Testkritische Analyse der Realkennzeichen nach Steller und Köhnken anhand von Daten aus Glaubhaftigkeitsgutachten

Wissenschaftliche Arbeit
zur Erlangung des Grades einer Diplom-Psychologin
im Fachbereich Psychologie
der Universität Konstanz

vorgelegt von

Domenica Schwind
Untertorstraße 24
70771 Leinfelden-Echterdingen

Erstgutachter: Professor Dr. Peter Steck
Zweitgutachter: Professor Dr. Max Hermanutz

Konstanz, im Oktober 2006

DANKSAGUNG

Ich danke der GWG München, insbesondere Herrn Dr. Joseph Salzgeber und Frau Dr. Monika Aymans für die zur Verfügung gestellten Glaubhaftigkeitsgutachten, ohne die die vorliegende Arbeit nicht möglich gewesen wäre. Darüber hinaus bedanke ich mich bei Frau Diana Goßmann für die Unterstützung und die zahlreichen Auskünfte.
Herrn Prof. Dr. Steck möchte ich für die nette Betreuung danken und Herrn Prof. Dr. Hermanutz dafür, dass er sich vor der Anmeldung der Diplomarbeit so kurzfristig für die Zweitbegutachtung dieser Diplomarbeit bereiterklärt hat.

INHALTSVERZEICHNIS

TABELLENVERZEICHNIS

1. EINLEITUNG

Die Frage nach der Glaubhaftigkeit von Zeugenaussagen, das heißt die Frage: ‚Stimmt die vorliegende Aussage in zuverlässiger Weise mit der Realität überein?' beschäftigt Juristen und Psychologen[1] schon seit geraumer Zeit. Gerade in solchen Fällen, in denen kaum Sachbeweise oder materielle Spuren zur Verfügung stehen, sind vor Gericht die Aussagen von Augenzeugen und deren Zuverlässigkeit von größter Bedeutung. Wenn darüber hinaus keine unbeteiligten Tatzeugen vorhanden sind und der Angeklagte die Vorwürfe abstreitet, wie es nach Undeutsch (1967, S. 26) häufig in Strafverfahren wegen „Verstößen gegen die Sittengesetzte" vorkommt, stellt die Aussage des mutmaßlichen Opfers oft die einzige Grundlage der Verurteilung dar, was der Frage nach dem Wahrheitsgehalt dieser Aussage besonderes Gewicht verleiht.

Es existieren heute verschiedene Ansätze mit deren Hilfe man versucht, den Wahrheitsgehalt von verbalen Äußerungen zu beurteilen; dazu zählen neben der in der vorliegenden Arbeit näher beleuchteten aussagepsychologischen Methode, die sich zum Großteil auf den Inhalt einer Aussage stützt, zum Beispiel auch Versuche, in der Physiologie oder im Verhalten der aussagenden Person Korrelate von Lügen bzw. wahren Angaben ausfindig zu machen. Bis auf die aussagepsychologische Glaubhaftigkeitsbegutachtung, wie sie in ihren grundlegenden Standards auch vom Bundesgerichtshof dargestellt wurde (BGH, 2000), konnte sich allerdings keines dieser Verfahren bisher als wirklich zuverlässig erweisen. Indessen liegt heute eine Vielzahl an empirischen Studien vor, welche die Zuverlässigkeit der aussagepsychologischen Methodik für unterschiedlichste Altersgruppen und bei einer weiten Bandbreite von Thematiken belegen (Köhnken, 2004).

Obwohl sich die Begutachtung durch einen aussagepsychologischen Sachverständigen als diagnostische Methode der Wahl zur Beurteilung des Wahrheitsgehaltes einer Aussage vor Gericht etabliert hat, ist zu beachten, dass sie in der täglichen Gerichtspraxis eher die Ausnahme als die Regel darstellt. Die Beurteilung einer Zeugenaussage obliegt nach höchstrichterlicher Rechtsprechung nämlich generell dem Richter (Aymans, 2005). Nur wenn die Beurteilung der Glaubhaftigkeit eine spezielle Sachkunde erfordert, über die ein Richter auch bei „spezifischer forensischer Vorerfahrung" nicht verfügt, ist er angehalten sich die Meinung eines aussagepsychologischen Sachverständigen einzuholen (Greuel, 2001, S. 7).

[1] Aus Gründen der Leseflüssigkeit wird in der vorliegenden Arbeit stets die männliche Form benutzt.

Nach Greuel (2001) lässt sich aus der Rechtsprechung verschiedener Strafgerichte eine Kasuistik ableiten, aus der ersichtlich wird, unter welchen Voraussetzungen eine richterliche Sachkunde nicht a priori unterstellt werden kann. Diese lauten wie folgt:

a) Die Zeugen sind die „einzigen Belastungszeugen … ohne daß zusätzlich ‚objektive' Sachbeweise vorliegen",
b) die Zeugen sind gleichzeitig auch Geschädigte (Opferzeugen), so dass „das Wirksamwerden potentieller Belastungsmotive nicht von vornherein ausgeschlossen werden kann",
c) die Tatvorwürfe sind mit „erheblicher Strafandrohung" verbunden und
d) es steht „Aussage gegen Aussage" (S. 8).

Die genannten Indikationen sind nach Greuel (2001) im Falle vermuteter sexueller (Gewalt-)Delikte besonders häufig erfüllt – eine Beobachtung, die sich im Großen und Ganzen mit den eingangs erwähnten Überlegungen von Undeutsch (1967) deckt. Die Beauftragung von psychologischen Sachverständigen erfolgt dementsprechend vorwiegend in Verfahren wegen mutmaßlicher Straftaten gegen die sexuelle Selbstbestimmung (§§ 176 bis 184 StGB) und bezieht sich auf die Aussagen meist kindlicher oder jugendlicher Opferzeugen, aber auch in solchen Fällen wird eine aussagepsychologische Begutachtung nicht immer in Auftrag gegeben. In der Regel müssen hierfür neben den genannten fallspezifischen noch weitere Einschränkungen vorliegen, die sich auf den aussagenden Zeugen beziehen. Die rechtlichen Vorgaben hierzu sind unterschiedlich restriktiv, je nachdem ob es sich um kindliche oder erwachsene Zeugen handelt (Greuel, 2001).

Ist der Hauptbelastungszeuge ein Kind, so wird die Hinzuziehung eines aussagepsychologischen Sachverständigen empfohlen, wenn der Zeuge sehr jung ist, die zu berichtenden Vorkommnisse schon sehr lange zurückliegen, Entwicklungsdefizite oder Beeinträchtigungen der kognitiven Leistungsfähigkeit vorliegen oder die Gefahr sekundärer Traumatisierungen durch unsachgemäße Befragung bzw. die Gefahr suggestiver Beeinflussungen im Vorfeld der Aussage besteht (Greuel et al., 1998; Greuel, 2001; Aymans, 2005).

Die aus der Rechtsprechung ableitbaren Vorgaben für die aussagepsychologische Begutachtung erwachsener Zeugen sind deutlich einschränkender, hier müssen erhebliche intellektuelle Defizite, psychische, psychosomatische oder neurologische Störungen oder eine manifeste Suchtproblematik vorliegen, um die Hinzuziehung eines Sachverständigen zu begründen (Greuel, 2001; Aymans, 2005). Generell sind die Gerichte also nur in solchen Fäl-

len angehalten, einen aussagepsychologischen Sachverständigen hinzuzuziehen, in denen es sich um „problematische“ Zeugen mit psychischen Auffälligkeiten handelt. Da psychologische Sachverständige allerdings auch unter anderen Umständen immer hinzugezogen werden *dürfen*, werden in der Gerichtspraxis nicht zwangläufig nur kognitiv-psychisch auffällige Zeugen begutachtet, in vielen Fällen wird der Begutachtungsauftrag auch – implizit oder explizit – durch schwierige motivationale Konstellationen begründet (Greuel, 2001).

Die Zahl der Verurteilungen wegen Straftaten gegen die sexuelle Selbstbestimmung ist in den vergangenen 15 Jahren in der Bundesrepublik Deutschland[2] fast stetig gestiegen, von 4779 im Jahre 1990 auf 7900 Verurteilungen 2004 (Statistisches Bundesamt, 2006). Da damit die Anzahl der Begutachtungen ebenfalls gestiegen sein dürfte[3], ist auch der Anspruch an die Zuverlässigkeit der aussagepsychologischen Methodik höher denn je. In der vorliegenden Arbeit soll ein Beitrag zur Bestimmung dieser Zuverlässigkeit geleistet werden, indem die Reliabilität einer der wichtigsten Komponenten der aussagepsychologischen Begutachtung, der Realkennzeichen nach Steller und Köhnken (1989) untersucht wird. Zunächst soll jedoch in Kapitel 2 kurz die geschichtliche Entwicklung der psychologischen Glaubhaftigkeitsbeurteilung bis hin zur heute aktuellen Konzeption dargestellt und diese im Anschluss genauer veranschaulicht werden. Dabei wird insbesondere auf ihre Voraussetzungen und Grenzen eingegangen sowie die bisher vorhandene empirische Stützung diskutiert. In Kapitel 3 erfolgt eine Präzisierung der Fragestellung der vorliegenden Untersuchung, deren Methodik in Kapitel 4 näher erläutert wird. Kapitel 5 umfasst die Darstellung der Ergebnisse der Untersuchung, die dann im anschließenden 6. Kapitel diskutiert werden.

[2] Da für die neuen Bundesländer noch keine flächendeckenden Angaben vorliegen, enthalten die Angaben des Statistischen Bundesamtes nur Daten für das frühere Bundesgebiet einschließlich Gesamt-Berlin.

[3] Die einzigen von der Autorin vorgefundenen Angaben zur Häufigkeit von Begutachtungen stammen aus dem Jahre 1982: Arntzen gibt dabei für die vorausgegangenen 30 Jahre eine Mindestanzahl von 30 000 aussagepsychologisch begutachteten Zeugen an (Arntzen, 1982).

2. THEORETISCHER HINTERGRUND

2.1 Die Anfänge der wissenschaftlichen Glaubhaftigkeitsbeurteilung

Eine genaue Festlegung der „Geburtstunde“ der wissenschaftlichen Glaubhaftigkeitsbeurteilung ist äußerst schwierig, nach Meinung einiger Autoren sogar unmöglich (Sporer, 1982). Nach Köhnken (1990) lässt sich ihre Entwicklung jedoch bis in zwei voneinander unabhängige Wurzeln zurückverfolgen – eine juristisch-kriminologische und eine allgemeinpsychologische.

Zunächst stellte sich die Frage nach der Glaubhaftigkeit von Zeugenaussagen allerdings weder für die Wissenschaft noch in der juristischen Praxis, da vom klassischen Altertum bis ins Mittelalter Glaubwürdigkeit als eine Eigenschaft angesehen wurde, welche man bestimmten Personengruppen generell zu- bzw. absprach, nicht aber einer konkreten Aussage. Zu den Personengruppen, denen über lange Zeit die Anerkennung ihrer allgemeinen Glaubwürdigkeit und daher ihrer Eignung als Zeugen verwehrt wurde, gehörten vor allem minderjährige und weibliche Zeugen (vgl. Undeutsch, 1967).

Diese Skepsis gegenüber der Brauchbarkeit von Zeugenaussagen durch Kinder und Frauen hielt sich vor allem auf juristisch-kriminologischer Seite bis ins 20. Jahrhundert hinein, wobei es hier auch gegen Zeugenaussagen als Beweismittel generell Vorbehalte gab. Entsprechende Meinungsäußerungen finden sich in den Veröffentlichungen verschiedenster Kriminalwissenschaftler, welche sich schon seit Beginn des 19. Jahrhunderts vereinzelt mit der Wahrnehmungs- und Gedächtnisfähigkeit von Zeugen befassten (z.B. Kleinschrodt, 1805; Mittermaier, 1834; Brauer 1834; Groß, 1898; alle zitiert nach Köhnken, 1990).

Parallel zu den Entwicklungen im juristisch-kriminologischen Bereich begannen sich seit Beginn des 20. Jahrhunderts auch Psychologen mit der Glaubhaftigkeit von Zeugen zu beschäftigen, wobei es sich zunächst hauptsächlich um eine experimentelle Herangehensweise handelte. Vorreiter hierbei waren vor allem Alfred Binet (1900) und William Stern (1902). Ihre Arbeiten standen am Anfang einer langen Reihe von Untersuchungen, in denen vorwiegend versucht wurde, Erinnerungsaussagen mit der experimentell manipulierten objektiven Wirklichkeit zu vergleichen und dadurch die Korrektheit der Erinnerungen zu beurteilen. Die Ergebnisse dieser Untersuchungen waren jedoch ernüchternd. Stern, der durch seine Bildversuche herausfinden wollte, „inwiefern die Durchschnittsaussage des normalen einwandfreien Zeugen als eine korrekte Wiedergabe des objektiven Thatbestandes betrach-

tet werden könne“ (Stern, 1902, S. 315) musste am Ende seiner Versuche konstatieren, fehlerlose Erinnerung sei wohl eher die Ausnahme als die Regel – und das obwohl seiner Meinung nach die Bedingungen für fehlerlose Erinnerungsleistungen im Experiment noch günstiger sind als im „praktischen Leben“ (ebd., S. 327).

Angesichts dieser experimentell festgestellten mangelhaften Zuverlässigkeit menschlicher Erinnerungen wurden von psychologischer Seite zunehmend Sorgen hinsichtlich all derjenigen Gerichtsverfahren geäußert, in denen das Urteil entscheidend von der Zeugenaussage des mutmaßlichen Opfers abhing. Zur Absicherung solcher Aussagen forderte erstmals William Stern 1903 grundsätzlich die Begutachtung durch einen psychologischen Sachverständigen (nach Undeutsch, 1967). Nachdem Stern selbst noch im selben Jahr als erster gerichtspsychologischer Sachverständiger zu einem Verfahren hinzugezogen wurde, weitete sich in den folgenden Jahren – unterbrochen durch den ersten Weltkrieg – der Einsatz psychologischer Sachverständiger vor Gericht immer mehr aus. Die Begutachtungen dieser ersten Generation gerichtspsychologischer Sachverständiger beschränkten sich aber weiterhin meist auf die Person des Zeugen, das heißt Zeugentüchtigkeit und Glaubwürdigkeit wurden weiterhin als stabile Persönlichkeitsmerkmale angesehen. Auch die wissenschaftlichen Grundlagen, auf die sich die Gutachter zu diesem Zeitpunkt stützten, waren aus heutiger Sicht eher fragwürdig, da sie außer auf Erkenntnisse aus der experimentellen Forschung und der damals noch jungen Disziplin der differentiellen Psychologie häufig nur auf persönliche Eindrücke und Einzelbeobachtungen zurückgreifen konnten.

Erst mit wachsender forensischer Erfahrung wurde eine systematischere und empirischere Herangehensweise möglich. Die entscheidende empirische Ausweitung der Aussagepsychologie setzte nach Arntzen (1983a) etwa ab 1948 ein, da nach dem zweiten Weltkrieg die Anforderung psychologischer Gutachten durch die Gerichte – auch aufgrund entsprechender Erlasse und Richtlinien – in ganz Europa immer mehr zur Regel wurde. Dementsprechend sieht auch Udo Undeutsch die Entwicklung der Aussagepsychologie in der Phase nach dem zweiten Weltkrieg gekennzeichnet durch den „Durchbruch der Erfahrung auf breiter Front“ (Undeutsch, 1967, S. 44). Durch die verstärkte Aktivität von psychologischen Gutachtern vor Gericht konnte eine breite Basis an Erfahrungsmaterial entstehen, welches zur empirischen Fundierung der Aussagepsychologie beitrug. Im Zuge dieser Entwicklung veränderte sich auch die vorher recht skeptische Einstellung gegenüber kindlichen Zeugenaussagen sowohl auf psychologischer als auch auf juristischer Seite zum positiven.

Die Untersuchungsmethoden der nach dem zweiten Weltkrieg tätigen Gutachter waren allerdings zunächst immer noch recht uneinheitlich. Zum Teil beschränkten sie sich auf das reine Aktenstudium, andere untersuchten weiterhin nur die Persönlichkeit des Zeugen, um dann dessen Glaubwürdigkeit im Allgemeinen zu beurteilen und bestenfalls einen kurzen Hinweis auf die spezielle Glaubwürdigkeit im vorliegenden Fall zu geben; eine eingehende Exploration zur Sache wurde anfangs nur von den wenigsten durchgeführt. Erst im Laufe der 50er Jahre setzte sich die Meinung Arnolds (1952) durch, der zwischen allgemeiner und der auf eine bestimmte Aussage bezogenen Glaubwürdigkeit unterschied und darauf hinwies, dass durchaus bei ein und dem selben Zeugen die eine vorhanden und die andere zu verneinen sein könne. Glaubwürdigkeit wurde zunehmend nicht mehr als stabiles Persönlichkeitsmerkmal, sondern als situationsabhängig gesehen, entsprechend verlagerte sich – wie vor allem von Undeutsch wiederholt gefordert – der Schwerpunkt der aussagepsychologischen Untersuchungen immer mehr von der Person des Aussagenden auf die Aussage selbst. Bei den meisten Gutachtern wurde daher eine Exploration zu den fraglichen Ereignissen zur Regel.

Folgerichtig konzentrierte sich auch die wissenschaftliche Aussagepsychologie der 1950er und 60er Jahre hauptsächlich auf die Frage, woran glaubwürdige Aussagen zu erkennen sind und wodurch sich glaubwürdige und unglaubwürdige Aussagen unterscheiden. So formulierte Undeutsch: „Der methodische Idealfall wäre, daß wahrheitsgemäße (mit den bekundeten Tatsachen übereinstimmende) Aussagen sich von wahrheitswidrigen (mit den bekundeten Tatsachen nicht übereinstimmenden) Aussagen in erkennbarer Weise unterscheiden, daß eine wahrheitsgemäße Darstellung gewissermaßen eine bessere Qualität hätte als eine wahrheitswidrige. Nach solchen Unterschieden galt es also zu fahnden“ (Undeutsch, 1967, S. 125). Im Anschluss daran präzisierte er genau diesen Gedanken in Form einer „heuristischen Hypothese“, welche diese Suche nach Qualitätsunterschieden leiten sollte und später als die so genannte *Undeutsch-Hypothese* in die Literatur einging: „Aussagen über selbsterlebte faktische Begebenheiten müssen sich von Äußerungen über nicht selbsterlebte Vorgänge unterscheiden durch Unmittelbarkeit, Farbigkeit und Lebendigkeit, sachliche Richtigkeit und psychologische Stimmigkeit, Folgerichtigkeit der Abfolge, Wirklichkeitsnähe, Konkretheit, Detailreichtum, Originalität und – entsprechend der Konkretheit jedes Vorfalles und der individuellen Erlebnisweise eines jeden Beteiligten – individuelles

Gepräge. Wer etwas erzählt, was er nicht selbst in der Realität erlebt hat, spricht unvermeidlich davon, ‚wie der Blinde von den Farben'" (S. 125f.).

Ausgehend von dieser Hypothese stellte Undeutsch seine unter 2.2.1 näher ausgeführten Glaubhaftigkeitsmerkmale zusammen, das heißt einen Katalog an spezifischen Merkmalen, welche er mit höherer Wahrscheinlichkeit in wahren als in falschen Aussagen erwartete. Er stützte sich dabei vor allem auf seinen eigenen Erfahrungen und Beobachtungen als praktisch tätiger Gutachter. Später entstanden auf ähnliche Weise weitere Merkmalssysteme anderer Autoren, die zum Teil auf der Arbeit Undeutschs aufbauten; auf diese Entwicklung soll im folgenden Abschnitt genauer eingegangen werden.

2.2 Entwicklung kriteriologischer Systeme in der Aussagepsychologie

Die heute gültige Kriteriologie zur Glaubhaftigkeitsbegutachtung nach Steller und Köhnken (1989), welche auch die Grundlage für die vorliegende Untersuchung bildet, stellt nur den Endpunkt einer längeren Entwicklung dar. Bereits im ersten Drittel des vergangenen Jahrhunderts wurden vereinzelt Merkmale beschrieben, mit deren Hilfe sich glaubhafte von unglaubhaften Aussagen unterscheiden lassen sollten; als erstes geschah dies mit einer gewissen Systematik 1930 durch Leonhart (Köhnken, 1990). Wie bereits kurz erwähnt, begannen später weitgehend unabhängig voneinander verschiedene praktisch tätige Gerichtspsychologen, vorwiegend aus dem deutschsprachigen Raum, aus ihrer Erfahrung heraus umfangreichere Kataloge von Glaubhaftigkeitsmerkmalen zusammenzustellen. Die vier wichtigsten und umfangreichsten Kriteriologien, die auf diese Weise entstanden, finden sich bei Undeutsch (1967), Trankell (1971), Arntzen (1970, 1983a), sowie bei Dettenborn, Fröhlich und Szewczyk (1984). Sie sollen im Folgenden kurz beschrieben werden, wobei sehr bald deutlich werden wird, wie stark sich die einzelnen Systeme überschneiden.

2.2.1 Das System der Glaubwürdigkeitskriterien nach Undeutsch

Udo Undeutsch, zweifelsohne eine der wichtigsten Persönlichkeiten in der Geschichte der wissenschaftlichen Aussagepsychologie, veröffentlichte 1967 im „Handbuch der Psychologie" die erste umfangreiche und systematische Zusammenstellung von Glaubwürdigkeitsmerkmalen. Dabei illustriert er diese mit zahlreichen Beispielen aus seiner eigenen Gutachterpraxis, aus welcher er die Merkmale auch größtenteils ableitete. Darüber hinaus

stützte er sich bei der Entwicklung seiner Kriteriologie auf die Vorarbeiten des Leipziger Landgerichtdirektors i. R. Carl Leonhart, den er als Vorläufer seiner Bemühungen bezeichnet. Außerdem bezieht er sich bei der Beschreibung einiger Merkmale auch auf entsprechende Weiterentwicklungen durch Arne Trankell, welcher Undeutschs erste Ansätze zur Definition von Glaubwürdigkeitsmerkmalen aufgenommen und unabhängig von ihm ausgebaut hatte.

Vor der Beschreibung der eigentlichen Glaubwürdigkeitskriterien führt Undeutsch weitere Faktoren aus, die bei der Beurteilung der Glaubhaftigkeit einer Zeugenaussage eine Rolle spielen, wie zum Beispiel Persönlichkeitsmerkmale, individueller Entwicklungsstand und Motivlage des Zeugen, Geschichte der Aussage und Verhalten des Zeugen während der Aussage. Insgesamt räumt er der Analyse der Aussage anhand der unten aufgeführten Kriterien aber eine herausragende Stellung ein, die Aussage in der aktuell vorliegenden Fassung stelle „das entscheidende Material“ dar (Undeutsch, 1967, S. 125).

Tabelle 1: Glaubwürdigkeitskriterien nach Undeutsch (1967)

1. Widerspruchslosigkeit zu anderweitig feststehenden Fakten
2. Realistik und Wirklichkeitsnähe
3. Konkretheit, Anschaulichkeit, Originalität und individuelle Durchzeichnung
4. Innere Stimmigkeit und Folgerichtigkeit
5. Eigentümliche oder ausgefallene Einzelheiten
6. Zeitliche und räumliche Verankerungspunkte
7. Außerhalb der Planungskapazität oder sogar des Verständnishorizontes der Erzählers liegende Details
8. Psychische Vorgänge des Täters und des Opfers
9. Charakteristische Entwicklungsdynamik der Beziehung zwischen Täter und Opfer, falls diese länger angedauert hat
10. Spontane Verbesserung der eigenen Aussage
11. Bericht fragmentarischer Handlungen
12. Unvorteilhafte Darstellung der eigenen Rolle, Selbstbelastungen
13. Einwände gegen die Richtigkeit der eigenen Aussage
14. Konstanz der Aussage, zumindest bezüglich des Kerngeschehens

Die von ihm postulierten Qualitätsunterschiede zwischen Berichten mit und ohne realen Erlebnishintergrund zeigen sich laut Undeutsch auf den in Tabelle 1 bezeichneten Dimensionen. Je zahlreicher und ausgeprägter die genannten Merkmale in einer Aussage vorkommen, desto größer wird nach Undeutsch die Beweiskraft dieser Aussage, unwahre Aussagen dagegen sind im Allgemeinen durch das Fehlen der Merkmale gekennzeichnet. Anders als zum Beispiel Arntzen (1983a) stellt Undeutsch aber keine Regel auf, nach der eine bestimmte Mindestanzahl von Kriterien vorhanden sein muss, um die Glaubwürdigkeit einer Aussage eindeutig feststellen zu können. Er weist allerdings bereits darauf hin, dass das Fehlen einiger Glaubwürdigkeitskriterien nicht automatisch als Unglaubwürdigkeit zu interpretieren, sondern die Aussagekraft der Aussageanalyse im Einzelfall stets unter Berücksichtigung von Persönlichkeit, Intelligenzniveau und Motivlage des Zeugen, den Eigenschaften des berichteten Erlebnisses und der Geschichte der Aussage zu bewerten sei.

2.2.2 Das aussagepsychologische Prozessmodell nach Trankell

Vom schwedischen Psychologen Arne Trankell stammt die einzige Systematik von Glaubwürdigkeitskriterien, die außerhalb des deutschen Sprachraumes veröffentlicht wurde. Die umfangreichste Beschreibung seines Ansatzes erschien 1970 in der schwedischen und 1971 in der von Undeutsch übersetzten deutschen Ausgabe seines Buches „Der Realitätsgehalt von Zeugenaussagen“, auf die sich auch die folgenden Ausführungen beziehen.

Trankells Ansatz unterscheidet sich insofern von dem Undeutschs, als er nicht nur eine weitgehend abgeschlossene Sammlung von Glaubhaftigkeitsmerkmalen zur Unterscheidung wahrer und falscher Aussagen darstellt, sondern eine umfassende Beschreibung des gesamten diagnostischen Prozesses der Glaubwürdigkeitsbeurteilung bietet. Im Gegensatz zu dem von Undeutsch, aber auch den anderen noch zu beschreibenden Autoren vertretenen „kriteriumsorientierten Ansatz“ wird er von Köhnken (1990, S. 105) dementsprechend als „prozessorientierter Ansatz“ bezeichnet. Darüber hinaus berücksichtigt Trankells Modell nicht wie die anderen Ansätze nur unwahre Aussagen, die auf absichtlich falschen Angaben beruhen, sondern auch solche, die durch unbewusste Irrtümer und Fehlleistungen zustande kommen, beispielsweise während der Wahrnehmung, Speicherung oder dem Abruf der Geschehnisse.

Stark vereinfacht beschreibt Trankell den aussagepsychologische Prozess als einen hypothesengeleiteten und auf Rückkopplungsmechanismen basierenden Vorgang: Zunächst werden auf Basis des vorhandenen Datenmaterials, das z.B. aus den Ermittlungsakten zu entnehmen ist, eine Realitäts- oder Nullhypothese und eine oder mehrere Alternativhypothesen formuliert. Das vorhandene Datenmaterial muss anschließend so lange analysiert und durch entsprechend der Hypothesen neu erhobene Daten ergänzt werden, bis es erschöpfend durch die Nullhypothese erklärt wird und es keine haltbaren Alternativhypothesen mehr gibt. Mit diesem hypothesengeleiteten Vorgehen griff Trankell bereits 1971 auf einen heute üblichen methodischen Standard vor.

Für die hypothesengeleitete Erhebung neuer Daten beschreibt Trankell verschiedene Methoden, worunter neben befragungspsychologischen und sozialpsychologischen Methoden auch die „Aussagepsychologische Realitätsanalyse“ fällt. Als deren wichtigstes Hilfsmittel bezeichnet er die aussagepsychologischen Realitätskriterien, die er in zwei Hauptgruppen einteilt: Die primären Realitätskriterien und die sekundären Kontrollkriterien (siehe Tabelle 2). Die primäre Realitätskriterien können dabei nicht generell auf beliebige Aussagen, sondern nur in gewissen Fällen angewendet werden, da nicht alle Aussagen Ansatzpunkte dafür bieten; die sekundären Kontrollkriterien dagegen sind allgemeiner anwendbar, können allerdings nur zur Ergänzung der primären Kriterien dienen, da sie für sich gesehen keine Funktion haben.

Wie in Tabelle 2 ersichtlich können die primären Kriterien wiederum in zwei Untergruppen gegliedert werden, wobei die eine Untergruppe auf die strukturellen Aspekten der Aussage abzielt, die andere auf ihren Inhalt, genauer gesagt „auf ihre inhaltliche Übereinstimmung mit den Erwartungen, die wir auf Grund unserer Kenntnisse über die Funktionsweise des Menschen hegen können“ (S. 146). Auch hinsichtlich der sekundären Kontrollkriterien unterscheidet Trankell zwei Untergruppen, die eine Gruppe dient zur formallogischen Kontrolle der Angaben, indem mögliche Alternativhypothesen überprüft werden, die andere bedient sich für die Validitätskontrolle einer empirischen Vorgehensweise.

Die in den verschiedenen Unterkategorien eingeordneten Kriterien, die von Trankell übrigens nur als Beispiele verstanden und durchaus für ausbaufähig gehalten wurden, sind hier in deutlich strengerem Sinne definiert als bei anderen Autoren. Für ihn ist ein Realitäts-*kriterium* nur als solches zu bezeichnen, wenn es ausschließlich in wahren Aussagen zu finden ist, nach dieser Definition müsste also schon ein einziges Kriterium ausreichen, um eine

Aussage als wahr zu klassifizieren. Von diesen Kriterien zu unterscheiden sind so genannte *Kennzeichen* realitätsbegründeter Schilderungen, die sowohl in wahren als auch – in abgeschwächter Form – in erfundenen Aussagen auftauchen können. Als Beispiel hierfür führt Trankell den Detailreichtum an: Erst durch eine besondere Qualität der berichteten Details kann dieses Kennzeichen zu einem „funktionstüchtigen" Kriterium wie z.B. dem bilateralen Emotionskriterium, dem Kompetenz- oder Einzigartigkeitskriterium werden (S. 123).

Tabelle 2: Systematik und Bedeutung der Realitätskriterien nach Trankell (1971)

PRIMÄRE REALITÄTSKRITERIEN

Strukturanalyse	**Inhaltsanalyse**
Bilaterales Emotionskriterium	Kompetenzkriterium
Ein Gefühlserlebnis, das in der Aussage geschildert wird, kann nicht allein durch die ebenfalls geschilderten Sinneseindrücke erklärt werden, sondern nur durch die gleichzeitige Bezugnahme auf die persönliche Situation des Zeugen zum Zeitpunkt der Beobachtung.	Die in der Aussage beschriebenen Geschehensabläufe und Details sind so beschaffen, dass es vermutlich die Kompetenz des Zeugen überstiegen hätte, sich diese selbst auszudenken.
Homogenitätskriterium	Einzigartigkeitskriterium
Verschiedene Details bestätigen sich gegenseitig und beschreiben unabhängig voneinander denselben Ablauf.	Die Aussage enthält so außergewöhnliche Details, dass es unabhängig von der Kompetenz des Zeugen unwahrscheinlich ist, dass sie erfunden sein könnten.
	Sequenzkriterium
	Mehrere Aussagen, die von ein und demselben Zeugen gemacht wurden, unterscheiden sich voneinander (nur) in gedächtnispsychologisch erwartbarer Weise.

SEKUNDÄRE KONTROLLKRITERIEN

Formallogische Kontrolle	**Empirische Validitätskontrolle**
Konsequenzkriterien	Isomorphiekriterium
Aus den bekannten Fakten lassen sich jeweils Alternativhypothesen ableiten, nach denen als Konsequenz bestimmte Merkmale in der Aussage zu erwarten wären; jedes dieser Merkmale ist ein Konsequenzkriterium, nach denen man in der Aussage suchen kann.	Die Aussage weist die gleiche formale Struktur auf wie frühere Aussagen desselben Zeugen, die mit Sicherheit falsch waren.

2.2.3 Die Glaubwürdigkeitskriterien nach Arntzen

Friedrich Arntzen, der sich auch schon zuvor mehrfach zu Themen der forensischen Aussagepsychologie geäußert hatte, veröffentlichte 1983 in der zweiten Auflage seines Buches „Psychologie der Zeugenaussage“ seine umfangreichste und detaillierteste Zusammenstellung von Glaubwürdigkeitsmerkmalen. Sie stellt eine Weiterentwicklung und Präzisierung der bereits in der ersten Auflage von 1970 ausgeführten Kriteriologie dar.

Statt generell von Glaubwürdigkeitsmerkmalen spricht Arntzen allerdings zunächst nur von „Aussageeigenarten, die zu Glaubwürdigkeitsmerkmalen werden können“ (1983a, S. 15). Aussageeigenarten an sich, zum Beispiel Konstanz oder Detaillierung finden sich seiner Meinung nach in einfacher Ausprägung auch in unglaubwürdigen Aussagen und ihr simples Vorliegen spricht deshalb weder für noch gegen die Glaubhaftigkeit einer Aussage. Erst wenn diese Aussageeigenheiten eine von Arntzen so benannte „Steigerungsform“, das heißt eine „bestimmte Steigerung ihrer Qualität“ (ebd., S. 20) aufweisen, werden sie zu Glaubwürdigkeitsmerkmalen. Solch eine Qualitätssteigerung kann durch verschiedene Faktoren bedingt werden: Zum einen durch einen großen Umfang oder das spontane, rasche Vorbringen der Aussage, zum anderen durch eine erschwerte Art der Befragung, die z.B. auf inhaltliche Vorhaltfragen verzichtet. Des Weiteren kann eine Steigerungsform auch aus Persönlichkeitseigenschaften und Kenntnissen des Zeugen resultieren, beispielsweise aus einer geistigen Minderbegabung, wenn man diese im Zusammenhang mit der Aussage sieht.

