

**Die Validität inhaltlicher Glaubhaftigkeitsmerkmale
unter suggestiven Bedingungen.
Eine empirische Untersuchung an Vorschulkindern.**

Dissertation
zu Erlangung des Doktorgrades
der Philosophischen Fakultät der Christian-Albrechts-Universität
zu Kiel

vorgelegt von
Dipl.-Psych. Petra Häner

Kiel 2007

Erstgutachter: Prof. Dr. Günther Köhnken

Zweitgutachter: Prof. Dr. Max Steller

Tag der mündlichen Prüfung: 25.01.2008

Durch den zweiten Prodekan, Prof. Dr. Steindorff

zum Druck genehmigt am: 30.01.2008

1.	Einleitung	5
2.	Methoden aussagepsychologischer Begutachtung.....	7
2.1	Zur Geschichte der Aussagepsychologie.....	7
2.2	Kriteriologien zur Beurteilung der Glaubwürdigkeit	11
2.3	Reliabilität inhaltsbezogener Kriterien	19
2.4	Validität inhaltsbezogener Kriterien.....	24
2.4.1	Feldstudien zur Validität inhaltlicher Realkennzeichen.....	25
2.4.2	Überlegungen zur Dienlichkeit von Feldstudien in der aussagepsycholog-.....	35
	ischen Forschung	35
2.4.3	Experimentelle Untersuchungen der inhaltlichen Realkennzeichen	39
2.4.4	Überlegungen zu den bisherigen experimentellen Untersuchungen.....	46
3.	Suggestion und Suggestibilität	50
3.1	Empirische Ergebnisse zur Suggestibilität.....	54
3.2	Suggestibilität im Kontext aussagepsychologischer Beurteilung.....	66
4.	Fragestellung und Hypothesen	72
4.1	Ableitung der Fragestellung	72
4.2.	Hypothesen der Untersuchung	80
5.	Methodik.....	81
5.1	Operationalisierung der Variablen	81
5.1.1	Unabhängige Variable	81
5.1.2	Abhängige Variable.....	82
5.1.3	Kontrollvariablen	83
5.1.4	Zusätzliche Variablen.....	83
5.2	Experimentelles Design.....	84
6.	Durchführung der Untersuchung	84
6.1.	Kaspertheater als Stimulusmaterial.....	84
6.2	Versuchspersonen.....	86
6.2.1	Rekrutierung der Versuchspersonen.....	86
6.2.2	Zuordnung der Versuchspersonen zu den Versuchsgruppen	87
6.3	Interviewer	88
6. 4	Rater	89

6.5	Ablauf der Untersuchung	90
6.5.1	Untersuchungsphase 1 : Erleben des Kaspertheaters.....	90
6.5.2	Untersuchungsphase 2: Suggestive Einflüsse.....	91
6.5.3	Untersuchungsphase 3: Erhebung der Aussage.....	92
6.5.4	Auswertung der Daten.....	94
6.5.5	Einschätzung der Realkennzeichen.....	96
7.	Darstellung der Ergebnisse.....	98
7.1	Kennwerte der Stichprobe.....	98
7.2	Interrater-Reliabilität.....	98
7.3	Einfluss der Kontrollvariablen auf den Summenscore der Realkennzeichen.....	100
7.4	Einfluss der Wortanzahl auf den Summenscore der Realkennzeichen.....	103
7.5	Auftreten der Realkennzeichen in den Aussagen.....	104
7.7	Ergebnisse zur Trennung der Versuchsgruppen mittels Realkennzeichen	108
7.8	Zusätzliche statistische Auswertungen.....	112
7.8.1	Trefferquoten.....	112
7.8.2	Urteilssicherheit	114
7.8.3	Variable Wortanzahl	116
8.	Diskussion der Ergebnisse.....	118
8.1	Allgemeine Feststellungen zur Untersuchung.....	119
8.2	Validität der kriterienorientierten Aussageanalyse bei Kindern im.....	123
	Vorschulalter	
	Literatur	131
	Anhang.....	143

1. Einleitung

Eine zentrale Frage in Gerichtsverfahren ist die Frage nach der Glaubhaftigkeit der Zeugenaussagen. Seit den fünfziger Jahren fungieren Psychologen regelmäßig bei der Klärung dieses Problems als Gehilfen des Gerichts. Straftaten gegen die sexuelle Selbstbestimmung stehen dabei im Vordergrund ihres Einsatzes. Anfangs wurde dabei auf durch Erfahrung erworbenes Wissen der psychologischen Sachverständigen zurückgegriffen, in den achtziger Jahren des letzten Jahrhunderts begann dann eine rege Forschungstätigkeit mit dem Ziel, das bisherige Vorgehen experimentell bestätigen zu können (vgl. Kapitel 2.1). Die Begutachtung des Realitätsgehaltes von Zeugenaussagen bezieht grundsätzlich drei Analysebereiche ein: Persönlichkeits-, Motiv- und Aussageanalyse. Die Aussageanalyse beinhaltet Inhaltsanalyse, Kompetenzprüfung und Rekonstruktion der Aussageentstehung. Im Mittelpunkt der Forschungsaktivitäten stand zunächst die Aussageanalyse, insbesondere die Inhaltsanalyse, die auch als Realkennzeichenanalyse, merkmals- oder kriterienorientierte Analyse bezeichnet wird. Von Interesse ist dabei die Frage, ob erlebnisbegründete Aussagen von Aussagen ohne Erlebnishintergrund unterschieden werden können. Bisherige Untersuchungen zur Validität der Realkennzeichen, die deren Brauchbarkeit überwiegend belegen, beziehen sich auf ältere Kinder und Erwachsene (vgl. Kapitel 2.4). In der forensischen Praxis wird aber gerade die Begutachtung junger Kinder gefordert, da aufgrund der noch nicht abgeschlossenen Reifung, z.B. der kognitiven und sprachlichen Kompetenzen, Unsicherheiten über die Glaubhaftigkeit bestehen. In der vorliegenden Arbeit soll daher überprüft werden, ob die kriterienorientierte Aussageanalyse auch bei Kindergarten- und Vorschulkindern eine geeignete Methode zur Feststellung des Realitätsgehaltes ihrer Aussage ist.

Ausgelöst durch spektakuläre Missbrauchsfälle in den siebziger Jahren in den USA und eine verstärkte öffentliche Auseinandersetzung mit dem sexuellen Kindesmissbrauch in Deutschland Ende der achtziger Jahre entwickelte sich eine rege Forschungsaktivität im Bereich von Suggestionen und Suggestibilität (vgl. Kapitel 3.1), die bis heute in zahlreichen Veröffentlichungen zu Tage tritt. Durch die empirische Suggestionsforschung konnte belegt werden, dass eine Vielzahl von Faktoren sowohl auf Seiten des Suggestors als auch auf Seiten des Rezipienten die Wahrnehmung, Speicherung und Reproduktion von

Erlebnissen beeinflussen können. Für die Glaubhaftigkeitsbeurteilung mit dem zentralen Element der inhaltsorientierten Aussageanalyse warfen diese Ergebnisse die Frage nach den Auswirkungen suggestiver Beeinflussung auf den Aussageinhalt auf. Die Rekonstruktion der Aussageentstehung bekam nunmehr einen besonderen Stellenwert. Die Vermutung, dass suggerierte Aussagen eine ähnliche Qualität aufweisen könnten wie erlebnisbegründete Aussagen findet zunehmend Bestätigung. Zusammenfassend zeigen die bisher vorliegenden Untersuchungen, dass sich suggestive Bedingungen nicht nur in Teilespekten eines tatsächlich erlebten Vorganges im Sinne von Erinnerungsveränderungen zeigen können, sondern sich auch in Aussagen über nicht selbst erlebte Vorgänge niederschlagen können (vgl. 3.2). Dabei steht in der praktischen Begutachtung weniger die gezielt beeinflusste Falschaussage im Vordergrund, sondern vielmehr die Übernahme von Vorgaben als Grundlage für eine irrtümliche Falschaussage. In der vorliegenden Arbeit soll daher überprüft werden, ob Kindergarten- und Vorschulkinder imstande sind, allein auf der Basis von verbalen Vorgaben Aussagen über ein selbst erlebtes Geschehen zu generieren. Weiterhin soll der Frage nachgegangen werden, ob derart induzierte Aussagen mittels der Realkennzeichenanalyse von Schilderungen unterschieden werden können, denen ein tatsächliches Erleben zugrunde liegt.

2. Methoden aussagepsychologischer Begutachtung

2.1 Zur Geschichte der Aussagepsychologie

Über den tatsächlichen Beginn wissenschaftlicher Aussageforschung herrscht keine Einigkeit: Während Undeutsch (1967) die Auffassung vertritt, das Buch „La Suggestibilité“ von Alfred Binet (1900) und die Arbeiten William Sterns (z.B. „Psychologie der Aussage“, 1902) markierten den Ursprung dieses Forschungsgebietes, sind andere Wissenschaftler der Meinung, den Zeitpunkt nicht genau bestimmen zu können. Sie begründen dies damit, dass einige allgemeinpsychologische Themen, welche schon vor der Jahrhundertwende abgehandelt wurden, später zum Gebiet der Aussagepsychologie gezählt wurden. In diesem Zusammenhang sind z.B. allgemeinpsychologische, experimentelle Arbeiten von Cattell und Binet zu nennen. Cattell (1895) führte Untersuchungen zum Gedächtnis beiläufiger Wahrnehmungen, Binet (1897) erstmals eine experimentelle Arbeit mit direktem Bezug zur Zeugenaussage durch.

Ein anderer Ursprung der wissenschaftlichen Aussagepsychologie findet sich in Arbeiten, die eher der juristischen und kriminologischen Richtung zuzuordnen sind. Grundlegende Arbeiten führten zu einer Reihe von Veröffentlichungen in denen es um Wahrnehmungs- und Gedächtnisfähigkeiten von Zeugen ging (Köhnken, 1990). Deutlich wird im historischen Rückblick, dass die Aussagepsychologie wesentliche Bezüge zu verschiedenen Gebieten der Psychologie hat, z.B. zur allgemeinen Psychologie (insbesondere zur gedächtnispsychologischen Forschung), der Sozialpsychologie (Entscheidungs- und Bewertungsprozesse im sozialen Kontext) oder in neuerer Zeit zur klinischen Psychologie (z.B. Traumatologie, dissoziative Störungen, Borderline- Persönlichkeit). Es ergeben sich aber auch fachübergreifende Bezüge, z.B. zu den Rechtswissenschaften, zur Medizin (Rechtsmedizin, Sexualmedizin), etwa bei Strafprozessen im Zusammenhang mit Sexualstraftaten.

Mit einem richtungweisenden Aufsatz von William Stern (1904) wurde Anfang des letzten Jahrhunderts eine entscheidende Konzeptualisierung der Aussagepsychologie geprägt, deren Wert bis heute unbestritten ist. Nach Stern ist eine Aussage als geistige

Leistung und als Verhörsprodukt aufzufassen, da sie sowohl von personalen als auch von situativen Merkmalen beeinflusst wird.

In der überwiegenden Anzahl der nachfolgenden Veröffentlichungen findet sich diese Konzeptualisierung. Aus ihr wird deutlich, dass eine Aussage keinesfalls ein getreues Abbild der Realität sein muss, sondern Einflüssen unterliegen kann, die in den verschiedenen Phasen der Aussageentstehung wirksam werden können (Steller & Volbert, 1997). Zu diesen Phasen zählen die Wahrnehmung des Sachverhaltes, die Speicherung und die Reproduktion (s.a. Wegener, 1981). So kann die Wahrnehmung durch Variablen wie Dauer der Wahrnehmung oder physikalische Bedingungen beeinflusst werden. Die sich anschließende Phase der Speicherung kann beispielsweise Einflüssen nachträglicher Informationen unterliegen. In der Reproduktionsphase können z.B. der zeitliche Abstand zwischen dem Ereignis und der Befragung, die Formulierung der Fragen oder die Vernehmungsatmosphäre inhaltliche Auswirkungen auf die Aussage haben.

