestellt, da Informationen zum Lernen am häufigsten textuell kodiert vorliegen (Rey, 2009). Anschließend erfolgt die Vorstellung von Gestaltungshinweisen zu Abbildungen. Im Anschluss daran werden Gestaltungsprinzipien für multimedial präsentiertes Lernmaterial vorgestellt.

#### *Prinzipien zur Textgestaltung*

Bei der Gestaltung von Texten zum Lernen spielt das Grundprinzip der Einfachheit eine große Rolle. Es bezieht sich vor allem auf die Wortwahl und den Satzbau (Langer, Schulz von Thun, & Tausch, 2006; Rey, 2009). Demnach sollten besonders anschauliche Begriffe verwendet (z.B. rhetorische Bilder, wie Metaphern; Ballstaedt, 1997) sowie Fremdwörter, Fachausdrücke oder ungeläufige Wörter vermieden oder im Vorfeld erklärt werden. Wenn viele Fachausdrücke und Fremdwörter eingesetzt werden müssen, weil sie z.B. erlernt werden sollen, ist es erforderlich, ein Glossar anzulegen (Ballstaedt, 1997). Auf komplizierte Darstellungen und verschachtelte Sätze sollte man wenn möglich verzichten.

Nach Ballstaedt (1997) ist es für das Lernen zudem förderlich, wenn Hilfen zu Verknüpfung neuer Informationen mit bereits vorhandenem Vorwissen bereitgestellt werden. Dazu eignet sich der Einsatz von Beispielen, Vergleichen und Exkursen.

Ein weiteres Prinzip zur Gestaltung von Texten zum Lernen ist das so genannte *Signaling Principle* (Mayer, 2005c). Wichtige Informationen werden hervorgehoben (z.B. durch Unterstreichung, Fettdruck), wobei keine zusätzlichen Informationen bereitgestellt, sondern lediglich die zum Verstehen essentiellen Aspekte markiert werden. Dieses Gestaltungsprinzip für Lerninhalte sollte jedoch nicht auf Prüfungsaufgaben übertragen werden, wenn das Auffinden relevanter Informationen in einer Darstellung selbst einen relevanten Teil der Aufgabe darstellt.

Für die Formulierung von textbasierten *Testaufgaben* bestehen einige Grundsätze, die wenn möglich immer berücksichtigt werden sollten (eine ausführliche Darstellung geben Jankisz & Moosbrugger, 2008).

Bei der Erstellung von Testaufgaben sollte besonders auf die *Verständlichkeit* der Formulierung geachtet werden, sodass der Testteilnehmer die Aufgaben in der Regel bereits beim ersten Durchlesen versteht (Jankisz & Moosbrugger, 2008). Eine hohe Verständlichkeit erreicht man durch eine positive Formulierung und Vermeidungen von Verneinung und besonders doppelter Verneinung. Auch die Vermeidung von komplizierten verschachtelten Satzgefügen fördert das Verständnis. Außerdem sollte auf ungeläufige Abkürzungen verzichtet werden. Die Angemessenheit der sprachlichen Formulierung an die jeweilige Zielgruppe (z.B. Kinder und Jugendliche) sollte überprüft werden.

Einen zweiten wichtigen Gesichtspunkt bei der Gestaltung von Testaufgaben stellt die *Eindeutigkeit* des Aufgabeninhalts dar (Jankisz & Moosbrugger, 2008). Zur Sicherung der Eindeutigkeit sollten anstelle von Universalausdrücken (z.B. immer, nie, keiner) mildere Formulierungen (z.B. häufig, selten, manche) eingesetzt werden. Weiterhin sollten Erklärungen und Definitionen, die man zum Verständnis der eigentlichen Frage benötigt, vor der Frage gegeben werden. Außerdem sollte der Aufgabeninhalt so präzise wie möglich formuliert werden, um den Interpretationsspielraum möglichst gering zu halten. Als relevant wird zudem betrachtet, dass eine Aufgabe nur eine Aussage beinhalten sollte, auf die sich die Frage eindeutig bezieht.

Bei der Formulierung von Testaufgaben sollte außerdem die *Aktualität der Begriffe* bedacht werden. Werden die Testaufgaben voraussichtlich über mehrere Jahre hinweg verwendet, sollten keine Formulierungen, die schnell veralten oder Beispiele aus dem Alltagsgeschehen, die später keine Relevanz oder Gültigkeit mehr aufweisen, gebraucht werden. Ein weiteres

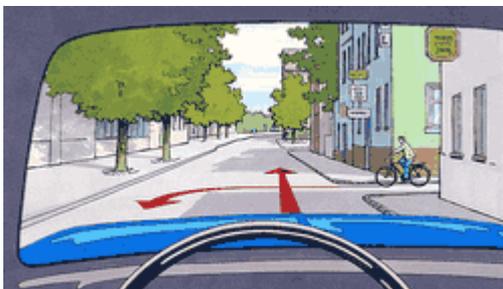
Prinzip bei der Formulierung von Testaufgaben stellt die Vermeidung von Wertungen und suggestiven Formulierungen dar.

### *Prinzipien zur Gestaltung von Abbildungen*

Häufig werden zum Lernen auch Abbildungen benutzt. Man kann nach Ballstaedt (1997) Abbilder nach ihrem Abstraktionsgrad in realistische Abbilder (Fotos, Gemälde), texturierte Abbilder (mit Schattierungen und Schraffuren), Linienabbilder (Strichzeichnungen) und schematische Abbilder (z.B. Verkehrsnetze) unterscheiden. Auch für die optimale Gestaltung solcher Abbildungen, deren zentrale Funktion die Veranschaulichung von Lerninhalten darstellt (Ballstaedt, 1997), wurden Richtlinien formuliert.

Realitäts- und detailtreue Bilder sind für das Lernen nicht immer optimal. Die Komplexität der Darstellungen sollte daher in der Regel reduziert werden. Die Reduktion der Komplexität vereinfacht es dem Lerner, relevante Bestandteile zu identifizieren.