Neben den Steigerungsformen müssen aber auch so genannte „Minderungsfaktoren“ beachtet werden, welche eine Aussageeigenart an Qualität und Aussagekraft verlieren lassen und dadurch verhindern können, dass sie den Status eines Glaubwürdigkeitsmerkmales erreicht. Solche Minderungsfaktoren werden von Arntzen allerdings nur beispielshalber benannt. So verliert die Aussageeigenart Konstanz an Wert, wenn die Bekundungen durch Stereotypie gekennzeichnet sind, die Aussageeigenart Detaillierung kann gemindert werden, wenn die vorgebrachten Details sachlich unwahrscheinlich oder gar widersprüchlich sind. Auch bestimmte Persönlichkeitseigenarten können zu Minderungsfaktoren werden, beispielsweise Schlagfertigkeit oder stark ausgeprägte Phantasie. Die in Tabelle 3 wiedergegebenen Glaubwürdigkeitskriterien wären nach Arntzen demnach nur eingeschränkt als solche zu bezeichnen und müssen vor dem Hintergrund der eben beschriebenen Konzeption gesehen werden.

Tabelle 3: Glaubwürdigkeitsmerkmale nach Arntzen (1983a, S. 16)

1. Glaubwürdigkeitskriterien, die sich aus dem *Aussageinhalt* ergeben
   a) Detaillierung und inhaltliche Besonderheiten – u.a. vom Zeugen wiedergegebene

      Gespräche und Interaktionen
      Eigenpsychische Vorgänge
      Phänomengebundene Beobachtungen
      Vielfältige Verflechtungen mit veränderlichen äußeren Umständen
      Negative Komplikationen
      Reaktionsketten
      Inhaltliche Verschachtelungen
      Ausgefallene, originelle Einzelheiten

   b) Homogenität der Aussage
      Schilderung einer dem Zeugen nicht bekannten Verhaltensmusters („Delikttypisch“)

2. Glaubwürdigkeitskriterien, die sich aus dem Verlauf der *Aussageentwicklung* ergeben
   a) Relative Konstanz und Inkonstanz einer Aussage in zeitlich auseinander liegenden Befragungen
   b) Ergänzbarkeit einer Aussage bei nachfolgenden Befragungen

3. Glaubwürdigkeitskriterien, die sich aus der *Aussageweise* ergeben
   a) Inkontinenz
   b) Nacherlebende Gefühlsbeteiligung
   c) Ungesteuerte Aussageweise

4. Kriterien aus dem *Motivationsumfeld* der Aussage

Aufgrund der Untersuchungen von Aussagen in Geständnisfällen, die seiner Meinung nach als „sicher glaubwürdige Aussagen“ zu werten sind, fordert Arntzen, „daß drei eindeutige Glaubwürdigkeitsmerkmale als Merkmalskomplex gegeben sein müssen, wenn die Glaubwürdigkeit einer Zeugenaussage als voll erwiesen gelten soll“ (Arntzen, 1983a, S. 22) und legt damit als einziger der hier aufgeführten Autoren einen Schwellenwert für die Anwendung der Glaubwürdigkeitsmerkmale fest. Wenn die drei Merkmale des Merkmalskomplexes darüber hinaus noch zu unterschiedlichen Kategorien gehören und deshalb mit verschiedenen Methoden erarbeitet wurden, sind sie laut Arntzen in besonderer Weise aussagekräftig und es können sogar Minderungsfaktoren einzelner Merkmale ausgeglichen werden.

2.2.4 Die Systematik der Glaubhaftigkeitskriterien nach Dettenborn, Fröhlich und Szewczyk

Auch in der damaligen DDR begannen sich Anfang der 70er Jahre Wissenschaftler auf Grundlage der Arbeiten von Undeutsch (1967), Arntzen (1970) und Trankell (1971) mit der Problematik der Glaubhaftigkeitsbeurteilung zu beschäftigen. Wichtige Vertreter waren hier vor allem Hans Szewczyk und Eckhard Littmann, die in diversen Publikationen zum Teil voneinander abweichende Zusammenstellungen von Kriterien zur Glaubhaftigkeitsbeurteilung sowie Validierungsstudien hierzu präsentierten (z.B. Szewczyk & Littmann, 1982; Littmann & Szewczyk, 1983). In ihrem 1984 erschienenen Lehrbuch „Forensische Psychologie“ versuchten Dettenborn, Fröhlich und Szewczyk, die bis dahin bekannten Merkmale glaubhafter Aussagen zu systematisieren.

Tabelle 4: Systematik der Glaubhaftigkeitsmerkmale nach Dettenborn, Fröhlich und Szewczyk (1984, S. 312ff.)

1. *Die Verankerung der Aussage in anderweitigen Tatsachen*
   a) Widerspruchslosigkeit zu anderweitig erhobenen feststehenden Tatsachen
   b) Übereinstimmungen mit (glaubhaften) Angaben anderer Zeugen oder des Beschuldigten
   c) Angaben sind nachprüfbar mit besonderen Lebensumständen oder Gewohnheiten des Täters verwoben.
   d) Entsprechung von Aussage und Zeugenpersönlichkeit im Niveau
2. *Aussagen über die Tat*
   a) Detailreichtum und Detailtreue
   b) Konstanz hinsichtlich der berichteten eigentlichen Sexualhandlung, dagegen schlechtere Erinnerung von Einzelheiten des Randgeschehens
   c) Wirklichkeitsnähe und Realismus
   d) Bericht eigentümlicher, ausgefallener Einzelheiten
   e) Berichte von Störungen des Handlungsablaufes
   f) Schilderungen kriminologisch spezifischer und sexuologischer Abweichungen vom normalen Sexualverhalten (sexuell abnormes Verhalten, perverse Praktiken)
   g) Berichtigungen und Präzisierungen der Aussage
   h) Beschreibung einer typischen Entwicklung der sexualbezogenen Handlungen
   i) Einwände gegen die Richtigkeit der eigenen Beobachtungen
   j) Bericht von Geschehnissen außerhalb des Verständnishorizontes
3. *Aussagen über das Taterleben*
   a) Hinweise auf die psychischen Vorgänge des Aussagenden
   b) Schilderung der beim Delikt erlebten sexuellen Vorgänge
   c) Spontane Äußerungen über reflexartige Reaktionen
4. *Das Aussageverhalten*
   a) Dem erlebten Geschehen adäquate gefühls- bzw. gemütsmäßige Beteiligung während der Aussage

Für die Anwendung des Kriteriensystems weisen die Autoren darauf hin, jedes Kriterium habe nur im Rahmen der Kriteriengesamtheit eine Bedeutung, der Stellenwert jedes einzelnen Kriteriums müsse daher empirisch geprüft werden. Da zum Zeitpunkt der Erscheinung des Buches noch kaum größere Validierungsstudien vorlagen, zitieren die Autoren erste Ergebnisse eigener Untersuchungen, auf Grund derer sie zu dem Schluss kommen, dass durchaus nicht alle genannten Kriterien signifikante Unterschiede zwischen glaubwürdigen und nicht glaubwürdigen Kindern zeigen. Folglich wenden sich die Autoren eindeutig gegen die Festlegung eines Schwellenwertes und ziehen ein recht nüchternes Fazit: „So bleiben die genannten Kriterien ein empirisches Material, das vor allem methodischen und didaktischen Wert hat, keinesfalls aber so gesehen werden darf, daß von einer bestimmten Zahl von erfüllten Kriterien an automatisch auf eine Glaubwürdigkeit rückgeschlossen werden darf“ (S. 317).

Neben der Analyse der Aussage anhand der oben aufgeführten Kriterien, die ihrer Meinung nach in erster Linie auf die Aussageehrlichkeit und die Aussagewilligkeit des Zeugen abzielt, erachten die Autoren auch eine Prüfung der Aussagefähigkeit, eine Analyse der zeitlichen Entwicklung der Anzeige, der Anzeigensituation und Anzeigenmotivation, sowie die genaue Betrachtung der Widersprüche in verschiedenen Aussagen desselben Zeugen als wichtige Aspekte der Glaubwürdigkeitsbeurteilung. Für die Analyse der Aussagefähigkeit findet sich in dem Lehrbuch ebenfalls ein Katalog an Kriterien, welche auf Grund von entwicklungs- und persönlichkeitsbedingten Aspekten und hinsichtlich des „sexuellen und erotischen Entwicklungsstandes“ für oder gegen die Richtigkeit der Aussage sprechen (Dettenborn et al., 1984, S. 306f.).

2.3 Die Realkennzeichen nach Steller und Köhnken

Aufbauend auf die eben dargestellten früheren Arbeiten veröffentlichten Max Steller und Günter Köhnken im Jahr 1989 eine neue und bis heute aktuelle Zusammenstellung von Glaubhaftigkeitskriterien für die Beurteilung von kindlichen Zeugenaussagen, bei der es sich im Wesentlichen um eine Überarbeitung und Systematisierung von Merkmalen aus den stark überlappenden bisherigen Kriteriologien handelt. Sie begründen die Notwendigkeit eines neuen Systems von Realkennzeichen mit dem Mangel an Systematik und an präzisen Definitionen und Unterscheidungen, der bei den bisherigen Ansätzen oft zu kritisieren war.

2.3.1 Darstellung und Definition der Realkennzeichen

Wie in Tabelle 5 dargestellt umfasst das System der Realkennzeichen für die Kriterien-orientierte Aussageanalyse nach Steller und Köhnken fünf Hauptkategorien mit insgesamt 19 Einzelkriterien, die im Folgenden kurz definiert werden sollen. Die Beschreibungen orientieren sich dabei eng am englischen Originaltext (Steller & Köhnken, 1989, S. 222-231), der unter dem Titel *Criteria-based statement analysis* erschien.

Tabelle 5: Realkennzeichen nach Steller & Köhnken (1989) in der deutschen Fassung nach Steller et al. (1992, S. 153)

Allgemeine Merkmale
1. Logische Konsistenz
2. Unstrukturierte Darstellung
3. Quantitativer Detailreichtum

Spezielle Inhalte
4. Raum-zeitliche Verknüpfungen
5. Interaktionsschilderungen
6. Wiedergabe von Gesprächen
7. Schilderung von Komplikationen im Handlungsverlauf

Inhaltliche Besonderheiten
8. Schilderung ausgefallener Einzelheiten
9. Schilderung nebensächlicher Einzelheiten
10. Phänomengemäße Schilderung unverstandener Handlungselemente
11. Indirekt handlungsbezogene Schilderungen
12. Schilderung eigener psychischer Vorgänge
13. Schilderung psychischer Vorgänge des Täters

Motivationsbezogene Inhalte
14. Spontane Verbesserung der eigenen Aussage
15. Eingeständnis von Erinnerungslücken
16. Einwände gegen die Richtigkeit der eigenen Aussage
17. Selbstbelastungen
18. Entlastung des Angeschuldigten

Deliktspezifische Inhalte
19. Deliktspezifische Aussageelemente

Allgemeine Merkmale

Diese Merkmale zielen auf die generellen Eigenschaften der Aussage als Ganzes ab und können beurteilt werden, ohne dass man Details des Inhaltes berücksichtigen muss.

1. **Logische Konsistenz** bezieht sich auf die Stimmigkeit der Aussage in sich, das heißt die geschilderten Zusammenhänge müssen logisch kohärent sein und dürfen keine Diskrepanzen und Ungereimtheiten aufweisen. Dieses Merkmal findet sich mit unter-

schiedlichen Bezeichnungen aber ähnlichen Definitionen in allen vorher aufgeführten Merkmalszusammenstellungen.

2. **Unstrukturierte Darstellung** bedeutet, dass die Ereignisse nicht logisch geordnet oder in chronologischer Reihenfolge vorgetragen werden – solch eine „auswendig gelernt" wirkende Darstellung würde man im Gegenteil eher bei einer erfundenen Geschichte erwarten. Stattdessen finden sich in der glaubhaften Aussage immer wieder Zeitsprünge, Einschübe und Abbrüche, wobei sich die einzelnen Elemente aber insgesamt immer zu einem stimmigen Ganzen zusammenfügen lassen müssen, sodass das Kriterium der logischen Konsistenz nicht verletzt wird. Es ist jedoch zu beachten, dass dieses Merkmal, das erstmals bei Arntzen (1983a) auftaucht, nur bei einem freien Vortrag der Aussage ohne zu starken strukturierenden Einfluss des Interviewers anwendbar ist.

3. **Quantitativer Detailreichtum** bezieht sich darauf, dass sich in einer wahren Aussage sehr viel mehr unterschiedliche Details finden als in erfundenen Aussagen, da es die meisten Zeugen überfordern würde, sich diese in großer Zahl auszudenken. Eine Vielzahl an Details zeigt sich z.B. in der Beschreibung von Orten, Personen oder der detaillierten, schrittweisen Abfolge von Ereignissen.

Spezielle Inhalte

Mit dieser zweiten Hauptkategorie der Realkennzeichen wird das Vorhandensein oder die Ausprägung von speziellen Inhalten der Aussage untersucht.

4. **Raum-zeitliche Verknüpfungen**, das heißt die Einbettung des Erzählten in einen äußeren Kontext, in den Alltag und die Lebensumstände des Aussagenden, werden schon bei Undeutsch (1967) als Glaubhaftigkeitsmerkmal gewertet, denn „reale Vorkommnisse hängen nicht beziehungslos zu Zeit und Raum in der Luft, sondern haben zeitliche und räumliche Verankerungspunkte" (S. 139). Solche Verankerungspunkte können z.B. Alltagsereignisse, Beziehungen, Gewohnheiten oder örtliche Besonderheiten sein, welche wechselseitig mit den eigentlichen Ereignissen in Verbindung stehen (siehe auch Arntzen, 1983a, S. 35).

5. Mit **Interaktionsschilderungen** ist die Wiedergabe von Handlungsketten gemeint, also von Abfolgen wechselseitiger Handlungen und Reaktionen, die sich zwischen dem aussagenden Zeugen und dem mutmaßlichen Täter abgespielt haben. Darunter können

auch Dialoge fallen, diese werden aber – wenn sie wörtlich wiedergegeben werden – zusätzlich unter dem Realkennzeichen *Wiedergabe von Gesprächen* gewertet.

6. **Wiedergabe von Gesprächen** bedeutet, wie oben bereits angesprochen, dass Gespräche bzw. Äußerungen zumindest teilweise wörtlich in der Aussage reproduziert werden. Eine rein inhaltliche Schilderung von Dialogen reicht demnach zur Erfüllung dieses Kriteriums nicht. Die beteiligten Personen sollten bei der Wiedergabe von Gesprächen erkennbar sein, was laut Steller und Köhnken dann in besonderem Maße erfüllt ist, wenn der Zeuge Vokabular des Angeklagten benutzt, welches für das Alter des Zeugen untypisch ist, Argumentationen des Angeklagten enthalten sind oder Gespräche wiedergegeben werden, die verschiedene Einstellungen des Täters und des Opfers bzw. des Zeugen deutlich machen.

7. **Schilderungen von Komplikationen im Handlungsverlauf** ist wie *Logische Konsistenz* ein Merkmal, über das relative Einigkeit herrscht und das daher in fast allen bisherigen Zusammenstellungen von Glaubhaftigkeitsmerkmalen auftaucht. Darunter fallen die Beschreibungen von Komplikationen jeglicher Art, Undeutsch (1967) nennt zum Beispiel missglückte Sexualhandlungen, das Auftreten unvorhergesehener Schwierigkeiten, Überraschtwerden oder den plötzlichen Abbruch der angelaufenen Handlungen.

Inhaltliche Besonderheiten

Diese Kategorie von Kriterien umfasst inhaltliche Bestandteile und individuelle Besonderheiten der Aussage, die ihre Konkretheit und Lebhaftigkeit und damit ihre Qualität erhöhen. Genauso wie bei der vorausgegangenen Kategorie der Realkennzeichen steht auch hier die kognitive Komponente im Vordergrund, das heißt die Frage, ob ein falschaussagender Zeuge geistig dazu in der Lage wäre, die geforderten Inhalte zu produzieren oder sich überhaupt bewusst zu sein, dass sie notwendig sein könnten.

8. Das Merkmal **Schilderung ausgefallener Einzelheiten** liefert laut Undeutsch „höchste Garantie für die Realität des Berichteten“ (Undeutsch 1967, S. 138) und bezieht sich auf die Wiedergabe von außergewöhnlichen oder einzigartigen Details, deren Auftreten so unwahrscheinlich ist, dass sie ein lügender Zeuge wohl kaum für geeignet halten würde, seine Aussage überzeugend darzustellen. Auch dieses Merkmal ist nicht nur Teil der Merkmalszusammenstellung von Undeutsch, sondern unter anderen Bezeich-

nungen wie „originelle“, „eigentümliche“ oder „außergewöhnliche“ Einzelheiten auch bei Trankell (1971), Arntzen (1983a) und Dettenborn, Fröhlich und Szewczyk (1984) zu finden.

9. **Schilderung nebensächlicher Einzelheiten** heißt, dass das aussagende Kind Details erwähnt, die für die eigentliche Anschuldigung überflüssig sind und nicht zur Erhärtung der Vorwürfe beitragen. Steller und Köhnken gehen wie Undeutsch (S. 135) davon aus, dass eine lügende Person sich keine irrelevanten, umständlichen Details ausdenken, sondern direkt auf die vorgeworfene Handlung zu sprechen kommen würden.

10. Das Kriterium **Phänomengemäße Schilderung unverstandener Handlungselemente** bedeutet, dass das Kind Handlungen oder Details berichtet, die ihm selbst zwar unverständlich sind, die es aber so konkret beschreiben kann, dass sich dem Interviewer ihre Bedeutung erschließt. Meist handelt es sich dabei um Schilderungen im Zusammenhang mit männlicher Erregung und Ejakulation, die vor allem den Verständnishorizont von kleineren Kindern noch deutlich übersteigen.

11. **Indirekt handlungsbezogene Schilderungen** liegen laut Steller und Köhnken dann vor, wenn der Zeuge während der Aussage Inhalte[4] vorbringt, die mit der Anschuldigung an sich nichts zu tun haben, aber mit ihnen thematisch in Beziehung stehen bzw. vom Zeugen damit assoziiert werden. Das Kriterium findet sich auch bei Arntzen, der berichtet, er habe eine solche Assoziation noch nie in einer Falschaussage gefunden (1983a, S. 38).

12. **Schilderung eigener psychischer Vorgänge** schließt die Beschreibung von eigenen Gefühlen, Empfindungen und Gedanken des Kindes ein, die es in der geschilderten Situation hatte. Dieses Merkmal wird sowohl von Undeutsch (1967), als auch von Arntzen (1983a) und Dettenborn et al. (1984) erwähnt.

13. Die **Schilderung psychischer Vorgänge des Täters**, wie sie diesem vom Kind zugeschrieben werden, gilt schon bei Undeutsch (S. 143) als Hinweis für die Glaubhaftigkeit einer Aussage und beinhaltet neben zugeschriebenen Gefühlen und Gedanken auch physiologische Reaktionen beim Täter.

---

[4] Im Originaltext von 1989 ist – wie auch bei Arntzen (1983a) – zunächst nur die Rede von Gesprächen, die der Zeuge mit den berichteten Anschuldigungen assoziiert und in der Aussage erwähnt . In späteren Publikationen (z.B. Steller, Wellershaus & Wolf, 1992; Steller & Volbert, 1999) fällt diese Einschränkung allerdings weg.

Motivationsbezogene Inhalte

Bei dieser vierten Gruppe von Kriterien wird im Gegensatz zu den vorausgegangenen Kategorien die motivationale Komponente betont, das heißt die Frage nach der Wahrscheinlichkeit, mit denen ein lügender Zeuge die folgenden Merkmale in seine Aussage aufnehmen würde.

14. **Spontane Verbesserung der eigenen Aussage** bedeutet, dass während der Exploration zur Sache spontan, d.h. ohne Nachfragen oder Suggestion durch den Sachverständigen die eigene Aussage verbessert wird, beziehungsweise neue oder klarere Erinnerungen vorgebracht werden. Dieses Kriterium geht auf Undeutsch zurück, der diesem Kriterium relativ viel Bedeutung beimisst. Dahinter steht die Überlegung, dass bewusst falschaussagende Personen wohl nicht freiwillig durch Verbesserungen oder Präzisierungen selbst Zweifel an der Glaubhaftigkeit ihrer Aussage wecken würden.

15. Das **Eingeständnis von Erinnerungslücken** wird aus ganz ähnlichen Gründen für ein Zeichen von Glaubhaftigkeit gehalten, da ein lügender Zeuge durch das Zugeben von Erinnerungslücken seine eigene Glaubhaftigkeit in Frage stellen würde, was wohl nicht sein Ziel sein kann.

16. **Einwände gegen die Richtigkeit der eigenen Aussage** wird ebenfalls schon von Undeutsch als Glaubhaftigkeitsmerkmal verwendet, da es wiederum nicht im Interesse eines lügenden Zeugen liegen kann, Zweifel an der Richtigkeit seiner eigenen Aussage aufzubringen. Auch Dettenborn et al. (1984) haben diese Merkmal in ihre Systematik aufgenommen.

17. **Selbstbelastungen** heißt in diesem Fall, dass der Zeuge durch bestimmte Äußerungen sich selbst und seine Rolle in der betreffenden Situation unvorteilhaft darstellt, sich eine Teilschuld an den Geschehnissen gibt oder zumindest das eigene Verhalten dem mutmaßlichen Täter gegenüber selbstkritisch als falsch oder unangebracht schildert. Auch ein solches Verhalten würde man von einem lügenden Zeugen kaum erwarten.

18. **Entlastung des Angeschuldigten** trifft dann zu, wenn der Zeuge das Verhalten des Angeschuldigten zu erklären oder zu entschuldigen versucht oder auch offensichtliche Möglichkeiten, den Angeschuldigten zusätzlich zu belasten, ungenutzt lässt.

Deliktspezifische Inhalte

Diese Kategorie umfasst Elemente der Aussage, die als typisch für das spezifische Verbrechen angesehen werden können, um das es geht. Um dies zu beurteilen ist ein fundiertes Wissen über die typischen Verläufe und Muster bestimmter Delikte nötig.

19. **Deliktspezifische Aussageelemente** sind besonders dann aussagekräftig, wenn sie zwar typisch für das Delikt sind, allerdings der allgemein verbreiteten Meinung widersprechen. Bereits Undeutsch (1967), Arntzen (1983a) und Dettenborn et al. (1984) erkannten die Wichtigkeit von Beschreibungen charakteristischer Entwicklungsdynamiken und Verhaltensmuster.

Die Beurteilung der Aussage soll nach Steller und Köhnken erfolgen, indem zusammenfassend die Ausprägung der Realkennzeichen bewertet wird, was zu einer bestimmten Wahrscheinlichkeitsaussage darüber führt, ob ein Zeuge einen berichteten Vorfall wirklich erlebt hat oder nicht. Dieses Vorgehen wird von den Autoren mit dem Begriff „criteria-based statement analysis“ („Kriterienorientierte Aussageanalyse“) bezeichnet, und ging später unter der Bezeichnung „criteria-based content analysis“ („Kriterienorientierte Inhaltsanalyse“), oder kurz CBCA in die englisch- und inzwischen auch deutschsprachige Literatur ein. Im Folgenden werden daher die Begriffe „(Kriterienorientierte) Aussageanalyse“, „(Kriterienorientierte) Inhaltsanalyse“, „CBCA“ oder einfach „Anwendung der Realkennzeichen“ synonym verwendet, wobei mit Realkennzeichen ab jetzt nur noch die in Tabelle 5 aufgeführten Merkmale in der Version von Steller und Köhnken (1989) bezeichnet sind.

Bezüglich der Aussagekraft der Realkennzeichen weisen Steller und Köhnken (1989) darauf hin, dass bis zu jenem Zeitpunkt keine formalisierte Entscheidungsregel oder ein Cut-Off-Wert existiert, mit dessen Hilfe man wahre von unwahren Aussagen eindeutig unterscheiden könnte. Die von Arntzen (1983a) vorgeschlagene Daumenregel, mindestens drei Kriterien müssten erfüllt sein, um eine Aussage als glaubhaft klassifizieren zu können, halten sie für irreführend, da die verschiedenen Kriterien bei der Bewertung des Wahrheitsstatus’ unterschiedlich stark ins Gewicht fallen. Darüber hinaus hätten außer dem Wahrheitsstatus der Aussage auch die kognitiven Fähigkeiten der aussagenden Person, die Länge der Aussage, sowie die Komplexität des berichteten Ereignisses Einfuß auf Anzahl und Ausprägung der vorhandenen Realkennzeichen.

In der soeben beschriebenen Fassung von Steller und Köhnken (1989) sind die Realkennzeichen seit nun fast 20 Jahren in Gebrauch und seit 1999 auch vom BGH anerkannt (BGH, 2000). Vor Veröffentlichung dieser Fassung fanden auch die zuvor beschriebenen älteren Merkmalssysteme in der Gerichtspraxis weit verbreitet Anwendung. Umso erstaunlicher ist es, dass diesem inhaltsanalytischen Ansatz während der gesamten Entwicklung letztlich keine wissenschaftliche Theorie zugrunde lag, seine Legitimation bezog er einzig aus dem Konsens der Fachleute. Erst die nachfolgende Forschung konnte die Annahmen der Experten, welche im folgenden Abschnitt 2.3.2 näher erläutert sind, im Nachhinein bestätigen und das vielfach beklagte Theoriedefizit in der Zwischenzeit scheinbar weitgehend ausgleichen. So sprechen Ergebnisse sowohl aus der Gedächtnispsychologie als auch aus dem Reality-Monitoring-Ansatz für eine wissenschaftliche Legitimierung der Realkennzeichen (Steck, 2006).

Aber nicht nur die theoretische Absicherung, auch die empirische Überprüfung der Gültigkeit der Glaubhaftigkeitsmerkmale im Sinne der Gütekriterien Validität, Reliabilität und Objektivität erfolgte erst nachträglich. Größere Studien hierzu begann man erst in den späten 80er Jahren durchzuführen, die merkmalsorientierte Inhaltsanalyse stellte also knapp drei Jahrzehnte lang ohne empirische Absicherung ihrer Güte die Grundlage für die psychologische Glaubhaftigkeitsbeurteilung an deutschen Gerichten dar (Greuel et al., 1998). Die Ergebnisse der Überprüfung der Realkennzeichen nach den Gütekriterien sind allerdings recht viel versprechend und werden unter 2.5 näher erläutert.

2.3.2 Zugrunde liegende Annahmen

Wie bereits angesprochen basiert das System der Glaubhaftigkeitskriterien nach Steller und Köhnken (1989) genauso wie die beschriebenen anderen Merkmalssysteme nicht auf einer grundlegenden Theorie, sondern lediglich auf einer Hypothese, genauer gesagt dem allgemeinen Postulat eines Qualitätsunterschiedes zwischen Berichten über Erfundenes und Berichten über Selbsterlebtes (Undeutsch-Hypothese). Diese allgemeine Hypothese wird in der Systematik von Steller und Köhnken (1989) durch zwei Überlegungen spezifiziert, die in den Definitionen der einzelnen Merkmale zum Teil bereits angeklungen sind.

Die eine Überlegung bezieht sich auf die ersten dreizehn Realkennzeichen, die primär inhaltliche Besonderheiten der Aussage beschreiben und ergibt sich aus der Begrenztheit der menschlichen Informationsverarbeitungskapazität. Während ein Wahraussagender nämlich

die Möglichkeit hat, seinen Bericht aus dem Gedächtnis zu rekonstruieren, muss der lügende Zeuge seine Aussage aus abstraktem Schemawissen heraus konstruieren und gegebenenfalls auch über mehrere Befragungen bzw. längere Zeiträume hinweg aufrechterhalten. Dies stellt „eine schwierige Aufgabe mit hoher Anforderung an die kognitive Leistungsfähigkeit des Zeugen dar", so dass „vor allem elementare, direkt zum Handlungsziel hinführende Handlungssequenzen" zu erwarten sind (Steller & Volbert, 1999, S. 51). Dagegen ist es – je nach gegebener Leistungsfähigkeit des Aussagenden – kognitionspsychologisch relativ unwahrscheinlich, dass daneben auch noch originelle Details, Handlungskomplikationen, Interaktionsketten und ähnliches in die Aussage eingebaut werden, da der lügende Zeuge bereits ein erhebliches Ausmaß seiner kognitiven Energie auf kreative Prozesse und Kontrollprozesse verwenden muss (Volbert, 2000).

Neben der verbalen Vermittlung falscher Information, welche von Köhnken (1990, S. 150) als „primäre Täuschung" bezeichnet wird, verfolgt ein erfolgreicher Lügner andererseits aber auch das Ziel, sich selbst als glaubwürdigen Kommunikator darzustellen, was Köhnken „sekundäre Täuschung" nennt. Um möglichst glaubwürdig zu erscheinen wird der Aussagende versuchen, Verhaltensweisen und Äußerungen, welche nach verbreiteter Meinung den Eindruck der Unglaubwürdigkeit erwecken, möglichst zu vermeiden. Aus dieser Überlegung heraus kommt der Komplex der motivationsbezogenen Kriterien nach Steller und Köhnken (1989) zu Stande, der gängige Anzeichen von Unglaubwürdigkeit umfasst.

2.4 Rahmenbedingungen der Kriterienorientierten Inhaltsanalyse in der Glaubhaftigkeitsdiagnostik

Die inhaltliche Analyse der Aussage anhand der oben genannten Realkennzeichen stellt nur einen – wenn auch zentralen – Bestandteil eines umfassenden hypothesengeleiteten Vorgehens zur Untersuchung der Erlebnisfundiertheit von Aussagen dar, das vor allem im englischen Sprachraum unter dem Begriff Statement Validity Analysis oder Statement Validity Assessment (SVA)[5] bekannt geworden ist. Die Ergebnisse der Kriterienorientierten Inhaltsanalyse dürfen daher niemals isoliert, sondern immer nur im Zusammenhang mit den anderen zu erhebenden Daten interpretiert werden.

[5] Der Ausdruck Statement Validity Analysis bzw. Assessment (SVA) wurde im englischen Sprachraum schon vor einiger Zeit eingeführt (Raskin & Esplin, 1991) und ist dort seitdem gebräuchlich. In den vergangenen Jahren setzte er sich allerdings immer mehr auch in deutschsprachigen Publikationen durch (z.B. Niehaus, 2001).

Das methodische Grundprinzip und die Bestandteile der SVA sowie die Stellung der Kriterienorientieren Inhaltsanalyse innerhalb dieses Verfahrens soll im Folgenden dargestellt werden. Anschließend werden Anwendung, Voraussetzungen und Grenzen der CBCA selbst aufgezeigt.

2.4.1 Einbettung der CBCA in einen umfassenden diagnostischen Entscheidungsprozess

Auch wenn sich diese Arbeit vornehmlich mit den Realkennzeichen, sowie ihrer Anwendung und Verlässlichkeit im Rahmen von Glaubhaftigkeitsbegutachtungen befasst, muss immer im Blick behalten werden, dass Schlussfolgerungen über den Erlebnisbezug einer Aussage keinesfalls allein aufgrund der Ergebnisse der Kriterienorientierten Inhaltsanalyse getroffen werden dürfen. Als Ergebnis der Inhaltsanalyse erhält man nämlich wie bei einem psychometrischen Test zunächst nur eine Art Rohwert, das heißt eine bestimmte Anzahl als erfüllt anzusehender Merkmale. Für sich gesehen sind die Daten jedoch in dieser Form bedeutungslos, da es immer darum geht, ob *dieser* bestimmte Zeuge *diese* bestimmte Aussage hätte erfinden können, eine stereotype Interpretation des Rohwertes ist demnach unzulässig (Köhnken, 2004). „Vielmehr erfolgt mit der merkmalsorientierten Inhaltsanalyse nur eine Einschätzung eines Aspektes der Qualität einer Aussage, zur Glaubhaftigkeitsbeurteilung ist diese Aussagequalität dann auf die personalen Voraussetzungen des Zeugen sowie auf die Entstehungs- und weitere Entwicklungsgeschichte der Aussage zu beziehen.“ (Steller & Volbert, 1999, S. 57f.). Auch Undeutsch (1967), Trankell (1971), Arntzen (1983a) und Dettenborn et al. (1984) hatten bei der Ausarbeitung ihrer Merkmalssysteme bereits auf die Wichtigkeit der Bezugnahme auf personale und situative Einflüsse auf die Aussage verwiesen.

In der deutschen Aussagepsychologie gilt das im obigen Zitat von Steller und Volbert (1999) nur grob umrissene Vorgehen längst als methodischer Standard für forensisch-psychologische Glaubhaftigkeitsgutachten, wie er z.B. im Standardwerk von Luise Greuel und Kollegen (1998) festgehalten ist. Des Weiteren zählen hierzu auch die Nachvollziehbarkeit und Transparenz des Begutachtungsprozesses sowie das hypothesengeleitete Vorgehen während der gesamten Begutachtung. Durch ein Urteil aus dem Jahre 1999 wurde der komplexe diagnostische Prozess, der durch die oben genannten Prinzipien bestimmt und im

folgenden genauer beschrieben wird, auch vom BGH als methodischer Mindeststandard für aussagepsychologische Gutachten im Gesetz verankert[6].