Erste experimentelle Untersuchungen zur Zeugenaussage nahm Stern (1902) vor, indem er Versuchspersonen detaillierte Bilder betrachten und sie diese unter verschiedenen Reproduktionsbedingungen und nach unterschiedlichen Intervallen wiedergeben ließ. Angeregt durch Sterns experimentelle Forschung begannen weitere Wissenschaftler tätig zu werden. Hier sind z.B. Max Wertheimer, der 1906 seine Dissertation zu Fragen der Glaubwürdigkeit unter dem Titel „Psychologische Tatbestandsdiagnostik“ einreichte und Hugo Münsterberg (1908), der in die Untersuchungen zur Zeugenaussage auch psychophysiologische Parameter einbezog, zu nennen. Ebenfalls anzuführen sind Varendonck (1911, zitiert nach Wegener, 1997), der nach zahlreichen Gedächtnisexperimenten mit jungen Mädchen die Unbrauchbarkeit von zwei Zeugenaussagen in einem Mordfall nachweisen konnte, und Edouard Claparède (1906, zitiert nach Wegener, 1997), der sich mit kindlichen Zeugenaussagen befasste.

Nach Steller (1988) und Greuel (1997) lässt sich die Entwicklung der Aussagepsychologie in verschiedene Phasen einteilen. Zu Beginn der Forschungstätigkeit im Bereich der Aussagepsychologie wurde angenommen, Menschen seien grundsätzlich in der Lage, Ereignisse unter bestimmten Bedingungen korrekt wiederzugeben (Köhnken, 1990). Die-

se experimentelle Frühphase aussagepsychologischer Forschung war durch eine Reihe laborexperimenteller Untersuchungen geprägt (z.B. Stern, 1904).

Nachdem zwischen den 30er und den 50er Jahren ein eher nachlassendes Interesse an aussagepsychologischer Forschung zu verzeichnen war (Latenzphase) schloss sich danach eine dritte Phase von Forschungsbemühungen an. Diese war im Gegensatz zum Laborexperiment durch die Feldbeobachtung geprägt. Dieses Vorgehen resultierte aus der Kritik an dem einseitig experimentellen Zugang zu den Problemen von Zeugenaussagen. Undeutsch (1967) sah darin eine negativistische Tendenz, da durch das Aufdecken immer neuer Fehlerquellen letztlich keine Hilfestellung für eine Verbesserung der Qualität von Zeugenaussagen oder zur Erstellung von Leitlinien zur gerichtlichen Befragung zu erwarten war. Die kritische Auseinandersetzung mit den früheren experimentellen Arbeiten führte auch zu einer inhaltlichen Verschiebung des Forschungsschwerpunktes. Anstatt nach den Einflüssen möglicher Fehlerquellen auf Zeugenaussagen zu suchen, stand in dieser Phase der Aussagepsychologie in Deutschland die Differenzierbarkeit von glaubwürdigen und unglaubwürdigen Aussagen im Vordergrund. 1955 erging ein Grundsatzurteil des Bundesgerichtshofs (BGHSt 1955,7, 82), das bestimmte, dass Sachverständige für die Begutachtung von Zeugenaussagen im Strafprozess zu bestellen sind, wenn die Zeugenaussagen von Kindern oder Jugendlichen die alleinigen oder wesentlichen Beweismittel darstellen. Das Urteil führte zu einem Anstieg praktischer Begutachtungsfälle und damit zu einem Zuwachs an forensischer Erfahrung für die Sachverständigen. In Folge dessen wurden in der Literatur zunehmend inhaltliche Merkmale von Aussagen beschrieben, die unter den Begriffen *Realkennzeichen* oder *Glaubwürdigkeitskriterien* zusammengefasst wurden. Diese Phase aussagepsychologischer Forschung zeichnete sich durch die fast ausschließliche Beschränkung auf Sexualdelikte, durch die Verwendung von Erfahrungsberichten praktisch tätiger Gutachter und durch ein nahezu völliges Fehlen von Laborexperimenten aus. Als Vertreter dieser Phase aussagepsychologischer Forschung sind Undeutsch, Arntzen (1982), Szewczyk (1973) und Trankell (1971) zu nennen.

Das Konzept der *Allgemeinen Glaubwürdigkeit*, das bisher der Zeugenbegutachtung zugrunde lag, wurde von dem Konzept der *Speziellen Glaubwürdigkeit* (Undeutsch, 1967) abgelöst. Befürworter des Konzeptes der Allgemeinen Glaubwürdigkeit vertraten die Auffassung, dass die Glaubwürdigkeit einer Person und ihrer Aussage mittels des relativ stabilen Persönlichkeitsmerkmals Wahrheitsliebe oder Wahrhaftigkeit beurteilt werden könne (Peters, 1972). Demgegenüber beinhaltet das Konzept der Speziellen Glaubwürdigkeit die Annahme, dass der Wahrheitsvorsatz nicht allein durch stabile Persönlichkeitsmerkmale erklärt werden kann, sondern dass die Motivation zu einer wahren oder falschen Zeugenaussage in erster Linie von der Situation bestimmt wird. Die Aussage selbst soll als die wesentliche Grundlage der Beurteilung der Glaubwürdigkeit betrachtet werden (Köhnken, 1990). Auch Undeutsch (1967) betonte nachdrücklich, dass nicht die *Glaubwürdigkeit* eines Zeugen, sondern die *Glaubhaftigkeit* der Aussage im Mittelpunkt des Begutachtungsprozesses stehen müsse.

Diese Erfahrungs- und Entwicklungsphase aussagepsychologischer Forschung (Greuel, 1997) war durch ein empirisch-kasuistisches Vorgehen geprägt und stand im Zeichen der Entwicklung systematischer Glaubwürdigkeitskriterien, der so genannten Kriteriologien. Es wurden verschiedene Kriteriensysteme entwickelt, die zwar einerseits inhaltliche Gemeinsamkeiten aufweisen, andererseits aber auch unterschiedliche Vorgehensweisen beinhalten. Die von Steller und Köhnken (1989) aus diesen Systemen entwickelte Gesamtkriteriologie gehört zur vierten Phase, in der sich die aussagepsychologische Forschung nach Steller (1988) seit Beginn der 80er Jahre befand. In dieser experimentellen Validierungsphase (Greuel, 1997) lag der Schwerpunkt auf systematischer empirischer Forschung. Ausgehend von Problemen der praktischen Zeugenbegutachtung befasste man sich besonders im anglo-amerikanischen Raum mit Fragen der Suggestibilität kindlicher Zeugenaussagen (Ceci & Bruck, 1993). Damit rückten auch Fragen nach der Aussageentstehung und nach der Geschichte der Aussage in den Mittelpunkt des Forschungsinteresses. Greuel (1997) definierte für die aussagepsychologische Forschung eine fünfte Phase, die von ihr als „Integrationsphase“ bezeichnet wurde, in der „das im Laufe eines Jahrhunderts entwickelte und validierte Expertenwissen endlich integriert werden kann und muss“ (Greuel, 1997, S.158). Neben

der Suggestionsforschung rücken in neuerer Zeit der Einfluss von Emotionen und Stress, so wie Persönlichkeitsstörungen auf die Erinnerungsleistungen und deren Auswirkungen auf die aussagepsychologische Beurteilung von Schilderungen traumatischer Ereignisse in den Blickpunkt aussagepsychologischer Forschung (Volbert, 2004)

2.2 Kriteriologien zur Beurteilung der Glaubwürdigkeit

Grundsätzlich lassen sich zwei Arten von Kriteriensystemen unterscheiden: Das prozessorientierte System und das kriterienorientierte System. Die kriterienorientierten Systeme, deren prominente Vertreter Undeutsch (1967), Arntzen (1982), Littmann und Szewczyk (1983) sind, beinhalten Zusammenstellungen von Kriterien zur Beurteilung der Glaubwürdigkeit. Demgegenüber liefert das prozessorientierte System von Trankell (1971) neben einer geringeren Anzahl an Glaubwürdigkeitskriterien Verfahrensregeln für den diagnostischen Prozess selbst. Glaubwürdigkeitskriterien werden in beiden Systemen als inhaltsbezogene Merkmale wie Detailreichtum, Folgerichtigkeit, Originalität und kontextuelle Zuordnungen verstanden. Synonyme für Glaubwürdigkeitskriterien sind Glaubwürdigkeitsmerkmale oder Realkennzeichen. Die verschiedenen Systeme sollen im Folgenden unter Berücksichtigung von Unterschieden und Gemeinsamkeiten dargestellt werden.

Das diagnostische Modell von Trankell (1971) beinhaltet ein prozessorientiertes Vorgehen. Schwerpunkt des Modells ist der Urteilsprozess, also die Frage, *wie* vom Vorhandensein bestimmter Merkmale auf die Glaubwürdigkeit geschlossen werden kann. Die Glaubwürdigkeitskriterien spielen dabei für Trankell eine eher untergeordnete Rolle. Der Autor stellt darüber hinaus Regeln auf, nach denen lediglich für den jeweiligen Einzelfall gültige Kriterien generiert werden können. Je nach Bedarf können somit Kriterien erzeugt werden, die nur unter den gegebenen Umständen des vorliegenden Falls Gültigkeit haben. Der Prozess der Begutachtung sieht dabei folgendermaßen aus: Auf der Basis vorhandener Kenntnisse werden eine Realitätshypothese und eine oder mehrere Alterna-

tivhypothesen aufgestellt. Die Realitätshypothese besagt, dass die Aussage auf einer wahren Erlebnisgrundlage beruht. Eine Alternativhypothese kann z.B. sein, dass der Aussage eine gezielte Falschbeziehungsabsicht zugrunde liegt. Die Datenerhebung und -analyse erfolgen solange, bis zwei so genannte *fundamentale Formalkriterien* als erfüllt betrachtet werden können. Das erste fundamentale Formalkriterium beinhaltet, dass die zu prüfende Hypothese durch die aus der Exploration des Zeugen gewonnenen und weitere Informationen erschöpfend erklärt werden kann. Das zweite Kriterium besagt, dass sich diese Hypothese als alleinige durch die vorhandenen Daten hinlänglich erklären lassen darf. Solange mehrere Hypothesen nebeneinander existieren und keine als alleinige durch die Informationen erklärt werden kann, sind die Datenerhebung und -analyse fortzusetzen. Trankell (1971) bezeichnet dieses Vorgehen als aussagepsychologische Realitätsanalyse. Er unterscheidet weiterhin zwischen aussagepsychologischen Realitätskriterien und Kennzeichen realistischer Schilderungen. Die aussagepsychologischen Realitätskriterien kommen ausschließlich in wahren Aussagen vor, die Kennzeichen realitätsbezogener Schilderungen können auch in Aussagen ohne Realitätsgrundlage vorhanden sein. Als ein Kennzeichen realitätsbezogener Schilderungen kann das Merkmal *Detailreichum* angesehen werden. Auch Aussagen ohne Erlebnisgrundlage können dieses Merkmal enthalten. Daher darf es nicht in jedem Fall als Realitätskriterium verwendet werden. Trankell unterscheidet außerdem zwischen primären und sekundären Realitätskriterien. Primäre Realitätskriterien lassen sich nur bei der Erfüllung bestimmter Voraussetzungen für die Beurteilung einer Aussage anwenden. Die sekundären Realitätskriterien haben eine allgemeinere Gültigkeit als die primären, dienen aber nur zu deren Ergänzung.

Zu den kriteriumsorientierten Systemen gehört das Kriteriensystem von Undeutsch (1967), der als einer der wichtigsten und richtungweisenden Protagonisten in der Entwicklungsgeschichte der Aussagepsychologie bezeichnet werden darf. Er führt verschiedene Bereiche an, die bei der Glaubwürdigkeitsbeurteilung zu prüfen sind: Persönlichkeitsmerkmale und individueller Entwicklungsstand des Zeugen, Motivlage des Zeugen, Geschichte und Entwicklung der Aussage, Tendenzen in der Aussage und die Aussage selbst (Undeutsch, 1993). Diese Bereiche sind von unterschiedlicher Relevanz für die Glaubwürdigkeitsbegutachtung. Die ersten vier Befundbereiche wertet Undeutsch (1993)

als "Vorstufe der Aussagebeurteilung" (S. 131). Zentrales Element der Aussagebeurteilung ist der fünfte Bereich - die Aussage selbst. Ausgangspunkt dieser Wertung war Undeutschs Feststellung, dass wahrheitsgemäße Darstellungen selbst erlebter Realität eine andere inhaltliche Qualität haben als nicht in der Realität begründete Darstellungen. Diese Überlegung wurde später von Steller (1988) als *Die Undeutsch Hypothese* in die wissenschaftliche Literatur eingeführt und ist seither die explizit formulierte Grundhypothese der Aussagebeurteilung.