Für Testaufgaben sollte man allerdings nicht unbedingt eine Komplexitätsreduktion vornehmen. Eine inhaltliche Abwägung sowie Anpassung an das jeweilig zu prüfende Lehrziel sollte zuvor erfolgen. Beispielaufgabe 5-1 zeigt zwei Abbildungen aus Aufgaben der derzeitigen Fahrerlaubnisprüfung. Aufgaben zu Vorfahrt und Vorrang (Beispielbild A) werden meist mit texturierten Abbildungen illustriert. Dies lenkt die Aufmerksamkeit auf die wichtigsten Bestandteile der Verkehrssituation, nämlich Straßenverlauf, Verkehrsteilnehmer und Verkehrszeichen. Diese Bestandteile der Abbildung reichen aus, um die Erreichung des Lehrziels „Anwendung der Vorfahrtsregeln“ zu überprüfen. Aufgaben zur Gefahrenlehre (Beispielbild B) werden durch realistische Abbildungen ergänzt. Nur so kann das Lehrziel „Identifikation von Gefahrenhinweisen“ valide abgeprüft werden, da diese auch im Realverkehr aus einer Vielzahl von Reizen herausgefiltert werden müssen.



A



B

**Beispielaufgabe 5-1 Abbildungen aus der derzeitigen theoretischen Fahrerlaubnisprüfung zu Aufgaben zu Vorfahrt und Vorrang (A) und zu Gefahrenlehre (B)**

Zusätzlich zur Komplexitätsreduktion kann man durch visuelles Hervorheben oder *Cueing*, die Aufmerksamkeit des Lernalers auf relevante Bildelemente lenken (Ballstaedt, 1997). Analog zum Signaling Principle, das für Textmaterial formuliert wurde, werden beim Cueing besonders wichtige Teile des Bildes in einer besonderen Form markiert. Zu diesem Zweck können farbliche Markierungen (z.B. rot in einer Schwarz-Weiß-Zeichnung), Hinweis Pfeile auf relevante Details, Überzeichnen (Vergrößerung wichtiger Bildausschnitte), unterschiedliche Strichstärken, Umrahmen und farbliche Unterlegung eingesetzt werden. Ballstaedt (1997) weist dabei darauf hin, dass nicht mehrere Mittel der Hervorhebung auf einem einzigen Bild eingesetzt werden sollen und dass man nicht zu viele Details hervorheben sollte.

Zur Veranschaulichung von Handlungen und Prozessen werden spezielle Gestaltungsempfehlungen gegeben (Ballstaedt, 1997). Es empfiehlt sich, Abfolgen von Teilhandlungen und Zuständen in der Form „Ausgangszustand – Handlung – Endzustand“ abzubilden. Dabei muss man, um eine Handlung verständlich darzustellen, einen repräsentativen Ausschnitt wählen. Zur Verdeutlichung können Bewegungspfeile, -striche oder übereinander gezeichnete Phasen hinzugefügt werden.

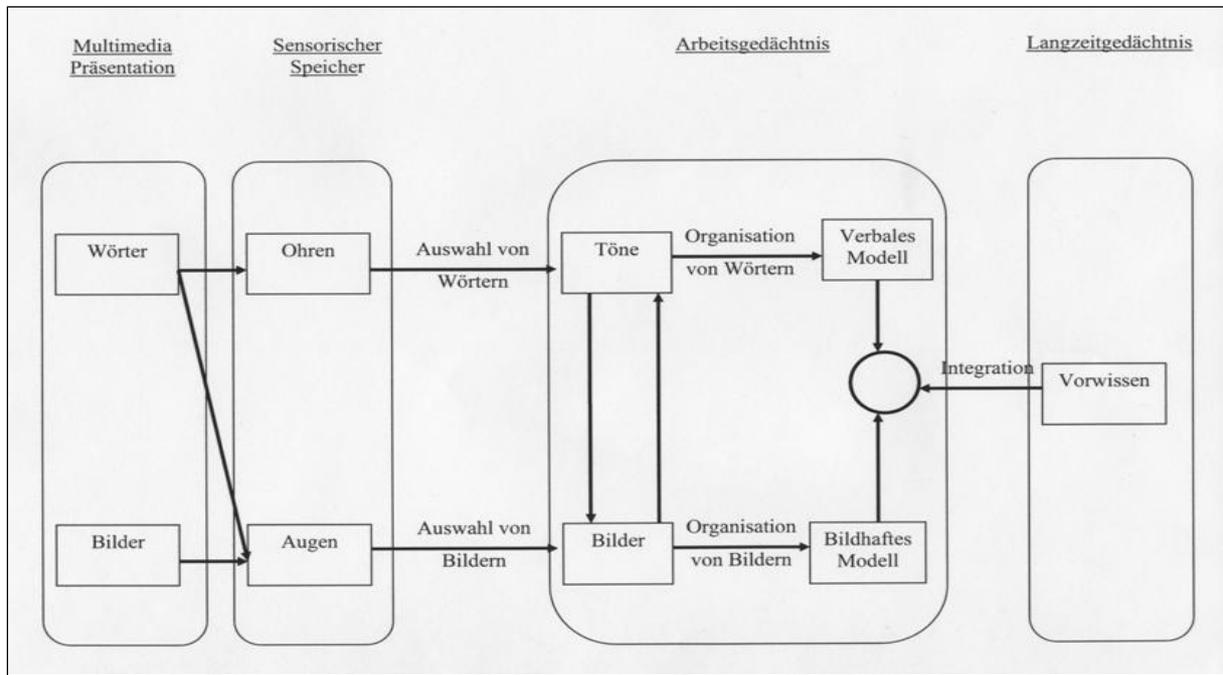
### *Prinzipien zur Gestaltung multimedialen Lernmaterials*

In Lehrbüchern und anderem Lernmaterial liegen oft Texte und Bilder zusammen vor, die sich gegenseitig aufeinander beziehen.

Eine Theorie, die sich mit der Wirksamkeit von solchem multimedialen Lernmaterial beschäftigt, ist die *Cognitive Load Theorie* (CLT) von Sweller (1988). Für das Lernen sind nach Sweller (2005) das Langzeitgedächtnis und das Arbeitsgedächtnis sowie deren Interaktion zentral. Aufgrund der CLT wird angenommen, dass die verfügbare kognitive Kapazität in einer Lernsituation begrenzt ist und abhängig von ihren Anforderungen auf unterschiedliche kognitive Prozesse aufgeteilt werden muss (Brünken, Plass, & Leutner, 2004). Die kognitive Belastung setzt sich aus drei verschiedenen Arten der Belastung zusammen. Die spezifische Anforderung, die von der Komplexität des Lerninhaltes selbst abhängt, wird als *intrinsic cognitive load* bezeichnet. Der intrinsic load von dynamischem Präsentationsmaterial hängt von der Komplexität seiner Darstellung sowie vom Vorwissen bzw. der spezifischen Expertise der Lerner ab (Sweller & Chandler, 1994). Instruktionale Bedingungen selbst, also Aspekte der äußerlichen Gestaltung von Lernmaterial, beeinflussen ebenfalls die Höhe der Beanspruchung. Diese Belastung der kognitiven Kapazität wird als *extraneous cognitive load* bezeichnet. Als *germane cognitive load* wird die kognitive Beanspruchung bezeichnet, die zu einer vertieften Auseinandersetzung des Lernenden mit

dem Lerninhalt führt und somit als lernförderlich gelten kann. Hinweise zur Gestaltung multimedialen Lernmaterials haben nach der CLT meist zum Ziel, den extraneous cognitive load zu reduzieren.