Als eine wesentliche und unerlässliche Komponente der Begutachtung wird vom BGH die Generierung von Hypothesen bzw. Annahmen über mögliche Quellen der Aussage angesehen, nach der sich das gesamte weitere Vorgehen richtet. Von den aufgestellten Hypothesen hängt es nämlich in erster Linie ab, welche diagnostischen Methoden (z.B. Testverfahren) angewendet werden und worauf bei der Erhebung der Daten der Schwerpunkt gelegt wird. Man unterscheidet bei der Hypothesenbildung zwischen der Realitäts- bzw. Wahrheitshypothese, die annimmt, dass die vorliegende Aussage auf eigenem Erleben beruht, und der globalen Unwahrhypothese, die nicht von einer Erlebnisfundiertheit der Aussage ausgeht. Diese sehr allgemeine Unwahrhypothese ist für den konkreten Fall noch in mehrere spezifischere Annahmen über die möglichen Quellen der Falschaussage zu untergliedern. Köhnken (2004, S. 45f.) nennt beispielsweise neben dem kompletten oder teilweisen Erfinden der Aussage den Transfer tatsächlicher Erlebnisse von anderen Personen auf den Beschuldigten, Instruktion durch Dritte, Suggestion oder mangelnde Unterscheidungsfähigkeit zwischen Realität und Phantasie aufgrund von psychischen Störungen als mögliche Alternativerklärungen. Er weist jedoch darauf hin, dass nicht in jedem Fall alle der genannten Alternativhypothesen in Frage kommen, jedoch immer sämtliche Alternativen betrachtet werden müssen, für die es Hinweise gibt. Solche Hinweise können meist aus den Akten des entsprechenden Falls entnommen werden, weshalb eine sorgfältige Aktenanalyse stets der erste Schritt einer Begutachtung sein sollte. Darüber hinaus müssen aber auch Hinweise berücksichtigt werden, die sich erst später in der Begutachtung ergeben. Die Phase der Hypothesenbildung ist also nicht statisch, „sondern prozeßhaft und reflexiv am gesamten Untersuchungsablauf orientiert“ (Greuel et al., 1998, S. 45).

Im weiteren Verlauf der Begutachtung verfolgt der Sachverständige das vom BGH benannte methodische Grundprinzip „einen zu überprüfenden Sachverhalt (hier: Glaubhaftigkeit der spezifischen Aussage) so lange zu negieren, bis diese Negation mit den gesammelten Fakten nicht mehr vereinbar ist“ (BHG, 2000, S. 167 f.). Konkret prüft man also zunächst alle gebildeten Unwahrhypothesen und verwirft diejenigen, die mit den gesammelten Fakten nicht mehr vereinbar sind. Dieser Prozess wird so lange fortgeführt, bis entweder

[6] BGH-Urteil vom 30.7.1999 – 1 StR 618/98 – LG Ansbach, veröffentlicht in der Entscheidungssammlung BGHSt 45, 164.

alle Hypothesen, die annehmen, die Aussage sei nicht wahr, verworfen sind und nur noch die Erlebnisfundiertheit als sinnvolle Erklärung zurück bleibt oder bis eine der Alternativhypothesen (z.B. Suggestion) aufgrund der verfügbaren Daten nicht mehr zurückgewiesen werden kann. In einem solchen Fall kann durch die Begutachtung nicht positiv bestätigt werden, dass die Quelle der Aussage wirklich in der eigenen Erfahrung des Zeugen liegt – dies wiederum bedeutet aber nicht, dass der Zeuge gelogen hat, sondern nur dass mit den Mitteln der SVA eine alternative Entstehung der Aussage nicht ausgeschlossen werden kann (Köhnken, 2004).

Zum Zwecke der Prüfung der aufgestellten Hypothesen hat der Sachverständige dem BGH-Urteil zufolge verschiedene Analysen vorzunehmen, deren Ergebnisse zur Verwerfung oder Bestärkung einzelner Hypothesen führen können.

Zunächst einmal werden die Angaben, die der Zeuge dem Sachverständigen gegenüber in der Exploration zum fraglichen Tatgeschehen macht, auf ihre Qualität und inhaltliche Konsistenz hin geprüft. Dies geschieht einerseits mit Hilfe der *Inhaltsanalyse* der Aussage anhand der Realkennzeichen, die vom BGH als „grundsätzlich empirisch überprüft“ (BHG, 2000, S. 171) angesehen werden. Andererseits wird die Stabilität der Aussageinhalte über verschiedene Befragungszeitpunkte hinweg mit Hilfe der so genannten *Konstanzanalyse* kontrolliert, sofern frühere Aussagen in den Akten gut dokumentiert sind.

Zusätzlich zur Prüfung der Qualität der Aussage ist die Zuverlässigkeit der Aussage vor dem Hintergrund personaler und situativer Besonderheiten des Falls zu beurteilen. Greuel et al. (1998) sprechen in Bezug auf die Untersuchung dieser Faktoren von „Validitätsüberprüfung“ der Aussage, in der englischsprachigen Literatur ist von der so genannten „Validity Checklist“ die Rede (Raskin & Esplin, 1991).

Die persönlichen Besonderheiten und Fähigkeiten des Zeugen sind laut BGH im Rahmen der so genannten *Kompetenzanalyse* zu beurteilen, welche „die Beurteilung der persönlichen Kompetenz der aussagenden Person, insbesondere seiner allgemeinen und sprachlichen intellektuellen Leistungsfähigkeit sowie seiner Kenntnisse in Bezug auf den Bereich, dem der erhobene Tatvorwurf zuzurechnen ist (z.B. Sexualdelikte)“ (BHG, 2000, S. 175), aber zum Beispiel auch die Abklärung besonderer Persönlichkeitseigenschaften umfasst. Mit Hilfe von Tests, Fragebögen, Beobachtung und Exploration lässt sich so einschätzen, ob der Zeuge kognitiv in der Lage wäre, eine Aussage mit der vorliegenden Qualität frei zu erfinden oder bewusst von einer Person auf eine andere, nämlich den Beschuldigten, zu über-

tragen. Auch Steller & Köhnken (1989) fordern am Rande der Beschreibung ihres Merkmalssystems, dass bei der Beurteilung der Qualität einer Aussage immer auch die intellektuellen und verbalen Fähigkeiten des Aussagenden berücksichtigt werden müssen.

Zusätzlich zu dieser vom BGH geforderten Kompetenzanalyse, also der Überprüfung individueller Leistungsbesonderheiten, sollte laut Greuel et al. (1998) auch eine Überprüfung der individuellen Aussagebesonderheiten durchgeführt werden, das heißt die Beurteilung des dem Zeugen eigenen Ausdrucksverhaltens und seines Berichtstils. In der Praxis hat sie hierbei bewährt, von der aussagenden Person zusätzlich zu den Angaben zur Sache auch einen Bericht zu einem anderen, möglichst fallneutralen Thema zu erheben und die dort sichtbar werdenden Ausdrucksbesonderheiten mit der in Frage stehenden Aussage zu vergleichen (Greuel, 2001; Hermanutz, Litzcke und Kroll, 2004).

Sonstige Größen, welche auf die Entstehung der Aussage Einfluss genommen haben könnten, werden in der so genannten *Fehlerquellenanalyse* untersucht. Zur Durchführung dieser Analyse betrachtet man die Entstehungsgeschichte der Aussage, das heißt z.B. die Umstände, unter denen es erstmals zu einer Aussage kam, die Reaktionen der Außenwelt auf die Aussage, die Zeit sowie die Ereignisse, die zwischen angeblicher Tat und Erstaussage lagen. Durch sie kann man zu einer Einschätzung darüber gelangen, ob eventuell fremdsuggestive Einflüsse auf die Erstaussage z.B. durch wiederholte, einseitig ausgerichtete Befragung oder auch am Ziel vorbeischießende Psychotherapie (Köhnken, 2004) in Erwägung zu ziehen sind.

Zur Fehlerquellenanalyse kann zusätzlich die *Motivationsanalyse* treten, welche die Feststellung möglicher Motive für eine unzutreffende Belastung des Angeschuldigten zum Ziel hat. Hierzu erweist sich oft die Beleuchtung der Beziehung zwischen mutmaßlichem Täter und Opfer sowie der möglichen Konsequenzen der Falschbeschuldigung für die Beteiligten oder für Dritte als sinnvoll. Der BGH weist jedoch darauf hin, dass eine festgestellte Belastungsmotivation nicht zwingend den Schluss auf eine Falschanschuldigung zulässt (BGH, 2000).

Vom BGH im besagten Urteil vom 30.7.1999 nicht explizit gefordert wird die Überprüfung der *Aussagetüchtigkeit* des Zeugen, obwohl ihr laut Greuel sogar die Rolle einer notwendigen Bedingung für die Glaubhaftigkeit zufällt, „da bei Negation der Aussagetüchtigkeit die Aussage selbst nicht mehr von forensischer Bedeutung ist“ (Greuel, 2001, S. 16). Das psychologische Konstrukt der Aussagetüchtigkeit stellt dementsprechend in der Kon-

zeption Greuels neben den Konstrukten der Aussagequalität und der Aussagezuverlässigkeit einen gleichwertigen Bestandteil der Glaubhaftigkeit im umgangssprachlichen Sinne dar (*aussagepsychologischer Konstrukt-Trias;* Greuel et al., 1998; Greuel, 2001) und sollte daher Gegenstand jeder aussagepsychologischen Begutachtung sein. Man versteht unter Aussagetüchtigkeit die Fähigkeit eines Menschen, „den der Zeugenaussage zugrunde liegenden Sachverhalt realistisch wahrzunehmen, im Gedächtnis zu speichern und in freier Rede oder in einer Befragung sachgerecht wiederzugeben“ (Steck, 2002, S. 16). Diese Fähigkeit kann bei einem Zeugen entweder generell eingeschränkt sein, z.B. durch Intelligenzminderung oder Hirnschädigungen, oder auch aktuell durch vorübergehende Störfaktoren wie Alkoholisierung oder Drogeneinwirkung beeinträchtigt werden. Das Vorliegen solcher Einschränkungen lässt sich über testpsychologische Methoden und eine gründliche biographische Anamneseerhebung bzw. eine gezielte Exploration und die Hinzuziehung objektiver Fakten (z.B. Blutalkoholwert) abklären.

Zum endgültigen Urteil über den Erlebnisbezug der Aussage ist anschließend eine systematische Integration der Ergebnisse aus der Abklärung der Aussagetüchtigkeit, der merkmalsorientierten Aussageanalyse und den zusätzlichen Validitätsüberprüfungen vorzunehmen (Greuel et al., 1998). Betrachtet man also die Ergebnisse der im Folgenden beschriebenen CBCA, so muss dies immer vor dem Hintergrund der übrigen Analysen geschehen.

2.4.2 Anwendung, Voraussetzungen und Grenzen der CBCA

Eine valide Anwendung der Realkennzeichen auf eine Zeugenaussage setzt zuallererst ein adäquates Vorgehen bei der Exploration zur Sache voraus, mit deren Hilfe das Material zur Analyse der Aussagequalität überhaupt erst erhoben wird. Bevor man direkte Fragen an ihn richtet, sollte der Zeuge erst einmal durch entsprechende Aufforderungen dazu gebracht werden, einen möglichst freien, zusammenhängenden Bericht zu produzieren, da ein solcher für die Inhaltsanalyse unerlässlich ist. Die anschließenden Fragen sollten so offen wie möglich sein und erst mit der Zeit spezifischer werden, ein Vorgehen, das von Steller und Volbert (1999, S. 63) als „Trichtertechnik“ bezeichnet wird. Suggestive Fragen haben in jeder Phase der Exploration zu unterbleiben. Um den möglichen Einfluss der verwendeten Berichtanstöße und Fragen abschätzen zu können, sollte die gesamte Exploration zur Sache auf

Ton- oder Videoband aufgezeichnet und zum Zwecke der Analyse – zumindest in ihren wesentlichen Teilen – wörtlich transkribiert werden.

Bei der anschließenden Beurteilung der einzelnen Realkennzeichen in der Aussage „geht es darum festzustellen, ob eine Aussageeigenart in einer Aussage quantitativ und/oder qualitativ so gut ausgeprägt ist, daß sie als Qualitätsmerkmal festzuhalten ist, das auf den Erlebnisbezug der Aussage verweist“ (Greuel et al., 1998, S. 160). Der erforderliche Qualitätsgrad muss dabei wie dargelegt von den individuellen Besonderheiten des Zeugen abhängig gemacht werden, z.B. von seinen kognitiven Voraussetzungen oder seinem persönlichen Erzählstil.

Nach Feststellung der Einzelmerkmale müssen diese zu einem Gesamturteil hinsichtlich der Aussagequalität integriert werden, wodurch die Beurteilung des Wahrheitsgehaltes deutlich mehr diagnostischen Wert erreicht als die zum Teil recht geringen Einzelvaliditäten der Merkmale (siehe 2.5.1) vermuten lassen. Dieser auf dem mathematisch und psychometrisch fundierten Prinzip der Aggregation beruhende Umstand wird auch vom BGH anerkannt und dem Gutachten von Fiedler und Schmid (1999) folgend zur Legitimierung der merkmalsorientierten Aussageanalyse in der Glaubhaftigkeitsbegutachtung angeführt (BGH, 2000). Durch die Aggregation heben sich nämlich die statistisch unabhängigen Fehleranteile der einzelnen Merkmale gegenseitig auf, die systematischen Anteile, die auf die gemeinsame zu erschließende Größe (hier: die tatsächliche Wahrheit der Aussage) zurückzuführen sind, werden dagegen verstärkt. Gerade wenn die einzelnen Merkmale für sich genommen von sehr begrenztem diagnostischem Wert sind, d.h. ihre Aussagekraft im Durchschnitt nur knapp über dem Zufall liegt, wirkt sich die Aggregation besonders stark aus (Fiedler & Schmid, 1999). Dies ist bei den Realkennzeichen der Fall, weshalb von einer guten Validität der aus ihnen gemeinsam abgeleiteten Schlüsse ausgegangen werden kann. Hierbei ergibt sich allerdings folgende Schwierigkeit: Einerseits muss durch die Integration der Einzelmerkmale zu einem Gesamturteil zwangsläufig eine bestimmte Anzahl erfüllter Merkmale als impliziter Grenzwert festgelegt werden, aufgrund dessen im Einzelfall die Bewertung der Aussage als inhaltsanalytisch glaubhaft oder nicht glaubhaft vorgenommen wird – ohne einen solchen Wert ist eine Entscheidung nicht möglich. Andererseits kann und darf aber nicht schematisch von einer bestimmten Anzahl erfüllter Merkmale im Sinne eines Schwellenwertes auf die Glaubhaftigkeit einer Aussage geschlossen werden (BHG, 2000, S. 171). Zwar gilt durchaus die Regel „Je mehr Realkennzeichen bei der Aussage erfüllt sind, umso

näher ist der Inhalt am real erlebten Geschehen; mit der Nähe zum Geschehen erhöht sich die Wahrscheinlichkeit, dass es sich um eine wahre Aussage handelt“ (Steck, 2002, S. 55f.). Wie viele und welche Kennzeichen in einer glaubhaften Aussage vorhanden sein müssen, ist aber eben nicht als allgemeiner Grundsatz festlegbar. Greuel et al. (1998) geben hier jedoch gewisse Anhaltspunkte, indem sie drei *Ausschlussmerkmale*, „die man im Sinne von *Mindestanforderungen* in jeder erlebnisgestützten Aussage erwarten kann“ (Logische Konsistenz, Detaillierungsgrad, Konstanz), sowie einige *Qualifizierungsmerkmale* mit besonders hoher diagnostischer Valenz aufzeigen, welche die Wahrscheinlichkeit des Erlebnisbezuges erhöhen (S. 161f.). Der Umkehrschluss auf eine Lüge bei Fehlen der Merkmale ist dagegen nicht gerechtfertigt, da dieses Fehlen auch durch andere Faktoren (z.B. Hemmungen, Angst, Gedächtnismangel) verursacht werden kann und somit die Inhaltsanalyse nicht als Methode zur „Lügendetektion“ missverstanden werden darf (Steller & Volbert, 1999).

Die Kriterienorientierte Aussageanalyse stößt allerdings, auch wenn sie richtigerweise als Methode zur Substantiierung des Erlebnisgehaltes von Aussagen verstanden wird, in gewissen Bereichen an ihre Grenzen.

Bietet eine Aussage etwa zu wenig Material und besteht beispielsweise nur aus einer einfachen Negation eines Sachverhaltes oder einer sehr knappen, nicht näher konkretisierbaren Behauptung, so ist eine Beurteilung des Erlebnisgehaltes mit aussagepsychologischen Mitteln kaum möglich. Dadurch ist die Anwendbarkeit der Aussageanalyse letztendlich auch von der Komplexität des inkriminierten Geschehens abhängig, da sehr kurze, einfache Vorgänge überhaupt erst nicht die Voraussetzungen für zahlreiche inhaltliche Qualitätsmerkmale in der Aussage bieten (Greuel et al., 1998).

Zum anderen ist hier anzuführen, „dass die Realkennzeichen ungeeignet sind, zur Unterscheidung zwischen einer wahren und einer suggerierten Aussage beizutragen“ (BHG, 2000, S. 171f.), da im Falle der Suggestion die aussagende Person subjektiv von der Wahrheit ihrer Aussage überzeugt ist und daher die den Realkennzeichen zugrunde liegenden Prozesse (siehe 2.3.2) nicht zum Tragen kommen. Um mögliche suggestive Einflüsse aufzudecken ist wie beschrieben die genaue Rekonstruktion der Aussagegenese wichtig.

Einige neuere Studien weisen darüber hinaus darauf hin, dass sich die (falschen) Aussagen von Personen, die mit den Realkennzeichen vertraut gemacht und angeleitet wurden, diese in ihre Angaben einzuarbeiten, mit Hilfe der CBCA nicht mehr von wahren Aussagen unterscheiden lassen (Vrij, Kneller & Mann, 2000; Vrij, Akehurst, Soukara & Bull, 2004).

Des Weiteren scheint sowohl die soziale Kompetenz der aussagenden Person (Vrij et al., 2004) als auch ihre Vertrautheit mit den geschilderten Vorgängen Einfluss auf den CBCA-Score zu haben, sodass zum Beispiel die Unterscheidung zwischen Aussagen mit wahrem Erlebnishintergrund und solchen mit lediglich aus den Medien entnommenem Hintergrund durch die CBCA erschwert ist (Pezdek et al., 2004).

Keine Einschränkung bezüglich der Anwendbarkeit der CBCA besteht dagegen hinsichtlich des Alters der begutachteten Zeugen und des Deliktes, das Thema der Aussage sein kann, auch wenn die Realkennzeichen ursprünglich für die Aussagen von Kindern zu mutmaßlichem sexuellem Missbrauche entwickelt worden waren. Die zugrunde liegende Idee der Kriterienorientierten Aussageanalyse ist genauso auf die Aussagen von Erwachsenen anwendbar und keineswegs auf Aussagen über sexuellen Missbrauch beschränkt (Köhnken, 2004; Aymans, 2005).

2.5 Empirische Stützung der Kriterienorientierten Inhaltsanalyse

2.5.1 Untersuchungen zur Validität

Da sich ausführliche Darstellungen der bisherigen Feld- und Laborstudien zur Validität der Realkennzeichen, also der Frage, wie gut die Realkennzeichen wirklich zur Unterscheidung von wahren und erfundenen Aussagen geeignet sind, zahlreich an anderer Stelle finden (einen guten Überblick bieten z.B. Greuel et al., 1998; Niehaus, 2001 und speziell für die englischsprachigen Studien Vrij, 2005), sollen sie hier nur zusammenfassend dargestellt werden.

Erste Versuche, die Realkennzeichen mit Daten aus dem „Feld“, d.h. der gutachterlichen Praxis zu validieren, wurden in Deutschland hauptsächlich über die nachträgliche Analyse forensisch-psychologischer Glaubhaftigkeitsgutachten unternommen, z.B. von Arntzen (1982, 1983a) oder Littmann und Szewczyk (1983), wobei die Autoren hierbei jeweils die Validität ihrer eigenen Glaubhaftigkeitsmerkmale prüften. Während Arntzen sich aufgrund seiner nicht genauer dokumentierten Untersuchungen von der Validität der Glaubhaftigkeitsmerkmale überzeugt zeigte (Arntzen, 1983a), ergab sich bei Littmann und Szewczyk (1983) ein differenzierteres Bild. Hier ermöglichten nur einige Merkmale eine gute Differenzierung zwischen glaubhaften und nicht glaubhaften Aussagen (z.B. *Realistik, Wirklichkeitsnähe* und *Originalität*), wohingegen andere sich nicht als valide erwiesen. Beide Arbei-

ten sind jedoch mit methodischen Problemen behaftet, die sie zur Validitätsprüfung der Merkmale wenig geeignet machen (vgl. Niehaus, 2001). Das größte Problem stellt dabei jeweils das Fehlen objektiver Außenkriterien dar; Arntzen (1983a) kann hier lediglich nachträgliche Geständnisse sowie die große Übereinstimmung der Ergebnisse der Begutachtungen mit den späteren Gerichtsurteilen anführen, Littmann und Szewczyk (1983) ziehen die von ihnen selbst getroffenen Glaubhaftigkeitsdiagnosen als Außenkriterium heran – ein Vorgehen, das notwendigerweise zu Zirkelschlüssen führt.

Die ersten größeren Feldstudien zur Untersuchung der Validität der Realkennzeichen nach Steller und Köhnken (1989) stammen hauptsächlich aus den USA. Als wichtigste sind hier die Arbeiten von Esplin et al. (1988), Boychuk (1991, beide zitiert nach Vrij, 2005), Lamb et al. (1997) sowie Parker und Brown (2000, zitiert nach Vrij, 2005) zu nennen, die bis auf die letztgenannte Studie alle die Aussagen von Kindern in Fällen des mutmaßlichen sexuellen Missbrauchs zum Gegenstand hatten. In Deutschland überprüften Krahé und Kundrotas (1992) die Validität der Realkennzeichen anhand von Aussagen angeblich oder tatsächlich vergewaltigter Frauen. Insgesamt ergaben sich in allen genannten Felduntersuchungen inhaltliche Unterschiede zwischen (wahrscheinlich) wahren und (wahrscheinlich) nicht wahren Aussagen im Sinne der Undeutsch-Hypothese, d.h. einige der Realkennzeichen kamen häufiger bzw. ausgeprägter in den wahren als in den falschen Aussagen vor. Allerdings konnten diese Unterschiede nicht für alle Realkennzeichen gefunden werden, vor allem die motivationsbezogenen Merkmale differenzierten nur sehr selten zwischen wahren und falschen Aussagen. In vier der fünf genannten Studien zeigten lediglich fünf Merkmale signifikante Unterschiede und damit eine hohe Validität, nämlich *Unstrukturierte Darstellung, Quantitativer Detailreichtum, Interaktionsschilderungen, Wiedergabe von Gesprächen* und *Schilderung eigener psychischer Vorgänge,* die Merkmale *Logische Konsistenz* und *Raum-zeitliche Verknüpfung* erwiesen sich zumindest in drei der fünf Studien als valide.

Bei der Interpretation dieser Ergebnisse sind neben spezifischer Schwächen einzelner Studien allerdings immer die beiden grundsätzlichen Nachteile von Feldstudien zur Validierung der Realkennzeichen im Hinterkopf zu behalten. Diese Nachteile stehen ihrem großen Vorteil, der hohen Realitätsnähe, gegenüber und bestehen zum einen in der starken Selektivität des Untersuchungsmaterials, zum anderen und vor allem aber im Fehlen von objektiven und unabhängigen Außenkriterien für den Wahrheitsstatus der jeweils betrachteten Aussagen (vgl. Köhnken, 1990, S. 115f.).

Diesen beiden Probleme kann man durch eine randomisierte Stichprobenauswahl und die kontrollierte Einteilung der Versuchspersonen in „Wahraussagende“ und „Lügner“ in experimentellen Laborstudien zur Validität der Realkennzeichen begegnen. Dafür besteht hier die besondere Schwierigkeit, „… ein Paradigma zu finden, das es erlaubt, Versuchspersonen zur Schilderung von Sachverhalten aufzufordern, für die nachweislich eine Erlebnisgrundlage gegeben ist oder nicht, wobei die emotionale und kognitive Bedeutung dieser Sachverhalte für den Aussagenden möglichst vergleichbar sein sollte mit dem Erleben eines Sexualdelikts, ohne dabei die Grenzen ethischer Zumutbarkeit für die Versuchspersonen zu überschreiten“ (Steller Wellershaus & Wolf, 1992, S. 161; siehe auch Arntzen, 1983b).

Die Autoren dieses Zitates selbst stellten sich als erste mit den Realkennzeichen nach Steller und Köhnken (1989) als Instrument zur Analyse der Aussagen dieser Herausforderung (Steller et al., 1992). Später folgten weitere Untersuchungen, z.B. von Wolf und Steller (1997) und Niehaus (2001) in Deutschland, sowie einige amerikanische Studien (z.B. Landry & Brigham, 1992), die sowohl mit Kindern als auch mit Erwachsenen als Versuchspersonen arbeiteten.

Zusammenfassend ergab sich vor allem in den drei genannten deutschen Studien eine gute Stützung der Undeutsch-Hypothese für die meisten der Realkennzeichen, nur die motivationsbezogenen Merkmale erwiesen sich auch hier als weniger valide. Für die englischsprachigen Validitätsstudien zieht Vrij (2005, S. 15) ein ähnliches Fazit: „… Criterion 3 (quantity of details) received the most support. … Unstructured production (Criterion 2), contextual embeddings (Criterion 4), and reproduction of conversation (Criterion 6) all received strong support as well. The so-called motivational criteria, Criterion 14 to 18, received less support than most cognitive criteria (1-13) “. Weiterhin ergab sich laut Vrij in elf von zwölf Studien eine Stützung der globalen Hypothese, dass in wahren Geschichten insgesamt mehr Realkennzeichen bzw. höhere CBCA-Scores vorzufinden sind als in falschen.

Eine gute Diskriminationsfähigkeit der motivationsbezogenen Merkmale zwischen wahren und falschen Geschichten konnte bisher lediglich in der Studie von Niehaus (2001) nachgewiesen werden. Dies ist vermutlich darauf zurückzuführen, dass es ihr mit ihrem Versuchsaufbau in besonderem Maße gelang, die oben zitierte Forderung von Steller et al. (1992) zu erfüllen und bei den teilnehmenden Kindern eine sehr realitätsnahe Motivation für das Hervorbringen möglichst glaubhafter Aussagen hervorzurufen. Neben dem Vergleich zwischen wahren und erfundenen Aussagen bezüglich der (modifizierten) Realkennzeichen

nach Steller & Köhnken (1989), der generell eine Bestätigung der Undeutsch-Hypothese ergab, wurde weiterhin ein Vergleich zwischen wahren Aussagen und solchen, die zumindest zum Teil auf einer wirklichen Wahrnehmungsgrundlage (über Berichte von anderen Personen oder über eigene, ähnliche Erfahrungen) basierten, vorgenommen. Hier trennten die motivationsbezogenen Merkmale sogar besonders gut, wohingegen die Diskriminationsfähigkeit der anderen Realkennzeichen unter dieser Bedingung abfiel[7].

Dieses Ergebnis konnte in einem neueren Replikationsversuch der Studie von Niehaus (2001) mit erwachsenen Probanden nicht bestätigt werden (Hettler, 2005). Zwar führte auch hier der Gruppenvergleich der wahren und der erfundenen Geschichten zu einem signifikanten Unterschied in der Häufigkeit der erfüllten Realkennzeichen, der Vorteil der motivationsbezogenen Merkmale bei der Unterscheidung von wahren und nacherzählten Aussagen wurde aber nicht gefunden.

2.5.2 Untersuchungen zur Reliabilität

Verglichen mit der Anzahl der Studien zur Validität der Realkennzeichen wurde der Überprüfung der Reliabilität – wie auch der der Objektivität – bisher nur sehr wenig Aufmerksamkeit gewidmet. Zunächst wurde die Messgenauigkeit meist nur am Rande von Validitätsstudien thematisiert und hierbei fast ausschließlich mittels des Verfahrens der Interrater-Reliabilität ermittelt, d.h. die Reliabilität wurde als Übereinstimmung der verschiedenen Beurteiler hinsichtlich ihrer Einschätzung der Realkennzeichen definiert. Sofern sie überhaupt angegeben werden, variieren die angegebenen Werte für die Beurteilerübereinstimmung in den verschiedenen Validitätsstudien sehr stark, wobei durch die Verwendung unterschiedlicher Reliabilitätsmaße (prozentuale Übereinstimmung, Kappa-Koeffizienten, Pearson-Korrelationen) eine direkte Vergleichbarkeit auch nicht immer gegeben ist. Relativ geringe Beurteilerübereinstimmungen wie z.B. bei Krahé und Kundrotas (1992), die von Kappa-Werten von nur .02 bis .35 für die einzelnen Realkennzeichen berichten, sind vermutlich größtenteils auf ein unzureichendes Training der Beurteiler und zu ungenaue Definitionen der einzelnen Merkmale zurückzuführen (Niehaus, 2001).

[7] Des Weiteren entwickelte Niehaus einen Katalog so genannter „Lügenmerkmale“, deren Vorliegen komplementär zu den Realkennzeichen gegen die Wahrheit einer Aussage spricht und überprüfte dies auf ihre Validität hin Auf diese soll aber hier nicht näher eingegangen werden.

Eine der ersten Studien, die explizit die Bestimmung der Reliabilität der CBCA zum Thema hatte, stammt von Anson, Golding & Gully (1993). Als Datenmaterial verwendeten die Autoren Videoaufzeichnungen der Aussagen von 23 kindlichen Zeugen, die Opfer eines sexuellen Missbrauchs geworden waren. Jeweils zwei von insgesamt vier trainierten Ratern beurteilten die Aussagen, die ausnahmslos durch umfassende Geständnisse gestützt wurden, hinsichtlich der Realkennzeichen, anschließend wurde die Interrater-Reliabilität berechnet. Die Autoren zogen hierzu drei unterschiedliche Maße heran: Die prozentuale Übereinstimmung der Beurteiler, den Kappa-Koeffizienten nach Cohen und – da Kappa durch deutlich von .5 abweichende Auftretenshäufigkeiten der Kriterien stark beeinflusst wird – Maxwell's RE-Koeffizienten[8].

Insgesamt lagen die Kappa-Werte für die einzelnen Realkennzeichen zwischen -.30 und 1, der Durchschnitt betrug .29; die Werte des RE-Koeffizienten nach Maxwell erreichten im Durchschnitt .49 und bewegten sich zwischen -.22 und 1. Neun Merkmale erreichten einen RE-Koeffizienten von über .50 und können daher laut Autoren als hinreichend reliabel gelten, namentlich *Logische Konsistenz, Quantitativer Detailreichtum, Wiedergabe von Gesprächen, Schilderung von Komplikationen im Handlungsverlauf, Phänomengemäße Schilderung unverstandener Handlungselemente, Schilderung psychischer Vorgänge des Täters, Einwände gegen die Richtigkeit der eigenen Aussage, Selbstbelastungen* und *Entlastung des Angeschuldigten.* Vier Merkmale, *Nebensächliche Einzelheiten, Raum-zeitliche Verknüpfungen, Ausgefallene Einzelheiten* und *Spontane Verbesserung der eigenen Aussage*, erreichten bezüglich des RE-Koeffizienten nach Maxwell Werte zwischen .30 und .50, was von den Autoren als gerade noch reliabel eingestuft wird. Als am wenigsten reliabel im Sinne der Beurteilerübereinstimmung erwies sich das Merkmal *Delikttypische Aussageelemente*, was eventuell auf die unterschiedliche und zum Teil unzureichende Erfahrung der Beurteiler mit den delikttypischen Merkmalen eines sexuellen Missbrauchs zurückzuführen sein könnte. Zu beachten ist bei dieser Untersuchung weiterhin, dass die Beurteilung nicht wie eigentlich gefordert anhand von Transkripten erfolgte, sondern Videos verwendet wurden, was die Beurteilung erschwert und somit zu einer Unterschätzung der Reliabilität geführt haben dürfte.

[8] Maxwell's random error coefficient of agreement.