Die Unterschiede zwischen realitätsbegründeten und phantasiegeleiteten Schilderungen manifestieren sich in inhaltlichen Merkmalen der Aussage, den Realitätskriterien. Diese unterscheidet Undeutsch (1993) in Grundlegende Kriterien, Sonderausprägungen dieser Kriterien und Sekundäre Kriterien. Grundlegende Kriterien sind beispielsweise Konkretheit, Detailreichtum und Innere Stimmigkeit oder Folgerichtigkeit. Unter Sonderausprägungen versteht Undeutsch u.a. die Wiedergabe eigenen Erlebens, Spontane Verbeserungen und Präzisierungen und Ergänzungen. Sekundäre Kriterien, die Undeutsch auch als negative Kontrollkriterien bezeichnet, fassen die Widerspruchsfreiheit der Aussage in sich zu Sachgesetzen und zu anderen Befunden und Tatsachen zusammen. Je mehr Kriterien in einer Aussage vorhanden sind und je ausgeprägter diese sind, desto höher ist nach Undeutsch der Beweiswert dieser Aussage.

Eine weitere kriteriumsorientierte Einteilung erstellte Arntzen (1982). Er unterscheidet darin *Aussageeigenarten* und *Glaubwürdigkeitskriterien*. Unter Aussageeigenarten versteht er Merkmale, die zwar eine gewisse Ähnlichkeit mit einem Glaubwürdigkeitskriterium aufweisen können, aber von geringerer Qualität sind und auch in falschen Aussagen vorkommen können. So kann eine Aussageeigenart auf den ersten Blick einem Glaubwürdigkeitskriterium wie z.B. Detailliertheit gleichen, aber dennoch nicht die erforderliche Qualität besitzen, um als ein solches zu gelten. Der Wert einer Aussage wird aufgrund bestimmter Merkmale derart verringert, dass eine Aussageeigenart nicht zum Glaubwürdigkeitskriterium werden kann. So können zum Beispiel eine auffallende Konstanz von aufeinander folgenden Aussagen oder eine ausgeprägte Phantasiefähigkeit des Probanden wertmindernd sein. Andererseits kann eine Aussageeigenart zum Glaubwürdigkeitskriterium aufgewertet werden, wenn beispielsweise die Aussage einen größeren

Umfang besitzt oder zügig erstattet wird, wenn die Explorationsbedingungen für den Zeugen erschwerend sind oder wenn bestimmte Persönlichkeitsmerkmale (z.B. kognitive Fähigkeiten) bei der Auswertung der Details berücksichtigt werden.

Die Glaubwürdigkeitskriterien von Arntzen (1982) lassen sich in vier Gruppen gliedern: Aussageinhalt, Aussageentwicklung, Aussageweise und motivationale Aspekte. Zu den Kriterien, die sich aus dem Aussageinhalt ergeben, gehören Detaillierung, inhaltliche Besonderheiten und Homogenität der Aussage. Hier sind sowohl quantitative (Detaillierung) als auch qualitative (inhaltliche Besonderheiten) Elemente integriert. Arntzen (1982) sieht die Detaillierung als grundsätzliche Voraussetzung einer wahren Aussage und misst der Schilderung inhaltlicher Besonderheiten mehr Bedeutung bei als der bloßen Anzahl von Details. Die Homogenität der Aussage unterteilt er in die innere Homogenität, die sich in einer integrierten, inhaltlich geschlossenen Darstellung äußert und in die äußere Homogenität, die die Vereinbarkeit der Aussage mit anderen Erkenntnissen über das Erlebte zeigt.

Die zweite Gruppe von Kriterien erschließt sich aus der Aussageentwicklung. Dazu gehören zum Beispiel die Konstanz zwischen zeitlich aufeinander folgenden Schilderungen einer Person und die Aussageergänzungen, die sich widerspruchsfrei in das schon Berichtete eingliedern lassen.

Zur dritten Gruppe der Glaubwürdigkeitskriterien zählen diejenigen Kriterien, die sich aus der Aussageweise ergeben wie Inkontinenz, nacherlebte Gefühlbeteiligungen und ungesteuerte Aussageweise. Unter dem Merkmal Inkontinenz ist eine sprunghafte, bruchstückhafte Schilderung des Sachverhaltes zu verstehen. Auch eventuelle emotionale Begleiterscheinungen, die sich als nacherlebte Gefühlbeteiligungen äußern, können in einer der Persönlichkeit des Zeugen angemessenen Form als Glaubwürdigkeitskriterium gewertet werden. Eine ungesteuerte Aussageweise, zu der z.B. rasches Sprechtempo und nonverbales Verhalten gezählt werden, kann zur Steigerung der Glaubwürdigkeit beitragen.

Die letzte Gruppe von Kriterien beinhaltet motivationspsychologische Aspekte. Dabei geht es z.B. um die Beziehung des Zeugen zum Beschuldigten.

Nach Arntzen (1982) darf vom Vorhandensein eines einzelnen Glaubwürdigkeitsmerkmals nicht auf die Glaubwürdigkeit der gesamten Aussage geschlossen werden. Erst beim Auftreten dreier eindeutiger Merkmale kann die Glaubwürdigkeit als erwiesen gelten. Arntzen setzt somit als einziger Vertreter der Kriteriensysteme einen quantitativen Wert als Entscheidungskriterium fest. Er weist auch darauf hin, dass das Vorliegen von Kriterien für die Glaubwürdigkeit der Aussage spricht, dass aber umgekehrt vom Fehlen von Kriterien nicht auf die Unglaubhaftigkeit einer Aussage geschlossen werden kann.

Zu der Gruppe der kriteriumsorientierten Systeme zählt ebenfalls die Kriteriologie von Szewczyk (1973), die zehn Jahre später von Littmann und Szewczyk (1983) in aktualisierter Form veröffentlicht wurde. Die Kriteriologie ist in sieben Gruppen unterteilt. Diese sind die Vorgeschichte der Aussage, das Aussageverhalten mit und ohne Bezug zum Inhalt, die Aussagegestaltung, der Erinnerungsprozess, inhaltliche Merkmale der Aussage und der intraindividuelle Aussagevergleich. Zur Vorgeschichte der aktuellen Aussage zählen Motive und Genese der Aussage. Unter dem Kriterium Aussageverhalten ohne Bezug zum Inhalt sind das nonverbale Verhalten während der Untersuchung und die Sprachfähigkeit zu verstehen. Zum Aussageverhalten mit Bezug zum Inhalt zählen die Mitteilungsbereitschaft, das Verhalten bei Vorhalt von Widersprüchen und emotionale Begleiterscheinungen. Die Aussagegestaltung beinhaltet Berichtigungen, Verbesserungen und Selbstbelastungen des Zeugen. Zum Erinnerungsprozess gehören der Reproduktionsstil, die Suggestibilität und das Zugeben von Erinnerungsgewissheiten. Detailreichtum, Wirklichkeitsnähe und das Schildern körperlich-psychischer Vorgänge zählen zu den inhaltlichen Merkmalen der Aussage. Beim intraindividuellen Aussagenvergleich sind Konstanz, Aussageergänzungen und Aussageeinschränkungen maßgeblich. Littmann und Szewczyk betrachten die genannten Kriterien lediglich als Hinweise für den Gutachter zur Analyse der Aussage und sprechen sich gegen bestimmte Grenzen aus, von denen an eine Aussage als glaubwürdig betrachtet werden kann. Sie argumentieren „die gutachterliche Entscheidungsfindung kann also keineswegs als „additives“ Auflisten oder „ungewichtiges“ Verarbeiten der Einzelkriterien verstanden werden“ (Littmann & Szewczyk, 1983, S.66). Allerdings bieten die Autoren keine Alternative für mögliche Entschei-

dungskriterien an (Köhnken, 1990), sondern verweisen auf künftige Untersuchungen mit multivariat- statistischen Analysen.

Zwischen dem prozessorientierten Ansatz von Trankell (1971) und den kriteriumsorientierten Systemen bestehen wesentliche Unterschiede. Im Gegensatz zu Kriteriensammlungen von Arntzen (1982), Undeutsch (1993) und Littmann und Szewczyk (1983), steht bei Trankell der *Prozess* der Begutachtung im Mittelpunkt. Trankell führt weniger Kriterien als die anderen Autoren an, er legt jedoch ein genaues diagnostisches Vorgehen für die Glaubwürdigkeitsprüfung fest. Über die kriteriumsorientierten Systeme hinaus bietet Trankell in seinem Modell die Möglichkeit, bei Bedarf fallspezifische Kriterien zu generieren. Ein weiteres Merkmal, das Trankells Modell von den drei anderen Modellen unterscheidet, ist das oben beschriebene Vorgehen bei der aussagepsychologischen Realitätsanalyse.

Die drei kriteriumsorientierten Ansätze lassen sich aufgrund ihrer inhaltlichen Ähnlichkeiten deutlich von Trankells (1971) Prozessmodell unterscheiden. Sie zeigen aber auch inhaltliche Unterschiede. Während Undeutsch (1967) die einzelnen Kriterien beschreibt und durch Beispiele erläutert, stellten Arntzen und Michaelis (1970) eine Systematik von Glaubwürdigkeitsmerkmalen auf, deren Einteilungsgesichtspunkte jedoch uneinheitlich sind. Dagegen nennt Szewczyk (1973) einzelne Kriterien, die aber nur unzureichend erläutert werden.

Vor dem Hintergrund der unterschiedlichen Systematiken, der Unterschiede in Ausführlichkeit und Auswahl der Kriterien, so wie der verschiedenartigen Definitionen und Abgrenzungen einzelner Merkmale der bis dahin vorliegenden Kriteriologien ergab sich für Steller und Köhnken (1989) der Anspruch, eine geordnete und überschaubare, integrative Systematik zu entwickeln. Sie beschränkten sich dazu auf die inhaltlichen Merkmale, die sich ausschließlich aus dem Wortinhalt einer Aussage ableiten lassen und unterteilten die herausgefilterten Glaubhaftigkeitsmerkmale oder Realkennzeichen in fünf Kategorien (s. Tabelle 1). Unter der Überschrift *Allgemeine Merkmale* werden jene Realkennzeichen zusammengefasst, die sich auf eine Zeugenaussage in ihrer Gesamtheit beziehen.

Diese Merkmale können ohne Kenntnis von Einzelheiten des Aussageinhalts beurteilt werden. Sie betreffen den Reproduktionsstil, die quantitative Detaillierung und die logische Stimmigkeit der Schilderung. Gleichzeitig ist ihr Vorhandensein die Grundvoraussetzung für eine weitere Analyse des Inhalts.

Tab.1:

Realkennzeichen in der Kategorisierung von Steller und Köhnken (1989)

Allgemeine Merkmale

1. logische Konsistenz und Widerspruchsfreiheit
2. Ungeordnete 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. Ausgefallene Einzelheiten
9. Nebensächliche Einzelheiten
10. Phänomengemäße Schilderung unverstandener Handlungselemente
11. Indirekt handlungsbezogene Schilderungen
12. Eigenpsychische Vorgänge
13. Fremdpsychische Vorgänge

Motivationsbezogene Inhalte

14. Spontane Verbesserung der eigenen Aussage
15. Zugeben von Erinnerungslücken
16. Unsicherheiten bei der Darstellung
17. Selbstbelastung
18. Entlastung des Angeschuldigten

Deliktspezifische Inhalte

19. Deliktspezifische Aussageelemente

Bei den *Speziellen Inhalten* und den *Inhaltlichen Besonderheiten* kommt es auf das Ausmaß der Konkretheit und der Anschaulichkeit einer Darstellung an. Zu diesen Kategorien zählen z.B. raum-zeitliche Verknüpfungen, die Wiedergabe von Gesprächen, die Schilderung ausgefallener Einzelheiten und unverstandener Handlungselemente. *Motivationsbezogene Inhalte* sollen Rückschlüsse auf die Motivlage des Aussagenden ermöglichen. Hierzu gehören z.B. spontane Verbesserungen der eigenen Darstellung, das Zugeben von Erinnerungsgewissheiten, Selbstbelastungen und die Schilderung von den Beschuldigten entlastenden Details. Die fünfte Gruppe bilden *Deliktspezifische Inhalte*, die sich aus dem empirisch-kriminologischem Wissen über typische Begehungsformen von Delikten ableiten lassen.