Die Grundannahme, die dafür spricht, dass überhaupt multimediale Lernsysteme bevorzugt verwendet werden sollten, um Wissen zu vermitteln, nennt sich „Multimedia-Prinzip“ (Mayer, 2001). Dieses Prinzip besagt, dass nachhaltiger gelernt wird und Lerninhalte tiefer verarbeitet werden, wenn der Lerner Texte und Bilder zum Lernen gebraucht. Das in Bild 5-1 dargestellte Modell integriert in das 3-Speicher-Modell des Gedächtnisses von Atkinson und Shiffrin (1969) Prozesse, die beim Lernen multimedial präsentierter Inhalte stattfinden. Mayers (2005b) Veranschaulichung seiner kognitiven Theorie multimedialen Lernens stellt ein Modell dar, wie Lerner mit multimedialen Lernmaterial lernen (Clark & Mayer, 2008). Demnach gelangen extern multimedial dargebotene Informationen in den sensorischen Speicher (*Sensory Memory*), bevor sie im Arbeitsgedächtnis (*Working Memory*) unter Zuhilfenahme relevanter Informationen aus dem Langzeitspeicher (*Long-Term Memory*) verarbeitet werden. Mayer zeigt in seinem Modell unterschiedliche Verarbeitungswege für visuelle und auditive Informationen auf, wobei Prozesse der *Selektion*, *Organisation* und *Integration* stattfinden. Selektion bedeutet, dass zunächst die relevanten Informationen aus dem aufgenommenen Material herausgefiltert werden. Die nachfolgende Organisation dieser wichtigen Informationen führt dazu, dass im verbalen Arbeitsgedächtnis ein textbasiertes mentales Modell (*Verbal Model*) und im imaginalen Arbeitsgedächtnis ein bildbasiertes mentales Modell (*Pictorial Model*) entsteht. Diese Modelle werden im nächsten Schritt integriert und dabei mit Vorwissen aus dem Langzeitspeicher verknüpft.



**Bild 5-1. Kognitive Theorie Multimedialen Lernens (nach Mayer, 2001)**

Eine Studie von Brünken et al. (2001) hat sich mit der Kodalität von Testaufgaben beschäftigt. Die Autoren konnten zeigen, dass bei der Verwendung von Bild und Text bei der Aufgabenpräsentation die Leistung bei prozeduralem Wissen besser war als bei den nur textuell dargebotenen Aufgaben. Dieser Effekt zeigte sich besonders, wenn die Präsentation des Lernmaterials ebenfalls durch Text und Bild erfolgt war. Beim Abruf von deklarativem Wissen konnte ein solcher Effekt nicht nachgewiesen werden. Die Ergebnisse sprechen dafür, eine Passung zwischen Informationsdarbietung (Wie wurde gelernt?) und Informationsabruf (Testen) anzustreben.

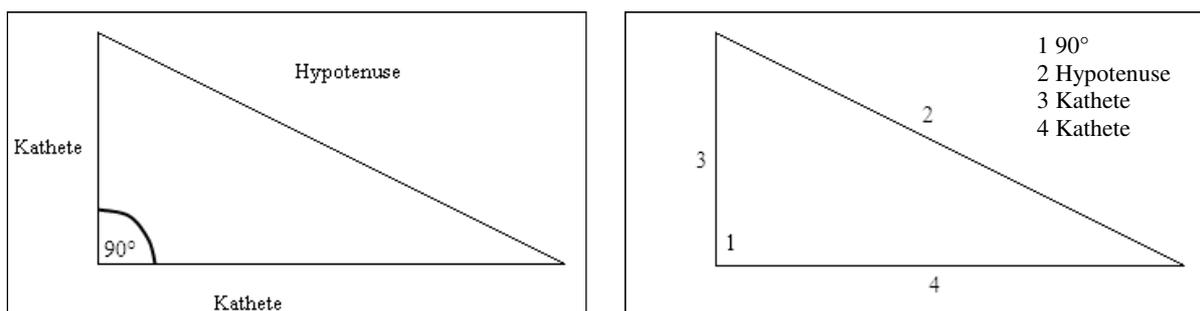
Beim Multimedia-Prinzip handelt es sich um ein generelles Gestaltungsprinzip für Lernmaterial. Um von seiner lernförderlichen Wirkung optimal profitieren zu können, sollten weitere Gestaltungsprinzipien, die im Besonderen für multimedial präsentiertes Material gelten, beachtet werden (Mayer, 2005c). Diese Prinzipien sollen dazu dienen, die kognitive Belastung für den Lerner, die von der Gestaltung der Instruktion verursacht wird, weiter zu reduzieren.

Das *Modalitäts-Prinzip* ist in dem sogenannten Modalitätseffekt begründet. Dieser Effekt besagt, dass es sich beim Lernen mit Text und Bild (bzw. Text und Animation) vorteilhafter für den Wissenserwerb erweist, den Text auditiv darzubieten als es bei einer rein visuellen Textdarbietung der Fall wäre (Brünken & Leutner, 2001; Brünken, Steinbacher, Plass, & Leutner, 2002). Das Prinzip, das sich hieraus ableitet, besagt, dass bei Lernmaterial, welches Bildmaterial enthält, der Text dazu auditiv dargeboten werden sollte.

Bisher sind jedoch keine Studien bekannt, die sich damit beschäftigen, ob das Modalitätsprinzip auch für die *Prüfung* von Lerninhalten Vorteile mit sich bringt. Man kann aber vermuten, dass es, dem Modalitätsprinzip zu Folge, die kognitive Belastung des Prüflings reduzieren könnte, wenn der Fragetext zu einem dargebotenen Bild auditiv anstatt visuell aufgenommen wird.