Diese Vermutung wird durch eine weitere Studie bestärkt, die sich ausschließlich auf die Reliabilität der CBCA konzentriert. Die Autoren Horowitz, Lamb, Esplin, Boychuk, Krispin und Reiter-Lavery (1997) überprüften hier neben der Interrater-Reliabilität auch die Test-Retest-Reliabilität der Realkennzeichen und ließen zu diesem Zweck die Transkripte von Interviews mit 100 Kindern, die mutmaßlich Opfer eines sexuellen Missbrauchs geworden waren, zu zwei verschiedenen Zeitpunkten von drei Beurteilern auf das Vorhandensein der 19 Realkennzeichen von Steller und Köhnken (1989) untersuchen. Alle Beurteiler hatten entweder schon langjährige Erfahrung in der Anwendung der Realkennzeichen oder durchliefen unter Anleitung eines Experten ein Training anhand von Transkripten, die nicht aus der Stichprobe der Studie stammten. Die Interrater-Reliabilitäten für die einzelnen Items wurden hier ebenfalls durch die Berechnung von Prozent- und Kappa-Werten, sowie dem RE-Koeffizienten nach Maxwell für beide Beurteilungszeitpunkte bestimmt. Die Kappa-Werte lagen in dieser Studie im Durchschnitt höher als bei Anson et al. (1993) und schwankten zwischen 0 und .71 zum ersten Zeitpunkt und zwischen .12 und .75 zum zweiten Zeitpunkt; die RE-Koeffizienten nach Maxwell variierten zum ersten Zeitpunkt zwischen .24 und .96, zum zweiten Zeitpunkt zwischen .33 und .95. Im Sinne der Interrater-Reliabililtät erwiesen sich dabei nur drei Kriterien durchgängig als nicht hinreichend reliabel, das heißt sie verfehlten zu beiden Zeitpunkten den kritischen Wert von .50 für den Maxwellschen RE-Koeffizienten und zwar *Nebensächliche Details*, *Spontane Verbesserung der eigenen Aussage* und *Eingestehen von Erinnerungslücken*. Die beiden Merkmale *Indirekt handlungsbezogene Schilderungen* und *Ausgefallene Details* waren zumindest zu einem der beiden Zeitpunkte nicht als ausreichend reliabel zu bezeichnen. Beachtenswert hierbei ist allerdings, dass bis auf *Spontane Verbesserungen der eigenen Aussage*, welches zum ersten Beurteilungszeitpunkt mit .24 den niedrigsten RE-Koeffizienten nach Maxwell überhaupt erreichte, alle anderen genannten Kriterien jeweils einen RE-Wert zwischen .30 und .50 aufwiesen und somit nach der Definition von Anson et al. (1993) zumindest als gerade noch reliabel zu bewerten wären. Insgesamt ergaben sich für den Summenscore der Realkennzeichen Interrater-Übereinstimmungen von $r = .78$ bis $r = .82$ zum ersten Beurteilungszeitpunkt und von $r = .86$ bis $r = .89$ zum zweiten Beurteilungszeitpunkt. Die erstmals betrachtete Test-Retest-Reliabilität zwischen den beiden Zeitpunkten lag für die drei verschiedenen Beurteiler zwischen $r = .85$ und $r = .91$. Alle angegebenen Reliabilitäten, die hier als Pearson-Korrelationen berechnet worden waren, wurden mit $p < .0001$ hochsignifikant.

Durch ein intensives Training der Beurteiler sowie präzise Operationalisierungen der Merkmale konnte Susanna Niehaus (2001) in ihrer Studie eine noch bessere Beurteilerübereinstimmung erreichen, die hinsichtlich des Summenscores aller betrachteten Realkennzeichen bei r = .96 lag (Pearson-Korrelation). Für die einzelnen Realkennzeichen erhielt sie Kappa-Übereinstimmungen zwischen .18 (39.4%) und 1 (99.4%) mit einem Median von .76. Die schlechtesten Werte erhielt sie dabei für die beiden Merkmale *Logische Konsistenz* und *Unstrukturierte Darstellung.*

Neben der Berechnung der Interrater- und der Test-Retest-Reliabilität ist in der klassischen Testtheorie die Bestimmung der Reliabilität auch über die so genannte innere Konsistenz möglich. Die Berechnung der inneren Konsistenz stellt in gewisser Weise eine Erweiterung der Testhalbierungs-Methode zur Reliabilitätsberechnung dar, wobei hier der Test in so viele „Untertests“ zerlegt wird, wie er Items hat. Aufgrund der Itemvarianzen und der Varianzen der Gesamtrohwerte wird dann ein Konsistenzkoeffizient berechnet, in der Regel Cronbachs Alpha (Bühner, 2004). Eine hohe innere Konsistenz spricht dafür, dass die Test-Items das gleiche zugrunde liegende Konstrukt messen und demnach ein Zusammenfassen der Items zu einer gemeinsamen Skala gerechtfertigt ist. Aus diesem Grund erscheint eine Realitätsprüfung im Sinne der inneren Konsistenz auch für die Systematik der Realkennzeichen sinnvoll, da für die Einschätzung der Erlebnisfundiertheit einer Aussage die Aggregation der Realkennzeichen notwendig ist (siehe 2.4.2), die psychometrischen Voraussetzungen hierfür allerdings bisher wissenschaftlich nicht abgesichert wurden (Steck, 2006). Durch die Berechnung der Trennschärfen der einzelnen Realkennzeichen kann darüber hinaus der Beitrag der einzelnen Merkmale zu dieser Skala und damit zur Unterscheidung zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen erfasst werden.

Ein erster Versuch in diese Richtung wurde von Hommers (1997) unternommen. Da er im weiteren Verlauf seiner Untersuchung eine psychometrische Anwendung der Realkennzeichen vornehmen wollte, war hierfür zunächst unter anderem mit Hilfe der Berechnung der inneren Konsistenz und einer Trennschärfenanalyse zu klären, wie gut sich die Realkennzeichen überhaupt mit hinreichender Reliabilität zur einem Summenscore zusammenfassen lassen, das heißt zur psychometrischen Identifikation von wahren und unwahren Aussagen eignen (S. 90). Er fasste dabei die Daten, die sich bei Anwendung der Realkennzeichen durch einen einzigen Beurteiler ergeben, als Messwiederholungen eines latenten

Wahrheitsstatus der Aussage auf – analog zu Items eines psychometrischen Tests, der den Wahrheitsstatus einer Aussage untersucht.

Als Datengrundlage diente ihm die bereits kurz erwähnte experimentelle Validierungsstudie von Steller, Wellershaus und Wolf (1992). In dieser Studie wurden die Realkennzeichen nach Steller & Köhnken (1989) an Schülern der ersten und vierten Klasse im Rahmen eines fiktiven Erzählwettbewerbes untersucht – die Schüler mussten je ein reales und ein fiktives Erlebnis aus einer von mehreren vorgegebenen Kategorien berichten, die den relevanten Sachverhalt des sexuellen Missbrauchs soweit vertretbar überzeugend simulierten, z.B. „Blut abgenommen bekommen", „von einem anderen Kind verhauen werden" oder „von einem Tier angefallen werden". Die Ausprägung der Kriterien in den Aussagen wurde auf einer Skala von 0 (nicht vorhanden) bis 3 (stark ausgeprägt) durch drei verschiedene Rater eingeschätzt, als Außenkriterium wurden die Angaben der Eltern herangezogen. Die drei Merkmale *Unstrukturierte Darstellung, Einwände gegen die Richtigkeit der eigenen Aussage* und *Deliktspezifische Aussageelemente* konnten aus versuchstechnischen Gründen nicht in die Auswertung einbezogen werden.

Nachdem er zunächst anhand einer Faktorenanalyse die grundsätzliche Annahme einer Skalenbildung positiv belegt hatte, führte Hommers mit diesen Daten drei Itemanalysen der Realkennzeichen nach der klassischen Testtheorie durch, und zwar einmal nur für die wahren, einmal nur für die unwahren und einmal für alle Geschichten. Die Itemanalysen ergaben insgesamt zufrieden stellende Trennschärfen, Schwierigkeiten und Alphakoeffizienten.

Betrachtet für alle Geschichten[9] und alle 16 betrachteten Realkennzeichen ergab sich ein Cronbachs Alpha von $\alpha = .77$. Die Schwierigkeiten variierten in der Regel unterhalb der Mitte von 4.5 auf der durch die Summierung über drei Rater entstandenen Skala von 0 bis 9. Die meisten Kriterien wurden demnach nur selten oder schwach ausgeprägt gefunden. Die Trennschärfen der einzelnen Kriterien, in diesem Fall r_{it} = part-whole-korrigierte Korrelationen des Kriteriums mit Summe der anderen, schwankte zwischen $r_{it} = .20$ und $r_{it} = .72$, wobei Kriterium 15 mit seiner negativer Trennschärfe eine Ausnahme bildete; der Mittelwert lag bei 0.41.

[9] Da in der vorliegenden Diplomarbeit keine simulierten Daten sondern Gutachten aus der gerichtspsychologischen Praxis verwendet wurden, bei denen eine definitive Beurteilung des Wahrheitsstatus' der Aussagen naturgemäß nicht möglich ist, konnten hier nur Analysen durchgeführt werden, die höchstwahrscheinlich sowohl wahre als auch unwahre Aussagen beinhalten. Bei der Darstellung der Ergebnisse von Hommers wird daher auf die Beschreibung der Analysen getrennt nach wahren und unwahren Geschichten verzichtet, da diese für die vorliegende Arbeit nicht relevant sind.

Die um das jeweilige Item reduzierten α- Werte der Kriteriensumme lagen in der Regel zwischen α = .70 und α = .77, wobei auch hier das Kriterium 15 *Zugeben von Lücken* durch die relativ starke Erhöhung von α beim Weglassen des Kriteriums negativ auffiel und daher für die nachfolgenden Berechnungen nicht berücksichtigt wurde. Ohne *Zugeben von Lücken* betrug das Cronbachs Alpha der Gesamtskala α = .81. Die Mittelwerte der Bewertungen der 15 verbleibenden Merkmale korrelierte mit dem Alter der Kinder, wobei die älteren Kinder insgesamt mehr Realkennzeichen produzierten als die jüngeren, und zwar sowohl in den wahren als auch in geringerem Umfang in den unwahren Geschichten.

Tabelle 6: Itemanalyse-Ergebnisse für alle Geschichten nach Hommers (1997, S. 93)

Kriterium	M	SD	r_{it}	Alpha
1. Logische Konsistenz	6.10	1.76	.41	.75
3. Detailreichtum	5.46	2.08	.72	.73
4. Raum-zeitliche Verknüpfungen	4.33	2.12	.60	.74
5. Interaktionsschilderungen	3.10	1.77	.62	.74
6. Gesprächswiedergaben	3.43	2.85	.40	.76
7. Handlungskomplikationen	1.76	2.02	.39	.75
8. Ausgefallene Details	2.22	1.95	.54	.74
9. Nebensächliche Details	4.67	2.22	.53	.74
10. Unverstandene Handlungselemente	1.10	1.58	.21	.77
11. Indirekt Handlungsbezogenes	0.56	1.30	.21	.77
12. Eigenpsychisches	4.40	2.73	.37	.76
13. Fremdpsychisches	1.19	1.91	.38	.76
14. Spontane Verbesserung	1.97	1.57	.39	.76
15. Eingestehen von Erinnerungslücken	2.68	2.59	-.15	.81
17. Selbstbelastung	0.88	1.37	.20	.78
18. Täterentlastung	1.19	1.61	.21	.77

M = Schwierigkeit des Kriteriums im Sinne der Testkonstruktion, S = Standardabweichung der Ratings um M, r_{it} = Trennschärfe (korrigierte Item-Summenscore-Korrelation), Alpha = Cronbachs Alpha der Summe der restlichen 15 Kriterien

Nachdem die Eignung der Realkennzeichen zur psychometrischen Anwendung grundsätzlich bestätigt war, nahm Hommers auf ihrer Basis eine Gruppenbildung vor und untersuchte so die differentielle Validität der Kriterien. Er kam dabei zu dem Ergebnis, dass sich die Kinder hinsichtlich ihrer „Lügenfähigkeit“ stark unterschieden und einige Realkennzeichen besonders leicht auch durch die schlechteren Lügner zu simulieren waren. Sehr leicht simulierbar und auch in den unwahren Aussagen schlechter Lügner oft vorhanden waren demnach die Kriterien *Logische Konsistenz, Interaktionsschilderungen, Spontane Verbesserung der eigenen Aussage* und *Selbstbelastungen*. Mit Vorsicht anzuwenden, da immer noch

recht leicht zu simulieren sind laut Hommers die Kriterien *Quantitativer Detailreichtum, Schilderung eigener psychischer Vorgänge, Schilderung psychischer Vorgänge des Täters* und *Eingeständnis von Erinnerungslücken*. Als weniger gut simulierbar erwiesen sich die Kriterien *Raum-zeitliche Verknüpfungen, Wiedergabe von Gesprächen, Schilderung ausgefallener Einzelheiten, Schilderung nebensächlicher Einzelheiten* und *Entlastung des Angeschuldigten*. Am schwierigsten zu simulieren waren allerdings die Kriterien *Phänomengemäße Schilderung unverstandener Handlungselemente, Schilderung von Komplikationen im Handlungsverlauf* und *Indirekt handlungsbezogene Schilderungen.* Selbst „Gute Lügner" hatten Schwierigkeiten, diese Merkmale in ihren unwahren Aussagen zu simulieren, wodurch sie als besonders valide gelten können.

Als weitere Feststellung führt Hommers an, dass die Validität der Kriterien auch abhängig vom Thema der berichteten Geschichten war. Bei der Bewertung der Validität der Realkennzeichen muss man demnach laut Hommers berücksichtigen, dass falsche Aussagen zu unterschiedlichen Themen unterschiedlich schwer produzierbar sind. Bei relativ leicht produzierbaren Aussagen stellt sich die Validität der Kriterien weder global noch im Einzelnen ein (Hommers, 1997, S. 99). Die leichtere Produzierbarkeit kann zum Beispiel. durch häufige Konfrontation mit dem Sachverhalt in der Realität oder in den Medien erreicht werden (siehe auch Pezdek et al., 2004).

Die Frage, ob die Zusammenfassung der Realkennzeichen zu einer gemeinsamen Skala zulässig ist und zu einem reliablen Summenscore führt, greift erstmals Lafrenz (2006) in ihrer Arbeit wieder auf. Basierend auf den Aussagen von 60 erwachsenen Personen, die im Rahmen einer Simulationsstudie erhoben worden waren und einen unterschiedlichen Wahrheitsgehalt aufwiesen, errechnete sie für die 17 untersuchten Realkennzeichen[10] als Wert für die Gesamt-Reliabilität ein Cronbachs Alpha von .566. Die Trennschärfen für die einzelnen Realkennzeichen bewegten sich dabei mit Ausnahme des Merkmals *Schilderung eigener psychischer Vorgänge,* das eine negative Trennschärfe aufwies, zwischen r_{it} = .065 und r_{it} = .423. Der Kernbestand der Realkennzeichen scheint also laut Autorin homogen zu sein, auch wenn die Skala in ihrer Untersuchung keine psychometrische Qualität erreichte; allerdings weist sie auch auf die relativ geringe Varianz innerhalb und zwischen den Versuchsgruppen hin, die zwangsläufig zu niedrigen Alpha-Werten führen muss.

[10] Die Realkennzeichen *Phänomengemäße Schilderung unverstandener Handlungselemente* und *Deliktspezifische Aussageelemente* wurden im Datenmaterial nicht vorgefunden bzw. nicht erhoben.

Am trennschärfsten erwiesen sich in ihrer Studie die Merkmale *Quantitativer Detailreichtum, Wiedergabe von Gesprächen* und *Nebensächliche Details*, als wenig trennscharf stellten sich die motivationsbezogenen Merkmale mit Ausnahme von *Spontane Verbesserung der eigenen Aussage* heraus. Sie trugen im Gegenteil sogar zu einer Verminderung der Reliabilität der Gesamt-Realkennzeichenskala bei, da durch ihre Selektion der Wert von Cronbachs Alpha leicht auf $\alpha = .585$ stieg. Durch die Selektion zweier weiterer wenig trennscharfer Items konnte die Reliabilität der Skala sogar auf $\alpha = .618$ verbessert werden, wobei sie dann nur noch zehn Realkennzeichen umfasste.

Neben der inneren Konsistenz der Realkennzeichen nach Steller und Köhnken (1989) überprüfte Lafrenz auch die Reliabilität von zehn so genannten Lügenmerkmalen, die zum Teil aus der Studie von Niehaus (2001) übernommen, zum Teil aus theoretischen Überlegungen selbst abgeleitet worden waren. Für diese ergab sich allerdings keine zufrieden stellende Reliabilität im Sinne der inneren Konsistenz, eine Skalenbildung wird hier als nicht zulässig beurteilt.

2.6 Die Suche nach einem Schwellenwert

Der Wunsch nach einem universell anwendbaren Schwellen- oder Cut-Off-Wert, der eine generelle Unterscheidung zwischen wahren und falschen Aussagen aufgrund einer bestimmten Anzahl erfüllter Realkennzeichen erlauben würde, ist nach Meinung der führenden aussagepsychologischen Autoren zwar in Hinblick auf Vereinfachung und Transparenz des Begutachtungsprozesses durchaus nachvollziehbar, methodisch aber eindeutig unzulässig (u.a. Dettenborn et al., 1984; Steller & Köhnken, 1989; Greuel et al., 1998; Steller & Volbert, 1999). Dieser Meinung ist auch der BGH in seinem Urteil aus dem Juli 1999 gefolgt (BHG, 2000). Die Berechnung bzw. Ableitung eines Schwellenwertes auf Grundlage experimentell oder aus dem Feld gewonnener Daten kann allerdings durchaus einen heuristischen und informativen Wert haben, auch wenn der ermittelte Wert dann keinesfalls als Richtlinie für zukünftige Begutachtungen missverstanden werden darf. Im Zusammenhang mit der Frage nach dem relativen Beitrag der einzelnen Realkennzeichen für die Gesamtdiagnose wurde ein solches Vorgehen sogar explizit angeregt (Steller, Wellershaus & Wolf, 1992, S. 167).

Wie bereits erwähnt hatte Arntzen (1983a) als einziger der genannten Autoren für das von ihm beschriebene System von Glaubwürdigkeitsmerkmalen einen Schwellenwert angegeben. Für ihn galt eine Aussage als „voll erwiesen“, wenn ein Komplex von drei eindeutigen Glaubwürdigkeitsmerkmalen vorlag (S. 22). Seine Feststellung stützte er auf nicht weiter ausgeführte systematische Untersuchungen von Aussagen, die später durch ein Geständnis bestätigt worden waren – in solchen oder anderweitig abgesicherten Aussagen waren laut Arntzen stets drei oder mehr Glaubwürdigkeitskriterien zu finden, wohingegen in als unglaubwürdig erwiesenen Fällen nie ein solcher Merkmalskomplex vorlag.

In einer neuen Untersuchung ermittelte Maier (2006) für 16 der 19 Realkennzeichen nach Steller und Köhnken (1989) einen Cut-Off-Wert von zehn erfüllten Merkmalen, ab dem in ihren Daten mit 95%-iger Wahrscheinlichkeit eine korrekte Klassifizierung als wahre Aussage möglich war. Ein Wert von acht oder weniger Realkennzeichen war dagegen ebenfalls mit einer Wahrscheinlichkeit von 95% je nach Bedingung entweder als frei erfundene oder als aus vorgegebenen Inhalten nacherzählte Aussage einzustufen. Bei ihrer Berechnung stützte sich Maier auf dieselbe Datengrundlage wie die erwähnten Studien von Lafrenz (2006) und Hettler (2005), es gingen demnach die Aussagen von 60 erwachsenen Personen in die Auswertung ein.

3. FRAGESTELLUNG

Das Ziel der vorliegenden Arbeit ist die Untersuchung der inneren Konsistenz und damit der Reliabilität der Realkennzeichen-Kriteriologie nach Steller und Köhnken (1989), wobei analog zum Vorgehen von Hommers (1997) und Lafrenz (2006) die Berechnung von Cronbachs Alpha sowie die Bestimmung der Trennschärfen der einzelnen Realkennzeichen erfolgt. Es soll dadurch die Frage geklärt werden, ob sich die Realkennzeichen psychometrisch zur Zusammenfassung zu einer gemeinsamen Skala eignen, welche hypothetisch den Wahrheitsgehalt einer Aussage misst und wenn ja, wie groß der Beitrag der einzelnen Realkennzeichen zu dieser Skala ist.

Grundsätzlich wurde diese Frage zwar bereits in den beiden genannten Studien sowohl für Kinder (Hommers, 1997) als auch für Erwachsene als Versuchspersonen (Lafrenz, 2006) positiv beantwortet, allerdings zogen die Autoren beider Arbeiten experimentell gewonnene Daten für ihre Analysen heran. In der vorliegenden Arbeit wurden dagegen zur Datengewinnung tatsächliche Glaubhaftigkeitsgutachten aus der aussagepsychologischen Praxis ausgewertet, die von der GWG – Gesellschaft für wissenschaftliche Gerichts- und Rechtspsychologie München, einem der größten Zusammenschlüsse forensisch-psychologischen Sachverständiger in Deutschland, zur Verfügung gestellt wurden. Die vorhandenen experimentellen Studien zur Überprüfung der Reliabilität der Realkennzeichen im Sinne der inneren Konsistenz sollen somit wie in der Forschung zur Validität durch eine Feldstudie ergänzt und mit diesen verglichen werden.

Entsprechend der Arbeit von Lafrenz (2006) wird neben der Bestimmung der Reliabilität der vollständigen Systematik der Realkennzeichen auch geprüft, ob durch die Selektion wenig trennscharfer Merkmale die Reliabilität des Kataloges als Ganzes erhöht werden kann, insbesondere durch den Wegfall der motivationsbezogenen Merkmale. Diese hatten sich in der genannten Untersuchung als wenig trennscharf erwiesen. Darüber hinaus soll durch weitere Trennschärfenanalysen die Reliabilität der Realkennzeichen gesondert für bestimmte Teilstichproben bestimmt werden. Auf diese Weise kann geklärt werden, ob die Realkennzeichen nach Steller und Köhnken (1989) speziell für bestimmte Gruppen von Zeugen verlässlich sind, die in der gutachterlichen Praxis häufig vorkommen, zum Beispiel für jüngere Zeugen oder solche, die mutmaßlich Opfer eines sexuellen Missbrauchs geworden sind.

Im zweiten Teil der Arbeit wird der Versuch unternommen, aus den Entscheidungen der Gutachter eine Art Schwellenwert für die Anzahl der für eine glaubhafte Aussage erforderlichen Merkmale abzuleiten, wie dies durch Maier (2006) bereits an experimentellen Daten vorgenommen wurde. Wie unter 2.6 dargelegt ist dabei zu beachten, dass dieser Wert lediglich die Entscheidungspraxis der aussagepsychologischen Gutachter der GWG widerspiegelt und keinesfalls als allgemeine Richtlinie für die Kriterienorientierte Aussageanalyse missverstanden werden darf. Unabhängig davon ist es ohnehin – wie ebenfalls ausführlich diskutiert – unzulässig, die Entscheidung über die Glaubhaftigkeit einer Aussage nur auf der Basis einer bestimmten Anzahl vorhandener Realkennzeichen zu fällen. Vielmehr muss z.B. auch der Grad der Ausprägung der einzelnen Merkmale, der hier zugunsten einer einfacher handhabbaren dichotomen Codierung vernachlässigt wurde, berücksichtigt und mit weiteren Faktoren wie der Persönlichkeit, den kognitiven Fähigkeiten oder dem Alter der Zeugen in Beziehung gesetzt werden.

Obwohl eine schematische Anwendung eines Schwellenwertes also eindeutig unzulässig ist, impliziert die vom BGH (2000) geforderte Anwendung des Aggregationsprinzipes und damit die Integration der Einzelmerkmale zu einem Gesamturteil jedoch immer auch die unausgesprochene Festlegung eines Cut-Off-Wertes – ohne ihn ist eine Entscheidung im Einzelfall nicht möglich. Die Frage nach dem impliziten Schwellenwert, aufgrund dessen erfahrene Aussagepsychologen entscheiden, ob eine vorliegende Aussage inhaltsanalytisch als erlebnisbasiert bzw. nicht erlebnisbasiert zu gelten hat, ist also durchaus bedeutsam und kann zumindest zur Orientierung äußerst hilfreich sein.

4. METHODIK

4.1 Darstellung der Datengrundlage

Als Datengrundlage dieser Arbeit dienten aussagepsychologische Glaubhaftigkeitsgutachten, die in den Jahren 2000 bis 2005 von Sachverständigen der GWG – Gesellschaft für wissenschaftliche Gerichts- und Rechtspsychologie München erstellt und der Autorin anonymisiert zur Auswertung zur Verfügung gestellt wurden.

Bei der GWG handelt es sich um einen Zusammenschluss von Diplom-Psychologen, die vor allem im Familienrecht und in der Aussagepsychologie als freiberufliche forensisch-psychologische Gutachter tätig sind. Aufgrund der freiberuflichen Tätigkeit liegt die Verantwortung für die einzelnen Gutachten allein beim jeweils vom Gericht beauftragen Sachverständigen, durch den Zusammenschluss profitieren die Sachverständigen aber von einer gemeinsamen Infrastruktur (z.B. Verwaltung, Räumlichkeiten, Bibliothek), einem breiten Weiterbildungsangebot und der Möglichkeit zur Supervision. Gegründet wurde die GWG 1982 von Prof. Dr. Michael Stadler und Dr. Joseph Salzgeber in München und hat sich seitdem stetig weiterentwickelt. Inzwischen arbeiten über 70 Diplom-Psychologen an mehr als 30 Orten in ganz Deutschland im Rahmen der GWG als psychologische Gutachter. Im zentralen Institut, das sich nach wie vor in München befindet, sind zurzeit fünf Sachverständige ausschließlich im aussagepsychologischen Bereich tätig[11].

Die zur Verfügung gestellten Gutachten lagen als anonymisierte Microsoft Word – Dateien vor, die im Netzwerk der GWG München archiviert und ausschließlich innerhalb des Institutes zugänglich sind. Die Selektion und Auswertung der Gutachten fand daher im Zeitraum von November 2005 bis Januar 2006 in den dortigen Räumlichkeiten statt. Die Autorin verpflichtete sich schriftlich, keine personenbezogenen Daten auf elektronischem oder sonstigen Weg aus dem Institut zu entfernen und wurde auf die Verschwiegenheitspflicht und die Pflicht zum Datenschutz bezüglich aller Daten, die nicht für die Öffentlichkeit bestimmt sind, hingewiesen.

[11] Für weitere Information bezüglich der GWG wird auf die Homepage des Institutes (www.gwg-insitut.com) verwiesen.

4.1.1 Gutachten-Stichprobe

Die aussagepsychologischen Sachverständigen der GWG München erhielten in den Jahren 2000 bis 2005 (Stand: 18. November 2005) insgesamt 408 Gutachtenaufträge, Auftraggeber waren dabei überwiegend Staatsanwaltschaften und Richter an Gerichten im südbayerischen Raum. In die Auswertung flossen Gutachten von 14 verschiedenen Sachverständigen ein, die im Laufe des betrachteten Zeitraumes im Rahmen der GWG München freiberuflich tätig waren. Da sich die Sachverständigen der GWG ausnahmslos nach den im BGH-Urteil vom 30.07.1999 festgelegten und unter 2.4 bereits beschriebenen methodischen Mindeststandards richten, können diese als vergleichbar gelten. Auch Aufbau und Strukturierung der Gutachten ist bei allen Sachverständigen einheitlich und orientiert sich eng an den Gutachtenrichtlinien des BDP sowie der einschlägigen Fachliteratur (z.B. Greuel et al., 1998). Dementsprechend beginnt ein schriftliches aussagepsychologisches Gutachten, wie es in der GWG typischerweise erstellt wird, mit der *Beschreibung des formalen Rahmens* der Begutachtung und der Darstellung des Sachverhaltes und der bisherigen Aussagen des Zeugen, soweit dies der *Akte* zu entnehmen ist (sog. „Anknüpfungstatsachen“). Anschließend werden *Logik und Methodik* der aussagepsychologischen Glaubhaftigkeitsbegutachtung erläutert, sowie auf Grundlage der juristischen Fragestellung und der Anknüpfungstatsachen *Hypothesen* formuliert, die den methodischen Standards laut BGH (2000) entsprechend den weiteren Verlauf der Begutachtung bestimmen. Nachdem Ablauf und Methodik der vorgenommenen *psychologischen Untersuchungen* kurz beschrieben wurden, folgt als nächstes der so genannte *Befund*, in dem die Ergebnisse der eigentlichen Begutachtung im Einzelnen ausgeführt werden. In der Regel wird dabei zunächst die *Aussagetüchtigkeit* und *Aussagegenauigkeit* des Zeugen, die *Entstehung der Aussage*, die *Motivlage* und das *Aussageverhalten* betrachtet, bevor anschließend das Kernstück der Begutachtung, die *Analyse der Aussage*, vorgenommen wird. Die Aussageanalyse setzt sich ihrerseits wie vom BGH (2000) gefordert aus der *Konstanzanalyse* und der *Inhaltsanalyse* anhand der Realkennzeichen nach Steller und Köhnken (1989) zusammen. Nach jedem der genannten Unterpunkte erfolgt meist eine Zusammenfassung der jeweiligen Ergebnisse, welche dann erst in einer abschließenden *Beantwortung der Beweisfrage* zu einem Gesamt-Urteil integriert werden.

Da in die Auswertung nur die Gutachten aufgenommen werden sollten, in denen eine vollständige Inhaltsanalyse der jeweiligen Aussage anhand der Realkennzeichen nach Stel-

ler und Köhnken (1989) durchgeführt worden war, musste zunächst eine Selektion der Fälle stattfinden. Im Zuge dieser Selektion wurden zunächst all diejenigen der 408 Fälle aussortiert, in denen zwar ein Begutachtungsauftrag an einen im Rahmen der GWG tätigen Sachverständigen erteilt wurde, die Begutachtung aber aus den verschiedensten Gründen überhaupt nicht erst zu Stande gekommen war. Dies war z.B. dann der Fall, wenn ein Zeuge auf sämtliche Versuche der Kontaktaufnahme nicht reagiert hatte oder die Anklage vorzeitig fallengelassen wurde. Ebenfalls nicht berücksichtigt werden konnten die Fälle, in denen in Absprache mit den Auftraggebern nur eine gutachterliche Stellungnahme oder ein Kurzgutachten ohne ausführliche Darstellung der einzelnen Analyseschritte erstellt worden war.

Aus den verbleibenden 163 Fällen wurden in einem zweiten Schritt wiederum all diejenigen ausgesondert, in denen zwar ein ausführliches Gutachten erstellt, aber keine komplette Inhaltsanalyse anhand der Realkennzeichen durchgeführt worden war, da die im Rahmen der Begutachtung gemachte Aussage zur Sache die für eine Inhaltsanalyse nötigen Grundanforderungen nicht erfüllt hatte. Einer Inhaltsanalyse nicht zugänglich sind zum Beispiel Aussagen, die nicht in Form eines ausreichend langen, zusammenhängenden Spontanberichts, sondern nur als knappe Antworten auf direkte Fragen vorliegen. Bei solchen kurzen, bruchstückhaften Aussagen findet die Inhaltsanalyse nicht genügend Ansatzpunkte, daher kann die Zuverlässigkeit der Aussage nicht eindeutig beurteilt werden (siehe auch 2.4.2). Häufig ist dies bei sehr jungen oder geistig retardierten Zeugen der Fall, wobei bei diesen beiden Zeugengruppen auch hinsichtlich der Aussagetüchtigkeit häufig Zweifel bestehen (Arntzen, 1983a).

Außerdem können Aussagen keiner Kriterienorientierten Analyse unterzogen werden, wenn sie sich insgesamt als zu knapp und nicht nachvollziehbar darstellen und somit die beiden Realkennzeichen *Logische Konsistenz* und *Quantitativer Detailreichtum* von Vornherein als eindeutig nicht erfüllt angesehen werden müssen. Da diese Merkmale aber als Voraussetzung für die Diagnose der Erlebnisbezogenheit einer Aussage gelten (z.B. Raskin & Esplin, 1991, S. 286; Greuel et al., 1998, S. 161), erübrigt sich bei ihrem gleichzeitigen Fehlen in den meisten Fällen eine weitere Analyse der Aussage[12].

[12] Manchmal wirken Aussagen allerdings auf den ersten Blick durchaus umfangreich und nachvollziehbar, der Mangel an Details und Folgerichtigkeit zeigt sich erst bei einer genaueren Analyse der Aussage – aus diesem Grund enthält die Auswertung in Anhang A dennoch Gutachten, bei denen eine komplette Inhaltsanalyse durchgeführt wurde, sich die genannten Merkmale dann aber als nicht gegeben herausstellten.

Alle 121 Gutachten, die nach diesem Selektionsprozess verblieben waren, wurden in die Auswertung aufgenommen. Es erfolgte also keine weitere Selektion, insbesondere nicht hinsichtlich des Tatvorwurfes oder des Alters der Zeugen.

4.1.2 Erfassung der relevanten Daten

Vor Beginn der eigentlichen Analyse wurden die Gutachten mit einem Code versehen, der aus einem Teil des jeweiligen Aktenzeichens bestand, aber keinen Rückschluss auf den Fall oder die Namen der Beteiligten zuließ. Dieser Code diente zur Kennzeichnung des jeweiligen Falles und wurde zusammen mit den anderen erhobenen Daten im Auswertungsschema vermerkt, das sich Anhang A befindet.