Ein Vorteil dieser integrativen Systematik liegt nach Ansicht der Autoren in der logischen Struktur des Merkmalskatalogs. Diese Struktur vereinfacht die Anwendung der Kriterien in der praktischen Begutachtung und in der Forschung. Das sich aus der Kriteriologie ergebende Vorgehen ausgehend von der Betrachtung der Gesamtaussage bis hin zu immer spezifischeren Einzelheiten der Aussage kann darüber hinaus auch die diagnostische Güte verbessern und subjektive Prozesse im Beurteilungskontext reduzieren.

Greuel (1997) äußerte sich kritisch über die einseitige Fokussierung auf den Aussageinhalt bei der Glaubwürdigkeitsbeurteilung. Sie hält die ausschließliche Verwendung inhaltsorientierter Glaubwürdigkeitskriterien für bedenklich und befürchtet eine „Eingenung des Forschungshorizontes“ (Greuel, 1997, S.218). Wichtige, mit der Aussage zusammenhängende Aspekte, wie z.B. Verhaltensbeobachtung und biographische Daten blieben oft unberücksichtigt und es bestehe die Gefahr, dass durch die ausschließliche Verwendung der 19 Realkennzeichen andere „in der Praxis als hochbedeutsam bewertete Kriterien wie das der Aussagepräzisierung“ unberücksichtigt blieben (Greuel, 1997, S. 218).

Hierzu ist anzumerken, dass die von Greuel kritisierte Beschränkung auf den von Steller und Köhnken (1989) entwickelten Merkmalskatalog und die Vernachlässigung anderer wesentlicher diagnostischer Befunde von den Autoren keinesfalls intendiert war.

Steller (1989) stellte die Anwendung des Merkmalskatalogs, der in der angloamerikanischen Literatur unter dem Begriff CBCA (Criteria Based Content Analysis) zu finden ist, in der praktischen Aussagebegutachtung als *ein*, wenn auch zentrales Element, des umfassenderen Statement Validity Assessment (SVA) dar.

Das SVA bezieht neben der inhaltsanalytischen Auswertung auch die Aussage begleitende Merkmale und Hintergrundinformationen in den Beurteilungsprozess ein. In einer *Validity Checklist* finden sich vier übergeordnete Kategorien. Dieses sind *psychological characteristics, interview characteristics, motivation* und *investigate questions*. Zu den psychological characteristics gehören Merkmale wie Angemessenheit der Sprache, des Affekts und die Empfänglichkeit für Suggestionen. Zu den interview characteristics zählen die Art der Fragen (z.B. suggestiv) und die Qualität der Befragung. Unter dem Aspekt motivation sind Merkmale wie Aussagemotivation, Umstände des erstmaligen Vorbringens der Beschuldigungen und mögliche Gründe für eine Falschaussage zu verstehen. In die Kategorie investigate questions fallen Merkmale wie Übereinstimmung mit Naturgesetzen sowie mit anderen Aussagen und Befunden. In der neueren Literatur zur Theorie und Praxis der forensisch-psychologischen Begutachtung wird diese integrative diagnostische Vorgehensweise als Standard der aussagepsychologischen Begutachtung beschrieben (Greuel et al. 1998). Dieses Vorgehen fand auch Eingang in die Rechtsprechung. Im BGH-Urteil vom 30.6.1999 (1 StR 618/98) wurden die wissenschaftlichen Anforderungen an Glaubhaftigkeitsgutachten formuliert und den Realkennzeichen individuelle Bedeutung für die Entscheidung ob die Angaben einer Person auf tatsächlichem Erleben beruhen zugesprochen.

2.3 Reliabilität inhaltsbezogener Kriterien

Zentrales Element der Beurteilung der Glaubhaftigkeit einer Aussage ist die Analyse der Aussagequalität anhand inhaltlicher Merkmale, den Realkennzeichen. Anzustreben ist bei der Anwendung dieser inhaltsbezogenen Kriterien, dass verschiedene Beurteiler (Rater) beim Einschätzen desselben Sachverhaltes unabhängig voneinander zu denselben Ergebnissen kommen. Ebenso bedeutungsvoll ist, dass ein Rater bei mehrmaliger Beur-

teilung desselben Sachverhaltes zu möglichst deckungsgleichen Einschätzungen kommt. Somit ist die Frage nach der inter- und intraindividuellen Übereinstimmung von Ratern beim Erkennen und Einschätzen des Ausprägungsgrades der inhaltlichen Merkmale ein wesentlicher Aspekt bei der Qualitätsbewertung der inhaltsbezogenen Aussageanalyse.

Untersuchungen zur Validität der Realkennzeichen erfassen in der Regel zusätzlich Ergebnisse zur Reliabilität, sowohl einzelner Kriterien als auch zur Gesamtheit der Kriterien. Dabei zeigt sich, dass die dort erreichten Reliabilitäten eine erhebliche Schwankungsbreite aufweisen.

So erzielten Wellershauß und Wolf (1989) in einer experimentellen Studie für verschiedene Kriterien sehr unterschiedliche Interrater- Reliabilitäten. Bei der Betrachtung der Übereinstimmung von 31 Ratern in vier Aussagen lagen die prozentualen Häufigkeiten zwischen 40% für die Kriterien Kontextuelle Einbettung und Ausgefallene Einzelheiten und 94% für das Kriterium Einwände gegen die Richtigkeit der eigenen Aussage. Die Übereinstimmungen von drei Ratern, die 176 Aussagen zu beurteilen hatten, lagen zwischen 52% für das Kriterium Wiedergabe von Interaktionen nonverbaler Art und 100% für das Kriterium Einwände gegen die Richtigkeit der eigenen Aussage. Auffallend ist dabei, dass eindeutig zu verstehende Kriterien wie das letztgenannte höhere Reliabilitätskoeffizienten erbrachten als möglicherweise weniger abgrenzbare inhaltliche Merkmale, so dass die unterschiedliche inhaltliche Plausibilität der Kriterien eine ursächliche Rolle zu spielen scheint. Diese Annahme ähnelt der von Tully (1999) geäußerten Vermutung, die Interrater- Reliabilität hänge davon ab, inwieweit das zu untersuchende Material eindeutig und prägnant sei.

Das Problem zeigte sich schon in einer Untersuchung von Köhnken und Wegener (1982, s. Kap. 2.4.2), zur Validität einzelner Glaubwürdigkeitsmerkmale, in der ebenfalls die Interrater- Reliabilität zwischen fünf trainierten Ratern überprüft wurde. Erst nach einer nachträglichen Präzisierung der Kriterien konnte eine befriedigende Übereinstimmung zwischen den Urteilern erreicht werden. Verwendete Kriterien waren Detailreichtum, Konstanz und Inkontinenz.

Krahé und Kundrotas (1992, s. Kap 2.4.1) berechneten in ihrer Feldstudie zur Glaubwürdigkeitsbeurteilung bei Vergewaltigungsanzeigen den Kappa- Koeffizienten, der das Ausmaß, in dem die Beurteilerübereinstimmung eine gemäß dem Zufall erwartete Übereinstimmung übersteigt, aufzeigt. Jedes Vernehmungsprotokoll wurde von vier Ratern eingeschätzt. Die Kriterien wurden auf einer dreistufigen Skala mit den Ausprägungen „nein“ „ja“ und „ja sehr“ beurteilt. Der mittlere Kappa- Koeffizient lag bei .14 und ist damit als sehr schwaches Ergebnis anzusehen. In diesem Zusammenhang verweisen die Autoren auf die Wichtigkeit eines angemessenen Trainings zum Gebrauch der Realkennzeichen.

Auch Sporer und Küpper (1995) betonen nach den Ergebnissen ihrer experimentellen Studie zu den Methoden der Glaubwürdigkeitsbeurteilung die Wichtigkeit präziser Operationalisierungen und ausreichender Rater-Schulungen. Die Autoren berechneten als Maß für die Interraterübereinstimmung Korrelationen zwischen den paarweisen Ratings von jeweils zwei von drei Ratern. Dabei fanden sich Werte von -.15 bis .98

Anson, Golding und Gully (1993) verwendeten in einer Untersuchung zur Reliabilität der kriterienorientierten Inhaltsanalyse als Ausgangsmaterial 23 Videos mit Aussagen von Opfern sexuellen Missbrauchs, bei denen davon ausgegangen wurde, die Schilderungen seien erlebnisbegründet. Die Aussagen wurden von vier Beurteilern bezüglich ihres Wahrheitsgehaltes bewertet. Die Autoren berechneten zur Überprüfung der Urteiler-Übereinstimmung verschiedene Maße. Der Kappa- Koeffizient lag im Mittel bei .29, die Werte des Maxwell's RE Koeffizienten lagen durchschnittlich bei .49 (mit einer Bandbreite von .13 bis 1.00). Schließlich wurde die durchschnittliche prozentuale Häufigkeit der Übereinstimmung der Ratings berechnet, welche bei 75 % lag. Analog zu den oben zitierten Autoren kommen auch Anson et al. zu dem Ergebnis, eine eindeutige Definition und Präzisierung der Kriterien sei unabdingbar, um zufrieden stellende Reliabilitätskoeffizienten zu erreichen.

Bestätigt wurde diese Folgerung durch eine Untersuchung von Petersen (1997) zur Konzeption und Evaluation eines Trainingprogramms in der Anwendung der Glaubwürdig-

keitskriterien. Dieses Programm wurde von Höfer (1995a) entwickelt und hatte zum Ziel durch effektives Training den Gütekriterien der klassischen Testtheorie (Objektivität, Reliabilität, Validität) weitmöglich Rechnung zu tragen. Insbesondere durch die Operationalisierungen der Kriterien konnte in der Untersuchung von Petersen erreicht werden, dass die Trainingsgruppe, die eine intensive dreiwöchige praktische Schulung durchlief, signifikant höhere Interrater- Reliabilitäten und Intrarater- Reliabilitäten erreichte als die Vergleichsgruppe, die lediglich theoretisch über die Methoden der Glaubhaftigkeitsdiagnostik informiert wurden. Auch bei kürzerer Trainingsdauer von elf Tagen zeigte sich in einer Folgeuntersuchung (Eggers, 2002), dass eine intensive Ausbildung in der Anwendung der Kriterien die Beurteilungsleistungen bedeutsam verbessern konnte. Insgesamt wurden Übereinstimmungen von über 70% über alle Kriterien ermittelt.

Neben der Fehlerquelle unzureichender Operationalisierung der inhaltsbezogenen Glaubhaftigkeitsmerkmale kann auch die Vielzahl der verwendeten Reliabilitätsmaße zu Fehleinschätzungen über die Zuverlässigkeit der Realkennzeichen führen. Petersen (1997) kritisierte, dass sich aus vielen Arbeiten nur begrenzt nachvollziehen lasse, welche Reliabilitätsmaße verwendet und wie die Reliabilitäten im Einzelnen bestimmt wurden. So werde bei der Berechnung der Reliabilität in vielen Untersuchungen eine gewisse Beliebigkeit und Willkür bezüglich der verwendeten Reliabilitätsmaße an den Tag gelegt. In der Regel wird die Reliabilität durch die Berechnung der Übereinstimmung von zwei oder mehreren Ratern hinsichtlich der Beurteilung inhaltlicher Merkmale dargestellt (Interrater- Reliabilität) und üblicherweise in Prozentwerten oder als Korrelationskoeffizient angegeben.