Ein weiteres Prinzip zur Gestaltung multimedialer Lernumgebungen – das *Redundanz-Prinzip* – besagt, dass keine redundanten Informationen angeboten werden sollen. Der Lerner benötigt zusätzliche kognitive Kapazität, um zu erkennen, dass die dargebotenen Informationen redundant sind, was als störend bei der Verarbeitung der relevanten Sachverhalte empfunden wird. Das Redundanz-Prinzip besagt im Hinblick auf die Gestaltung multimodaler Lernaufgaben, dass gesprochener und gleichzeitig visuell dargebotener Text nicht exakt die gleiche Information bereitstellen sollen, da die wahrgenommene kognitive Belastung dadurch erhöht wird (Kalyuga, 2000). Diese Störeffekte konnten bislang nicht gefunden werden, wenn ein Bild und ein gesprochener Text redundante Informationen präsentieren (Kalyuga, 2000). Das Redundanz-Prinzip kann auch auf Prüfungsaufgaben übertragen werden, da die Aufgabenbearbeitung nicht durch Redundanzen gestört werden sollte.

Das *Prinzip der räumlichen Nähe* besagt, dass Informationen oder Informationsquellen, die sich inhaltlich aufeinander beziehen, in räumlicher Nähe zueinander präsentiert werden sollten (Mayer, 2005c). So sollte beispielsweise ein erklärender Text in eine Zeichnung integriert werden anstatt ihn separat zu präsentieren (vgl. Bild 5-2), um es dem Lerner zu erleichtern, Rückschlüsse auf die Kohärenz der Informationen zu ziehen.



**Bild 5-2 Darstellung eines rechtwinkligen Dreiecks mit integriertem Text (links) und separat präsentiertem Text (rechts).**

Analog dazu besagt das *Prinzip zeitlicher Nähe*, dass Informationen oder Informationsquellen, die sich inhaltlich aufeinander beziehen, in zeitlicher Nähe zueinander präsentiert werden sollten (Mayer, 2005c).

Der Vorteil multimodal präsentierter Informationen kommt besonders zum Tragen, wenn visuelle und auditive Informationen simultan dargeboten werden, da bei einer zeitlich

aufeinanderfolgenden Darbietung durch die Integration der Informationen zu viele Ressourcen benötigt werden (Mayer, 2001).

Die Prinzipien räumlicher und zeitlicher Nähe sollten wenn möglich auch bei Testmaterial verfolgt werden, um es dem Prüfling zu erleichtern, Bezüge zwischen Bild und Text herzustellen, insofern dies nicht Teil der Prüfungsleistung ist.

### **5.2.2 Prinzipien zur Gestaltung dynamisch präsentierten Lernmaterials**

Neuere Lernmaterialien, die in elektronischer Form und zuweilen auch online angeboten werden, beinhalten häufig dynamisch präsentiertes Bildmaterial. Besonders zur Darstellung von Handlungen ist dynamisches Material hilfreich (Clark & Lyons, 2011). Selten wird jedoch ausschließlich dynamisches Bildmaterial dargeboten. Meist wird gesprochener oder geschriebener Text integriert. Auch für die Gestaltung von Lernmaterial, das dynamische Darstellungen beinhaltet, wurden Richtlinien formuliert. Während Betrancourt (2005) nur die fünf Prinzipien *Apprehension Principle*, *Congruence Principle*, *Interactivity Principle*, *Attention-Guiding Principle* und das *Flexibility Principle* aufführt, werden bei Mayer (2008) zehn Prinzipien aufgestellt, für die empirische Evidenz vorliegt. Mayer leitet die Wirkung dieser Prinzipien aus der *Cognitive Theory of Multimedia Learning* ab. Nach Mayer (2008) ist animiertes Lernmaterial für den Lerner dann besonders anspruchsvoll, wenn es inhaltlich komplex ist, die Animation schnell abläuft und der Lerner mit dem Material nicht vertraut ist. Er verbindet drei Ziele mit den zehn aufgestellten Gestaltungsprinzipien für dynamisches Lernmaterial: (1) Reduktion extrinsischer Belastung, (2) Bewältigung intrinsischer Belastung und (3) Förderung generativer Verarbeitung. Die theoretische Grundlage für diese dreigeteilte Zielsetzung liefert die Cognitive Load Theory (Sweller, 1988, 2005; Sweller & Chandler, 1994).

Im Folgenden werden die wichtigsten Gestaltungsprinzipien beim Lernen mit dynamisch präsentiertem Bildmaterial dargestellt. Zudem wird diskutiert, inwieweit es angemessen erscheint, diese Gestaltungsprinzipien bei der Erstellung von Prüfungsmaterial, das Animationen enthält, zu berücksichtigen. Dies geschieht vor dem Hintergrund der aktuellen Diskussion, inwieweit dynamische Präsentationen von Verkehrssituationen in die theoretische Fahrerlaubnisprüfung sinnvoll integriert werden können.

#### *Gestaltungsprinzipien zur Reduktion extrinsischer Belastung*

Lernmaterial erzeugt extrinsische Belastung, wenn es durch seine Gestaltung verarbeitungsintensiv für den Lerner ist. Die folgenden Prinzipien dienen demnach der Reduktion dieser gestaltungsbedingten Belastung durch das Lernmaterial.

Das *Coherence Principle* (Mayer, 2008) bzw. *Apprehension Principle* (Betrancourt, 2005) besagt, dass kein Material präsentiert werden sollte, das nicht unmittelbar zum Verständnis des Lerninhaltes beiträgt. Nach Betrancourt (2005) bedeutet dies für die Vermittlung von Lernstoff auch, dass auf realitätsgetreue Komplexität verzichtet werden kann, wenn eine einfachere Darstellung das Lernen erleichtern könnte. Auch bei Prüfungsmaterial sollten die eingesetzten Animationen keine irrelevanten oder rein dekorativen Elemente beinhalten, da diese möglicherweise das Instruktionsverständnis gefährden, wenn zum Beispiel nicht deutlich wird, worum es bei der Aufgabe geht. Dadurch besteht die Gefahr der Erhöhung des Messfehlers und damit einhergehend eine Einschränkung der Genauigkeit der Messung. Dieses Gestaltungsprinzip steht in Verbindung mit dem *Seductive Details-Effekt*, der besagt, dass es einen negativen Einfluss auf das Lernen hat, wenn das Lernmaterial, statt nur relevante Fakten zu liefern, mit Informationen angereichert ist, die für den Lerner zwar interessant, aber zur Lösung der Aufgaben irrelevant sind. Diese interessanten Details können das Lernen erschweren, indem sie das Interesse des Lerners auf sich lenken und damit die Aufmerksamkeit vom eigentlichen Lerngegenstand abziehen (Kintsch, 1996). Moderiert wird der Effekt laut Garner et al. (1991) unter anderem durch das Vorwissen. Bisher wurde der Effekt der Seductive Details lediglich für das Lernen untersucht (z.B. Park, 2010).