Im Zuge der Gutachten-Analyse wurde zunächst bei jedem Gutachten aus den Anknüpfungstatsachen entnommen, wie alt und welchen Geschlechts die aussagende Zeugin bzw. der aussagende Zeuge war, wie der Tatvorwurf lautete und in welcher Beziehung der Beschuldigte zum Zeugen stand. Die Erhebung dieser beiden letzten Variablen diente ausschließlich deskriptiven Zwecken, um so ein genaueres Bild von der Art der untersuchten Fälle zu gewinnen. Anschließend wurde jeweils der Abschnitt zur Inhaltsanalyse der Aussage dahingehend geprüft, welche der Realkennzeichen nach Steller & Köhnken (1989) vom begutachtenden Sachverständigen als erfüllt angesehen worden waren und welche nicht, wobei das Vorhandensein eines Kennzeichens mit „1“, sein Nichtvorhandensein mit „0“ codiert wurde. Durch diese dichotome Codierung fanden zwar Abstufungen in der Intensität bzw. Eindeutigkeit der Kennzeichen keine Berücksichtigung, eine Definition von klar abgegrenzten Intensitäts-Kategorien wäre aber schon alleine aufgrund der ungleichen Schreibstile und Ausdrucksgewohnheiten der verschiedenen Sachverständigen nur schwer umsetzbar gewesen.

Zusätzlich zu den einzelnen Realkennzeichen wurde zuletzt noch das Urteil des Sachverständigen aufgrund der Kriterienorientierten Aussageanalyse erhoben, das heißt seine Einschätzung der Aussage als wahrscheinlich erlebnisfundiert oder wahrscheinlich nicht erlebnisfundiert aufgrund des Vorliegens bestimmter Kombinationen von Realkennzeichen. Dies war zur späteren Berechnung des Schwellenwertes nötig. Es soll an dieser Stelle allerdings noch einmal darauf hingewiesen werden, dass dieses Urteil aufgrund der Inhaltsanalyse bei der Beurteilung der Glaubhaftigkeit einer Aussage immer nur ein vorläufiges sein kann, da es die anderen untersuchten Aspekte wie Aussagemotivation oder Persönlichkeit

des Zeugen nur indirekt berücksichtigt. Erst in der endgültigen *Beantwortung der Beweisfrage* werden alle erhobenen Befunde zu einem Gesamturteil integriert, dieses kann sich daher vom Zwischenfazit nach der Inhaltsanalyse durchaus unterscheiden. Bei Greuel et al. (1998) heißt es hierzu: „Da die Ergebnisse der Validitätsüberprüfung die Resultate der Aussageanalyse im engeren Sinne immer relativieren können, sind durchaus Fälle denkbar und in der forensischen Praxis nicht selten anzutreffen, in denen die aussageimmanente Analyse einer Aussage zwar eine hohe inhaltliche, für den Erlebnisbezug der Aussage sprechende *Qualität* der Aussage bestätigt, massive Störfaktoren, insbesondere der Aussageentwicklung, jedoch starke Zweifel an der *Zuverlässigkeit* der Aussage begründen, daß letztendlich mit aussagepsychologischen Mitteln eine Wirklichkeitsbasis der entsprechenden Aussage nicht mehr zu belegen ist“ (S. 48, Hervorhebungen im Original).

In Zusammenhang mit der Erfassung der Daten ergab sich die Schwierigkeit, dass in manchen Gutachten nicht nur eine, sondern mehrere Aussagen einer kompletten Inhaltsanalyse unterzogen worden waren und daher nicht jedem Gutachtencode eindeutig nur eine Inhaltsanalyse zugeordnet werden konnte. Einerseits fanden sich nämlich in einigen Gutachten die Aussagen mehrerer verschiedener Zeugen, andererseits wurde in anderen Gutachten die Aussage eines einzigen Zeugen in Aussagenteile zu bestimmen abgrenzbaren Vorfällen unterteilt, die dann getrennt einer Inhaltsanalyse unterzogen und daher wie mehrere einzelne Aussagen behandelt wurden.

Der erstere Fall, also das Vorliegen von Aussagen mehrerer Zeugen, traf auf 15 Gutachten zu, beispielsweise wenn einem Täter Vergehen an mehreren Kindern zur Last gelegt wurden und diese dann bei demselben Gutachter zur selben Sache aussagten. Hier wurden die Aussagen der verschiedenen Zeugen einzeln in die Auswertung aufgenommen und durchnummeriert, indem an den jeweiligen Gutachtencode „-1“, „-2“ usw. angehängt wurde (siehe Anhang A). Insgesamt fanden sich in 13 Gutachten die Aussagen von zwei verschiedenen und in zwei Gutachten die Aussagen von drei verschiedenen Zeugen. Damit erhöhte sich die Anzahl der in die Datenanalyse aufgenommenen Aussagen auf 138.

Der zweite geschilderte Sachverhalt gründet sich darauf, dass die Sachverständigen der GWG in Fällen von fortgesetztem und zum Teil über Jahre andauerndem sexuellen Missbrauch dem so genannten „Individuierungsgebot“ des BGH folgen. Der große Senat für

Strafsachen des BGH hatte in einem Urteil vom 03.05.1994[13] festgelegt, dass in Fällen, in denen ein sexueller Missbrauch von Schutzbefohlenen oder Kindern nach §§ 174 und 176 StGB sich über Monate oder gar Jahre erstreckt, dieser nicht als „fortgesetzte Handlung" und somit als eine Tat angesehen werden kann, sondern alle unterscheidbaren Einzeltaten getrennt betrachtet und beurteilt werden müssen. Wörtlich heißt es: „Die Verbindung mehrerer Verhaltensweisen, die jede für sich eine Straftatbestand erfüllen, zu einer fortgesetzten Handlung setzt voraus, daß dies, was am Straftatbestand zu messen ist, zur sachgerechten Erfassung der verwirklichten Unrechts und der Schuld unumgänglich ist. Jedenfalls bei den Tatbeständen der §§ 173, 174, 176 und 263 ist dies nicht der Fall" (BGH, 1994, S. 138).

In den Gutachten zu entsprechenden Fällen wird von den Sachverständigen der GWG daher bei der Exploration und Analyse der Aussage ebenfalls eine Unterscheidung in abgrenzbare Einzelsituationen vorgenommen, sodass in 16 Gutachten mehrere Aussageteile zu finden waren, die jeweils jede für sich einer vollständigen Inhaltsanalyse unterzogen worden waren. Aus den in diesen 16 Gutachten vorliegenden Einzelaussagen musste eine bestimmte für die Analyse ausgewählt werden, um das mehrfache Einfließen desselben Zeugen in die Stichprobe zu vermeiden. Es wurde dabei jeweils diejenige Einzelaussage in die Datenanalyse aufgenommen, die von den Sachverständigen als am substantiiertesten und umfangreichsten bezeichnet wurde und somit in der Regel auch die meisten Realkennzeichen beinhaltete.

4.1.3 Codierungsregeln

In den Gutachten wurden von Seiten der Sachverständigen stets alle Realkennzeichen auf ihr Vorhandensein in der Aussage überprüft, jedoch nur die aufgeführt, die als erfüllt gelten konnten oder für die es Anhaltspunkte gab. Es konnten daher alle Realkennzeichen, die im Text nicht auftauchten, von vornherein mit „0" codiert werden. Mit „0" codiert wurden weiterhin alle Realkennzeichen, für die der Sachverständige zwar gewisse Anhaltspunkte feststellen konnte, die er aber unter den gegebenen Umständen trotzdem als nicht erfüllt gewertet hatte. Dies war beispielsweise der Fall, wenn ein Realkennzeichen in Bezug zum sonstigen Aussageverhalten bei nicht in Frage stehenden Ereignissen (Aussagebaseline)

[13] BGH-Urteil vom 03.05.1994 – GSSt 2 und 3/93 – LG Mainz, LG Wuppertal, veröffentlicht in der Entscheidungssammlung BGHSt 40, 138.

oder zur intellektuellen Kapazität des Zeugen zu schwach ausgeprägt war. Schließlich erhielten alle Realkennzeichen eine „0“-Codierung, die aus verschiedenen Gründen auf die betreffende Aussage nicht anwendbar waren. Als Beispiel für diesen Fall eignet sich das Realkennzeichen *Phänomengemäße Wiedergabe unverstandener Handlungselemente*, das sich definitionsgemäß nur auf die Aussagen von kleinen Kindern und geistig retardierten Zeugen anwenden lässt, bei denen man von Unverständnis für bestimmte Handlungselemente ausgehen kann.

Alle anderen Formulierungen, die das Vorhandensein eines Merkmals ausdrückten, wurden mit „1“ codiert, und zwar auch dann, wenn sie mit Einschränkungen verbunden waren. Beispiele für solche Einschränkungen sind „das Merkmal ist hier nur schwach ausgeprägt“, „das Merkmal ist zwar als erfüllt anzusehen, seine Aussagekraft wird in diesem Fall aber eingeschränkt durch...“, „das Merkmal ist im weiteren Sinne erfüllt“, „das Merkmal ist – vor dem Hintergrund der kognitiven Ausstattung der Zeugin – gegeben“.

4.2 Methodik der Datenauswertung

4.2.1 Trennschärfenanalyse und Itemselektion

Die Trennschärfenanalyse ist ein wichtiger Bestandteil der so genannten Itemanalyse, die klassischerweise bei der Test- und Fragebogenkonstruktion Anwendung findet. Die Itemanalyse umfasst neben der Analyse der Trennschärfen auch die Bestimmung der Schwierigkeitsindizes und der Homogenität der Test-Items. Mit ihrer Hilfe wird in der Entwicklungsphase eines Tests überprüft, ob die vorgesehenen Test-Items überhaupt der Test-Absicht entsprechen. Als zusätzliches Reliabilitäts-Maß für den Gesamt-Test wird häufig noch Cronbachs Alpha berechnet, welches das gängigste Maß für die innere Konsistenz eines Tests darstellt. In seine Berechnung fließen neben der Anzahl der Test-Items auch die Itemvarianzen und die Varianzen der Gesamtrohwerte ein. Dabei fällt Alpha generell höher aus, je mehr Items im Test vorhanden sind und je niedriger die Itemvarianzen im Vergleich zur Varianz der Gesamt-Testwerte sind (Bühner, 2004). Je heterogener eine Stichprobe also ausfällt, desto höhere Werte kann Cronbachs Alpha annehmen (siehe auch Lafrenz, 2006).

Nach Abschluss der Itemanalyse erfolgt in der Regel auf Grundlage der berechneten Kennwerte die Selektion ungeeigneter Items mit dem Ziel, Cronbachs Alpha und damit die Reliabilität des endgültigen Tests zu maximieren.

Die Trennschärfe eines Items stellt rein rechnerisch die (part-whole-) korrigierte Korrelation des Items mit dem Gesamt-Testscore, also der Summe aller Item-Rohwerte dar. Die part-whole-Korrektur wird notwendig, da das betreffende Item sonst mit in den Gesamt-Testscore, mit dem es korreliert wird, eingehen und die Trennschärfe damit überschätzt würde (Bühner, 2004).

Inhaltlich drückt die Trennschärfe eines Items aus, „wie gut ein Item eine Skala die aus den restlichen Items gebildet wird widerspiegelt bzw. wie prototypisch ein Item für die Skala ist“ (Bühner, 2004, S. 87). Das bedeutet, dass trennscharfe Items meist nur von denjenigen Probanden richtig gelöst bzw. erfüllt werden, die auch insgesamt einen hohen Testscore aufweisen und daher die vom Test gemessene Eigenschaft oder Fähigkeit sehr wahrscheinlich in hohem Ausmaß besitzen („gute“ Probanden). Eine niedrige Trennschärfe haben dementsprechend diejenigen Items, die von „guten“ und „schlechten“ Probanden gleichermaßen oft gelöst bzw. erfüllt werden und demnach nicht gut zwischen diesen beiden Probandengruppen trennen. Als hoch würde man nach Fisseni (1997, S. 124) Trennschärfen über .50 bezeichnen, Trennschärfen zwischen .30 und .50 gelten als mittel, und erst Werte unter .30 als niedrig.

Ob die ermittelten Trennschärfen unter Berücksichtigung der Stichprobengröße überhaupt signifikant von null abweichen, kann analog zur Signifikanzprüfung bei Korrelationen mit folgender Formel ermittelt werden:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

Für Stichproben des Umfangs n > 3 kann man zeigen, dass der Ausdruck mit df = n – 2 t-verteilt ist (Bortz, 1999, S. 207). Löst man diese Gleichung nach r auf, so können für verschiedene Signifikanzniveaus und Freiheitsgrade kritische Korrelationen berechnet und diese mit den vorliegenden Trennschärfen verglichen werden:

$$r_{krit} = \sqrt{\frac{1}{1+\frac{n-2}{t^2}}}$$

In der vorliegenden Arbeit soll wie bei Hommers (1997) und Lafrenz (2006) die Trennschärfenanalyse auf die Zusammenstellung der Realkennzeichen nach Steller und Köhnken (1989) angewendet werden, um zu prüfen, wie gut sie das hypothetisch zugrunde liegende Konstrukt „Realitätsnähe“ repräsentieren. In diesem Fall entsprechen die einzelnen Realkennzeichen den Test-Items, die in den jeweiligen Aussagen entweder „gelöst“ d.h. vorhanden sein oder „nicht gelöst“, d.h. nicht vorhanden sein können. Im Folgenden werden daher die Begriffe „Realkennzeichen“ und „Item“ gleichbedeutend verwendet. Als Testscore wird hier aufgrund der dichotomen Codierung die Summe der in einer Aussage gelösten, d.h. vorhandenen und damit mit „1“ bewerteten Merkmale verwendet. Er kann daher für jede Aussage Werte zwischen null und maximal 19 annehmen.

Die Prüfung der Signifikanz der ermittelten Trennschärfen erfolgt in der vorliegenden Arbeit auf einem Signifikanzniveau von $\alpha < .05$ und $\alpha < .01$. Die nach obiger Formel berechneten kritischen Korrelationen für die jeweilige Stichprobengröße finden sich jeweils direkt vor der Darstellung der Ergebnisse der betreffenden Trennschärfenanalyse.

Als erstes erfolgt hierbei die Darstellung der Trennschärfenanalyse aller 19 Realkennzeichen, die unter Berücksichtigung sämtlicher 138 ausgewerteter Aussagen durchgeführt wurde. Durch diese recht große und heterogene Feld-Stichprobe kann ein guter Überblick über die Reliabilität und Trennschärfe der Realkennzeichen, wie sie in der täglichen aussagepsychologischen Praxis Anwendung finden, gewonnen werden.

Wie bereits beschrieben zeigte sich allerdings in verschiedenen Validierungsstudien der Realkennzeichen-Kriteriologie nach Steller und Köhnken (1989) eine mangelnde Validität der motivationsbezogenen Merkmale (z.B. Steller et al., 1992). Aus diesem Grund wird in der nächsten dargestellten Analyse untersucht, ob eventuell ein Mangel an Reliabilität für die fehlende Validität dieser Merkmale verantwortlich ist, da nach Fisseni (1997, S. 66) Reliabilität eine Voraussetzung für Validität darstellt. In diesem Fall müsste sich die Gesamt-Reliabilität der Realkennzeichen-Kriteriologie durch das Weglassen der motivationsbezogenen Merkmale verbessern. Entsprechend dem Vorgehen von Lamb et al. (1997), sowie Hettler (2005), Maier (2006) und Lafrenz (2006), wurde dabei allerdings das Merkmal *Spontane Verbesserung* aus der Kategorie der motivationsbezogenen Merkmale ausgenommen, da das spontane Einfügen von Verbesserungen auch als inhaltliche Eigenschaft der Aussage gesehen werden könnte, welche eher Hinweise auf kognitive Vorgänge und weniger auf die Motivlage bei der aussagenden Person gibt.

Nachdem die Realkennzeichen anfänglich nur für die Anwendung bei kindlichen Zeugen und hier hauptsächlich für Opfer von sexuellem Missbrauch konzipiert worden waren und auch von Steller und Köhnken vor allem für diesen Zweck empfohlen werden (Steller & Köhnken, 1989, S. 233), finden sich im Anschluss Darstellungen weiterer Trennschärfenanalysen, welche dieser ursprünglichen Konzeption Rechnung tragen. Zum einen wird die Stichprobe nach dem Alter geteilt, so dass die Trennschärfen der Realkennzeichen vergleichend für ältere und jüngere Zeugen betrachtet werden können. Zum anderen wird eine Trennschärfenanalyse durchgeführt, in die ausschließlich die Aussagen hinsichtlich sexuellen Missbrauchs einfließen. Mit den Ergebnissen dieser Analyse werden genauere Aussagen über die Verlässlichkeit der Realkennzeichen in ihrem ursprünglichsten und wohl auch immer noch häufigsten und wichtigsten Anwendungsgebiet möglich.

Zusätzlich zu den Trennschärfen wird bei jeder Analyse das Cronbachs Alpha für die Gesamt-Reliabilität der Realkennzeichen-Skala im Sinne der inneren Konsistenz angegeben, darüber hinaus die Auftretenshäufigkeit der einzelnen Items sowie das Cronbachs Alpha, das sich durch das Weglassen des jeweiligen Items ergeben würde. Durch diesen Wert wird eine Beurteilung dahingehend möglich, ob ein Realkennzeichen überhaupt zur Reliabilität der Kriteriologie als Ganzes beiträgt oder diese im Gegenteil eher vermindert. In der auf alle Trennschärfenanalysen folgenden Itemselektion wird daher vor jedem Selektionsschritt geprüft, für welches Item der Wert für Cronbachs Alpha durch sein Weglassen am höchsten würde und genau dieses Item wird anschließend selektiert. Dieser Vorgang wird so oft wiederholt, bis kein Item mehr vorhanden ist, dessen Selektion zu einer Erhöhung von Cronbachs Alpha führen würde und somit die Reliabilität der Skala ihren maximalen Wert erreicht hat.

4.2.2 Bestimmung eines Schwellenwertes: Diskriminanzanalyse und Logistische Regression

Will man empirisch der Frage nachgehen, anhand welcher Einflussgrößen man zwei (oder mehr) Gruppen am besten unterscheiden kann, welche Eigenschaften ein Fall haben muss, um einer bestimmten Gruppe zugeordnet zu werden und mit welcher Wahrscheinlichkeit dies geschieht, stehen prinzipiell zwei statistische Verfahren zur Wahl: Die Diskriminanzanalyse und die logistische Regression. Beide Verfahren sind besonders dann angebracht, wenn eine kategoriale, nominalskalierte abhängige Variable mit zwei Ausprägungen

(0/1-Ereigniss) vorliegt, was auch als Zwei-Gruppen-Fall interpretiert werden kann (Backhaus et al., 2003). Bei beiden Verfahren kann der Schwellenwert für die unabhängigen Variablen spezifiziert werden, auf Grund dessen die Einteilung in die eine oder die andere Gruppe erfolgt.

Im vorliegenden Fall sollen die einzelnen Zeugenaussagen auf der Grundlage der Anzahl der vorhandenen Merkmale entweder der Gruppe der glaubhaften oder der nicht glaubhaften Aussagen zugeordnet werden. Es handelt sich hier demnach um eine binäre abhängige Variable „Urteil der Gutachters“, welche die beiden Ausprägungen „glaubhaft“ und „nicht glaubhaft“ annehmen kann. Anders als in den meisten Fällen üblich soll hier der Einfluss nur einer unabhängigen Variable, das heißt einer Größe, auf die sich die Zuteilung zu einer bestimmten Gruppe begründet, untersucht werden, und zwar die Anzahl der vorhandenen Realkennzeichen. Für diese Problemstellung wären Diskriminanzanalyse und logistischer Regression gleichermaßen geeignet, denn bei Vorliegen der jeweiligen Grundvoraussetzungen liefern sie vergleichbare prädiktive und klassifikatorische Ergebnisse und arbeiten mit ähnlichen diagnostischen Maßen (Hair et al., 1998). In den genannten Grundvoraussetzungen aber liegt der entscheidende Unterschied zwischen den beiden Verfahren: Im Vergleich zur Diskriminanzanalyse ist die logistische Regression an weniger Prämissen geknüpft und somit als wesentlich robuster anzusehen. So setzt die Diskriminanzanalyse z.B. Normalverteilung der unabhängigen Variablen sowie gleiche Varianzen in den betrachteten Gruppen voraus, wohingegen für die logistische Regression solche Voraussetzungen nicht nötig sind. Allerdings hätte die Diskriminanzanalyse gegenüber der logistischen Regression den Vorteil, dass aufgrund des zugrunde liegenden linearen Modells für den berechneten Schwellenwert ein Konfidenzintervall gebildet werden kann, was seine Aussagekraft absichern würde. In einem ersten Schritt soll demnach das Vorliegen der genannten Voraussetzungen für die Diskriminanzanalyse geprüft werden, da ihr bei einem positiven Ergebnis dieser Überprüfung der Vorzug zu geben wäre.

Die ***Diskriminanzanalyse*** ist – wie oben bereits beschrieben – ein Verfahren, mit dessen Hilfe ein bestimmter Fall aufgrund von Merkmalen (unabhängigen Variablen) einer von zwei oder auch mehreren fest vorgegebenen Gruppen zugeordnet werden kann. Zentraler Teil der Diskriminanzanalyse ist die Aufstellung der so genannten Diskriminanzfunktion

$$d = b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_n \cdot x_n + a$$

Dabei sind x_1 bis x_n fallweise die Werte der einbezogenen Variablen, b_1 bis b_n, sowie die Konstante a die von der Analyse zu ermittelnden Koeffizienten[14]. Die Ermittlung der Koeffizienten soll in der Weise erfolgen, dass sich die Werte der Diskriminanzfunktionen beider Gruppen maximal unterscheiden und somit die Gruppen so gut wie möglich getrennt werden. Rechnerisch geschieht dies, indem die Varianz zwischen den Gruppen (between-group variance) gegenüber der Varianz innerhalb der Gruppen (within-group variance) maximiert wird (Hair et al., 1998).

Wie gut die Trennung gelungen ist, kann zum einen über eine Korrelation zwischen den berechneten Werten der Diskriminanzfunktion und der Gruppenzugehörigkeit überprüft werden (kanonische Korrelation) – je höher dieser Wert, desto besser. Darüber hinaus wird über die Testgröße Wilks' Lambda, die annähernd χ^2-verteilt ist, geprüft, ob sich die mittleren Werte der Diskriminanzfunktion in den beiden Gruppen signifikant unterscheiden.

Der Schwellenwert für die unabhängige Variable berechnet sich aus dem gewichteten Mittelwert der beiden Gruppen, das Konfidenzintervall erhält man über die gepoolte Varianz der Gruppen, mit deren Hilfe man die Standardabweichung des Schwellenwertes bestimmen kann.

Zur Überprüfung der Voraussetzungen für die Diskriminanzanalyse bieten sich folgende Testverfahren an: Die Frage, ob die unabhängige Variable „Anzahl der erfüllten Realkennzeichen" innerhalb der beiden empirischen Gruppen normalverteilt ist, kann hier mit Hilfe des Shapiro-Wilk-Tests sowie des Lilliefors-Tests geklärt werden, welcher eine Modifikation des Kolmogorov-Smirnov-Tests darstellt. Beide Verfahren testen die Nullhypothese, dass die in den Daten vorliegende Verteilung sich nicht von der Normalverteilung unterscheidet. Der Vergleich zweier Stichprobenvarianzen kann über den so genannten Levene-Test erfolgen, welcher von der Nullhypothese ausgeht, dass die beiden Varianzen gleich sind und mögliche Varianzunterschiede nur stichprobenbedingt bzw. zufällig sind. Sofern die genannten Verfahren für das Vorliegen der Voraussetzungen sprechen, soll mit den Daten eine Diskriminanzanalyse durchgeführt werden, ansonsten wäre eine logistische Regression angezeigt.

[14] Da im vorliegenden Fall nur eine unabhängige Variable in die Analyse einbezogen werden soll, würde die Diskriminanzfunktion hier verkürzt $d = b \cdot x + a$ lauten.

Bei der ***logistischen Regression*** werden über den Regressionsansatz die Gewichte bestimmt, mit denen die betrachteten Einflussgrößen als unabhängige Variablen die Wahrscheinlichkeit dafür beeinflussen, dass ein realer Fall zu einer bestimmten Gruppe gehört. Im Unterschied zur linearen Regressionsanalyse versucht die logistische Regression also nicht, Schätzungen für die Beobachtungen der binären abhängigen Variablen vorzunehmen, sondern die Eintrittswahrscheinlichkeiten dieser Beobachtungswerte abzuleiten und zwar unter Verwendung der logistischen Funktion. Die Wahrscheinlichkeit für das Eintreten des Ereignisses bei einem Fall wird dabei nach folgendem Ansatz berechnet:

$$p(y=1) = \frac{1}{1+e^{-z}} \qquad \text{mit} \qquad z = b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_n \cdot x_n + a$$

Die logistische Funktion stellt also eine Wahrscheinlichkeitsbeziehung zwischen einem bestimmten Ereignis y = 1 (hier: Fall wird in Gruppe *nicht glaubhaft* eingeordnet) und den unabhängigen Variablen x_n her[15]. Die durch die logistische Funktion geschätzten Regressionskoeffizienten b_n spiegeln dabei die Einflussstärke der jeweils betrachteten unabhängigen Variablen x_n auf die Höhe der Wahrscheinlichkeitsbeziehung wieder. Die Schätzung der Parameter b_n durch die logistische Funktion erfolgt in der Weise, dass die Wahrscheinlichkeit („Likelihood"), die in der Stichprobe beobachteten Erhebungsdaten zu erhalten, maximiert wird.

Zur Überprüfung der Güte des logistischen Modells kann auf verschiedene Arten von Gütekriterien zurückgegriffen werden. In der vorliegenden Arbeit sollen der Likelihood Ratio-Test, die Wald-Statistik und Nagelkerke-R^2 zur Abschätzung des model-fits verwendet werden, da sie gegenüber anderen Gütekriterien verschiedene Vorteile bieten.

Beim *Likelihood-Ratio-Test* wird das durch das Modell maximierte -2fache des logarithmierten Likelihood (-2LL) verglichen mit demjenigen -2LL-Wert, der sich ergibt, wenn alle Regressionskoeffizienten der unabhängigen Variablen auf Null gesetzt werden und nur noch der konstante Term betrachtet wird (sog. „Null-Modell"). Ist die absolute Differenz zwischen dem -2LL-Wert des Null-Modells und dem des vollständigen Modells klein, so tragen die unabhängigen Variablen anscheinend nur wenig zur Unterscheidung der betrachteten Gruppen bei (Backhaus et al., 2003).

[15] Auch hier kann die Gleichung aufgrund des Vorliegens nur einer unabhängigen Variablen zu $z = b \cdot x + a$ vereinfacht werden.

Der Likelihood-Ratio-Test testet also folgende Nullhypothese:

H_0: Alle Regressionskoeffizienten sind gleich Null, die unabhängigen Variablen haben keinen bedeutenden Einfluss.

H_1: Alle Regressionskoeffizienten sind ungleich Null, die unabhängigen Variablen haben einen bedeutenden Einfluss.

Als Testgröße dient die absolute Differenz zwischen dem -2LL des Null-Modells und dem des vollständigen Modells, die mit J Freiheitsgraden (J = Zahl der unabhängigen Variablen, also in diesem Fall J = 1) asymptotisch χ^2-verteilt ist. Ist die Differenz signifikant von Null verschieden, so muss die Nullhypothese verworfen werden.

Die Nullhypothese, dass nicht alle, sondern nur ein bestimmtes b_n gleich Null ist[16] und daher die zugehörige Variable keinen Einfluss auf die Trennung der Gruppen hat, kann durch die so genannte *Wald-Statistik* überprüft werden, welche ebenfalls asymptotisch χ^2-verteilt ist.

Nagelkerke-R^2 zählt zu den so genannten Pseudo-R^2-Statistiken, die versuchen, den Anteil der erklärten Varianz des logistischen Regressionsmodells zu quantifizieren. Dabei wird auch bei den Pseudo-R^2-Statistiken auf das Verhältnis zwischen dem logarithmierten Likelihood des Nullmodells und dem des vollständigen Modells zurückgegriffen. Laut Backhaus et al. (2003, S. 448) sind Werte ab $R^2 = 0.4$ als gut und Werte ab $R^2 = 0.5$ als sehr gut im Sinne der Varianzaufklärung durch das Modell zu interpretieren.

[16] Da nur eine unabhängige Variable betrachtet wird, messen der Likelihood-Ratio-Test und die Wald-Statistik in diesem Fall dasselbe.

5. ERGEBNISSE

5.1 Deskriptive Ergebnisse

Insgesamt konnten aus den 121 verwertbaren Gutachten die Aussagen von 138 Personen extrahiert werden, da wie unter 4.2.1 bereits erwähnt in 15 Gutachten die Aussagen mehrerer Zeugen einer Inhaltsanalyse unterzogen wurden.

Bei den aussagenden Zeugen handelte es sich in der überwiegenden Mehrheit (88%) um Mädchen oder Frauen, nur 17 der 138 Zeugen waren männlich. Das Alter der Zeugen zum Zeitpunkt der Aussage reichte von fünf bis 62 Jahren, wobei sich der Großteil der Zeugen im Pubertäts- bis frühen Erwachsenenalter befand – 72% von ihnen waren zwischen zwölf und 25 Jahren alt. Der Altersdurchschnitt betrug dementsprechend 17.32 Jahre mit einer Standardabweichung von 8.832.

Betrachtet man die Tatvorwürfe, die Anlass für die Erstellung der Gutachten und somit Gegenstand der jeweiligen Aussagen waren, zeigte sich, dass es sich fast ausschließlich um Straftaten gegen die sexuelle Selbstbestimmung (§§ 174 bis 184 StGB) handelte. In deutlich mehr als der Hälfte der Fälle (65%) erfolgte die Begutachtung wegen eines Verdachtes auf sexuellen Missbrauch, wobei unter sexuellem Missbrauch in der vorliegenden Arbeit die Tatbestände sexueller Missbrauch Schutzbefohlener nach § 174 StGB, sowie sexueller und schwerer sexueller Missbrauch von Kindern nach §§ 176 und 176a StGB zusammengefasst wurden. In der Häufigkeit folgten Vorwürfe der sexuellen Nötigung und der Vergewaltigung nach § 177 StGB, die mit jeweils 12% gleich oft vorkamen, sowie körperliche Misshandlung Schutzbefohlener nach § 225 StGB mit 7%. Weitere Tatvorwürfe sind Tabelle 7 zu entnehmen.

Tabelle 7: Tatvorwürfe in der Gutachten-Stichprobe

Tatvorwurf	Relative Häufigkeit[a]
Sexueller Missbrauch	65%
Sexuelle Nötigung	12%
Vergewaltigung	12%
Misshandlung	7%
Sexuelle Handlungen vor Kindern	2%
Gefährliche Körperverletzung	1%
Erpressung	1%
Beleidigung	1%

[a] Die Abweichung der Summe von 100% ist rundungsbedingt

Die mutmaßlichen Täter stammten in etwas mehr als der Hälfte der Fälle aus dem direkten familiären Umfeld der Zeugen, dagegen handelte es sich in 49% der Fälle um außerfamiliäre Tatverdächtige wie Nachbarn, Bekannte der Eltern oder Lehrer.

Von den 138 Aussagen wurden aufgrund der Inhaltsanalyse 57% als wahrscheinlich erlebnisfundiert (glaubhaft) und 43% als wahrscheinlich nicht erlebnisfundiert (nicht glaubhaft) beurteilt.

5.2 Trennschärfenanalysen und Itemselektion

Im Folgenden werden die Ergebnisse der Trennschärfenanalysen und Itemselektionen dargestellt. Die kritischen Korrelationen wurden basierend auf einem Signifikanzniveau von $\alpha < .05$ bzw. $\alpha < .01$ mit der unter 4.2.1 angegebenen Formel berechnet. Alle anderen Berechnungen und Analysen wurden mit SPSS, Version 12.0 durchgeführt.

5.2.1 Einbeziehung aller Items und aller Zeugen

Zunächst wurden alle 19 Realkennzeichen unter Berücksichtigung der Aussagen sämtlicher 138 Zeugen einer Trennschärfenanalyse unterzogen. Die Ergebnisse sind in Tabelle 8 zusammengefasst, welche analog zu Tabelle 1 bei Hommers (1997, S. 93) neben den Trennschärfen der Kriterien (r_{it}: part-whole-korrigierte Korrelation des Kriteriums mit der Summe der anderen) auch die Mittelwerte der Bewertung der einzelnen Kriterien, sowie die jeweiligen α-Koeffizienten bei Weglassen des betreffenden Kriteriums zeigt. Auf die Darstellung der Standardabweichung konnte in diesem Fall verzichtet werden, da die Bewertung der Kriterien nicht wie bei Hommers bzw. Steller et al. (1992) auf einer vierstufigen Skala, sondern nur dichotom mit „0“ oder „1“ erfolgte und bei dichotomen Items die Itemstreuung rechnerisch vollkommen durch die Itemschwierigkeit determiniert wird (Bühner, 2004). Aufgrund der dichotomen Bewertung ist der Mittelwert der Bewertungen hier auch gleichbedeutend mit der durchschnittlichen Häufigkeit des Auftretens des Items in den Aussagen bzw. mit dem Schwierigkeitsindex p in der klassischen Itemanalyse. Niedrige Werte für p stehen dementsprechend für ein seltenes Auftreten und im Sinne der klassischen Itemanalyse für eine hohe Schwierigkeit des jeweiligen Realkennzeichens.

Bei einem Signifikanzniveau von 1% und einer Stichprobengröße von 138 liegt die kritische Korrelation bei $r_{krit\ (df = 136;\ \alpha = 1\%)} = .198$, das heißt alle Trennschärfen, die größer als dieser Wert sind, sind auf dem 1%-Niveau signifikant. Für ein Signifikanzniveau von 5% gilt die kritische Korrelation von $r_{krit\ (df = 136;\ \alpha = 5\%)} = .141$.