Horowitz, et al. (1997) untersuchten in einer Feldstudie neben der Interraterreliabilität auch die Retest- Reliabilität. Drei geschulte Beurteiler unterzogen jeweils 100 Transkripte von Aussagen über sexuelle Übergriffe einer Inhaltsanalyse bezüglich der 19 Realkennzeichen und wiederholten dies nach einem Zeitabstand von längstens drei Monaten. Die Aussagen stammten von 21 männlichen und 79 weiblichen Kindern und Jugendlichen im Alter von 2 bis 19 Jahren (wobei der Median 7; 7 Jahre und der Mittelwert 8; 7 Jahre betrug). Der Wahrheitsstatus wurde bei 79 Aussagen durch medizinische Befunde, Ges-

tändnisse, Polygraph-Ergebnisse, Widerrufe sowie die Aussagen der Opfer und Abweichungen in deren Aussagen durch einen weiteren Beurteiler auf einer siebenstufigen Skala eingeschätzt. Dabei wurden 5% der Aussagen als eher nicht erlebnisbegründet bewertet. Für 21 der 100 Aussagen waren keine Informationen über den Wahrheits-Hintergrund der Aussage vorhanden. Die Berechnungen ergaben bezüglich des erhobenen Gesamtscores der Realkennzeichen pro Aussage sehr gute Reliabilitäten. Die Interrater- Reliabilität variierte zwischen .78 und .82 zum ersten Untersuchungszeitpunkt und von .86 bis .89 zum zweiten. Die Retest- Reliabilitäten lagen zwischen .85 und .91. In einer weiteren Analyse (Generalizibility) zeigte sich, dass Unterschiede in den Interrater- Reliabilitäten hinsichtlich einzelner Kriterien bestanden. Das Eingestehen von Erinnerungslücken, die Schilderungen nebensächlicher Einzelheiten und Spontane Verbesserungen der Aussagen erbrachten deutlich geringere Interrater- Reliabilitäten als Einwände gegen die Richtigkeit der eigenen Aussage, Selbstbelastung, Entlastung des Angeschuldigten, Phänomengemäße Schilderungen, Logische Konsistenz. Die Überlegung, dass mehr Worte und damit längere Aussagen der Kindern verbunden mit einer möglicherweise größeren Auftretenswahrscheinlichkeit von Realkennzeichen einen Einfluss auf die Reliabilitätsmaße haben könnten, führte zu einer weiteren Analyse, in der die Untersuchungsgruppe nach dem Median in zwei Gruppen unterteilt wurde. Erwartungsgemäß produzierten ältere Kinder mehr Wörter als jüngere ($r = .65$, $p < .05$). Es ergaben sich aber keine signifikanten Unterschiede in den Interrater- und Retest- Reliabilitäten.

Die Autoren folgern aus ihrer Studie, dass die erhaltenen Reliabilitäten die Brauchbarkeit der Realkennzeichen unterstützen, allerdings insbesondere Studien mit einer größeren Anzahl von unwahren Aussagen erforderlich sind, um aufzuzeigen, ob die Kriterien auch bei fehlendem Erlebnishintergrund inter- und intraindividuell ausreichend übereinstimmend signiert werden.

Überwiegend zeigen die stark unterschiedlichen Reliabilitäten in den dargestellten Untersuchungen, dass die inhaltlichen Kriterien grundsätzlich zwar ein zuverlässiges Instrumentarium der Aussagenanalyse sein können aber offensichtlich eine erhebliche Bedeutung darin liegt, möglichst wenig Interpretationsspielraum für die Definitionen der Kriterien zu geben. Das bedeutet, dass die Schulung der Rater in der Anwendung der Real-

kennzeichen sowohl für empirische Arbeiten als auch in der forensischen Praxis grundlegende Bedeutung erhält. Ein weiterer im empirischen Bereich zu berücksichtigender Aspekt ist, dass je nach Rahmenbedingungen der Studien unterschiedliche Reliabilitätsmaße zu unterschiedlichen Einschätzungen führen können, somit kann der beliebige Einsatz von Reliabilitätsmaßen deren Aussagekraft erheblich reduzieren. (Horowitz et al. 1997; Petersen, 1997).

2.4 Validität inhaltsbezogener Kriterien

Seit Jahrzehnten wurden Glaubwürdigkeitskriterien in der psychologischen Glaubhaftigkeits-Beurteilung von Aussagen im Rahmen der forensischen Tätigkeit von praktisch tätigen Psychologen verwendet, ohne dass diese Kriterien einer systematischen Validierung unterzogen wurden. Inzwischen liegen mehrere Studien (z.B. Boychuk, 1991, Anson et al. 1993, Wolf & Steller, 1997) vor, in denen die Validität der inhaltslichen Realkennzeichen im Hinblick auf ihre Differenzierungsfähigkeit zwischen selbst erlebten und erfundenen Schilderungen überprüft wurde. Die eingesetzten Methoden und Strategien der Validitätsüberprüfung waren dabei Feldstudien und Simulationsstudien. In Feldstudien werden authentische Zeugenaussagen aus Gerichtsakten ausgewählt und einer kriterienorientierten Glaubhaftigkeitsbeurteilung unterzogen. In Simulationsstudien wird die Variable *Erlebter vs. nicht erlebter Sachverhalt* experimentell variiert, um so die Hypothese, ob die inhaltslichen Realkennzeichen in erlebnisbegründeten Schilderungen häufiger auftreten als in Darstellungen nicht erlebter Ereignisse, zu überprüfen. Die zu Beginn der vierten Phase der Aussagepsychologie geführte Diskussion über die methodische Überlegenheit des Experiments gegenüber der Feldstudie (Arntzen, 1983; Köhnken & Wegener, 1985; Steller, 1988) wird inzwischen als überholt angesehen (Greuel, 1997). Vielmehr stehen in der nun postulierten fünften Phase der Aussagepsychologie integrative Überlegungen und die Einbeziehung von praktisch relevanten Variablen und Merkmalskombinationen im Vordergrund des Forschungsinteresses. Dennoch bilden die bisher durchgeführten

Studien die Basis für Überlegungen und Forschungsvorhaben und sollen daher im Folgenden dargestellt werden.

2.4.1 Feldstudien zur Validität inhaltlicher Realkennzeichen

Littmann und Szewczyk (1983) werteten 173 Gutachten aus dem Bereich des sexuellen Missbrauchs aus. Neben verschiedenen untersuchten Merkmalskategorien wie Deliktart, Milieu und Anamnese der Zeugen, Täter-Opfer-Beziehung, Genese der Anzeige und Aussage, Anzeige und Aussagemotivation wurde auch eine formale und inhaltliche Aussageanalyse durchgeführt. Verwendete Kriterien, die mit dem inhaltlichen Kriterienkatalog von Steller und Köhnken (1989) übereinstimmen, waren: Spontane Verbesserungen der eigenen Aussage, Selbstbelastung durch die Aussagen, Zugeben von Erinnerungsgewissheiten, Schilderung unverstandener Handlungen, Detailreichtum, Schilderungen unvorhergesehener Komplikationen im Tatgeschehen, Realistik, Wirklichkeitsnähe, Originalität und Schilderungen körperlich-psychischer Vorgänge des Zeugen/Beschuldigten bei der Tat. Die verwendete Stichprobe setzte sich aus 73 glaubwürdigen Aussagen und 60 Aussagen, deren Glaubwürdigkeit erheblich angezweifelt bzw. verneint wurde, zusammen. 40 Aussagen galten als eingeschränkt glaubwürdig. Die prozentualen Merkmalshäufigkeiten der Kriterien zeigten, dass in den 73 glaubwürdigen Aussagen die Kriterien Berichtigungen, Verbesserungen der Aussage, Selbstbelastung, Zugeben von Erinnerungsgewissheiten, Detailreichtum und Schilderung körperlich-psychischer Vorgänge des Beschuldigten bei der Tat signifikant häufiger auftraten als in der Gruppe der nicht glaubwürdigen Aussagen. Die Autoren wiesen daraufhin, dass für die Gruppe der angezweifelten Aussagen – wegen deren heterogener Zusammensetzung – keine verlässlichen Kriterien für den Realitätsgehalt dieser Aussagen vorhanden waren. Diese Kritik ist allerdings auch auf die Gruppe der wahren Aussagen auszudehnen, da auch hier keine objektiven Sachverhalte wie Tätergeständnisse oder medizinische Befunde vorlagen, sondern nur das positive Begutachtungsergebnis als Außenkriterium verwendet wurde. Die methodischen Mängel dieser Studie, insbesondere die unzuverlässige Abgrenzung der untersuchten Gruppen, erlauben keine Aussage über die Brauchbarkeit und Trennfä-

higkeit der verwendeten Merkmale, so dass diese Studie insgesamt keinen Beitrag zur Frage der Validität der Glaubwürdigkeitskriterien leistet.

Michaelis- Arntzen (1987) ließ 100 Geständnisaussagen und 74 Falschaussagen durch Beurteiler analysieren und fand einen deutlichen Unterschied in der Verteilung der Realkennzeichen in realitätsbezogenen und falschen Aussagen. Kritisch anzumerken ist, dass die Beurteiler über den Wahrheitsgehalt der Aussagen informiert waren und somit nicht auszuschließen ist, dass dieses Wissen die Analyse und Klassifizierung im Sinne der Hypothese beeinflusst hat. Steller (1988) kritisierte außerdem die mangelnde Transparenz in der Publikation von Michaelis- Arntzen , in der methodische und quantitative Angaben fast vollständig fehlten.

Esplin, Boychuk und Raskin (1988) untersuchten 40 Aussagen von Kindern im Alter von drei bis fünfzehn Jahren in Fällen sexuellen Missbrauchs. In dieser Untersuchung wurden erstmalig authentische Aussagen analysiert, die Aussagen der Kinder wurden von Tonbandaufnahmen transkribiert. In 20 Fällen war der Wahrheitsgehalt der Aussagen durch Geständnisse und/oder medizinische Befunde bestätigt worden, während in den übrigen 20 nicht erlebnisbegründeten Aussagen Befunde vorlagen, die Zweifel am Wahrheitsgehalt der Anschuldigungen bedingten (z.B. Widerruf der Anschuldigung, Abweisung der Anklage, Freispruch in der Gerichtsverhandlung). Die Aussagen wurden von einem Beurteiler, der den Wahrheitsgehalt der Aussagen nicht kannte, analysiert. Das Auftreten der neunzehn Kriterien wurde auf einer 3-Punkte Skala (abwesend, vorhanden, stark vorhanden) markiert. Für jede Aussage wurde ein Summenscore über alle Kriterien gebildet. Es zeigte sich ein signifikanter Unterschied zwischen erlebnisbegründeten und zweifelhaften Aussagen. In wahren Aussagen wurden signifikant mehr Kriterien gefunden als in unwahren Aussagen.

Bei der Betrachtung der einzelnen Realkennzeichen zeigte sich, dass sieben der neunzehn Kriterien (Wiedergabe von Gesprächen, Komplikationen im Handlungsverlauf, Ausgefaltete Einzelheiten, Indirekt handlungsbezogene Schilderungen, Psychische Vorgänge des Beschuldigten, Einwände gegen die Richtigkeit der eigenen Aussage, Selbstbelastungen) ausschließlich in den erlebnisbegründeten Aussagen vorhanden waren und zehn Kriterien

(Logische Konsistenz, Ungeordnete Darstellung, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Nebensächliche Einzelheiten, Eigenpsychische Vorgänge, Spontane Verbesserungen, Erinnerungslücken, Deliktspezifische Einzelheiten) in der Gruppe der wahren Aussagen häufiger auftraten als in der Gruppe der zweifelhaften Aussagen. Die Autoren werteten ihre Ergebnisse als Beleg für die Undeutsch-Hypothese, nach der qualitative Unterschiede zwischen wahren und falschen Aussagen vorhanden sind. Die darüber hinausgehende Schlussfolgerung, dass in erlebnisbegründeten Aussagen grundsätzlich das Vorhandensein der meisten der Kriterien zu erwarten ist, während es einige Kriterien gibt, die in falschen Aussagen nicht auftreten, ist als Überinterpretation zu bewerten, die sich in keiner späteren Untersuchung bestätigte. Anzumerken ist, dass keinen der nachfolgenden experimentellen Untersuchungen eine so große Anzahl von Kriterien mit einer trennenden Funktion zwischen wahren und falschen Aussagen aufzeigte.

Wells und Loftus (1991) Kritik, dass bei der oben beschriebenen Studie die zweifelhaften Aussagen möglicherweise irrtümlich als unglaublich eingeschätzt und Alters- und Entwicklungsvariablen nicht berücksichtigt wurden, führte zu einer Reanalyse der Befunde aus dieser Studie (Raskin & Esplin, 1991), die allerdings keine neuen, bzw. weiterreichenden Ergebnisse erbrachte. Zwei Versuchspersonen aus der Gruppe der zweifelhaften Aussagen wurden eliminiert und es wurde eine Kombination der Außenkriterien vorgenommen, indem mindestens zwei von drei Kriterien (Vorhandensein medizinischer Befunde, Widerruf der Anschuldigung, negatives Polygraphergebnis des Beschuldigten) erfüllt sein mussten, um in die Gruppe der zweifelhaften Aussagen aufgenommen zu werden.

In einer Feldstudie von Boychuk (1991) wurden 75 transkribierte Aussagen von Kindern im Alter zwischen vier und sechzehn Jahren untersucht. 50 Aussagen wurden der Kategorie der erlebnisbegründeten Aussagen zugeordnet. Diese Gruppe wurde unterteilt, indem bei einem Teil der Gruppe die Außenkriterien Geständnisse und Arztbefunde erfüllt sein mussten, während der andere Teil lediglich aufgrund von Geständnissen in diese

Gruppe eingeordnet wurde. In den übrigen 25 Fällen war der Vorwurf des sexuellen Missbrauchs zuvor widerlegt worden.