Aber auch bei der Erstellung von Prüfungsmaterial könnte der Einbau solcher interessanter aber ablenkender Details von Bedeutung sein. Durch die Anreicherung des Prüfungsmaterials mit Seductive Details wäre ein Einfluss auf die allgemeine Schwierigkeit der Aufgaben denkbar. Möglicherweise verändert sich auch der Einfluss von Seductive Details in Abhängigkeit von dem Vorwissen des Prüflings.

Ein weiteres Gestaltungsprinzip wird *Signaling Principle* (Mayer, 2008) bzw. *Attention-Guiding Principle* (Betrancourt, 2005) bzw. *Cueing* (Plass, Homer, & Hayward, 2009) genannt. Da es in der Natur der Animation liegt, dass durch die dargestellte Veränderung das vorangegangene Bild nicht wieder aufrufbar ist, könnten durch einen kurzen Moment der Unaufmerksamkeit wichtige Details übersehen werden. Daher sollte der Lerner auf relevante Merkmale, die sich innerhalb der Darstellung verändern, aufmerksam gemacht werden. Gerade die Novizen in einer Domäne können noch nicht zwischen relevanten und irrelevanten Aspekten unterscheiden und könnten beispielsweise durch verbale Äußerungen darauf aufmerksam gemacht werden. Beim Prüfen mithilfe von Animationen ist dieses Gestaltungsprinzip anzuwenden, wenn das Instruktionsverständnis gefördert werden soll. Es ist allerdings zu überprüfen, ob nicht gerade die Prüfungsaufgabe darin besteht, das Wichtige

eigenständig vom Unwichtigen zu trennen. In diesem Fall sollten keine Hilfestellungen gegeben werden.

Dem *Redundancy Principle* (Mayer, 2008) zufolge sollten keine redundanten Informationen gegeben werden. Das Abstimmen mit den anderen Informationen, um festzustellen ob die gegebenen Informationen tatsächlich übereinstimmen, nimmt kognitive Kapazität ein. Daher sollte man z.B. Text, der auditiv zu einer Animation begleitend eingespielt wird, nicht zusätzlich visuell auf dem Bildschirm zeigen. Dieses Prinzip ist auch auf das Prüfen übertragbar, da auch dabei die Informationsverarbeitung nicht durch redundante Informationen gestört werden soll.

Dem *Spatial Contiguity Principle* (Mayer, 2008) zufolge sollten relevante Informationen in der Animation in geringem räumlichen Abstand zueinander präsentiert werden, um eine Integration zu erleichtern. Wird in einer Animation auch Text visuell dargeboten, sollte dieser eher innerhalb der Animation präsentiert werden und möglichst nahe bei der Bezugsregion auftauchen (auch *Split Attention principle*; Plass, Homer, & Hayward, 2009). Dieses Gestaltungsprinzip ist auch auf Prüfungsaufgaben anzuwenden, wenn es sich dabei nicht gerade um Such- und Zuordnungsaufgaben handelt, bei denen eigenständige Integration von Informationen verlangt wird.

Analog dazu besagt das *Temporal Contiguity Principle*, dass inhaltlich zusammengehörige Informationen in Animationen auch in geringem zeitlichen Abstand zueinander gezeigt werden sollten (Mayer, 2008). Besonders gesprochener Text, der zur Erklärung der Animation präsentiert wird, sollte zeitlich an die Animation angepasst werden. Das was gerade auf dem Bildschirm erscheint, sollte unmittelbar kommentiert werden. Dieses Prinzip ist auch auf Animationen, die zu Prüfungszwecken eingesetzt werden, zu übertragen.

### *Gestaltungsprinzipien zur Bewältigung intrinsischer Belastung*

Die intrinsische Belastung hängt von der Schwierigkeit bzw. der Komplexität des Lernmaterials selbst ab. Die intrinsische Belastung kann daher durch die Gestaltung des Lernmaterials nicht beeinflusst werden. Man kann jedoch durch eine günstige Gestaltung Einfluss darauf nehmen, wie gut mit der intrinsischen Belastung umgegangen wird.

Laut dem *Segmenting Principle* (Mayer, 2008) sollte das animierte Lernmaterial in sinnvoller Weise in Lernsegmente unterteilt werden. Das Lernen wird zusätzlich unterstützt, wenn Lerner die Darbietung diese Segmente selbst kontrollieren können. Um eine Überlastung der Teilnehmer zu verhindern, erscheint es auch bei Prüfungsaufgaben sinnvoll, das animierte Prüfungsmaterial in bedeutungsvolle Einheiten zu unterteilen und dem Prüfling die Möglichkeit zu geben, selbst zu entscheiden, wann das nächste Segment abgespielt wird.

Dem *Pre-training Principle* zufolge ist es für das Lernen förderlich, wenn die Namen und Merkmale der Hauptkomponenten der Animation bereits bekannt sind (Mayer, 2008). Diese Wissensinhalte sollten daher zuvor erworben werden, damit sich der Lerner auf die in der dargebotenen Animation dargestellten Prozesse konzentrieren kann. Auch für die Güte von Prüfungsaufgaben kann es von Vorteil sein, wenn sichergestellt wird, dass die Prüflinge die Namen und Hauptmerkmale der wichtigsten Komponenten der Animation kennen. Dadurch wird gewährleistet, dass die Leistung in Prüfungsaufgaben, die eigentlich darauf abzielen, zu überprüfen, ob dynamische Prozesse verstanden werden, nicht vom deklarativen Wissen bzgl. auftretender Komponenten abhängt. Zudem sollte auch der generelle Umgang mit den Testaufgaben, die dynamische Präsentationen beinhalten, trainiert werden.

Das *Modality Principle* besagt, dass Text bei der Präsentation einer Animation eher auditiv als visuell dargeboten werden sollte (Mayer, 2008). Dadurch wird eine Überlastung des visuellen Kanals bei der Informationsverarbeitung vermieden. Das Modalitätsprinzip sollte auch bei der Prüfung Anwendung finden, um den Prüfungsteilnehmern die Möglichkeit zu geben, alle wichtigen Informationen, die sie zur Lösung der Aufgaben benötigen, aufzunehmen zu können, ohne, dass es zu einer Überlastung kommt.