Tabelle 8: Ergebnisse der Trennschärfenanalyse unter Berücksichtigung aller Items und aller Zeugen

Kriterium	p	r_{it}	α
1. Logische Konsistenz	.64	.705**	.826
2. Unstrukturierte Darstellung	.30	.471**	.838
3. Detailreichtum	.67	.677**	.828
4. Raum-zeitliche Verknüpfung	.59	.550**	.834
5. Interaktionsschilderungen	.55	.528**	.835
6. Gesprächswiedergaben	.41	.517**	.836
7. Handlungskomplikationen	.41	.482**	.838
8. Ausgefallene Details	.38	.315**	.846
9. Nebensächliche Details	.28	.283**	.847
10. Unverstandene Handlungen	.15	.293**	.845
11. Indirekt Handlungsbezogenes	.01	.063	.850
12. Eigenpsychisches	.60	.593**	.832
13. Fremdpsychisches	.26	.548**	.835
14. Spontane Verbesserung	.09	.291**	.845
15. Eingestehen von Erinnerungslücken	.14	.212**	.848
16. Einwände gegen eigene Aussage	.04	.188*	.848
17. Selbstbelastung	.40	.408**	.841
18. Täterentlastung	.44	.350**	.844
19. Deliktspezifisches	.41	.538**	.835

Anmerkungen: p = Schwierigkeit/Auftretenshäufigkeit des jeweiligen Merkmals, r_{it} = part-whole korrigierte Item-Summenscore-Korrelation, α = Cronbachs Alpha der Summe der restlichen 18 Kriterien bei Weglassen des jeweiligen Kriteriums

* $p < .05$ bei $r_{krit} = .141$ mit $t_{(df = 136;\ \alpha = 5\%)} = 1.656$
** $p < .01$ bei $r_{krit} = .198$ mit $t_{(df = 136;\ \alpha = 1\%)} = 2.354$

Das Cronbachs Alpha als Maß für die innere Konsistenz und damit Gesamt-Reliabilität der Skala der 19 Realkennzeichen betrug $\alpha = .847$. Im Übrigen lassen sich die Ergebnisse der Trennschärfenanalyse wie folgt zusammenfassen: Die Trennschärfen variierten von $r_{it} = .063$ beim Kriterium *Indirekt Handlungsbezogenes* bis $r_{it} = .705$ beim Kriterium *Logische Konsistenz.* Der Mittelwert der Trennschärfen betrug $r_{it} = .422$ und unterschied sich somit nicht wesentlich vom Trennschärfen-Mittelwert bei Hommers, der mit $r_{it} = .41$ angegeben wurde. Allerdings war dort ein Realkennzeichen durch eine negative Trennschärfe aufgefallen, was im vorliegenden Fall nicht vorzufinden war. Dagegen stachen hier die sehr geringe

und auch auf dem 5%-Niveau nicht signifikante Trennschärfe und das extrem seltene Auftreten, d.h. im Sinne der Itemanalyse die hohe Schwierigkeit, des Merkmales *Indirekt Handlungsbezogenes* heraus. Bis auf *Einwände gegen die eigene Aussage* erwiesen sich ansonsten alle Merkmale auf dem 1%-Niveau als trennscharf.

Insgesamt lagen die Schwierigkeiten der Items zwischen p = .01 bei *Indirekt Handlungsbezogenes* und p = .67 bei *Detailreichtum*, der Mittelwert lag bei p = .36. Die Mehrheit der Merkmale kam also eher selten vor, lediglich fünf der 19 Merkmale waren in mehr als der Hälfte der Aussagen zu finden.

Die α-Werte bei Weglassen des jeweiligen Kriteriums schwankten zwischen α = .826 und α = .850, eine bedeutsame Steigerung der Reliabilität war also durch das Weglassen eines einzelnen Realkennzeichens nicht zu erreichen. Allerdings veränderte sich auch durch die Selektion der fünf Realkennzeichen mit den höchsten Werten für α bei Weglassen des jeweiligen Kriteriums die Gesamt-Reliabilität nur geringfügig und ging insgesamt über α = .854 nicht hinaus.

Tabelle 9: Cronbachs Alpha und Anzahl der verbleibenden Items auf verschiedenen Stufen der Itemselektion unter Berücksichtigung aller Items und aller Zeugen

	(Zusätzlich) Selektiertes Item	Cronbachs α	Anzahl Items
Ausgangswert	-	.847	19
1. Selektionsschritt	11. Indirekt Handlungsbezogenes	.850	18
2. Selektionsschritt	16. Einwände gegen eigene Aussage	.851	17
3. Selektionsschritt	15. Eingestehen von Erinnerungslücken	.852	16
4. Selektionsschritt	9. Nebensächliche Details	.853	15
5. Selektionsschritt	14. Spontane Verbesserung	.854	14

Anmerkung: Weggelassen wurde jeweils das Item, bei dem sich der Wert für Cronbachs Alpha für die restlichen Kriterien dadurch am meisten erhöhte.

In obiger Tabelle 9 sind die genauen Ergebnisse der Itemselektion festgehalten, wobei jeweils nur das Realkennzeichen aufgeführt ist, welches im genannten Selektionsschritt erstmals weggelassen wurde; die zuvor bereits selektierten Kriterien sind selbstverständlich ebenfalls nicht mehr berücksichtigt, was auch an der Anzahl der verbleibenden Items ersichtlich ist. Die Selektion wurde nach dem fünften Selektionsschritt beendet, da eine weitere Optimierung von α durch Weglassen anderer Items nicht mehr möglich war.

5.2.2 Trennschärfenanalyse ohne motivationsbezogene Realkennzeichen

Bei einer zweiten Trennschärfenanalyse wurden mit Ausnahme des Merkmals *Spontane Verbesserung* die motivationsbezogenen Realkennzeichen ausgeschlossen. Insgesamt flossen also in die in Tabelle 10 dargestellte Trennschärfenanalyse nur 15 Realkennzeichen ein, wobei gegenüber der ersten Analyse die Kriterien *Eingestehen von Erinnerungslücken, Einwände gegen die eigene Aussage, Selbstbelastung* und *Entlastung des Täters* fehlten. Da sich der Stichprobenumfang der Aussagen mit n = 138 nicht von dem der ersten Analyse unterschied, fanden auch hier die kritischen Korrelationen von r_{krit} = .141 für α < .05 und .198 für α < .01 Anwendung.

Tabelle 10: Ergebnisse der Trennschärfenanalyse ohne die Realkennzeichen 15 bis 18

Kriterium	p	r_{it}	α
1. Logische Konsistenz	.64	.706**	.819
2. Unstrukturierte Darstellung	.30	.468**	.834
3. Detailreichtum	.67	.681**	.821
4. Raum-zeitliche Verknüpfung	.59	.588**	.827
5. Interaktionsschilderungen	.55	.511**	.832
6. Gesprächswiedergaben	.41	.496**	.833
7. Handlungskomplikationen	.41	.451**	.836
8. Ausgefallene Details	.38	.349**	.842
9. Nebensächliche Details	.28	.285**	.845
10. Unverstandene Handlungen	.15	.306**	.842
11. Indirekt Handlungsbezogenes	.01	.056	.848
12. Eigenpsychisches	.60	.601**	.826
13. Fremdpsychisches	.26	.552**	.830
14. Spontane Verbesserung	.09	.258**	.844
19. Deliktspezifisches	.41	.548**	.829

Anmerkungen: p = Schwierigkeit/Auftretenshäufigkeit des jeweiligen Merkmals, r_{it} = part-whole korrigierte Item-Summenscore-Korrelation, α = Cronbachs α der Summe der restlichen 14 Kriterien bei Weglassen des jeweiligen Kriteriums

* p < .05 bei r_{krit} = .141 mit $t_{(df = 136;\ \alpha = 5\%)}$ = 1.656
**p < .01 bei r_{krit} = .198 mit $t_{(df = 136;\ \alpha = 1\%)}$ = 2.354

Entgegen den Erwartungen lag die Gesamt-Reliabilität dieser verkürzten Skala unter der der vollständigen Skala und betrug α = .844. Die Trennschärfen lagen zwischen r_{it} = .056 für *Indirekt Handlungsbezogenes* und r_{it} = .706 für das Merkmal *Logische Konsistenz* und waren bis auf das Merkmal *Indirekt Handlungsbezogenes* allesamt auf dem 1%-Niveau signifikant. Der Mittelwert der Trennschärfen betrug r_{it} = .457 und war damit etwas höher als bei Einbeziehung aller Realkennzeichen. Die um das jeweilige Kriterium reduzierten

Alpha-Werte variierten zwischen α = .819 und α = .848. Durch eine weitere Selektion von Realkennzeichen mit geringen Trennschärfen (*Indirekt Handlungsbezogenes, Nebensächliche Details, Spontane Verbesserung der eigenen Aussage*) konnte für die Reliabilität ein Wert von α = .852 erreicht werden. Dieser Wert liegt aber immer noch unter dem bei der vorausgegangenen Itemselektion maximal erreichten, insgesamt erbrachte das Weglassen der motivationsbezogenen Kriterien also keine Steigerung der Gesamt-Reliabilität.

5.2.3 Getrennte Trennschärfenanalysen nach Altersgruppen

Um herauszufinden, ob die Reliabilität der Realkennzeichen auch abhängig vom Alter der Zeugen zum Zeitpunkt der Exploration ist, wurden Trennschärfenanalysen getrennt für Zeugen über bzw. unter einer bestimmten Altersgrenze berechnet („Alters-Cut"). Ausführlicher dargestellt werden sollen an dieser Stelle nur die Ergebnisse für einen Alters-Cut bei 14 Jahren, da hier die Gruppen annähernd gleich groß und die Unterschiede hinsichtlich der Reliabilität im Vergleich zu einem Cut bei 13 oder bei 15 Jahren relativ stark ausgeprägt waren. Die Gruppengrößen und Cronbachs Alpha für andere Möglichkeiten zur Splittung der Stichprobe hinsichtlich des Alters finden sich in Tabelle 11.

Tabelle 11: Gruppengröße und Cronbachs Alpha für unterschiedliche Splittungen der Stichprobe nach dem Alter

Alters-Cut bei…		**n**	**α**
12 Jahren	Zeugen ≤ 12	33	.815
	Zeugen > 12	105	.857
13 Jahren	Zeugen ≤ 13	50	.848
	Zeugen > 13	88	.849
14 Jahren	**Zeugen ≤ 14**	**66**	**.852**
	Zeugen > 14	**72**	**.843**
15 Jahren	Zeugen ≤ 15	81	.849
	Zeugen > 15	57	.841

Anmerkungen: n = Gruppengröße, α = Cronbachs Alpha als Gesamt-Reliabilität der Skala

Wie erkennbar ergaben sich durch die Aufteilung der Stichprobe bei einem Alter von 14 Jahren zwei Teilstichproben, die 72 bzw. 66 Zeugen umfassten. In der jüngeren Teilstichprobe lag das Alter im Durchschnitt bei 11.74 Jahren mit einer Standardabweichung von 2.207; 54 der 66 Zeugen (82%) waren weiblich. Der Tatvorwurf lautete mit 83% zum überwiegenden Teil sexueller Missbrauch, gefolgt von Misshandlung Schutzbefohlener mit

6%. In der Teilstichprobe der über 14-jährigen waren dagegen nur noch 49% sexuelle Missbrauchsfälle enthalten, hier spielten sexuelle Nötigung (21%) und Vergewaltigung (19%) eine größere Rolle. Das Durchschnittsalter der Zeugen lag in dieser Gruppe bei 22.43 mit einer Standardabweichung von 9.521. Diese im Vergleich zur jüngeren Teilstichprobe deutlich höhere Standardabweichung lässt sich leicht durch die weitaus größere Varianz der Alterswerte in dieser Gruppe erklären, die zwischen 15 und 62 Jahren lagen.

Die Ergebnisse der Trennschärfenanalysen getrennt für die beiden Altersgruppen (14 Jahre und jünger bzw. über 14 Jahre) sind in Tabelle 12 aufgeführt. Für die jüngere Stichprobe mit einem Umfang von 66 Zeugen ergab sich bei $\alpha < .05$ eine kritische Korrelation von $r_{krit\ (df = 64;\ \alpha = 5\%)} = .204$ und $r_{krit\ (df = 64;\ \alpha = 1\%)} = .286$ bei $\alpha < .01$. Für die ältere Stichprobe, die 72 Zeugen umfasste, war $r_{krit\ (df = 70;\ \alpha = 5\%)} = .195$ bei $\alpha < .05$ und $r_{krit\ (df = 70;\ \alpha = 1\%)} = .274$ bei $\alpha < .01$.

Tabelle 12: Ergebnisse der Trennschärfenanalysen getrennt nach Zeugen unter und über 14 Jahren

	Zeugen ≤ 14 Jahre (n = 66)			**Zeugen > 14 Jahre (n = 72)**		
Kriterium	**p**	**r_{it}**	**α**	**p**	**r_{it}**	**α**
1. Logische Konsistenz	.59	.779**	.828	.69	.633**	.826
2. Unstrukturierte Darstellung	.35	.432**	.846	.26	.536**	.831
3. Detailreichtum	.68	.720**	.832	.67	.646**	.825
4. Raum-zeitliche Verknüpfung	.58	.725**	.831	.61	.395**	.838
5. Interaktionsschilderungen	.55	.427**	.846	.56	.623**	.826
6. Gesprächswiedergaben	.38	.487**	.843	.44	.540**	.830
7. Handlungskomplikationen	.38	.579**	.838	.44	.393**	.838
8. Ausgefallene Details	.38	.380**	.848	.39	.258*	.845
9. Nebensächliche Details	.29	.193	.856	.28	.371**	.839
10. Unverstandene Handlungen	.18	.402**	.847	.13	.200*	.844
11. Indirekt Handlungsbezogenes	.00	.000	.855	.03	.075	.846
12. Eigenpsychisches	.58	.598**	.837	.63	.587**	.828
13. Fremdpsychisches	.20	.488**	.843	.32	.593**	.828
14. Spontane Verbesserung	.06	.343**	.849	.11	.252*	.843
15. Eingestehen von Erinnerungslücken	.06	.142	.854	.21	.248*	.844
16. Einwände gegen eigene Aussage	.02	.223*	.853	.07	.177	.844
17. Selbstbelastung	.30	.282*	.852	.49	.511**	.832
18. Täterentlastung	.45	.354**	.850	.43	.353**	.840
19. Deliktspezifisches	.44	.570**	.839	.38	.528**	.831

Anmerkungen: p = Schwierigkeit/Auftretenshäufigkeit des jeweiligen Merkmals, r_{it} = part-whole korrigierte Item-Summenscore-Korrelation, α = Cronbachs Alpha der Summe der restlichen 18 Kriterien bei Weglassen des jeweiligen Kriteriums

* $p < .05$ bei $r_{krit} = .204$ mit $t_{(df = 64;\ \alpha = 5\%)} = 1.669$ für die Jüngeren und $r_{krit} = .195$ mit $t_{(df = 70;\ \alpha = 5\%)} = 1.667$ für die Älteren
** $p < .01$ bei $r_{krit} = .286$ mit $t_{(df = 64;\ \alpha = 1\%)} = 2.386$ für die Jüngeren und $r_{krit} = .274$ mit $t_{(df = 70;\ \alpha = 1\%)} = 2.381$ für die Älteren

Die Auftretenshäufigkeit, d.h. die Schwierigkeit der Merkmale lag in der jüngeren Stichprobe der unter 14-jährigen zwischen p = .00 und p = .68 mit einem Mittelwert von p = .34, in der Stichprobe der über 14-jährigen zwischen p = .03 und p = .69 bei einem Mittelwert von p = .38. Die Aussagen der älteren Zeugen enthielten also tendenziell mehr Realkennzeichnen als die der jüngeren.

Da das Realkennzeichen *Indirekt Handlungsbezogenes* in der Gruppe der unter 14-jährigen Zeugen überhaupt nicht vorkam, betrug seine Trennschärfe r_{it} = .000; abgesehen davon zeigten die Merkmale *Eingestehen von Erinnerungslücken* mit r_{it} = .142 und *Nebensächliche Details* mit r_{it} = .193 die geringsten Trennschärfen, beide wurden auch auf dem 5%-Niveau nicht signifikant. Das Signifikanzniveau von 1% verfehlten die Realkennzeichen *Einwände gegen die eigene Aussage* und *Selbstbelastungen.* Die höchste Trennschärfe erreichte das Merkmal *Logische Konsistenz* mit einem hochsignifikanten Wert von r_{it} = .779.

Bei den über 14-jährigen Zeugen schwankten die Trennschärfen zwischen r_{it} = .075 und r_{it} = .646, wobei hier erstmals nicht das Merkmal *Logische Konsistenz* die höchste Trennschärfe aufwies, sondern das Merkmal *Detailreichtum*. Die geringste und neben der des Merkmals *Einwände gegen die eigene Aussage* als einzige nicht signifikante Trennschärfe zeigte allerdings wiederum das Merkmal *Indirekt Handlungsbezogenes*. Als lediglich auf dem 5%-Niveau trennscharf erwiesen sich vier Merkmale, nämlich *Ausgefallene Details*, *Unverstandene Handlungen*, *Spontane Verbesserungen* sowie *Eingestehen von Erinnerungslücken*. Im Durchschnitt lagen die Trennschärfen in der Gruppe der Zeugen unter 14 mit r_{it} = .428 ganz leicht über dem der Gruppe der über 14-jährigen von r_{it} = .417.

Der maximale Wert für Cronbachs Alpha bei Weglassen eines Merkmales erreichte α = .856 bei *Nebensächliche Details* für die jüngere und α = .846 bei *Indirekt Handlungsbezogenes* für die ältere Zeugengruppe. Cronbachs Alpha als Maß für die Gesamt-Reliabilität der Realkennzeichen betrug – wie bereits in Tabelle 11 dargestellt – für die unter 14-jährigen α = .852 und war damit um .009 höher als für die älteren Zeugen, sowie um .005 höher als für die Gesamtstichprobe.

Obwohl sich weder hinsichtlich der Gesamt-Reliabilität noch im Mittelwert der Schwierigkeiten oder der Trennschärfen extrem auffällige Abweichungen zwischen den beiden Altersgruppen ergaben, fielen bei genauerer Betrachtung der einzelnen Merkmale einige Besonderheiten auf. Hinsichtlich der Auftretenshäufigkeit bzw. Schwierigkeit der Items zeigten sich vor allem bei den motivationsbezogenen Merkmalen Altersunterschiede.

Diese kamen mit Ausnahme der *Täterentlastung* in den Aussagen der über 14-jährigen Zeugen zum Teil über doppelt so häufig vor wie in denen der jüngeren Zeugen, wenn auch auf insgesamt eher niedrigem Niveau. Ebenfalls sehr viel häufiger in den Aussagen der älteren Zeugen zu finden war das Merkmal *Fremdpsychisches*, wohingegen eine *Unstrukturierte Darstellungsweise* eher für die Aussagen jüngerer Zeugen kennzeichnend war.

Bezüglich der Trennschärfen zeigten sich neben der bereits erwähnten Tatsache, dass bei den älteren Zeugen erstmals das Merkmal *Detailreichtum* und nicht *Logische Konsistenz* die höchste Trennschärfe insgesamt erreichte, besonders bei *Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Handlungskomplikationen, Nebensächliche Details, Unverstandene Handlungen* und *Selbstbelastung* klare Unterschiede zwischen den über und den unter 14-jährigen. Während sich *Logische Konsistenz, Raum-zeitliche Verknüpfungen, Handlungskomplikationen* und *Unverstandene Handlungen* bei den jüngeren Kindern als besonders trennscharf erwiesen, erreichten die Merkmale *Interaktionsschilderungen*, *Nebensächliche Details* und *Selbstbelastung* bei den über 14-jährigen Jugendlichen und Erwachsenen deutlich höhere Trennschärfen.

Entsprechend fielen auch die in den Tabellen 13 und 14 getrennt voneinander dargestellten Ergebnisse der für beide Altersgruppen durchgeführten Itemselektionen zur Optimierung der Reliabilität unterschiedlich aus. Zwar waren in beiden Gruppen auch gleiche Merkmale von der Selektion betroffen, allerdings trug z.B. das Weglassen des Merkmals *Unverstandene Handlungen* nur bei den älteren Zeugen zu einer Reliabilitätssteigerung bei, während es bei den jüngeren zu den trennschärfsten Kriterien zählte. Genau umgekehrt verhielt es sich mit dem Merkmal *Nebensächliche Details*.

Tabelle 13: Cronbachs Alpha und Anzahl der verbleibenden Items auf verschiedenen Stufen der Itemselektion für Zeugen ≤ 14 Jahre

	(Zusätzlich) Selektiertes Item	Cronbachs α	Anzahl Items
Ausgangswert		.852	19
1. Selektionsschritt	9. Nebensächliche Details	.856	18
2. Selektionsschritt	11. Indirekt Handlungsbezogenes	.859	17
3. Selektionsschritt	15. Eingestehen von Erinnerungslücken	.862	16
4. Selektionsschritt	16. Einwände gegen eigene Aussage	.864	15
5. Selektionsschritt	17. Selbstbelastung	.866	14
6. Selektionsschritt	18. Täterentlastung	.868	13
7. Selektionsschritt	14. Spontane Verbesserung	.869	12

Tabelle 14: Cronbachs Alpha und Anzahl der verbleibenden Items auf verschiedenen Stufen der Itemselektion für Zeugen > 14 Jahre

	(Zusätzlich) Selektiertes Item	Cronbachs α	Anzahl Items
Ausgangswert		.843	19
1. Selektionsschritt	11. Indirekt Handlungsbezogenes	.846	18
2. Selektionsschritt	8. Ausgefallene Details	.848	17
3. Selektionsschritt	10. Unverstandene Handlungen	.850	16
4. Selektionsschritt	16. Einwände gegen eigene Aussage	.852	15
5. Selektionsschritt	15. Eingestehen von Erinnerungslücken	.853	14
6. Selektionsschritt	14. Spontane Verbesserung	.856	13

Anmerkung: Weggelassen wurde bei beiden Itemselektionen jeweils das Item, bei dem sich der Wert für Cronbachs Alpha für die restlichen Kriterien dadurch am meisten erhöhte

Wie ersichtlich konnten durch Weglassen der sieben am wenigsten trennscharfen Items für die Gruppe der unter 14-jährigen Zeugen für Cronbachs Alpha ein Wert von $\alpha = .869$ erreicht werden. Bei den über 14-jährigen wurde ein Wert von $\alpha = .856$ erzielt, wobei hierfür sechs Items eliminiert werden mussten.

5.2.4 Trennschärfenanalyse für die Fälle mit dem Tatvorwurf sexueller Missbrauch

Da die Realkennzeichen ursprünglich für die Analyse von Aussagen über sexuellen Missbrauch entwickelt worden waren, wurden in eine letzte Trennschärfenanalyse nur diejenigen Fälle einbezogen, in denen der Tatvorwurf auf sexuellen Missbrauch oder schweren sexuellen Missbrauch nach §§ 174, 176 oder 176a StGB lautete.

Die Stichprobe umfasste hier 90 Zeugenaussagen, wobei 80 (89%) von weiblichen und 10 (11%) von männlichen Zeugen stammten. Das Alter der Zeugen zum Zeitpunkt der Aussage war mit 15.27 im Durchschnitt etwas niedriger als in der Gesamtstichprobe und schwankte zwischen 5 und 37 Jahren (SD = 6.127). Das Verhältnis von Aussagen, die aufgrund der Inhaltsanalyse als glaubhaft beurteilt wurden, zu denjenigen, die als nicht glaubhaft beurteilt wurden, betrug ähnlich wie in der Gesamtstichprobe 56 zu 44 Prozent.

Für das Signifikanzniveau von 5% und einem Stichprobenumfang von n = 90 berechnete sich für die kritische Korrelation ein Wert von $r_{krit\ (df = 88;\ \alpha = 5\%)} = .174$. Für $\alpha < .01$ wurde die kritische Korrelation $r_{krit\ (df = 88;\ \alpha = 1\%)} = .245$.

Wie in Tabelle 15 zu sehen ist, variierten die Schwierigkeiten der Items auch hier wie in der Gesamtstichprobe zwischen p = .01 und p = .67, der Mittelwert betrug p = .35. Die Trennschärfen der Realkennzeichen lagen mit r_{it} = .438 im Mittel leicht über denen für die Gesamtstichprobe und reichten von r_{it} = .060 bei *Indirekt Handlungsbezogenes* bis r_{it} = .755 bei *Logische Konsistenz*.

Bis auf zwei Ausnahmen waren die Trennschärfen aller Merkmale zumindest auf dem 5%-Niveau signifikant, nicht signifikant wurden *Eingestehen von Erinnerungslücken* und – wie auch in allen anderen Analysen – das Merkmal *Indirekt Handlungsbezogenes*. Auffällig waren die im Vergleich zur Gesamtstichprobe relativ hohen Trennschärfen der Merkmale *Unstrukturierte Darstellung*, *Raum-zeitliche Verknüpfungen*, *Unverstandene Handlungen* und *Selbstbelastung.*

Tabelle 15: Ergebnisse der Trennschärfenanalyse für die Fälle mit Tatvorwurf sexueller Missbrauch (n = 90)

Kriterium	p	r_{it}	α
1. Logische Konsistenz	.63	.755**	.839
2. Unstrukturierte Darstellung	.31	.556**	.848
3. Detailreichtum	.67	.721**	.840
4. Raum-zeitliche Verknüpfung	.58	.631**	.844
5. Interaktionsschilderungen	.53	.488**	.851
6. Gesprächswiedergaben	.34	.526**	.849
7. Handlungskomplikationen	.39	.547**	.848
8. Ausgefallene Details	.38	.292**	.860
9. Nebensächliche Details	.26	.287**	.859
10. Unverstandene Handlungen	.19	.409**	.854
11. Indirekt Handlungsbezogenes	.01	.060	.862
12. Eigenpsychisches	.54	.609**	.845
13. Fremdpsychisches	.23	.520**	.850
14. Spontane Verbesserung	.08	.260**	.859
15. Eingestehen von Erinnerungslücken	.07	.146	.861
16. Einwände gegen eigene Aussage	.01	.183*	.861
17. Selbstbelastung	.37	.333**	.858
18. Täterentlastung	.49	.407**	.855
19. Deliktspezifisches	.47	.601**	.846

Anmerkungen: p = Schwierigkeit/Auftretenshäufigkeit des jeweiligen Merkmals, r_{it} = part-whole korrigierte Item-Summenscore-Korrelation, α = Cronbachs Alpha der Summe der restlichen 18 Kriterien bei Weglassen des jeweiligen Kriteriums

* p < .05 bei r_{krit} = .174 mit $t_{(df = 88;\ \alpha = 5\%)}$ = 1.662
**p < .01 bei r_{krit} = .245 mit $t_{(df = 88;\ \alpha = 1\%)}$ = 2.369

Die Gesamt-Reliabilität der Realkennzeichen-Skala lag für die Fälle des sexuellen Missbrauchs bei α = .859. Cronbachs Alpha bei Weglassen eines einzelnen Merkmales nahm Werte zwischen α = .839 bei *Logische Konsistenz* und α = .862 bei *Indirekt Handlungsbezogenes* an; durch die Selektion mehrerer Items konnte sogar eine Reliabilität von α = .872 erreicht werden, allerdings würde die Kriteriologie dann nur noch zwölf Realkennzeichen umfassen. Von der Selektion betroffen sind neben *Indirekt Handlungsbezogenes* hauptsächlich die motivationalen Kriterien mit Ausnahme von *Täterentlastung.* Auch das Weglassen der Merkmale *Nebensächliche* und *Ausgefallene Details* trägt hier positiv zur Reliabilität der Skala bei (siehe Tabelle 16).

Tabelle 16: Cronbachs Alpha und Anzahl der verbleibenden Items auf verschiedenen Stufen der Itemselektion für die Fälle mit Tatvorwurf sexueller Missbrauch (n = 90)

	(Zusätzlich) Selektiertes Item	Cronbachs α	Anzahl Items
Ausgangswert		.859	19
1. Selektionsschritt	11. Indirekt Handlungsbezogenes	.862	18
2. Selektionsschritt	15. Eingestehen von Erinnerungslücken	.864	17
3. Selektionsschritt	16. Einwände gegen eigene Aussage	.867	16
4. Selektionsschritt	14. Spontane Verbesserung	.868	15
5. Selektionsschritt	9. Nebensächliche Details	.870	14
6. Selektionsschritt	8. Ausgefallene Details	.871	13
7. Selektionsschritt	17. Selbstbelastung	.872	12

Anmerkung: Weggelassen wurde jeweils das Item, bei dem sich der Wert für Cronbachs Alpha für die restlichen Kriterien dadurch am meisten erhöhte.

5.3 Bestimmung eines Schwellenwertes

Für die Bestimmung eines hypothetischen Schwellenwertes für die Anzahl der vorliegenden Realkennzeichen, welcher optimal zwischen den als erlebnisfundiert bzw. nicht erlebnisfundiert befundenen Aussagen trennt, kam aufgrund der binären und normalskalierten abhängigen Variable und der metrischen unabhängigen Variable sowohl eine Diskriminanzanalyse als auch eine logistische Regression in Frage. In einem ersten Schritt wurde zunächst das Vorliegen der für die Diskriminanzanalyse notwendigen, strengeren Voraussetzungen geprüft, da dieses Verfahren den Vorteil eines Konfidenzintervalls für den Schwellenwert bietet. Bei der Prüfung auf Normalverteilung ergab sich allerdings sowohl für die Gruppe der glaubhaften als auch für die der nicht glaubhaften Aussagen eine überzufällige Abweichung der Verteilungen der unabhängigen Variable von der Normalverteilung, die Nullhypothese musste also für beide Gruppen verworfen werden.

Tabelle 17: Tests auf Normalverteilung der unabhängigen Variablen „Anzahl der vorliegenden Realkennzeichen“

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistik	**df**	**Sig.**	**Statistik**	**df**	**Sig.**
glaubhaft	.165	79	.000	.957	79	.010
nicht glaubhaft	.140	59	.006	.939	59	.005

[a] Signifikanzkorrektur nach Lilliefors. Weitere Anmerkungen: df = Freiheitsgrade; Sig. = Signifikanz

Auch die Varianzen der unabhängigen Variable in beiden Stichproben unterscheiden sich im Levene-Test signifikant voneinander, wenn auch nur auf dem 5%-Niveau ($df_1 = 1$, $df_2 = 136$; $p = .017$); die Voraussetzung der Varianzhomogenität kann dennoch ebenfalls nicht als erfüllt angesehen werden.

Tabelle 18: Test auf Homogenität der Varianzen der unabhängigen Variablen „Anzahl der vorliegenden Realkennzeichen“

	Levene-Statistik	**df_1**	**df_2**	**Signifikanz**
Basiert auf dem Mittelwert	5.805	1	136	.017
Basiert auf dem Median	5.835	1	136	.017
Basierend auf dem Median mit angepassten df	5.835	1	134, 105	.017
Basiert auf dem getrimmten Mittel	6.072	1	136	.015

Anmerkungen: df = Freiheitsgrade

Aufgrund der fehlenden Voraussetzungen für die Diskriminanzanalyse wurden die Daten einer logistischen Regression unterzogen, wobei als abhängige Variable die Zugehörigkeit zu der Gruppe der als glaubhaft bzw. nicht glaubhaft beurteilten Aussagen diente, als Kovariate die Anzahl der als erfüllt angesehenen Realkennzeichen.

Die Gruppe der aufgrund der Inhaltsanalyse als glaubhaft beurteilten Aussagen umfasste n = 79, die der als nicht glaubhaft beurteilten Aussagen n = 59 Fälle. Im Durchschnitt enthielten die glaubhaften Aussagen 10.00 (SD = 2.402), die nicht glaubhaften Aussagen 2.53 (SD = 1.804) Realkennzeichen; dieser Unterschied erwies sich im t-Test mit $p < .001$ als hochsignifikant.

Tabelle 19: Variablen der logistischen Regressionsfunktion

	b	SE	Wald	df	Signifikanz
Summenscore	1.932	.479	16.257	1	.000
Konstante	-11.483	2.969	14.960	1	.000

Anmerkungen: b = Regressionskoeffizient, SE = Standardfehler, Wald = annähernd χ^2 -verteilter Wald-Koeffizient, df = Freiheitsgrade; Sig. = Signifikanz

Für die Funktion der logistischen Regression konnten folgende in Tabelle 19 ersichtliche Werte berechnet werden: Für die unabhängige Variable „Summenscore der Realkennzeichen“ ergab sich ein Regressionskoeffizient von b = 1.932 mit einem Standardfehler von SE = .479. Aus dem Wald-Koeffizienten von 16.257 ist zu schließen, dass b signifikant von Null verschieden ist und die unabhängige Variable somit sehr gut zwischen den beiden Gruppen trennt (df = 1; p < .001). Als Konstante wurde ein Wert von -11.483 berechnet, sodass sich folgende Funktion der logistischen Regression (siehe S. 57) ergibt:

$$p\,(y = \text{nicht glaubhaft}) = \frac{1}{1+e^{-z}} \qquad \text{mit } z = 1.932 \cdot x - 11.483$$

Mit Hilfe dieser ermittelten Funktion lässt sich anschließend der Schwellenwert bestimmen, es handelt sich dabei um den Wert für x (Anzahl der Realkennzeichen), bei dem die berechnete Wahrscheinlichkeit für die Zugehörigkeit zur Gruppe *nicht glaubhaft* erstmals unter .50 fällt und somit durch das Modell eine Klassifizierung in die Gruppe *glaubhaft* vorhergesagt wird. Dies ist bei x = 6 der Fall, da hier p (y = nicht glaubhaft) = .47, während die Wahrscheinlichkeit für x = 5 noch bei .86 liegt (vergleiche auch Klassifikation aller Fälle aufgrund des Regressionsmodells in Anhang B). Auf Grundlage des ermittelten logistischen Regressionsmodells würden also alle Fälle mit einer Anzahl von sechs oder mehr Realkennzeichen der Gruppe der glaubhaften Aussagen zugeordnet, alle Fälle mit weniger als sechs Realkennzeichen der Gruppe der nicht glaubhaften.