Zwölf der neunzehn Kriterien (Logische Konsistenz, Ungeordnete Darstellung, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Wiedergabe von Gesprächen, Komplikationen im Handlungsverlauf, Indirekt handlungsbezogene Schilderungen, Eigenpsychische Vorgänge, Spontane Verbesserungen, Deliktspezifische Inhalte) traten in wahren Aussagen signifikant häufiger auf als in den unwahren Schilderungen.

Als weitere Variablen wurden das Alter der Kinder, die Missbrauchsdauer und die Anzahl der durchgeführten Befragungen untersucht. Es zeigte sich, dass die Kriterien Interaktionsschilderungen und Eingeständnis von Erinnerungslücken signifikant häufiger in den Schilderungen älterer Kinder auftraten. Weiter wurde festgestellt, dass mit zunehmender Befragungshäufigkeit die Anzahl kriterienerfüllender Textstellen abnahm und dass in den Aussagen mehrfach missbrauchter Kinder bestimmte Realkennzeichen (Logische Konsistenz, Ungeordnete Darstellung, Detailreichtum, Raum-zeitliche Verknüpfungen, „Indirekt handlungsbezogene Schilderungen, Spontane Verbesserungen, „Deliktspezifische Inhalte) signifikant häufiger auftraten als bei einmaligem Missbrauch.

Vrij (2002) kritisierte an dieser Untersuchung, dass zwar versucht wurde, Schwächen der früheren Untersuchung von Esplin et al. (1988) zu vermeiden, dass aber durch den partiellen Verzicht von medizinischen Befunden in der Gruppe der missbrauchten Kinder und durch die gemeinsame statistische Verrechnung beider Untergruppen die angestrebte Versuchsbedingung nicht ausreichend realisiert wurde.

Alterseffekte in Bezug auf die Realkennzeichen wurden auch in einer Studie von Lamers-Winkelmann, (1992a, 1992b, 1995) gefunden. Die Aussagen von 103 Kindern im Alter zwischen zwei und elf Jahren, bei denen der Verdacht des sexuellen Missbrauchs bestand, wurden in drei Gruppen unterteilt. Bei 36 Kindern wurde der sexuelle Missbrauch aufgrund von Tätergeständnissen und/oder der Überführung des Täters als erlebnisbegründet eingestuft. Bei 61 Kindern wurde ein sexueller Missbrauch als sehr wahrscheinlich und bei sechs Kindern als unbegründet angenommen. Diese Einordnung erfolgte aufgrund der Schilderungen der Kinder, medizinischer Befunde, Verhaltensauffälligkeiten

und des Vorhandenseins von Zeugen. Für jedes transkribierte Interview wurden Auftretenshäufigkeiten und Gesamtscores aus den Ausprägungen der Realkennzeichen in den Aussagen der Kinder gebildet. Während sich im Gesamtscore keine Unterschiede über die drei Gruppen zeigten und auch keine erfolgreiche Trennung der Gruppen mittels der Kriterien gelang, konnten jedoch bedeutsame Alterseffekte festgestellt werden. Bei älteren Kindern, bei denen der sexuelle Missbrauch als belegt galt, wurden mehr Kriterien sowie höhere Gesamtscores gefunden als bei jüngeren Kindern dieser Gruppe. In den Schilderungen acht- bis elfjähriger Kinder traten die Kriterien Interaktionsschilderungen, Wiedergabe von Gesprächen, Nebensächliche Einzelheiten und Raum-zeitliche Verknüpfungen signifikant häufiger auf als bei den vier- bis fünfjährigen Kindern. Insgesamt ist aus dieser Studie, die durch das Fehlen einer zahlenmäßig ausreichend besetzten Kontrollgruppe methodisch zu kritisieren ist und daher eher weniger Aussagekraft für die Überprüfung der Validität der Realkennzeichen besitzt, zu folgern, dass das Alter der Versuchspersonen als zu kontrollierende Variable in Studien zwingend berücksichtigt werden sollte.

Anson et al. (1993) überprüften neben der oben genannten Reliabilität auch das Auftreten der neunzehn Realkennzeichen. Sie untersuchten 23 Aussagen von vier- bis dreizehnjährigen Kindern, bei denen der Vorwurf des sexuellen Missbrauchs durch Geständnisse bestätigt wurde. Basismaterial für die Untersuchung waren Videoaufnahmen von Interviews. Vier in der Anwendung der inhaltlichen Aussagenanalyse trainierte Rater waren blind hinsichtlich des Wahrheitsgehaltes der Aussagen und der Anzahl wahrer Schilderungen. Jede Aussage wurde von zwei Ratern beurteilt. Die Auftretenshäufigkeiten der gefundenen Kriterien zeigten eine Streubreite von 0% (Selbstbelastung) bis 91% (Logische Struktur), der Durchschnitt lag bei 41%.

Insgesamt waren die in dieser Studie gefundenen Kriterien deutlich weniger ausgeprägt als in der „wahren“ Gruppe der Untersuchung von Esplin et al. (1988). Die Autoren begründeten dies u.a. mit der Art des Basismaterials und den verschiedenen Interviewstrategien, die innerhalb ihrer Untersuchung bei der Erhebung der Aussagen angewendet wurden sowie mit der Beschränkung auf bestätigte Fälle. Die Ergebnisse zeigten ebenfalls einen Alterseffekt: Sechs Kriterien (Logische Konsistenz, Raum-zeitliche Verknüp-

funzen, Interaktionsschilderungen, Wiedergabe von Gesprächen, Entlastung des Beschuldigten und Deliktspezifische Inhalte) korrelierten positiv mit dem Alter der Probanden und mit dem Gesamtscore der gefundenen Kriterien. Allerdings besitzt diese Studie für die Frage der Validität der Glaubwürdigkeitskriterien aufgrund des Fehlens einer Kontrollgruppe, bzw. einer Gruppe ohne Erlebnishintergrund wenig Aussagekraft.

Krahé und Kundrotas (1992) führten eine Feldstudie mit erwachsenen Probanden durch. Dabei wurden dreißig polizeiliche Vernehmungsprotokolle von Vergewaltigungsanzeigen von zwei Beurteilergruppen bewertet. Fünfzehn dieser Aussagen wurden zuvor aufgrund der Widerrufe der Anzeigenden als unglaublich eingestuft und 15 Aussagen wurden aufgrund von Täter-Geständnissen und Sachbeweisen als glaubhaft eingestuft. In der ersten Beurteilergruppe schätzten 31 Polizeibeamte den Wahrheitsgehalt von jeweils vier Aussagen mithilfe des Kriterienkatalogs von Steller und Köhnken (1989) ein, ohne eine spezielle Schulung in der Anwendung dieser Kriterien erhalten zu haben. Die einzelnen Kriterien wurden auf einer dreistufigen Skala eingeschätzt (mit den Ausprägungen: „nein“, „ja“, „ja sehr“) und die Beurteiler hatten ein globales Urteil über den Wahrheitsgehalt der Aussagen abzugeben. In der zweiten Gruppe urteilten 22 Beamte intuitiv lediglich aufgrund der Vernehmungsinhalte über den Wahrheitsgehalt des behaupteten Geschehens.

In die Auswertung gingen 120 Beurteilungen ein, nachdem jedes der 30 Protokolle von vier Beurteilern bewertet wurde. Die Kriterien Logische Konsistenz und Schilderungen eigener psychischer Vorgänge traten signifikant häufiger in wahren Aussagen auf. Das Kriterium Schilderung ausgefallener Einzelheiten wurde dagegen signifikant häufiger in der Gruppe der Falschaussagen gefunden. Neun Kriterien waren zwar nicht signifikant aber deutlich häufiger in den Falschaussagen zu finden, als in wahren Aussagen. Für die Beurteilung des Wahrheitsgehaltes der gesamten Aussagen ergaben sich keine statistisch bedeutsamen Unterschiede bei Anwendung des Kriterienkataloges bzw. ohne dessen Hilfe. Die Trefferquote unter Verwendung der Kriterien betrug 74,2 % gegenüber 63,3% bei intuitivem Urteil.

Die diskriminanzanalytische Verrechnung der Daten zeigte, dass auf der Grundlage der Real kennzeichen 80,8 % der Fälle korrekt gemäß den a- priori- Klassifikationen „wahr“

oder „falsch“ zugeordnet werden konnten. Stufenweise Diskriminanzanalysen zeigten darüber hinaus, dass das Vorhandensein der Kriterien Logische Konsistenz, Ausgefallene Einzelheiten, Selbstbelastung, Fremdpsychische Vorgänge, Deliktspezifische Inhalte, Erinnerungslücken und Komplikationen im Handlungsverlauf die Zuordnung in die wahre Gruppe ausmachten, während in der falsch aussagenden Gruppe die Kriterien Logische Konsistenz, Unverstandene Handlungselemente, Selbstbelastung, Komplikationen im Handlungsverlauf und Spontane Verbesserungen die Zuordnung bestimmten.

Gruppiert nach wahrem vs. falschem Urteil ergab sich nach diskriminanzanalytischer Berechnung eine Kombination von acht Variablen (Logische Konsistenz, Ausgefallene Einzelheiten, Unverstandene Handlungselemente, Komplikationen im Handlungsverlauf, Spontane Verbesserungen, Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen), die ausschlaggebend für das Urteil waren. Lediglich fünf dieser Kriterien (Logische Konsistenz, Ausgefallene Einzelheiten, Komplikationen im Handlungsverlauf, Spontane Verbesserungen, Interaktionsschilderungen) wurden auch bei der Analyse der a- priori klassifizierten Aussagen gefunden.

Die Folgerung der Autorinnen, dass es ohne nennenswerten Informationsverlust möglich sei, den ursprünglichen Katalog von 19 Realkennzeichen auf eine sparsamere Anzahl zu reduzieren, erscheint jedoch wenig sinnvoll, da die inhaltlichen Kriterien vom Basismaterial abhängig sind und somit je nach Aussageinhalt unterschiedliche Kriterien relevant werden können. Der Nutzen einer Reduktion des Kriterienkataloges, der für die praktische Handhabung bereits einen überschaubaren Umfang hat, erscheint ebenfalls zweifelhaft. Die Aussagekraft dieser Untersuchung ist aufgrund gravierender methodischer Mängel ohnehin erheblich eingeschränkt. So ist das Basismaterial der Untersuchung (polizeiliche Vernehmungsprotokolle) für eine Analyse nach inhaltsbezogenen Kriterien kaum geeignet, da es sich nicht um Wortprotokolle handelt, sondern um zusammengefasste Vernehmungen, bei denen erfahrungsgemäß erhebliche Informationsverluste und Missverständnisse zwischen Fragenden und Aussagenden nicht ausgeschlossen werden können. Die Wahl von eingestandenen Falschaussagen als einziges Außenkriterium kann darüber hinaus auch zu erheblichen Vermischungen der Untersuchungsgruppen geführt haben, da es sich um falsche Aussagewiderrufe handeln

könnte (Volbert, 1992), oder aber auch teilweise wahr und teilweise falsch berichtet worden sein könnte.

Als Anregung für weitere Untersuchungen ist der Einsatz diskriminanzanalytischer Verfahren im Hinblick auf die fehlende theoretische Begründung der Wirksamkeit bzw. Validität der Kriterien zu sehen. Möglicherweise kann das Auffinden von Kriterienkombinationen die Hypothesenbildung hinsichtlich der Klärung inhaltlicher Gründe für die Brauchbarkeit der Realkennzeichen anregen.

Eine Feldstudie mit 98 Aussagen von Kindern im Alter zwischen vier und dreizehn Jahren ($M=8,72$ Jahre, $S=2,35$) zur Validität der inhaltlichen Realkennzeichen wurde von Lamb et al.(1997) durchgeführt.

In die inhaltliche Analyse wurden 14 Kriterien aufgenommen. Diese waren: Logical structure and coherence, Unstructured production with spontaneous digressions, Quantity of details, especially regarding time, place, persons, and events, Embedding of events in temporal and spatial context, Descriptions of interactions, Reproduction of conversation, Unexpected complications or interruptions, Unusual details that are meaningful, Superfluous or peripheral details, Accurately-reported details misunderstood, References to other sexually toned events occurring outside the specific incident, References to one's feelings or thoughts during the incident, Attributions of thoughts, feelings or motivations to the perpetrator, Spontaneous corrections or additions.