#### *Prinzipien zur Unterstützung generativer Verarbeitung*

Generative Verarbeitung beinhaltet das Auswählen relevanter Information aus dem Lernmaterial, die mentale Organisation in bildliche und verbale mentale Modelle, das Verbinden dieser Modelle mit dem Vorwissen und der Modelle untereinander. Generative Verarbeitung führt zu tiefem Verständnis, das sich zum Beispiel in der Fähigkeit zum Transfer zeigt. Welchen Effekt generative Verarbeitung auf Testleistungen hat, ist bisher nicht geklärt. Die Vermutung, dass der Aufbau eines tieferen Verständnisses die Leistung befördert, liegt jedoch nahe. Die beiden von Mayer (2008) angeführten Gestaltungsprinzipien zur Unterstützung generativer Verarbeitung sollten daher wenn möglich auch in Prüfungsmaterial berücksichtigt werden.

Laut dem *Personalization Principle* lernen Individuen dann besser aus Animationen und entwickeln ein tieferes Verständnis der gezeigten Prozesse, wenn sie im Begleittext sprachlich direkt angesprochen werden und dieser unterstützende Text eher einer Konversation als einem formalen Text gleicht (Mayer, 2008).

Das *Voice Principle* besagt ergänzend dazu, dass Text, der während einer Animation auditiv präsentiert wird, zur Förderung generativer Verarbeitung nicht durch eine computergenerierte Stimme oder eine Stimme mit ausländischem Akzent dargeboten werden sollte (Mayer, 2008).

Bisher wurden noch keine Prinzipien für die Gestaltung von Testaufgaben entwickelt, die dynamische Präsentationen einschließen. Die angeführten Gestaltungsprinzipien für die dynamische Darbietung von Lerninhalten lassen sich auch nicht unmittelbar auf die Prüfung von Lerninhalten anwenden. Bedingungen, die das Lernen erleichtern, wie zum Beispiel die Reduktion der Komplexität, sollten zur Folge haben, dass Prüfungsaufgaben lediglich vereinfacht werden und zudem an Validität verlieren. An Aufgaben der theoretischen Fahrerlaubnisprüfung wird aber der Anspruch gestellt, dass sie möglichst gut zwischen Könnern und Nichtkönnern trennen. Hinsichtlich einer konzeptionellen und empirischen Trennung von kognitiver Belastung einerseits und Aufgabenschwierigkeit andererseits besteht derzeit noch ein erheblicher Forschungsbedarf (Brünken, Plass, & Leutner, 2003).

### ***5.3 Implikationen für die Erstellung neuer Aufgabenformate für die Fahrerlaubnisprüfung***

Multimediales Material kann durch fünf Dimensionen klassifiziert werden: Kodalität, Modalität, Dynamik/Animation, Interaktivität/Adaptivität und Immersion. Bisher variieren die Aufgaben der theoretischen Fahrerlaubnisprüfung nur auf der Dimension der Kodalität, da es reine Textaufgaben und Textaufgaben mit statischen Bildern gibt. Neue Aufgabeformate könnten Variationen auf den übrigen Dimensionen mit sich bringen.

Für die Erstellung von Lern- und Prüfungsmaterial wurden Gestaltungshinweise vorgestellt. Dabei wurde zwischen Textmaterial, Bildmaterial und Multimediamaterial unterschieden.

Um das Lernen mit multimedialem Lernmaterial zu optimieren, wurden theoretisch und empirisch fundierte Gestaltungsempfehlungen formuliert. Da es angestrebt wird, in die theoretische Fahrerlaubnisprüfung auch dynamisches Bildmaterial zu implementieren, wurden Gestaltungsprinzipien hierfür besonders eingehend behandelt.

Bei Testmaterial, das dynamische Präsentationen enthält, sollte Text auditiv präsentiert werden. Zudem sollten Redundanzen vermieden werden. Text und Animation sollten in räumlicher und zeitlicher Nähe zueinander präsentiert werden. Animationen sollten in bedeutungsvolle Einheiten unterteilt werden und als solche vom Testteilnehmer selbstständig abgespielt werden können. Vor dem Test sollte die Benennung der Bestandteile der dynamischen Präsentationen erlernt werden und der Umgang mit den Animationen trainiert werden.

## 6 Zusammenfassung

Es wird angestrebt, die theoretische Fahrerlaubnisprüfung durch den Einsatz neuer Aufgabenformate zu optimieren. Dabei sollten die drei Hauptgütekriterien Objektivität, Reliabilität und Validität gewahrt und die Bedeutsamkeit der Nebengütekriterien Ökonomie und Testfairness für die theoretische Fahrerlaubnisprüfung berücksichtigt werden.

Die theoretische Fahrerlaubnisprüfung ist ein kriteriumsorientierter Test. Die Erstellung neuer Aufgabenformate sollte daher nach den Prinzipien der mehrschrittigen Konstruktion kriteriumsorientierter Tests erfolgen.

Bei der Wahl neuer Antwortformate haben sich handlungsnahe Formate als vielversprechend erwiesen. Insbesondere die Reaktionszeiterfassung hat sich in ihrer Güte bewährt und wird daher als Prüfungsformat für die theoretische Fahrerlaubnisprüfung empfohlen.

In die Fahrerlaubnisprüfung, die die Erreichung von Lehrzielen überprüfen soll, die sich auf das Verhalten in einem dynamischen Umfeld beziehen, sollen dynamische Präsentationen integriert werden, so dass sowohl ein handlungsnahes Antwortverhalten als auch eine realitätsnahe Darstellung der Inhalte erfolgen kann.

Bei der Gestaltung von Testmaterial, das heißt der Formulierung von Aufgaben, der Erstellung von Abbildern und Animationen sowie der Zusammenstellung dieser drei Arten sollten bestimmte theoretisch und empirisch fundierte Gestaltungshinweise befolgt werden.

Für Textaufgaben gilt dabei, dass sie verständlich, eindeutig und frei von Wertungen und kurzlebigen Begriffen formuliert sein sollten. Abbildungen sollten in ihrer Komplexität an das Lehrziel angepasst sein. Dynamische Präsentationen im Testmaterial sollten durch auditiv präsentierten Text begleitet werden, während man Redundanzen vermeiden sollte. Förderlich erscheint zudem eine räumliche und zeitliche Nähe von Animation und zugehörigem Text. Testteilnehmern sollte die Möglichkeit gegeben werden, Animationen, in bedeutungsvolle Einheiten unterteilt, selbstständig abspielen zu können. Die Begriffe, die wichtig zum Verständnis einer eingesetzten Animation sind, sollten im Vorhinein erlernt werden. Den Testteilnehmern sollte vor dem eigentlichen Test die Möglichkeit gegeben werden, den Umgang mit Animationen, die in einer Testung zum Einsatz kommen, zu trainieren.