Die Anpassung des eben dargestellten Regressionsmodells an die Daten war insgesamt als sehr gut zu bezeichnen. Auf Grundlage des Schwellenwertes von sechs Realkennzeichen konnten 97.1% der Fälle richtig klassifiziert werden, nur bei vier Fällen wich die vom Modell vorhergesagte Klassifikation von der tatsächlichen ab (vgl. Tabelle 20). Drei dieser Fälle wurden dabei fälschlicherweise aufgrund der enthaltenen Anzahl an Realkennzeichen (in einem Fall sechs und in zwei Fällen sieben Realkennzeichen) als glaubhaft klassifiziert, obwohl das Urteil der Gutachter auf nicht glaubhaft lautete.

Tabelle 20: Klassifikationsmatrix aufgrund des logistischen Regressionsmodells

	Vorhergesagt durch logistische Regression		
Beobachtet in Gutachten	**Glaubhaft**	**Nicht glaubhaft**	**Richtig klassifiziert**
Glaubhaft	78	1	98.7 %
Nicht glaubhaft	3	65	94.9 %
Gesamt			**97.1 %**

Auch die übrigen berechneten Gütemaße deuteten auf eine gute Aufklärung der vorliegenden Daten durch das Modell hin. Durch die Hinzunahme der unabhängigen Variable verbessert sich der -2LL-Wert um 166.45 gegenüber dem Null-Modell, das nur die Konstante enthält, auf 21.949; dieser Wert war im χ^2-Test hochsignifikant (df = 1, p < .001). Die gute Anpassung des Modells wurde auch durch die Pseudo-R^2-Statistik nach Nagelkerke bestätigt, die mit R^2 = .941 beachtlich hoch ausfiel. Demnach kann die Varianzaufklärung durch die logistische Regression näherungsweise auf 94,1% beziffert werden.

6. DISKUSSION

Im folgenden Abschnitt werden die im vorausgegangenen Kapitel ausführlich dargestellten Ergebnisse der Trennschärfenanalysen, Itemselektionen und der Schwellenwertberechnung nochmals im Überblick zusammengefasst und ihre Bedeutung erörtert. Im Anschluss folgt eine kritische Würdigung des gewählten methodischen Ansatzes sowie der Einschränkungen, die sich daraus für die Interpretation und Übertragbarkeit der vorliegenden Erkenntnisse ergeben. Abschließend folgen das Fazit und ein Ausblick auf mögliche weitere Forschungsansätze.

6.1 Zusammenfassung und Interpretation der Ergebnisse

6.1.1 Gesamt-Reliabilität der Realkennzeichen

Verglichen mit den Reliabilitäten, wie sie sich in vorausgegangenen Laborstudien ergeben hatten, konnte den Realkennzeichen nach Steller und Köhnken (1989) in der vorliegenden Arbeit mit Datenmaterial aus der Gutachtenpraxis eine sehr hohe Reliabilität im Sinne der inneren Konsistenz bescheinigt werden. Für alle 19 Realkennzeichen und über die Gesamtstichprobe von 138 Aussagen betrachtet lag sie bei $\alpha = .847$ (Cronbachs Alpha) und erreichte somit schon fast psychometrische Qualität. Hommers (1997) hatte für die von ihm betrachteten 16 Realkennzeichen ein Cronbachs Alpha von .77 berechnet, Lafrenz (2006), in deren Analyse 17 Merkmale eingingen, erhielt $\alpha = .566$.

Die sichtbaren Unterschiede in den Reliabilitäten können zum Teil wohl auf die im Vergleich zu den genannten experimentellen Studien hier sehr heterogene Stichprobe zurückgeführt werden, was zwangsläufig höhere Korrelationen erlaubt als eine geringe Varianz innerhalb der Stichprobe (Steck, 2006). Zum anderen steigt Cronbachs Alpha auch mit der Anzahl der im Test enthaltenen Items an (Bühner, 2004), sodass das Einbeziehen von 19 statt 16 oder 17 Realkennzeichen in die Analyse ebenfalls zu einem höheren Alpha-Wert geführt haben könnte. Dass darüber hinaus die Realkennzeichen-Kriteriologie nach Steller und Köhnken (1989) bei der Beurteilung von Aussagen zu realen Vorfällen durch ausgebildete Sachverständige zu zuverlässigeren Schlüssen führt, als dies im Labor unter experimentellen Aussagebedingungen und mit oft weniger erfahrenen Beurteilern der Fall ist, wäre zwar nahe liegend und in gewisser Weise auch wünschenswert, kann aufgrund einer einzigen Feld-

studie allerdings noch nicht beurteilt werden. Zumindest weist die hohe Reliabilität jedoch auf die Eindeutigkeit des den Realkennzeichen zugrunde liegenden Konstruktes der Erlebnisbasiertheit bei den Gutachtern hin. Das von Hommers (1997, S. 99) und Lafrenz (2006, S. 63) für die jeweils von ihnen betrachteten Stichproben gezogene Fazit, die Zusammenfassung der Realkennzeichen zu einer gemeinsamen Skala sei gerechtfertigt, erfährt demnach durch die vorliegenden Daten deutlich Unterstützung.

Durch die Selektion der fünf am wenigsten trennscharfen Items konnte die Reliabilität der Realkennzeichen für die Gesamtstichprobe nur geringfügig von $\alpha = .847$ auf $\alpha = .854$ verbessert werden. Insbesondere das Weglassen der motivationsbezogenen Merkmale mit Ausnahme von *Spontane Verbesserung der Aussage* erbrachte entgegen der ersten Vermutung und den Ergebnissen von Lafrenz (2006) keine Steigerung, sondern im Gegenteil eine Verminderung der Gesamt-Reliabilität der Skala. Aus Sicht des vorliegenden Datenmaterials erscheint daher eine Korrektur des derzeit gültigen Kataloges nach Steller und Köhnken (1989) insgesamt nicht notwendig, da er sich auch in seiner derzeitigen Zusammenstellung psychometrischer gut zur Unterscheidung von wahren und unwahren Aussagen eignet. Im Besonderen Maße trifft dies offenbar für die Aussagen von jüngeren Zeugen und Aussagen über sexuellen Missbrauch zu, die Reliabilitäten fielen für diese beiden Zeugengruppen mit $\alpha = .852$ bzw. $\alpha = .859$ noch höher aus, als für die Gesamtstichprobe. Auch für die über 14-jährigen Zeugen ergab sich allerdings noch eine gute Gesamt-Reliabilität von $\alpha = .834$.

6.1.2 Trennschärfenanalysen und Itemselektion

Betrachtet man die Trennschärfen der Realkennzeichen genauer, so waren diese im Durchschnitt über alle Aussagen als recht zufrieden stellend anzusehen, allerdings gab es zwischen einzelnen Merkmalen große Unterschiede. Acht Merkmale wiesen eine hohe Trennschärfe von über .50 auf, das heißt sie repräsentieren das zu Grunde liegende Konstrukt der Realitätsnähe besonders gut. Es handelte sich dabei um die Merkmale *Logische Konsistenz, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Gesprächswiedergaben, Eigenpsychisches, Fremdpsychisches* und *Deliktspezifisches*. Eine nach Fisseni (1997) immerhin noch mittlere Trennschärfe erreichten die Merkmale *Unstrukturierte Darstellung, Handlungskomplikationen, Ausgefallene Einzelheiten, Selbstbelastung* und *Täterentlastung*, sie können daher ebenfalls als Facetten des betrachteten Konstruktes Realitätsnähe gelten. Besonders schlecht im Sinne der Trennschärfe schnitt dagegen das

Merkmal *Indirekt Handlungsbezogenes* ab, aufgrund seiner extrem hohen Schwierigkeit konnte es kaum zur Unterscheidung zwischen glaubhaften und nicht glaubhaften Aussagen beitragen. Seine Selektion führte im Gegenteil sogar zu einer relativ deutlichen Erhöhung der Reliabilität. Da dieses Realkennzeichen auch in verschiedenen anderen Studien nur sehr selten in Aussagen gefunden wurde (z.B. Steller et al., 1992; Lamb et al., 1997; Hettler, 2005) stellt sich die Frage, ob sein seltenes Auffinden nicht durch die recht schwer verständliche und zum Teil bei verschiedenen Autoren auch uneinheitliche Definition (siehe S. 19, Fußnote 4) mitverursacht wird. In diesem Fall könnte durch eine Präzisierung der Definition dieses Merkmals eventuell eine Verbesserung seiner Trennschärfe erreicht werden.

Ebenfalls recht hohe Schwierigkeiten und geringe Trennschärfen zeigten einige der motivationsbezogenen Merkmale. Dies kann aber nicht auf den gesamten Komplex der motivationsbezogenen Merkmale generalisiert werden, da die beiden Merkmale *Selbstbelastung* und *Entlastung des Täters* mit Trennschärfen von r_{it} = .408 bzw. r_{it} = .350 und mittleren Schwierigkeitswerten vergleichsweise gut abschnitten. Dies ist vermutlich auch der Grund dafür, dass die Trennschärfenanalyse der Realkennzeichen ohne die motivationsbezogenen Merkmale außer *Spontane Verbesserung der eigenen Aussage* nicht zu der erwarteten Steigerung, sondern zu einer Minderung der Reliabilität der verbleibenden Skala führte. Aus der Sicht der vorliegenden Studie sind demnach im Gegensatz zu *Selbstbelastung* und *Entlastung des Täters* nur die drei motivationalen Merkmale *Eingestehen von Erinnerungslücken, Einwände gegen eigene Aussage* und *Spontane Verbesserung* kaum zur Unterscheidung zwischen glaubhaften und nicht glaubhaften Aussagen geeignet. Sie könnten mit positiven Auswirkungen auf die Gesamt-Reliabilität auch aus dem Katalog der Realkennzeichen gestrichen werden; da die zu erwartende Steigerung der Reliabilität jedoch nur sehr gering ausfallen würde, besteht zu einem solchen Schritt letztlich keine Notwendigkeit.

Bei geringer bis mittlerer Schwierigkeit waren auch die Realkennzeichen *Nebensächliche Details* und *Unverstandene Handlungen* insgesamt gesehen nur wenig trennscharf, beide waren aber zumindest für bestimmte Gruppen von Zeugen brauchbare Kriterien für die Unterscheidung zwischen glaubhaften und nicht glaubhaften Aussagen; mit dem Wissen um ihre differentielle Reliabilität scheint es also durchaus sinnvoll, sie im Katalog der Realkennzeichen zu belassen. So war die *Phänomengemäße Schilderung unverstandener Handlungselemente* vor allem – was aufgrund der Definition dieses Merkmals auch kaum überrascht – in den Aussagen von jüngeren Kindern und in Aussagen über angeblichen sexuellen

Missbrauch ein hochsignifikant zuverlässiges Kennzeichen für erlebnisfundierte Aussagen, wohingegen es bei jugendlichen und erwachsenen Zeugen einen sehr viel geringeren Hinweiswert aufwies. Die Schilderung von *Nebensächlichen Details* war umgekehrt bei der Beurteilung von Aussagen Jugendlicher und Erwachsener sogar auf dem 1%-Niveau signifikant trennscharf, während seine Trennschärfe bei den Zeugen unter 14 Jahren auch das 5%-Niveau der Signifikanz verfehlte. Dieser Befund entspricht den Ergebnissen von Lafrenz (2006), die bei ihrer Analyse einer Stichprobe mit erwachsenen Zeugen ebenfalls für das Merkmal *Überflüssige Details* – als eines der wenigen Merkmale – eine auf dem 5%-Niveau signifikante Trennschärfe errechnete.

Im Vergleich zwischen den Aussagen von jüngeren Zeugen unter 14 Jahren und jugendlichen bzw. erwachsenen Zeugen ab 15 Jahren zeigten sich bezüglich der Aussagekraft der Realkennzeichen weitere deutliche Unterschiede. Speziell für ältere Zeugen als trennscharf erwiesen sich neben *Nebensächliche Details* auch *Interaktionsschilderungen* und *Selbstbelastungen.* Darüber hinaus zeigte bei den erwachsenen Zeugen nicht wie bei den jüngeren Zeugen das Merkmal *Logische Konsistenz,* sondern *Detailreichtum* die höchste Trennschärfe, ein Ergebnis, das sich ebenfalls mit den Beobachtungen von Lafrenz (2006) deckt. Insgesamt wiesen die Realkennzeichen allerdings für die kindlichen Zeugen eine höhere Reliabilität auf, im Speziellen gilt dies für die Merkmale *Logische Konsistenz, Raum-zeitliche Verknüpfungen, Handlungskomplikationen* und, wie bereits erwähnt, *Unverstandene Handlungen*. Offensichtlich sind diese Realkennzeichen speziell in den Aussagen von jüngeren Kindern von großem Hinweiswert, da sie hier aufgrund der geringeren kognitiven Fähigkeiten der Kinder das zugrunde liegende Konzept Realitätsnähe besser zu repräsentieren scheinen als in den Aussagen von älteren Kindern und Erwachsenen. Hinsichtlich der übrigen Realkennzeichen ergaben sich keine nennenswerten Unterschiede zwischen den beiden Altergruppen, die Trennschärfen entsprechen in etwa den anhand der Gesamtstichprobe errechneten.

Entsprechend ihrer ursprünglichen Konzeption zeigten die Realkennzeichen mit $\alpha = .859$ eine besonders hohe Reliabilität für diejenigen Fälle, in denen die betrachteten Aussagen einen angeblichen sexuellen Missbrauch zum Thema hatten. In diesen Aussagen gewinnen vor allem die Merkmale *Unstrukturierte Darstellung, Raum-zeitliche Verknüpfungen, Handlungskomplikationen, Unverstandene Handlungen, Selbstbelastung* und *Deliktspezifisches* an Bedeutung. Daneben sind die bereits in der Gesamtstichprobe sehr trennscharfen

Merkmale *Logische Konsistenz, Detailreichtum, Gesprächswiedergaben,* sowie *Eigen-* und *Fremdpsychisches* auch hier zur Unterscheidung zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen gut geeignet.

6.1.3 Berechnung eines Schwellenwertes

Da aufgrund der fehlenden Voraussetzungen (Normalverteilung und Varianzhomogenität der unabhängigen Variablen in den beiden Gruppen) keine Diskriminanzanalyse durchführbar war, wurde der hypothetische Schwellenwert für die Anzahl der vorhandenen Realkennzeichen mit Hilfe einer logistischen Regression berechnet. Dabei ergab sich aus den vorliegenden Daten eine optimale Trennung zwischen den als erlebnisbasiert und den als nicht erlebnisbasiert bewerteten Aussagen bei einem Wert von sechs in der Aussage vorhandenen Realkennzeichen – bei sechs oder mehr Realkennzeichen erfolgte in der Regel eine Beurteilung als glaubhaft, mit fünf oder weniger Realkennzeichen als nicht glaubhaft. Auf der Grundlage dieses Schwellenwertes konnten 98% der von den Sachverständigen als glaubhaft beurteilten Aussagen sowie 95% der als nicht glaubhaft beurteilten Aussagen korrekt klassifiziert werden. Diese hohen Trefferquoten liegen deutlich über den von Maier (2006) berichteten und sprechen zusätzlich zu den übrigen Gütemaßen für eine sehr gute Erklärung der Daten durch die berechnete Regressionsgleichung.

Der im Vergleich zu Maier (2006) niedrig liegende Schwellenwert könnte wiederum auf die unterschiedliche Zusammensetzung der analysierten Stichproben zurückzuführen sein. Die Stichprobe in der experimentellen Studie von Maier bestand ausschließlich aus Studenten, denen man ein hohes Bildungsniveau und gute soziale Kompetenzen unterstellen kann, an die Aussagen dieser Zeugen ist also von Vornherein ein recht hoher Anspruch hinsichtlich der Qualität zu stellen (siehe auch Vrij et al., 2004). Dagegen stellen die in der forensischen Praxis begutachteten Zeugen, aus denen sich die Stichprobe der hier vorliegenden Arbeit zusammensetzt, eine sehr heterogene Gruppe hinsichtlich des Alters und der sozialen Schichten dar, wobei vermutet werden kann, dass die meisten Zeugen ein eher geringes Bildungsniveau aufweisen. Unter Berücksichtigung ihrer kognitiven, sprachlichen und sozialen Fähigkeiten genügen bei diesen Zeugen offensichtlich weniger erfüllte Merkmale, um ihre Aussage als glaubhaft bewerten zu können.

6.2 Kritische Würdigung des methodischen Vorgehens

Am in Kapitel 4 ausgeführten methodischen Vorgehen lassen sich einige Punkte kritisieren, welche die Übertragbarkeit der ausgeführten Ergebnisse einschränken könnten. Diese sollen im Folgenden diskutiert werden.

Zum einen könnte die Codierung der einzelnen Realkennzeichen als *vorhanden* oder *nicht vorhanden* durch die Autorin als subjektiv bezeichnet werden, da keine Objektivierung z.B. durch einen Zweitrater und die Berechnung von Interrater-Reliabilitäten vorgenommen wurde. Allerdings war im Zuge dieser Codierung ja keine eigene Einschätzung bezüglich des Vorliegens oder Nichtvorliegens der einzelnen Merkmale in der Aussage gefordert, sondern es musste lediglich die durch den Sachverständigen getroffene Entscheidung erfasst werden. Die Schwierigkeit bestand allein darin, bestimmte einschränkende Formulierungen (z.B. „das Merkmal XY ist zwar vorhanden, kann aber unter den gegebenen Umständen … nicht als Beleg für die Glaubhaftigkeit der Aussage gelten“) dahingehend zu deuten, ob das betreffende Merkmal nun als in der Aussage vorhanden gewertet werden sollte oder nicht. Durch die Festlegung und konsequente Anwendung expliziter Codierungsregeln (siehe 4.1.3) für solche Fälle konnte aber ein ausreichendes Maß an Objektivität gewährt werden.

Ähnlich subjektiv erscheinen kann auch die Auswahl einer bestimmten Teilaussage für die Stichprobe in den Fällen, in denen aufgrund des Individuierungsgebotes die Aussage eines Zeugen vom Sachverständigen in mehrere abgrenzbare Einzelsituationen unterteilt analysiert wurde. Aber auch hier erfolgte die Entscheidung stets auf der Grundlage der vom Sachverständigen getroffenen Feststellung, welche der vorliegenden Einzelaussagen als die umfangreichste und am meisten substantiierte gelten kann. Eine solche Feststellung lag in allen 16 betroffenen Gutachten vor.

Auch die Berechnung eines Schwellenwertes für die Anzahl der Realkenneichen, die in einer Aussage mindestens vorhanden sein müssen, um sie als glaubhaft bzw. erlebnisfundiert zu bezeichnen, mag zunächst fragwürdig erscheinen, zumal in der Fachwelt die Zulässigkeit der Anwendung eines universellen Schwellenwertes in der Kriterienorientierten Aussageanalyse generell verneint wird (siehe 2.6). Auf die Tatsache, dass der gefundene Schwellenwert lediglich zur Orientierung dienen kann und nicht als allgemeine Richtlinie generalisierbar ist, wurde jedoch mehrfach hingewiesen. Er spiegelt lediglich die Entscheidungspraxis der Sachverständigen wieder, deren Gutachten in die vorliegende Analyse eingegangen sind und ist schon allein deshalb nicht übertragbar, weil das Verfahren der logisti-

schen Regression nicht von der Normalverteilung der unabhängigen Variable ausgeht, sondern den Berechnungen direkt die in der Stichprobe vorgefundene Verteilung zugrunde legt, die aber keinesfalls der Verteilung in der Gesamtpopulation entsprechen muss.

Für eine deskriptive Wiedergabe der gängigen Entscheidungspraxis bei aussagepsychologischen Begutachtungen ist der errechnete Schwellenwert aber sehr wohl geeignet und aufgrund der durch das Aggregationsprinzip implizierten Notwendigkeit zur Festlegung eines solchen Wertes für die Entscheidung im Einzelfall auch als Orientierungswert bedeutsam. Durch den hohen Aufklärungswert des logistischen Modells (R^2 = .941) wird deutlich, dass die Sachverständigen in ihren Urteilen nur wenig vom berechneten Schwellenwert von sechs Realkennzeichen abwichen und dieser daher als sehr robust angesehen werden kann.

Weiterhin ist zu beachten, dass auch die von Köhnken (1990) kritisierte Selektivität der Stichprobenauswahl als allgemeines Problem der aussagepsychologischen Feldforschung natürlich auch auf die vorliegende Studie zutrifft. Nicht die Untersucher, sondern Staatsanwaltschaften und Gerichte entscheiden darüber, welche Fälle überhaupt begutachtet werden, weshalb eine Repräsentativität des Materials nicht unbedingt gegeben ist. Wie bereits in der Einleitung angesprochen erfolgt die Hinzuziehung eines aussagepsychologischen Sachverständigen nur unter sehr eng umgrenzten Bedingungen, wobei es sich in der Regel um Verfahren wegen Sexualdelikten handelt und fast immer Opferzeugen mit ihren Aussagen einen mutmaßlichen Täter belasten. Es ist demnach nicht auszuschließen, dass die in der vorliegenden Arbeit gefundenen Reliabilitäten und Trennschärfen der Realkenneichen nach Steller und Köhnken (1989) nur für den begrenzten Ausschnitt der betrachteten Fälle Gültigkeit besitzen und z.B. nicht auf die entlastende Aussage eines unbeteiligten Dritten in einem Zivilprozess übertragbar sind. Da die Beauftragung eines aussagepsychologischen Sachverständigen zu solchen Fragestellungen heute jedoch (noch) nicht üblich ist, bleibt fraglich, ob die Sicherstellung einer solchen Übertragbarkeit überhaupt sinnvoll und notwendig ist. Dagegen konnte in der vorliegenden Studie durch die Art der Stichprobenziehung sichergestellt werden, dass die vorgestellten Ergebnisse als repräsentativ für die in der täglichen aussagepsychologischen Gutachtenpraxis faktisch relevanten Anwendungsbereiche gelten können. Alle in einem Zeitraum von fast fünf Jahren bei einem renommierten Sachverständigeninstitut eingegangenen Gutachtenaufträge wurden in die Analyse aufgenommen, es erfolgte lediglich eine Selektion der Gutachten hinsichtlich der grundsätzlichen Voraussetzung für die Analyse (Vorliegen einer komplett durchgeführten CBCA).

6.3 Fazit und Ausblick

Als Fazit dieser Arbeit bleibt festzuhalten, dass die Realkennzeichen in der heute angewandten Form nach Steller und Köhnken (1989) insgesamt eine recht gute Reliabilität im Sinne der inneren Konsistenz erreichen, was dafür spricht, dass mit ihrer Hilfe die Messung eines gemeinsamen zugrunde liegenden Konstrukts, nämlich der Realitätsnähe einer Aussage, zuverlässig möglich ist. Gleichzeitig weisen die vorliegenden Daten aber auch darauf hin, dass diese Zuverlässigkeit noch gesteigert werden könnte, indem zum Beispiel das sehr selten auftretende und daher wenig zur Messung des Konstruktes Realitätsnähe beitragende Merkmal *Indirekt Handlungsbezogene Schilderungen* aus dem Katalog entfernt würde. Das Gleiche gilt für die motivationsbezogenen Merkmale *Spontane Verbesserung der eigenen Aussage*, *Eingestehen von Erinnerungslücken* und *Einwände gegen die Richtigkeit der eigenen Aussage*, denen neben ihrer geringen Trennschärfe und dem ohnehin seltenen Auftreten auch mangelnde Validität vorgeworfen werden muss. Allerdings sind die durch diese Itemselektion zu erwartenden Steigerungen der Reliabilität so gering, dass insgesamt eine Korrektur des vorliegenden Kataloges nicht notwendig erscheint. Allenfalls könnte eine Präzisierung der Definition des Merkmales *Indirekt handlungsbezogene Schilderungen* erwogen werden, um ein häufigeres Auffinden zu ermöglichen, seine Schwierigkeit zu senken und damit eventuell die Trennschärfe zu erhöhen.

Durchgängig für alle Gruppen von Zeugen als besonders trennscharf erwiesen sich die Merkmale *Logische Konsistenz, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionschilderungen, Wiedergabe von Gesprächen, Eigenpsychisches, Fremdpsychisches* und *Deliktspezifisches*. Diese Realkennzeichen dürfen aufgrund der vorliegenden Daten als beste Kennzeichen für die Realitätsnähe einer Aussage gelten und ihrem Vorliegen in einer Aussage somit besondere Bedeutung beigemessen werden. Daneben sollte im Einzelfall beachtet werden, dass einige Merkmale für bestimmte Gruppen von Zeugen besonders an Trennschärfe gewinnen oder verlieren, was sich auch auf die Gesamt-Reliabilität des Realkennzeichen-Katalogs speziell für diese Zeugengruppen auswirkt. Generell kann von einer besonders hohen Reliabilität für jüngere Zeugen und mutmaßliche Opfer von sexuellem Missbrauch ausgegangen werden.

So viel versprechend die vorliegenden Ergebnisse in Bezug auf die Zuverlässigkeit der aussagepsychologischen Methode auch zu sein scheinen, darf bei ihrer Interpretation eines nicht vergessen werden: Es wurde hier – wie in so gut wie allen Studien, die sich mit dieser Thematik beschäftigen – lediglich die Güte eines Bestandteiles der aussagepsychologischen Methode untersucht, nämlich die der Kriterienorientierten Inhaltsanalyse. Über die Verlässlichkeit und Genauigkeit des gesamten Begutachtungsprozesses mit all seinen beschriebenen Komponenten kann dagegen allein aufgrund der vorliegenden und ähnlicher Studien, die sich nur mit den Gütekriterien der Realkennzeichen beschäftigen, nicht geschlossen werden. Genau das wird, allerdings unter umgekehrten Vorzeichen, in einigen vorwiegend amerikanischen Publikationen (z.B. auch Lamb et al., 1997; Vrij et al., 2000; Pezdek et al., 2004; Vrij et al., 2004) dennoch getan. Die Autoren stellen hier unzulässigerweise aufgrund der einschränkenden Ergebnisse hinsichtlich der CBCA die forensische Brauchbarkeit der SVA insgesamt in Frage, indem diese im Wesentlichen auf die Anwendung der Realkennzeichen beschränkt wird. Steller und Volbert (1999) schreiben hierzu: „So finden sich in vielen englischsprachigen Publikationen im einleitenden Text durchaus Hinweise, daß die dort ‚Criteria-Based Content Analysis' (CBCA) genannte merkmalsorientierte Inhaltsanalyse lediglich ein Teil einer umfassenden und die spezifischen Voraussetzungen der Person berücksichtigenden Methode zur Beurteilung der Glaubhaftigkeit einer Aussage (‚Statement Validity Assessment' (SVA)) sei, dennoch wird in der Regel lediglich geprüft, inwieweit die merkmalsorientierte Inhaltsanalyse Testkriterien erfüllt, um auf der Basis dieser Ergebnisse Bewertungen abzugeben, inwieweit das gesamte Verfahren wissenschaftlich abgesichert ist, um vor Gericht angewandt zu werden" (S. 83). Um solchen unzulässigen Folgerungen zu begegnen ist es aus Sicht der Autorin notwendig, in zukünftigen Studien den Versuch zu unternehmen, die Validität des aussagepsychologischen Begutachtungsprozesses als Ganzes zu bestimmen. Da dabei allerdings sehr viele Faktoren zu berücksichtigen sind, bleibt die Realisierung einer solchen Studie wohl sehr schwierig.

Allerdings besteht auch auf der Ebene der Realkennzeichen durchaus noch Forschungsbedarf. Da sich in der vorliegenden Studie deutliche Unterschiede zwischen den Aussagen von Kindern und denen von Jugendlichen bzw. Erwachsenen gezeigt haben und von verschiedener Seite die verstärkte Hinzuziehung von aussagepsychologischen Sachverständigen zur Begutachtung Erwachsener gefordert wird (z.B. Greuel, 2001; Aymans, 2005), sollte auch an die Ausformulierung eines speziellen Realkennzeichenkataloges für Erwachsene

gedacht werden. Auf diese Weise könnte die Gesamt-Reliabilität der Realkennzeichen für erwachsene Zeugen optimiert werden, die sich in der Studie von Lafrenz (2006) als vergleichsweise niedrig erwies. Auch in der vorliegenden Studie fiel die Reliabilität der Gesamtskala für jugendliche und erwachsene Zeugen im Vergleich zu Kindern etwas ab. Für eine solche Modifizierung des Realkennzeichen-Kataloges ist es allerdings nötig, die Besonderheiten der Aussagen Erwachsener noch deutlicher herauszuarbeiten als es in der vorliegenden Arbeit sowie der Arbeit von Lafrenz (2006) bisher geschehen ist.

7. ZUSAMMENFASSUNG

Das System der Realkennzeichen nach Steller und Köhnken (1989) hat sich in Deutschland als wichtiger Bestandteil der aussagepsychologischer Begutachtungsmethodik etabliert und kann aufgrund zahlreicher Studien als empirisch gesichert angesehen werden. Analog zu zwei bereits existierenden Studien von Hommers (1997) und Lafrenz (2006) wurden die Realkennzeichen in der vorliegenden Arbeit testkritisch analysiert. Über Trennschärfenanalysen und die Berechnung von Cronbachs Alpha als Maß für die Reliabilität der Gesamtskala der Realkennzeichen sollte insbesondere untersucht werden, ob sie zuverlässig ein gemeinsames Konstrukt, nämlich die Realitätsnähe der analysierten Aussage, abbilden. Dies würde ihre Zusammenfassung zu einem Summenscore und somit auch eine psychometrische Anwendung legitimieren. Im Gegensatz zu den oben genannten Arbeiten handelte es sich hier um eine Feldstudie, die zugrunde liegenden Daten wurden also nicht experimentell erhoben, sondern stammen aus realen aussagepsychologischen Sachverständigengutachten. Diese wurden von der GWG – Gesellschaft für wissenschaftliche Gerichts- und Rechtspsychologie München zur Verfügung gestellt. Bei der Analyse aller 138 Aussagen, die aus den dort in den Jahren 2000 bis 2005 erstellten Gutachten entnommen werden konnten, ergab sich für die Gesamt-Reliabilität der 19 Realkennzeichen ein Cronbachs Alpha von $\alpha = .847$. Durch die Selektion von fünf wenig trennscharfen Items konnte dieser Wert geringfügig auf $\alpha = .854$ erhöht werden. Die Selektion aller motivationsbezogener Realkennzeichen bis auf *Spontane Verbesserung der eigenen Aussage* erbrachte dagegen, anders als bei Lafrenz (2006), keine Verbesserung der Reliabilität. Bei einer Differenzierung der Analysen nach bestimmten Zeugengruppen erwiesen sich die Realkennzeichen entsprechend ihrer ursprünglichen Konzeption besonders für jüngere Kindern und bei Aussagen über mutmaßlichen sexuellen Missbrauch als reliabel. Hinsichtlich der Trennschärfen der einzelnen Realkennzeichen zeigten *Logische Konsistenz, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionschilderungen, Wiedergabe von Gesprächen, Eigenpsychisches, Fremdpsychisches* und *Deliktspezifisches* durchgängig für alle Zeugengruppen hohe Werte. Diese Merkmale trugen also am meisten zur Reliabilität der Skala bei und repräsentieren somit das zugrunde liegende Konstrukt Realitätsnähe besonders gut. Als kaum repräsentativ für dieses Konstrukt und daher eher ungeeignet zur Unterscheidung zwischen glaubhaften und nicht glaubhaften Aussagen

stellte sich das Merkmal *Indirekt Handlungsbezogene Schilderungen*, sowie die drei motivationsbezogenen Realkennzeichen *Spontane Verbesserung der eigenen Aussage, Eingestehen von Erinnerungslücken* und *Einwände gegen die Richtigkeit der eigenen Aussage* heraus. Durch ihr Weglassen konnte die Reliabilität der Skala als Ganzes erhöht werden. Des Weiteren unterschieden sich die Trennschärfen einzelner Merkmale je nach betrachteter Altersgruppe, sodass insgesamt der Gedanke an die Formulierung einer leicht abgeänderten Version des Realkennzeichen-Kataloges speziell für erwachsene Zeugen nahe liegt. Mit Hilfe einer logistischen Regression konnte ein Schwellenwert von sechs in einer Aussage vorhandenen Realkennzeichen ermittelt werden, der in den vorliegenden Daten optimal zwischen glaubhaften und nicht glaubhaften Aussagen trennte. Da dieser Schwellenwert jedoch lediglich die Urteilspraxis der GWG-Sachverständigen widerspiegelt, kann er nur als Orientierungswert dienen, seine Anwendung im Sinne einer allgemeingültigen Entscheidungsregel ist unzulässig.