Die Beschränkung auf diese 14 „memory- related criteria“ (S.256) begründen die Autoren mit Voruntersuchungen, in denen sich gezeigt habe, dass bei den fünf ausgelassenen Kriterien nur geringe Interraterreliabilitäten vorhanden gewesen seien. Um die Probleme früherer Feldstudien mit der Validität in Bezug auf das Außenkriterium Wahrheitsgehalt der Anschuldigung zu reduzieren, wurde eine Merkmalskombination aus mehreren Außenkriterien erstellt, die Aussage des Kindes wurde dabei nicht einbezogen. Die Aussagen der Kinder wurden auf einer fünfstufigen Skala die Wahrscheinlichkeit betreffend, dass ein sexueller Missbrauch stattgefunden hat, eingestuft. 76 Aussagen wurden als glaubhaft bzw. sehr glaubhaft beurteilt, 13 Aussagen als unwahrscheinlich bzw. sehr unwahrscheinlich und 9 Aussagen als fraglich.

Die Auswertung zeigte signifikante Mittelwertunterschiede bei der Betrachtung der Gesamtscores der Realkennzeichen, die auf einer zweistufigen Skala (vorhanden vs. nicht vorhanden) signiert wurden. Die Kriterien Unstructured production with spontaneous digressions, Quantity of details, Embedding of events in temporal and spatial context, Descriptions of interactions und Reproduction of conversation, traten signifikant häufiger in der Gruppe der erlebnisbegründeten Aussagen auf. Logical structure and coherence trat in allen Gruppen gleichermaßen auf. Es zeigte sich zudem ein signifikanter Alterseffekt: In den Schilderungen älterer Kinder fanden sich mehr Kriterien als in den Ausführungen jüngerer Kinder.

Die Autoren werteten ihre Ergebnisse als ernüchternd für die praktische Anwendung der inhaltlichen Analyse in der forensischen Praxis. Entgegen ihren Erwartungen, die sich aus früheren Feldstudien (Boychuk, 1991; Raskin & Esplin, 1991) ableiteten, in denen sieben von neunzehn Kriterien ausschließlich in der Gruppe der wahren Aussagen auftraten und zehn weitere Kriterien in dieser Gruppe signifikant häufiger auftraten als in unwahren Aussagen, waren die hier gefundenen Kriterien auch in der nicht erlebnisbegründeten Gruppe ausgeprägt vorhanden. Offensichtlich waren die Autoren überzeugt, einen *cut-off-Wert* des Gesamtscores festlegen zu können, der erlaube, eine Klassifizierung aufgrund dieses Zahlenwertes vorzunehmen. Eingeschränkt wird die Aussagekraft dieser Untersuchung durch das zahlenmäßige Ungleichgewicht zwischen erlebten (Gruppengröße: 76) und nicht erlebten (Gruppengröße: 13) Darstellungen, das für eine vergleichende statistische Analyse eine ungünstige Voraussetzung bietet.

Craig, Scheibe, Raskin, Kirchner und Dodd (1999) untersuchten neben dem Einfluss verschiedener Interviewtechniken (offene vs. direkte Fragen) auf das freie Berichten von sexuellen Übergriffen auch das Auftreten von Realkennzeichen. Achtundvierzig Aussagen von Kindern zwischen drei und sechzehn Jahren wurden von in der Anwendung der Realkennzeichen trainierten Polizeibeamten analysiert. Außenkriterien für die Feststellung des Erlebnisgehaltes der Aussagen waren das Ergebnis des Beschuldigten im Polygraphtest, das Vorliegen eines Geständnisses oder die Rücknahme der Beschuldigung seitens des Kindes. In erlebnisbegründeten Schilderungen wurden signifikant mehr Realkennzeichen gefunden als in Berichten ohne Erlebnishintergrund.

Es zeigte sich wie in den vorangegangenen Studien zudem ein deutlicher Alterseffekt: In den Aussagen zehn bis sechzehnjähriger Kinder wurden signifikant mehr Kriterien gefunden als in den Schilderungen drei bis neunjähriger Kinder. Deutlich wurde auch die Überlegenheit von offenen Fragen zum Erreichen von Schilderungen gegenüber der Verwendung von direkten Fragen. In Darstellungen, die auf offene Fragen erfolgten, wurden signifikant mehr Kriterien gefunden als in den übrigen. Dabei konnte auch signifikant zwischen erlebnisbegründeten und nicht erlebnisbegründeten Schilderungen unterschieden werden, es fanden sich mehr Kriterien in den erlebnisbasierten Aussagen.

Buck, Warren, Betman und Brightman (2002) untersuchten in ihrer Feldstudie mit 104 Kinder zwischen zwei und vierzehn Jahren das Auftreten der 19 Realkennzeichen in Abhängigkeit vom Alter der Aussagenden, die in sechs Altersbereiche unterteilt wurden. Es zeigten sich signifikante Unterschiede in den Gruppen. So war der Gesamtwert über alle Kriterien bei den jüngeren Kindern signifikant niedriger als bei den älteren. Auch hinsichtlich des Auftretens einzelner Kriterien gab es Unterschiede: In Schilderungen der zwei- bis drei- und vierjährigen Kindern wurden signifikant weniger Kriterien gefunden als in den Aussagen der sieben- bis vierzehn- bzw. neun- bis vierzehnjährigen. Dabei waren einige Kriterien zudem in den Aussagen älterer Kinder teilweise überhaupt nicht vorhanden, wie Selbstbelastung, Entlastung des Angeschuldigten, Einwände gegen die Richtigkeit der eigenen Aussage und Schilderung psychischer Vorgänge des Täters. Die Autoren folgern aus diesem Ergebnis, dass die Anwendung der Inhaltsanalyse möglicherweise jüngere Kinder benachteilige, indem ihnen fälschlicherweise die Erlebnisbasis abgesprochen werde. Dies könne u.a. auch auf der Tatsache beruhen, dass sie in einem freien Interview weniger umfangreiche Schilderungen lieferten als ältere Kinder und daher die Auftretenswahrscheinlichkeit von Kriterien niedriger sei. Allerdings führen die Autoren einen gravierenden Mangel ihrer Studie selbst an: Ihnen fehlten Informationen über den tatsächlichen Wahrheitsgehalt der untersuchten Aussagen, so dass mit einer noch größeren Unsicherheit als in der Studie von Boychuk (1991) offen bleiben muss, ob die angestrebte Untersuchungsbedingung (vorhandene Erlebnisbasis) erreicht werden konnte. Das bedeutet, dass das Auftreten bzw. Fehlen von Kriterien möglicherweise durch eine unter-

schiedliche tatsächliche Erlebnisbasis (vorhanden vs. nicht vorhanden) verursacht war und nicht durch das Alter der Versuchspersonen.

2.4.2 Überlegungen zur Dienlichkeit von Feldstudien in der aussagepsychologischen Forschung

Feldstudien gehen im Bereich der aussagepsychologischen Beurteilung mit dem grundlegenden Problem einher, dass ein valides, objektives Außenkriterium zur Bestimmung der Erlebnisgrundlage einer Aussage in der Regel fehlt. Geständnisse der Täter als Außenkriterium wirken zwar auf den ersten Blick überzeugend, die nähere Betrachtung von Geständnisfällen zeigt aber, dass es auch zu falschen Geständnissen kommen kann (Peters, 1972; Wegener, 1992; Littmann, 1992; Volbert, 1992).

Vrij (2002) problematisiert die Abhängigkeit eines Geständnisses von der Qualität der Aussage. So führt unabhängig vom tatsächlichen Wahrheitsstatus eine geringe Aussagequalität zu einer geringeren Auftretenswahrscheinlichkeit von Geständnissen als eine höhere inhaltliche Qualität, da taktische Erwägungen auf Seiten des Beschuldigten eine Rolle spielen können, z.B. im Falle einer inhaltlich wenig differenzierten Aussage trotz Schuld keinesfalls ein Geständnis abzulegen, bzw. trotz Unschuld ein Geständnis abzulegen, um eine geringere Strafe zu erhalten.

Einen Lösungsansatz bietet die oben beschriebene Untersuchung von Lamb et. al. (1997), in der als mehrdimensionales Außenkriterium u.a. medizinische Befunde, Geständnisse, Täter-Opfer-Beziehungen und Polygraph-Ergebnisse erhoben wurden, um eine Zuordnung zu den Versuchsgruppen vornehmen zu können. Allerdings sind gerade medizinische Befunde, die augenscheinlich eine hohe Validität haben, zum einen nur selten in sexuellen Missbrauchsfällen vorhanden, zum anderen sind sie oft mehrdeutig und bieten keinen eindeutigen Ursache-Befund- Zusammenhang.

Ein weiterer Nachteil von Feldstudien ist, dass die Bedingungen nicht oder nur ansatzweise kontrolliert werden können. So fällt auf, dass mögliche Moderatorvariablen wie kognitive und verbale Fähigkeiten oder das Geschlecht in die bisherigen Feldstudien

nicht ausreichend einbezogen wurden. Auch ist in einigen Feldstudien zwar ein sehr weiter Altersbereich vorhanden, einzelne Altersgruppen sind allerdings nur sehr gering vertreten und schränken damit die Generalisierbarkeit der Ergebnisse deutlich ein. Dies betrifft insbesondere die Vorschul- und Kindergartenkinder.

Weitere Kritikpunkte an den dargestellten Feldstudien beziehen sich auf die Datenerhebung, z.B. auf die Befragungstechniken, mit denen die Aussagen erhoben werden. Mehrere Untersuchungen im Bereich der Suggestionsforschung belegen, dass u.a. die Anzahl der Befragungen und die verwendete Fragetechnik einen erheblichen Einfluss auf den Aussageinhalt haben können. So stellte Boychuk (1991) in ihrer Feldstudie fest, dass mit zunehmender Befragungshäufigkeit weniger kriterienerfüllende Textstellen gefunden wurden. Anson et al. (1993) sehen in dem Einsatz verschiedener Interviewtechniken in ihrer Untersuchung ebenfalls eine mögliche Ursache für ihre von früheren Studien abweichenden Ergebnisse. Auch Lamb et al. (1994) nannten Qualitätsunterschiede der in den verschiedenen Untersuchungen zur Validität der Aussageanalyse durchgeführten Interviews als mögliche Ursache für die Tatsache, dass nicht in allen Untersuchungen signifikante Unterschiede zwischen wahren und nicht erlebnisbegründeten Schilderungen vorhanden waren. Buck et al. (2002) versuchten in ihrer Untersuchung unterschiedliche Befragungsstile mit einzubeziehen, indem sie das Auftreten der Realkennzeichen unterschiedlich bewerteten, je nachdem, ob ein Realkennzeichen nach einer offenen oder nach einer direkten Frage auftrat. Es zeigte sich, dass bei der Verwendung von direkten Fragen sowohl bei jüngeren als auch bei älteren Kindern mehr Kriterien signiert wurden.

Weiterhin ist auch das verwendete Stimulusmaterial zu beachten. In den Studien von Esplin et al. (1988) und Boychuk (1991) wurden Wortprotokolle der Aussagen als Basismaterial verwendet. Dieses Vorgehen ist als zwingend notwendige Voraussetzung für die Aussagekraft und Generalisierbarkeit der Ergebnisse zu sehen, da, wie in der Studie von Krahé und Kundrotas (1992), andernfalls von konfundiertem Ausgangsmaterial auszugehen ist. So sind z.B. Kriterien wie das Zugeben von Erinnerungslücken und das Schildern nebensächlicher Einzelheiten in zusammengefassten Protokollen wahrscheinlich nicht zu finden, da sie für die polizeiliche Rekonstruktion des Tatgeschehens aufgrund ihres vermeintlich geringen Informationsgehaltes nicht in das Protokoll aufgenommen werden.

Ungeachtet der Einschränkungen der Aussagekraft von Feldstudien, haben sie den Vorteil der Authentizität des Aussagematerials. Diesem Vorteil kommt gerade im forensischen Bereich eine besondere Bedeutung zu. Die Anwendung der Kriterien in der praktischen Begutachtung erstreckt sich in der überwiegenden Zahl der Fälle auf Schilderungen aus dem Bereich des sexuellen Missbrauchs. Somit kann das in experimentellen Studien verwendete Aussagematerial naturgemäß nur eine Annäherung an die Realität sein mit der Folge, dass die Generalisierbarkeit der Ergebnisse eingeschränkt ist.