## 7 Checklisten zur Gestaltung von Testaufgaben

Die folgenden Checklisten enthalten in übersichtlicher, expliziter Form die Gestaltungsempfehlungen zur Erstellung von Testaufgaben, die im vorliegenden Manual vorgestellt wurden. Anhand dieser Checklisten kann überprüft werden, inwieweit alle relevanten Gestaltungsregeln für Text, Abbildungen, Animationen und Multimediamaterial bei der selbstständigen Konstruktion von Testaufgaben erfüllt werden.

### 7.1 Checkliste Gestaltung von Texten

1. Werden anschauliche Begriffe (z.B. Metaphern) verwendet?
2. Wird auf Fremdwörter, Fachausdrücke und ungeläufige Begriffe verzichtet oder wurden diese zuvor definiert?
3. Sind die Sätze einfach strukturiert (keine Schachtelsätze)?
4. Sind Fragen und Aussagen positiv formuliert (keine Verneinungen)?
5. Ist die Formulierung an die Zielgruppe angepasst?
6. Werden Universalausdrücke (z.B. immer, nie, keiner) abgemildert (z.B. häufig, selten, manche)?
7. Sind die Aufgabeninhalte präzise formuliert?
8. Beinhaltet die Aufgabe nur eine einzige Aussage, auf die sich die Frage eindeutig bezieht?
9. Wird auf *moderne* Formulierungen verzichtet, die schnell veralten können?
10. Wird auf die Verwendung von Beispielen aus dem aktuellen Tagesgeschehen verzichtet?
11. Sind die Aussagen wertfrei und nicht suggestiv?
12. Werden die Teilnehmer im Text direkt angesprochen?
13. Wird auditiv dargebotener Text durch eine natürliche, nicht computergenerierte Stimme vorgetragen?
14. Wird auditiv dargebotener Text akzentfrei vorgetragen?

### 7.2 Checkliste Gestaltung von Abbildungen

1. Ist die Komplexität der Abbildung an das abzurufende Lehrziel angepasst (ggf. reduziert)?
2. Sind einzelne wichtige Bilddetails hervorgehoben?

3. Sind Handlungsvorgänge in der Form: Ausgangszustand - Handlung - Endzustand abgebildet?
4. Sind repräsentative Ausschnitte eines Prozesses abgebildet?
5. Sind Mittel zur Verdeutlichung von Bewegung (z.B. Pfeile, sich überlagernde Phasen) eingefügt?

### **7.3 Checkliste Gestaltung von Animationen**

1. Enthält die Animation keine rein dekorativen Elemente?
2. Ist die Komplexität der Animation (Realitätsnähe) an das abzurufende Lehrziel angepasst?
3. Wird während der Animation auf relevante Aspekte hingewiesen<sup>3</sup>?
4. Ist die Animation in bedeutungsvolle Segmente unterteilt?
5. Kann die Präsentation der einzelnen Segmente der Animation vom Teilnehmer selbstständig gestartet werden?
6. Sind die Hauptkomponenten der Animation den Teilnehmern bekannt?
7. Wurde der Umgang mit den Animationen bereits trainiert?

### **7.4 Checkliste Gestaltung von Multimediamaterial**

1. Wird Begleittext zu einer Abbildung oder einer Animation auditiv dargeboten?
2. Werden redundante Informationen vermieden?
3. Befinden sich Abbildung und Begleittext in räumlicher Nähe zueinander?
4. Befindet sich die Beschriftung in räumlicher Nähe zu dem zugehörigen Ausschnitt der Abbildung?
5. Werden Abbildung oder Animation und auditiv dargebotener Begleittext möglichst simultan dargeboten?
6. Ist die Geschwindigkeit des auditiv dargebotenen Begleittexts an die in der Animation dargestellten Prozesse angepasst?

---

<sup>3</sup> Nur, wenn es sich nicht explizit um die Anforderung der Aufgabe handelt, relevante Informationen innerhalb der Animation zu identifizieren.

## 8 Literaturverzeichnis

- Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Atkinson, R. C., & Shiffrin, R. M. (1969). Storage and retrieval processes in short-term memory. *Psychological Review*, 76, 179-193.
- Ballstaedt, S.-P. (1997). *Wissensvermittlung. Die Gestaltung von Lernmaterial*. Weinheim: Beltz-PVU.
- Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287-296). New York: Cambridge University Press.
- Bloom, B. S. (1974). *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz.
- Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals (Handbook I. Cognitive Domain)*. New York Longman.
- Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder Aufmerksamkeitsfokussierung? Empirische Ergebnisse zur "Split-Attention-Hypothese" beim Lernen mit Multimedia. *Unterrichtswissenschaft*, 29, 357-366.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38, 53-61.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: auditory load and modality effects. *Instructional Science*, 32, 115-132.
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, 49(2), 109-119.
- Brünken, R., Steinbacher, S., Schnotz, W., & Leutner, D. (2001). Mentale Modelle und Effekte der Präsentations- und Abrufkodalität beim Lernen mit Multimedia. *Zeitschrift für pädagogische Psychologie*, 15(1), 16-27.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Chapman, P. R., & Underwood, G. (1998). Visual search of dynamic scenes: Event types and the role of experience in viewing driving situations. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 369-393). Oxford: Elsevier.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard Setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, California: SAGE Publications.
- Clark, R. C., & Lyons, C. (2011). *Graphics for learning. Proven guidelines for planning designing, and evaluating visuals in training materials* (2 ed.). San Francisco: Pfeiffer.
- Clark, R. C., & Mayer, R. E. (2008). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (2 ed.). San Francisco: Pfeiffer.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dave, R. H. (1968). Eine Taxonomie pädagogischer Ziele und ihre Beziehung zur Leistungsmessung. In K. Ingenkamp & T. Marsolek (Eds.), *Möglichkeiten und Grenzen der Testanwendung in der Schule* (pp. 225-238). Weinheim: Verlag Julius Beltz.
- Deery, H. A. (1999). Hazard and risk perception among young novice drivers. *Journal of Safety Research*, 30(4), 225-236.