8. LITERATURVERZEICHNIS

ANSON, D. A., GOLDING, S. L. & GULLY, K. J. (1993). Child sexual abuse allegations: Reliability of criteria-based content analysis. *Law and Human Behavior, 17,* 331-341.

ARNOLD, W. (1952). Die Psychologische Begutachtung der Zeugentüchtigkeit und Glaubwürdigkeit bei Kindern und Jugendlichen. *Psychologische Rundschau, 3,* 265-280.

ARNTZEN, F. (1970). *Psychologie der Zeugenaussage. Einführung in die forensische Aussagepsychologie* (1. Auflage). Göttingen: Hogrefe.

ARNTZEN, F. (1982). Die Situation der Forensischen Aussagepsychologie in der Bundesrepublik Deutschland. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 107-119). Stockholm: P.A. Norstedt & söners.

ARNTZEN, F. (1983a). *Psychologie der Zeugenaussage. System der Glaubwürdigkeitsmerkmale* (2. Auflage). München: Beck.

ARNTZEN, F. (1983b). Die Grenzen experimenteller Verfahren in der Forensischen Aussagepsychologie. *Zeitschrift für experimentelle und angewandte Psychologie, 30 (4),* 523-528.

AYMANS, M. (2005). *Die Qualität sachverständigen Handelns bei der aussagepsychologischen Begutachtung von Zeugenaussagen*. München: Herbert Utz Verlag.

BACKHAUS, K., ERICHSON, R., PLINKE, W. & WEIBER, R. (2003). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 10. Auflage. Berlin: Springer.

BUNDESGERICHTSHOF (1995). *Entscheidungen des Bundesgerichtshofes in Strafsachen – BGHSt 40, 138*. Köln/Berlin: Carl Heymanns Verlag.

BUNDESGERICHTSHOF (2000). *Entscheidungen des Bundesgerichtshofes in Strafsachen – BGHSt 45, 164*. Köln/Berlin: Carl Heymanns Verlag.

BINET, A. (1900). *La suggestibilité*. Paris: Schleicher.

BORTZ, J. (1999). *Statistik für Sozialwissenschaftler*. 5. Auflage. Heidelberg: Springer.

BÜHNER, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Person Studium.

DETTENBORN, H., FRÖHLICH, H.-H. & SZEWCZYK, H. (1984). *Forensische Psychologie*. Berlin: VEB Deutscher Verlag der Wissenschaften.

FIEDLER, K. & SCHMID, J. (1999). Gutachten über Methodik und Bewertungskriterien für Psychologische Glaubwürdigkeitsgutachten. *Praxis der Rechtspsychologie, 9 (2),* 5-45.

FISSENI, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik.* 2. Auflage. Göttingen: Hogrefe.

GREUEL, L. (2001). *Wirklichkeit – Erinnerung – Aussage.* Weinheim: Psychologie Verlags Union.

GREUEL, L., OFFE, S., FABIAN, A., WETZELS, P., FABIAN, T., OFFE, H. & STADLER, M. (1998). *Glaubhaftigkeit der Zeugenaussage – Theorie und Praxis der forensisch-psychologischen Begutachtung*. Weinheim: Psychologie Verlags Union.

HAIR, J. F. (JR.), ANDERSON, R. E., TATHAM, R. L. & BLACK, W. C. (1998). *Multivariate Data Analysis*. Fifth Edition. New Jersey: Prentice Hall.

HERMANUTZ, M., LITZCKE, S. M. & KROLL, O. (2004). Das Projekt ALiBi. Aussageanalyse: Lügen in Befragungen identifizieren. Befragen, Vernehmen und Glaubhaftigkeit einschätzen lernen. *Polizei und Wissenschaft, 1,* 2-7.

HETTLER, S. (2005). *Evaluation eines erweiterten Kanons inhaltlicher Kennzeichen wahrer und falscher Zeugenaussagen.* Unveröffentlichte Diplomarbeit, Universität Konstanz. URL: http://www.ub.uni-konstanz.de/kops/volltexte/2005/1526/.

HOMMERS, W. (1997). Die aussagepsychologische Kriteriologie unter kovarianzstatistischer und psychometrischer Perspektive. In L. Greuel, T. Fabian & M. Stadler (Hrsg.), *Psychologie der Zeugenaussage* (S. 87-100). Weinheim: Psychologie Verlags Union.

HOROWITZ, S., LAMB, B., ESPLIN, P., BOYCHUK, T., KRISPIN, O. & REITER-LAVERY, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology, 2,* 11-21.

KÖHNKEN, G. (1990). *Glaubwürdigkeit – Untersuchungen zu einem psychologischen Konstrukt*. Weinheim: Psychologie Verlags Union.

KÖHNKEN, G. (2004). Statement Validity Analysis and the ‚detection of the truth'. In P. A. Granhag & L. A. Strömwall (Eds.), *The Detection of Deception in Forensic Contexts* (pp. 41-63). Cambridge, UK: University Press.

KRAHÉ, B. & KUNDROTAS, S. (1992). Glaubwürdigkeitsbeurteilung bei Vergewaltigungsanzeigen: Ein aussageanalytisches Feldexperiment. *Zeitschrift für experimentelle und angewandte Psychologie, 39 (4),* 598-620.

LAFRENZ, B. (2006). *Trennschärfeanalyse der so genannten Realkennzeichen in der aussagepsychologischen Diagnostik*. Unveröffentlichte Diplomarbeit. Universität Konstanz.

LAMB, M. E., STERNBERG, K. J., ESPLIN, P. W., HERSHKOWITZ, I., ORBACH, Y. & HOVAV, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse & Neglect, 21 (3),* 255-264.

LANDRY, K. L. & BRIGHAM, J. C. (1992). The effect of training in criteria-based content analysis on the ability of detect deception in adults. *Law and Human Behavior, 16,* 663-676.

LITTMANN, E. & SZEWCZYK, H. (1983). Zu einigen Kriterien und Ergebnissen forensisch-psychologischer Glaubwürdigkeitsbegutachtung von sexuell mißbrauchten Kindern und Jugendlichen. *Forensia, 4,* 55-72.

MAIER, B. (2006). *Diskriminanzanalytische Untersuchung inhaltlicher Kennzeichen wahrer und falscher Zeugenaussagen*. Unveröffentlichte Diplomarbeit, Universität Konstanz.

NIEHAUS, S. (2001). *Zur Anwendbarkeit inhaltlicher Glaubhaftigkeitsmerkmale bei Zeugenaussagen unterschiedlichen Wahrheitsgehaltes*. Frankfurt/M: Peter Lang.

PEZDEK, K., MORROW, A., BLANDON-GITLIN, I., GOODMAN, G. S., QUAS, J. A., SAYWITZ, K. J., BIDROSE, S., PIPE, M.-E., ROGERS, M. & BRODIE, L. (2004). Detecting deception in children: Event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology, 89 (1),* 119-126.

RASKIN, D. C. & ESPLIN, P. W. (1991). Statement validity assesment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment, 13*, 265-291.

SPORER, S. L. (1982). A brief history of the psychology of testimony. *Current Psychological Reviews, 2*, 323-340.

STATISTISCHES BUNDESAMT (2006). *Lange Reihen zur Strafverfolgung. Verurteilte nach ausgewählten Straftaten, Geschlecht und Altersgruppen (Früheres Bundesgebiet einschl. Berlin-West, seit 1995 einschl. Gesamt-Berlin).* Wiesbaden: Statistisches Bundesamt. www.destatis.de.

STECK, P. (2002). Die Beurteilung der Glaubwürdigkeit von Zeugenaussagen in der polizeilichen Vernehmung. In Zentraler Psychologischer Dienst der Bayrischen Polizei (Hrsg.), *Berichtband der Fachtagung vom 04.10.2000. Aktiencheck Polizeipsychologie – Kaufen, Halten, Verkaufen?* (S. 50-64). München: Polizeipräsidium.

STECK, P. (2006). Aussageanalyse. In M. Hermanutz & S. M. Litzcke (Hrsg.), *Vernehmung in Theorie und Praxis: Wahrheit – Irrtum – Lüge* (S. 169-184). Stuttgart: Boorberg.

STELLER, M. & KÖHNKEN, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods for investigation and evidence* (pp. 217-245). New York: Springer.

STELLER, M. & VOLBERT, R. (1999). Wissenschaftliches Gutachten. Forensisch-aussagepsychologische Begutachtung (Glaubwürdigkeitsbegutachtung). *Praxis der Rechtspsychologie, 9 (2),* 46-101.

STELLER, M., WELLERSHAUS, P. & WOLF, T. (1992). Realkennzeichen in Kinderaussagen: Empirische Grundlagen der Kriterienorientierten Aussageanalyse. *Zeitschrift für experimentelle und angewandte Psychologie, 39 (1),* 151-170.

SZEWCZYK, H. & LITTMANN, E. (1982). Untersuchungen zur Glaubwürdigkeitsbeurteilung kindlicher Zeugen. In A. Trankell (Ed.), *Reconstructing the past – The role of psychologists in criminal trials* (pp. 73-103). Stockholm: P. A. Norstedt & söners.

TRANKELL, A. (1971). *Der Realitätsgehalt von Zeugenaussagen. Methodik der Aussagepsychologie*. Göttingen: Vandenhoek & Ruprecht.

UNDEUTSCH, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Hrsg.), *Handbuch der Psychologie, Bd. 11: Forensische Psychologie* (S. 26-181). Göttingen: Hogrefe.

VOLBERT, R. (2000). Standards der psychologischen Glaubhaftigkeitsdiagnostik. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren. Indikationen, Methoden und Qualitätsstandards* (S. 113-123). Darmstadt: Steinkopff.

VRIJ, A. (2005). Criteria-based content analysis. A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11 (1),* 3-41.

VRIJ, A., AKEHURST, L., SOUKARA, S. & BULL, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science, 36 (2),* 113-126.

VRIJ, A., KNELLER, W. & MANN, S. (2000). The effect of informing liars about criteria-based content analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology, 5,* 57-70.

WOLF, P. & STELLER, M. (1997). Realkennzeichen in Aussagen von Frauen. Zur Validierung der Kriterienorientierten Aussageanalyse für Zeugenaussagen von Vergewaltigungsopfern. In L. Greuel, T. Fabian & M. Stadler (Hrsg.), *Psychologie der Zeugenaussage* (S. 121-130). Weinheim: Psychologie Verlags Union.

ANHANG

Anhang A: Auswertung

Anhang B: Klassifikation der Fälle aufgrund des Regressionsmodells

Anhang A: Auswertung

Realkennzeichen

Nr.	Gutachten Code	Alter Zeuge	m/w	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Gesamt-score	Urteil Gutachter	Innerfamiliär?	Tatvorwurf
1	35171	35	w	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	nicht glaubhaft	Adoptivvater	sexueller Missbrauch
2	03790	13	w	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	nicht glaubhaft	Vater; Stiefmutter	Misshandlung
3	14165	12	w	1	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	6	glaubhaft	außerfamiliär	sex. Handlungen vor Kindern
4	16894	13	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	außerfamiliär	Vergewaltigung
5	9974	12	w	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
6	15801-1	15	w	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	7	glaubhaft	außerfamiliär	sexueller Missbrauch
7	15801-2	15	w	1	0	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
8	125448	15	w	1	0	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	7	glaubhaft	außerfamiliär	Vergewaltigung
9	318569	16	w	1	1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	7	glaubhaft	außerfamiliär	sexuelle Nötigung
10	216116	21	w	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	0	0	1	1	13	glaubhaft	außerfamiliär	sexueller Missbrauch
11	5761	14	w	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	5	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
12	9992	18	w	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0	5	nicht glaubhaft	außerfamiliär	sexuelle Nötigung
13	216509	33	w	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	3	nicht glaubhaft	Exfreund	Vergewaltigung
14	188311	14	w	1	1	1	0	1	1	1	0	1	0	0	1	0	0	0	0	1	1	0	10	glaubhaft	Lebensgefährte Mutter	sexueller Missbrauch
15	103528	15	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Cousine Mutter + Freund	sexueller Missbrauch
16	134941	16	w	0	1	1	0	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	11	glaubhaft	Exfreund	Vergewaltigung
17	1743	16	w	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1	8	glaubhaft	Vater	sexueller Missbrauch
18	11872	23	m	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	1	0	1	12	glaubhaft	außerfamiliär	Erpressung
19	10662	16	w	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0	1	1	10	glaubhaft	Stiefvater	sexueller Missbrauch
20	310279	16	w	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	0	1	1	1	15	glaubhaft	außerfamiliär	sexueller Missbrauch
21	22483	11	m	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0	0	1	1	1	13	glaubhaft	Lebensgefährte Mutter	sexuelle Nötigung
22	13485	18	w	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	8	glaubhaft	Vater	Misshandlung
23	135681	15	w	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
24	142514	13	w	1	0	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	1	11	glaubhaft	Lebensgefährte Mutter	sexueller Missbrauch
25	144355	21	w	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	1	12	glaubhaft	Stiefvater	sexueller Missbrauch
26	25137	13	w	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	Lebensgefährte Mutter	sexueller Missbrauch
27	3000	22	w	1	0	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	7	glaubhaft	außerfamiliär	sexueller Missbrauch
28	18910	41	w	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	1	1	0	10	glaubhaft	Exfreund	sexuelle Nötigung

29	3530	8	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Vater	sexueller Missbrauch
30	4388	13	w	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	4	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
31	116614	23	w	1	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	1	1	1	11	glaubhaft	Onkel	sexueller Missbrauch
32	2333-1	12	w	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
33	2333-2	12	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
34	103171-1	22	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Vater	sexueller Missbrauch
35	103171-2	24	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2	nicht glaubhaft	Vater	sexueller Missbrauch
36	09114	14	w	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	8	glaubhaft	Vater	Misshandlung
37	301293	24	m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
38	18506	10	w	1	0	1	1	1	0	1	1	0	1	0	1	1	0	0	0	1	0	1	11	glaubhaft	Opa	sexueller Missbrauch
39	41360	17	w	1	0	1	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	9	glaubhaft	außerfamiliär	sexuelle Nötigung
40	24168-1	11	m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	außerfamiliär	sex. Handlungen vor Kindern
41	24168-2	10	m	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sex. Handlungen vor Kindern
42	1276	8	w	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	1	0	5	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
43	4840	13	w	1	0	1	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1	1	8	glaubhaft	Stiefvater	sexueller Missbrauch
44	119304	17	w	1	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	1	1	1	11	glaubhaft	Vater	sexueller Missbrauch
45	29946-1	7	w	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	4	nicht glaubhaft	Opa	sexueller Missbrauch
46	29946-2	9	w	1	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	1	0	10	glaubhaft	Opa	sexueller Missbrauch
47	17278	11	w	1	1	1	1	0	1	0	0	1	1	0	1	0	0	0	0	1	0	1	10	glaubhaft	außerfamiliär	sexueller Missbrauch
48	3186	15	w	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	1	0	1	13	glaubhaft	außerfamiliär	sexueller Missbrauch
49	148367-1	11	w	1	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1	1	1	12	glaubhaft	außerfamiliär	sexueller Missbrauch
50	148367-2	9	w	1	0	1	1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	9	glaubhaft	außerfamiliär	sexueller Missbrauch
51	4093	13	w	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
52	19860	15	w	1	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	6	glaubhaft	außerfamiliär	sexuelle Nötigung
53	1313	12	m	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	4	nicht glaubhaft	Stiefoma	sexueller Missbrauch
54	124557	20	w	1	0	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	0	0	12	glaubhaft	außerfamiliär	Vergewaltigung
55	116766	16	w	1	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	1	1	1	11	glaubhaft	Exfreund	sexuelle Nötigung
56	106732	21	w	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	nicht glaubhaft	Exfreund	Vergewaltigung
57	13333	14	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Stiefvater	sexueller Missbrauch
58	18058	14	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
59	123942	16	w	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexuelle Nötigung
60	32507-1	15	w	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	1	1	11	glaubhaft	Stiefmutter	Misshandlung
61	32507-2	16	w	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1	15	glaubhaft	Stiefmutter	Misshandlung
62	32507-3	19	m	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	16	glaubhaft	Stiefmutter	Misshandlung

63	7414-1	20	w	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	3	nicht glaubhaft	außerfamiliär	Vergewaltigung
64	7414-2	20	w	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	nicht glaubhaft	außerfamiliär	Vergewaltigung
65	128875	33	w	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	1	8	glaubhaft	Exfreund	sexuelle Nötigung
66	104318	15	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Lebensgefährte Mutter	sexueller Missbrauch
67	21046	17	w	1	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	1	11	glaubhaft	außerfamiliär	sexueller Missbrauch
68	108152	7	w	1	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0	0	1	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
69	3496	14	w	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	0	1	0	1	13	glaubhaft	außerfamiliär	sexueller Missbrauch
70	14595	15	w	1	0	1	1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	0	10	glaubhaft	Exfreund	Vergewaltigung
71	11219	18	w	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	Onkel	sexueller Missbrauch
72	127391-1	12	w	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	10	glaubhaft	Vater	sexueller Missbrauch
73	127391-2	14	m	1	0	1	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	8	glaubhaft	Vater	sexueller Missbrauch
74	38964	11	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Vater	sexueller Missbrauch
75	16329	5	w	1	1	1	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	1	10	glaubhaft	Vater	sexueller Missbrauch
76	107273	14	w	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	1	1	1	11	glaubhaft	Bruder	sexueller Missbrauch
77	00473	34	w	1	0	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	9	glaubhaft	Vater	sexueller Missbrauch
78	105621	27	w	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	nicht glaubhaft	Pflegevater	sexueller Missbrauch
79	9897	13	w	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1	1	1	13	glaubhaft	Vater	sexueller Missbrauch
80	140047	25	w	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	nicht glaubhaft	mehrere außerfam. Täter	sexueller Missbrauch
81	107350	19	w	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	3	nicht glaubhaft	Stiefvater	sexueller Missbrauch
82	11686	11	w	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	7	glaubhaft	außerfamiliär	sexueller Missbrauch
83	210717	62	w	1	1	1	1	1	0	0	1	1	1	0	1	1	0	1	0	0	1	0	12	glaubhaft	außerfamiliär	Vergewaltigung
84	37006-1	11	w	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	0	1	1	12	glaubhaft	außerfamiliär	sexueller Missbrauch
85	37006-2	13	w	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
86	25019	30	w	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	nicht glaubhaft	Exfreund	sexueller Missbrauch
87	32165	13	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	nicht glaubhaft	Stiefvater	sexueller Missbrauch
88	27437	23	w	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	3	nicht glaubhaft	Vater	sexueller Missbrauch
89	23582-1	14	m	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	nicht glaubhaft	Pflegeeltern	Misshandlung
90	23582-2	15	m	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	5	nicht glaubhaft	Pflegeeltern	Misshandlung
91	122877	16	w	1	1	1	1	1	1	1	0	1	0	0	1	1	0	0	0	1	1	1	13	glaubhaft	Vater	sexueller Missbrauch
92	7042	10	w	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	4	nicht glaubhaft	Stiefvater	sexueller Missbrauch
93	115742	46	w	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	3	nicht glaubhaft	Ehemann	sexuelle Nötigung
94	307771	8	w	1	1	1	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	8	glaubhaft	Vater	Vergewaltigung
95	2766	14	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
96	13057	40	w	1	0	1	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	8	glaubhaft	Exfreund	sexuelle Nötigung

97	188312	13	m	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	3	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
98	18299	15	w	1	0	1	0	1	1	1	1	0	0	0	1	0	0	1	0	1	0	0	9	glaubhaft	außerfamiliär	sexuelle Nötigung
99	17489	13	w	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	3	nicht glaubhaft	außerfamiliär	Misshandlung
100	28610	8	w	1	0	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
101	28695	9	w	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	4	nicht glaubhaft	Vater	sexueller Missbrauch
102	304110-1	15	m	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
103	304110-2	13	m	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
104	382911	21	w	1	0	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1	1	0	11	glaubhaft	Stiefvater	sexuelle Nötigung
105	300254	14	w	1	0	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0	1	1	11	glaubhaft	Lebensgefährte Mutter	sexueller Missbrauch
106	6090	11	w	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	1	1	15	glaubhaft	Stiefvater	sexueller Missbrauch
107	217235	25	w	1	0	1	0	1	1	1	0	1	0	0	1	1	0	0	0	1	1	0	10	glaubhaft	Partner	gefährl. Körperverl.
108	10134	37	w	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	6	glaubhaft	außerfamiliär	sexueller Missbrauch
109	20157-1	11	w	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	1	1	0	10	glaubhaft	Onkel; Opa	sexueller Missbrauch
110	20157-2	9	w	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1	1	0	8	glaubhaft	Onkel; Opa	sexueller Missbrauch
111	201601	43	w	1	0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	8	glaubhaft	Exfreund	Vergewaltigung
112	304469	12	w	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	8	glaubhaft	außerfamiliär	sexueller Missbrauch
113	146091	20	w	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	1	15	glaubhaft	Partner	Vergewaltigung
114	129390	13	w	1	0	1	1	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	12	glaubhaft	außerfamiliär	sexueller Missbrauch
115	319623	10	w	1	0	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	8	glaubhaft	außerfamiliär	sexueller Missbrauch
116	5536	18	w	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	0	1	1	0	7	nicht glaubhaft	außerfamiliär	Vergewaltigung
117	16860	17	w	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	4	nicht glaubhaft	Stiefvater	sexueller Missbrauch
118	140615	14	w	1	1	1	1	0	1	1	1	0	0	0	1	1	0	0	0	1	1	1	12	glaubhaft	außerfamiliär	sexueller Missbrauch
119	2408	15	w	1	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	6	nicht glaubhaft	Lebensgefährte Mutter	sexuelle Nötigung
120	6154	17	w	1	0	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1	1	1	12	glaubhaft	außerfamiliär	sexueller Missbrauch
121	142168	34	w	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	1	1	1	1	12	glaubhaft	Halbbruder	sexuelle Nötigung
122	382912	21	w	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	5	glaubhaft	außerfamiliär	sexueller Missbrauch
123	25612	50	w	1	0	1	1	1	1	1	0	1	0	0	1	0	0	0	0	1	0	0	9	glaubhaft	außerfamiliär	Vergewaltigung
124	21273	14	w	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	1	0	1	7	nicht glaubhaft	außerfamiliär	sexuelle Nötigung
125	8998	13	w	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	1	12	glaubhaft	Vater	sexueller Missbrauch
126	11448	12	m	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	1	11	glaubhaft	außerfamiliär	sexueller Missbrauch
127	302293	14	w	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
128	12345	13	w	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	8	glaubhaft	Stiefvater	sexueller Missbrauch
129	123456	23	w	1	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	8	glaubhaft	außerfamiliär	Beleidigung
130	106023	20	w	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexueller Missbrauch

131	311706	18	w	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	2	nicht glaubhaft	Ehemann	Vergewaltigung
132	23617-1	24	w	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	1	8	glaubhaft	außerfamiliär	sexueller Missbrauch
133	23617-2	25	w	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	5	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
134	2347	15	w	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	nicht glaubhaft	außerfamiliär	sexuelle Nötigung
135	308218	8	w	1	1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	1	9	glaubhaft	außerfamiliär	sexuellen Missbrauch
136	4419-1	13	m	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
137	4419-2	14	m	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
138	4419-3	14	m	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	4	nicht glaubhaft	außerfamiliär	sexueller Missbrauch
	Summe	17,31884		89	42	93	82	76	57	57	53	39	21	2	83	36	12	19	6	55	61	56	939			

Anhang B: Klassifikation der Fälle aufgrund des Regressionsmodells

	Code	Beobachtete Gruppe	Summen-score	p (y = nicht glaubhaft)	Vorhergesagte Gruppe
1	24168-1	nicht glaubhaft	0	.99999	nicht glaubhaft
2	103528	nicht glaubhaft	0	.99999	nicht glaubhaft
3	3530	nicht glaubhaft	0	.99999	nicht glaubhaft
4	103171-1	nicht glaubhaft	0	.99999	nicht glaubhaft
5	13333	nicht glaubhaft	0	.99999	nicht glaubhaft
6	18058	nicht glaubhaft	0	.99999	nicht glaubhaft
7	104318	nicht glaubhaft	0	.99999	nicht glaubhaft
8	38964	nicht glaubhaft	0	.99999	nicht glaubhaft
9	32165	nicht glaubhaft	0	.99999	nicht glaubhaft
10	16894	nicht glaubhaft	0	.99999	nicht glaubhaft
11	2333-2	nicht glaubhaft	1	.99993	nicht glaubhaft
12	301293	nicht glaubhaft	1	.99993	nicht glaubhaft
13	105621	nicht glaubhaft	1	.99993	nicht glaubhaft
14	2766	nicht glaubhaft	1	.99993	nicht glaubhaft
15	304110-1	nicht glaubhaft	1	.99993	nicht glaubhaft
16	304110-2	nicht glaubhaft	1	.99993	nicht glaubhaft
17	4419-1	nicht glaubhaft	1	.99993	nicht glaubhaft
18	23582-1	nicht glaubhaft	2	.99951	nicht glaubhaft
19	24168-2	nicht glaubhaft	2	.99951	nicht glaubhaft
20	123942	nicht glaubhaft	2	.99951	nicht glaubhaft
21	2347	nicht glaubhaft	2	.99951	nicht glaubhaft
22	135681	nicht glaubhaft	2	.99951	nicht glaubhaft
23	25137	nicht glaubhaft	2	.99951	nicht glaubhaft
24	103171-2	nicht glaubhaft	2	.99951	nicht glaubhaft
25	11219	nicht glaubhaft	2	.99951	nicht glaubhaft
26	37006-2	nicht glaubhaft	2	.99951	nicht glaubhaft
27	302293	nicht glaubhaft	2	.99951	nicht glaubhaft
28	106023	nicht glaubhaft	2	.99951	nicht glaubhaft
29	4419-2	nicht glaubhaft	2	.99951	nicht glaubhaft
30	106732	nicht glaubhaft	2	.99951	nicht glaubhaft
31	311706	nicht glaubhaft	2	.99951	nicht glaubhaft
32	03790	nicht glaubhaft	3	.99662	nicht glaubhaft
33	17489	nicht glaubhaft	3	.99662	nicht glaubhaft
34	115742	nicht glaubhaft	3	.99662	nicht glaubhaft
35	140047	nicht glaubhaft	3	.99662	nicht glaubhaft
36	107350	nicht glaubhaft	3	.99662	nicht glaubhaft
37	25019	nicht glaubhaft	3	.99662	nicht glaubhaft
38	27437	nicht glaubhaft	3	.99662	nicht glaubhaft
39	188312	nicht glaubhaft	3	.99662	nicht glaubhaft
40	216509	nicht glaubhaft	3	.99662	nicht glaubhaft
41	7414-1	nicht glaubhaft	3	.99662	nicht glaubhaft
42	7414-2	nicht glaubhaft	3	.99662	nicht glaubhaft
43	35171	nicht glaubhaft	4	.97713	nicht glaubhaft
44	4388	nicht glaubhaft	4	.97713	nicht glaubhaft
45	29946-1	nicht glaubhaft	4	.97713	nicht glaubhaft
46	4093	nicht glaubhaft	4	.97713	nicht glaubhaft

	Code	Beobachtete Gruppe	Summen-score	p (y = nicht glaubhaft)	Vorhergesagte Gruppe
47	1313	nicht glaubhaft	4	.97713	nicht glaubhaft
48	7042	nicht glaubhaft	4	.97713	nicht glaubhaft
49	28695	nicht glaubhaft	4	.97713	nicht glaubhaft
50	16860	nicht glaubhaft	4	.97713	nicht glaubhaft
51	4419-3	nicht glaubhaft	4	.97713	nicht glaubhaft
52	23582-2	nicht glaubhaft	5	.86088	nicht glaubhaft
53	9992	nicht glaubhaft	5	.86088	nicht glaubhaft
54	5761	nicht glaubhaft	5	.86088	nicht glaubhaft
55	1276	nicht glaubhaft	5	.86088	nicht glaubhaft
56*	382912	glaubhaft	5	.86088	nicht glaubhaft
57	23617-2	nicht glaubhaft	5	.86088	nicht glaubhaft
58	14165	glaubhaft	6	.47266	glaubhaft
59	19860	glaubhaft	6	.47266	glaubhaft
60*	2408	nicht glaubhaft	6	.47266	glaubhaft
61	10134	glaubhaft	6	.47266	glaubhaft
62	318569	glaubhaft	7	.11491	glaubhaft
63*	21273	nicht glaubhaft	7	.11491	glaubhaft
64	15801-1	glaubhaft	7	.11491	glaubhaft
65	3000	glaubhaft	7	.11491	glaubhaft
66	11686	glaubhaft	7	.11491	glaubhaft
67	125448	glaubhaft	7	.11491	glaubhaft
68*	5536	nicht glaubhaft	7	.11491	glaubhaft
69	123456	glaubhaft	8	.01846	glaubhaft
70	13485	glaubhaft	8	.01846	glaubhaft
71	09114	glaubhaft	8	.01846	glaubhaft
72	128875	glaubhaft	8	.01846	glaubhaft
73	13057	glaubhaft	8	.01846	glaubhaft
74	9974	glaubhaft	8	.01846	glaubhaft
75	15801-2	glaubhaft	8	.01846	glaubhaft
76	1743	glaubhaft	8	.01846	glaubhaft
77	2333-1	glaubhaft	8	.01846	glaubhaft
78	4840	glaubhaft	8	.01846	glaubhaft
79	108152	glaubhaft	8	.01846	glaubhaft
80	127391-2	glaubhaft	8	.01846	glaubhaft
81	28610	glaubhaft	8	.01846	glaubhaft
82	20157-2	glaubhaft	8	.01846	glaubhaft
83	304469	glaubhaft	8	.01846	glaubhaft
84	319623	glaubhaft	8	.01846	glaubhaft
85	12345	glaubhaft	8	.01846	glaubhaft
86	23617-1	glaubhaft	8	.01846	glaubhaft
87	307771	glaubhaft	8	.01846	glaubhaft
88	201601	glaubhaft	8	.01846	glaubhaft
89	18299	glaubhaft	9	.00272	glaubhaft
90	41360	glaubhaft	9	.00272	glaubhaft
91	308218	glaubhaft	9	.00272	glaubhaft
92	148367-2	glaubhaft	9	.00272	glaubhaft
93	00473	glaubhaft	9	.00272	glaubhaft
94	25612	glaubhaft	9	.00272	glaubhaft

	Code	Beobachtete Gruppe	Summen-score	p (y = nicht glaubhaft)	Vorhergesagte Gruppe
95	217235	glaubhaft	10	.00039	glaubhaft
96	18910	glaubhaft	10	.00039	glaubhaft
97	188311	glaubhaft	10	.00039	glaubhaft
98	10662	glaubhaft	10	.00039	glaubhaft
99	29946-2	glaubhaft	10	.00039	glaubhaft
100	17278	glaubhaft	10	.00039	glaubhaft
101	127391-1	glaubhaft	10	.00039	glaubhaft
102	16329	glaubhaft	10	.00039	glaubhaft
103	20157-1	glaubhaft	10	.00039	glaubhaft
104	14595	glaubhaft	10	.00039	glaubhaft
105	32507-1	glaubhaft	11	.00006	glaubhaft
106	116766	glaubhaft	11	.00006	glaubhaft
107	382911	glaubhaft	11	.00006	glaubhaft
108	142514	glaubhaft	11	.00006	glaubhaft
109	116614	glaubhaft	11	.00006	glaubhaft
110	18506	glaubhaft	11	.00006	glaubhaft
111	119304	glaubhaft	11	.00006	glaubhaft
112	21046	glaubhaft	11	.00006	glaubhaft
113	107273	glaubhaft	11	.00006	glaubhaft
114	300254	glaubhaft	11	.00006	glaubhaft
115	11448	glaubhaft	11	.00006	glaubhaft
116	134941	glaubhaft	11	.00006	glaubhaft
117	11872	glaubhaft	12	.00001	glaubhaft
118	142168	glaubhaft	12	.00001	glaubhaft
119	144355	glaubhaft	12	.00001	glaubhaft
120	148367-1	glaubhaft	12	.00001	glaubhaft
121	37006-1	glaubhaft	12	.00001	glaubhaft
122	129390	glaubhaft	12	.00001	glaubhaft
123	140615	glaubhaft	12	.00001	glaubhaft
124	6154	glaubhaft	12	.00001	glaubhaft
125	8998	glaubhaft	12	.00001	glaubhaft
126	124557	glaubhaft	12	.00001	glaubhaft
127	210717	glaubhaft	12	.00001	glaubhaft
128	22483	glaubhaft	13	.00000	glaubhaft
129	216116	glaubhaft	13	.00000	glaubhaft
130	3186	glaubhaft	13	.00000	glaubhaft
131	3496	glaubhaft	13	.00000	glaubhaft
132	9897	glaubhaft	13	.00000	glaubhaft
133	122877	glaubhaft	13	.00000	glaubhaft
134	32507-2	glaubhaft	15	.00000	glaubhaft
135	310279	glaubhaft	15	.00000	glaubhaft
136	6090	glaubhaft	15	.00000	glaubhaft
137	146091	glaubhaft	15	.00000	glaubhaft
138	32507-3	glaubhaft	16	.00000	glaubhaft

* falsch klassifiziert aufgrund der logistischen Regressionsfunktion.