- Frey, A., & Seitz, N.-N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies In Educational Evaluation*, 35(2-3), 89-94.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: W. Kohlhammer.
- Garner, R., Alexander, P. A., Gillingham, M. G., Kulikowich, J. M., & Brown, R. (1991). Interest and learning from text. *American Educational Research Journal*, 28, 643-659.
- Grau, O. (2003). *Virtual Art: From Illusion to Immersion*. Cambridge: Mass MIT Press.
- Harrow, A. (1972). *A Taxonomy of the Psychomotor Domain: A Guide for Developing Behavioral Objectives*. New York: McKay.
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722-738.
- Hull, M. A., & Christie, R. J. (1993). *The hazard perception test: The Geelong trial and future developments* (VicRoads Report GR 93-13).
- Jankisz, E., & Moosbrugger, H. (2008). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 27-72). Berlin: Springer.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human -- Computer Studies*, 66(9), 641-661.
- Jude, N., & Wirth, J. (2007). Neue Chancen bei der technologiebasierten Erfassung von Kompetenz. In J. Hartig & E. Klieme (Eds.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (pp. 49-56). Berlin: Bundesministerium für Bildung und Forschung.
- Jurecka, A., & Hartig, J. (2007). Computer- und netzwerkbasierendes Assessment. In B. f. B. u. F. (BMBF) (Ed.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (pp. 37-48).
- Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology*, 16(2), 161-172.
- Kerres, M. (2001). *Multimediale und telemediale Lernumgebungen*. München: Oldenbourg.
- Kintsch, W. (1996). Lernen aus Texten. In J. Hoffmann & W. Kintsch (Eds.), *Enzyklopädie der Psychologie. Lernen* (pp. 503-528). Göttingen: Hogrefe.
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Klauer, K. J., & Leutner, D. (2007). *Lehren und Lernen. Einführung in die Instruktionspsychologie*. Weinheim: Beltz-PVU.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1975). *Taxonomie von Lernzielen im affektiven Bereich* (H. Dreesmann, Trans.). Weinheim und Basel: Beltz.
- Langer, I., Schulz von Thun, F., & Tausch, R. (2006). *Sich verständlich ausdrücken* (Vol. 8). München: Reinhardt.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, UK: Cambridge University Press.
- Mayer, R. E. (2005a). *Cambridge Handbook of Multimedia Learning*. Santa Barbara, USA: Cambridge University Press.
- Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning*. Cambridge, MA: Cambridge University Press.
- Mayer, R. E. (2005c). Principles for reducing extraneous processing in multimedia learning: Coherence, Signaling, Redundancy, Spatial Contiguity and Temporal Contiguity. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 183-200). New York: Cambridge University Press.

- Mayer, R. E. (2008). Research-based principles for learning with animation. In R. Lowe & W. Schnotz (Eds.), *Learning with animation. Research implications for design* (pp. 30-48). Cambridge: Cambridge University Press.
- Mayer, R. E., & Chandler, P. (2001). When Learning Is Just a Click Away. Does Simple User Interaction Foster Deeper Understanding of Multimedia Messages? *Journal of Educational Psychology*, 93(2), 390-397.
- Mayhew, D. R., & Simpson, H. M. (1995). *The role of driving experience: Implications for the training and licensing of new drivers*. Toronto, Ontario: Insurance Bureau of Canada.
- McKenna, F. P., & Crick, J. L. (1991). Experience and expertise in hazard perception. In G. B. Grayson & J. F. Lester (Eds.), *Behavioural research in road safety*. Crowthorne, Berkshire: Transport and Road Research Laboratory.
- McKenna, F. P., & Crick, J. L. (1994). *Developments in hazard perception. Final Report: Department of Transport (UK)*.
- Mills, K. L., Hall, R. D., McDonald, M., & Rolls, G. (1998). *The effects of hazard perception training on the development of novice driver skills* (Road Safety Report No. 4). London: Department of the Environment, Transport and the Regions.
- Moosbrugger, H., & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 7-26). Berlin: Springer.
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19.
- Nußbaum, A. (1987). Das Modell der Generalisierbarkeitstheorie. In K. J. Klauer (Ed.), *Kriteriumsorientierte Tests* (pp. 114-136). Göttingen: Hogrefe.
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.
- Park, B. (2010). Testing the Additivity Hypothesis of Cognitive Load Theory. Saarbrücken: Dissertation, Digitale Bibliothek der Universität des Saarlandes, [http://scidok.sulb.uni-saarland.de/volltexte/2010/3478/pdf/Dissertation\\_BP\\_2010.pdf](http://scidok.sulb.uni-saarland.de/volltexte/2010/3478/pdf/Dissertation_BP_2010.pdf).
- Pelz, D. C., & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident Analysis and Prevention*, 6, 45-58.
- Plass, J. L., Homer, B. D., & Hayward, E. O. (2009). Design Factors for Educationally Effective Animations and Simulations. *Journal of Computing in Higher Education*, 21(1), 31-61.
- Quimby, A. R., Maycock, G., Carter, I. D., Dixon, R., & Wall, J. G. (1986). *Perceptual abilities of accident involved drivers* (TRL Research Report 27). Crowthorne, Berkshire TRL Limited.
- Quimby, A. R., & Watts, G. R. (1981). *Human factors and driving performance* (TRL Report LR1004). Crowthorne, Berkshire: Transport and Road Research Laboratory.
- Renkl, A., & Atkinson, R. K. (2007). Interactive learning environments: contemporary issues and trends. An introduction to the special issue. *Educational Psychology Review*, 19, 235-238.
- Rey, G. D. (2009). *E-Learning. Theorien, Gestaltungsprinzipien und Forschung*. Bern: Huber.
- Rost, J. (2004). *Lehrbuch Testtheorie- Testkonstruktion* (Vol. 2., vollständig überarbeitete und erweiterte Auflage). Bern: Verlag Hans Huber.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction: The Journal of the European Association for Research on Learning and Instruction*, 14(3).
- Shrock, S., & Coscarelli, W. (2000). *Criterion-referenced test development*. Silver-Spring, MD: ISPI.

- Siegrist, S. (Ed.). (1999). *Driver Training, testing and licensing - towards theory-based management of young drivers' injury risk in road traffic. Results of EU-Project GADGET, Work Package 3*. Bern.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19-30). Cambridge, MA: Cambridge University Press.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 186-233.
- Tversky, B., Morrison, J.-B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human Computer Studies*, 57, 247-262.