In einer experimentellen Untersuchung von Wellershaus und Wolf (1989) wurde diese Problematik aufgegriffen. Die Autoren versuchten, die einem sexuellen Missbrauch zu Grunde liegenden Variablen wie Kontrollverlust, negative emotionale Tönung etc. zu simulieren, um eine weitgehende Annäherung an das reale Erleben zu erreichen. In einer anderen Studie (Wolf & Steller, 1997) wurden die eben angeführten Variablen durch Berichte über tatsächliches Geburtserleben von Frauen zumindest hinsichtlich des Kontrollverlusts und der starken affektiven Komponente des Geschehens eindrucksvoll abgebildet.

Bei der Überlegung, welchen Wert und welche Folgerungen sich aus den dargestellten Untersuchungen für die Validität der kriterienbezogenen Inhaltsanalyse ableiten lassen, ist kein einheitliches Ergebnis zu erlangen. Zum einen zeigen die Resultate der Feldstudien eine weite Spanne vom deutlichen Auftreten der Kriterien in erlebnisbegründeten (Boychuk, 1991) bis zum signifikant häufigerem Auftreten in falschen Aussagen (Krahé & Kundrotas, 1992). Zum anderen zeigt die kritische Auseinandersetzung mit den Studien, dass die erzielten Ergebnisse aufgrund der stark unterschiedlichen methodischen Grundlagen der Feldstudien, die von methodisch sehr sorgfältig angelegten Studien (z.B. Anson et al., 1993) reichen bis hin zu Studien, die gravierende Mängel z.B. bezüglich der Realisierung der zu untersuchenden Variablen (wie bei Krahé & Kundrotas, 1992) aufweisen, hinsichtlich ihrer Aussagekraft sehr unterschiedlich zu bewerten sind.

Abzuleiten ist aus den dargestellten Studien, dass die aus der praktischen Begutachtung entwickelten Realkennzeichen in ihrer Bandbreite bei systematischem Einsatz offensichtlich Bestand haben. Hier zeigt sich die partielle Überlegenheit von Feldstudien gegenüber experimentellen Studien hinsichtlich der ökologischen Validität. Die Anwendung der inhaltlichen Aussagenanalyse erfolgt überwiegend im Bereich des sexuellen Missbrauchs und damit in einem Bereich, der in experimentellen Untersuchungen aus ethischen Gründen nicht simuliert werden kann. Somit bieten Feldstudien hier eine wichtige inhaltliche Ergänzung. Die ethischen Grenzen führen aber auch in Feldstudien zu einer weiteren Einschränkung der methodischen Güte: Zumindest für den Altersbereich der Kinder können keine Kontrollgruppen eingesetzt werden, da kein Stimulusmaterial aus dem Bereich des sexuellen Missbrauchs geboten werden kann (wie z. B. das Betrachten eines sexuell getönten Filmes). Möglich ist allerdings die Verwendung von erwiesenen Falschbeschuldigungen als Kontrollgruppe.

Eindeutig zeigen die Ergebnisse der beschriebenen Feldstudien, dass erhebliche Alterseffekte im Zusammenhang mit den Kriterien auftreten. In Aussagen von jüngeren Kindern wurden weniger Kriterien gefunden als bei älteren. Die Folgerung, dass die Methode der inhaltsbezogenen Analyse für jüngere Kinder daher nicht geeignet sei, greift aber zu kurz, da im Einzelfall auch *ein* signiertes Kriterium erhebliche Aussagekraft haben und neben anderen Informationsquellen eine Beurteilung des Wahrheitsgehaltes ermöglichen kann. Unbestritten bleibt der Verdienst der ersten Feldstudien (Esplin et al., 1988; Boychuk, 1991) in diesem Bereich, da sie die für die praktische Aussagebeurteilung zentrale Undeutsch-Hypothese, nach der qualitative Unterschiede zwischen erlebnisbegründeten und phantasiegeleiteten Schilderungen bestehen, erstmals systematisch untersuchten und deren Gültigkeit, zwar eingeschränkt durch methodische Mängel, prinzipiell bestätigen konnten.

2.4.3 Experimentelle Untersuchungen der inhaltlichen Realkennzeichen

Eine der ersten Experimentalstudien zur Überprüfung der Validität der Glaubwürdigkeitskriterien wurde von Köhnken und Wegener (1982) durchgeführt. Aus der Kriteriensammlung von Arntzen (1982) wurden die Kriterien Detaillierung und inhaltliche Besonderheiten, Konstanz der Aussage über mehrere, zeitlich auseinander liegende Befragungen und Inkontinenz im Sinne einer unzusammenhängenden, sprunghaften Aussageweise ausgewählt. Es wurde geprüft, ob ihr Ausprägungsgrad in wahren Aussagen größer ist als in unwahren Darstellungen. Die Stichprobe bestand aus 37 weiblichen Versuchspersonen im Alter von sechzehn bis siebzehn Jahren. Der Erlebnisgruppe wurde ein Film gezeigt mit der expliziten Aufforderung, dessen Inhalt später wiederzugeben. Thema des Films waren verbale Auseinandersetzungen zwischen Familienmitgliedern beim Abendessen. Die Vergleichsgruppe erhielt kurze verbale Informationen über den Filminhalt mit dem Auftrag, daraus eine zusammenhängende Geschichte zu bilden. Die erhobenen Aussagen wurden von geschulten Ratern analysiert. Lediglich das Kriterium Detailreichtum konnte zwischen den auf Beobachtung und den nur auf verbalen Informationen beruhenden Aussagen differenzieren. Das Kriterium Unzusammenhängende, sprunghafte Darstellungsart, bzw. Inkontinenz hingegen trat häufiger in der Vergleichsgruppe auf. Auch hinsichtlich der Konstanz der Darstellungen über zwei Befragungen ergaben sich keine Unterschiede zwischen den Gruppen. Es bestanden lediglich niedrige Interrater-Reliabilitäten zwischen den fünf Ratern. Die Generalisierbarkeit der Ergebnisse wird durch die Wahl des Untersuchungsparadigmas in Frage gestellt, das keinen Vergleich bzw. keine Übertragung der Ergebnisse auf reale Missbrauchsfälle erlaubt. Zum einen wurden in der Erlebnisgruppe keine tatsächlich erlebten Sachverhalte gewählt, sondern lediglich die Beobachtung eines Geschehens in Form eines Films, zum anderen ist anzunehmen, dass Schilderungen über einen Familienstreit von den Versuchspersonen auch aufgrund eigener Erfahrungen gegeben werden konnten. Daher ist nicht davon auszugehen, dass in dieser Gruppe Personen ohne einen tatsächlichen Erlebnishintergrund repräsentiert sind und damit für die Überprüfung der Annahme qualitativer Unterschiede zwischen Schilderungen selbst erlebter Ereignisse und nicht selbst erfahrener Geschehnisse geeignet sind.

Rüth- Bemelmans (1984) experimentelle Untersuchung mit Schülern einer siebten Klasse (ca. 12 – 13-Jährige) sollte die Validität der Glaubwürdigkeitskriterien anhand eines Merkmalskatalogs von sechzehn Kriterien überprüfen. Dreiundfünfzig Versuchspersonen hatten die Aufgabe, sich in einer Befragung als Katzenbesitzer darzustellen. Etwa die Hälfte der Versuchspersonen waren tatsächlich Katzenbesitzer und hatten somit einen realen Erlebnishintergrund für ihre Schilderungen. Neben den 16 Kriterien wurden die Variablen Gesamteindruck und etwaige Verhaltensauffälligkeiten des Aussagenden erhoben. Die transkribierten Interviews wurden hinsichtlich des Auftretens der inhaltlichen Kriterien untersucht und jede Aussage wurde nach ihren Wahrheitsgehalt(„wahr“ vs. „unwahr“ klassifiziert. Außerdem gingen der Verhaltens- und Gesamteindruck in die Einschätzung mit ein. Es wurden 92,5% der Aussagen richtig klassifiziert. Kritisch ist in dieser Studie die mögliche Konfundierung von Verhaltensvariablen mit den inhaltlichen Realkennzeichen zu sehen. Somit kann nicht sicher geschlossen werden, ob vorrangig die inhaltlichen Merkmale die überwiegend richtige Zuordnung ermöglichten oder der subjektive Eindruck des Beurteilenden.

Eine experimentelle Studie mit siebzehn vier- bis zehnjährigen Kindern führten Honts, Peters, Dewitt und Amato (1992) durch. Die Versuchspersonen wurden "angestiftet", nach dem Erleben eines tatsächlich beobachteten bzw. verbal übermittelten Diebstahls das Geheimnis über die Person des Diebes zu wahren bzw. eine fälschliche Beschuldigung zu erheben. Die Beurteilung der Aussagen anhand der Realkennzeichen ergab hinsichtlich des Gesamtscores signifikante Unterschiede zwischen den Gruppen. Die Aussagen der Kinder über den real erlebten Diebstahl wiesen höhere Summenscores für die Realkennzeichen auf, als Aussagen in denen eine falsche Beschuldigung vorgenommen wurde. Dieses Ergebnis unterstützt die Validität der kriterienorientierten Analyse.

Yuille (1988, zitiert nach Horowitz, 1991) untersuchte das Vorhandensein von Realkennzeichen und den Wahrheitsstatus von wahren und falschen Aussagen bei 49 Erst- und Drittklässlern im Alter von sechs bis neun Jahren. Die Kinder hatten die Aufgabe, jeweils eine wahre und eine falsche Geschichte zu schildern. Diese wurden von Ratern, die den Wahrheitsstatus der Schilderungen nicht kannten, auf das Vorhandensein von

Realkennzeichen überprüft und hinsichtlich ihres eingeschätzten Wahrheitsstatus klassifiziert. Es ergaben sich Trefferquoten von 90,9% für die Einordnung der wahren Geschichten und von 74,4% für die Identifizierung der ausgedachten Berichte. In dieser Studie wurde für die Beurteilung des Wahrheitsgehaltes ein Cut-off- Wert verwendet. Eine Schilderung wurde als wahr beurteilt, wenn die ersten drei Kriterien des Merkmalskataloges von Steller und Köhnken (1989) Logische Konsistenz, Ungeordnet sprunghafte Darstellung, Quantitativer Detailreichtum und zwei weitere beliebige Realkennzeichen signiert wurden. Die Ergebnisse bestätigten die Validität der Realkennzeichen. Außerdem zeigten die Ergebnisse jedoch, dass bei der Anwendung der kriterienorientierten Aussagenanalyse die Tendenz besteht, nicht erlebnisbegründete Schilderungen irrtümlich als erlebt einzuordnen. Der Fehler, Darstellungen mit Erlebnishintergrund als unwahr einzustufen, trat dagegen deutlich seltener auf.

Joffe und Yuille (1992) überprüften in einer Untersuchung mit 142 Kindern im Alter von acht und zehn Jahren die Differenzierungsleistung der Realkennzeichen. Wahre Aussagen wurden mit falschen, instruierten Schilderungen verglichen, wobei die Instruktionen für diese Schilderungen variierten. Als weitere Variable wurde das Alter einbezogen. Die Realkennzeichen konnten lediglich zwischen den wahren und partiell instruierten Aussagen bei älteren Kindern trennen. Sie konnten jedoch nicht zwischen wahren und vollständig instruierten Schilderungen unterscheiden.

Ziel der experimentellen Untersuchung von Wellershaus und Wolf (1989) war die Überprüfung der Hypothese von Undeutsch (1967), nach der qualitative Unterschiede zwischen wahren und phantasiegeleiteten Schilderungen bestehen. Achtundachtzig Kinder der ersten und vierten Grundschulklasse erzählten jeweils eine wahre und eine erdachte Geschichte aus verschiedenen Bereichen, die hinsichtlich der emotionalen und kognitiven Bedeutung der Sachverhalte mit dem Erleben eines Sexualdeliktes vergleichbar sein sollten. Die Eigenbeteiligung des Berichtenden, die negative emotionale Tönung des Geschehens und ein weitgehender Kontrollverlust über die Situation seitens des Betroffenen wurden dabei als Kriterien für die Auswahl der inhaltlichen Bereiche, der in der Studie zu berichtenden Erlebnisse gefordert (z.B. Blut abgenommen bekommen,

eine Operation als Patient erfahren und von einem Tier angefallen werden). Die Ausprägungen der Realkennzeichen in den wahren und unwahren Aussagen wurden auf einer vierstufigen Skala (nicht vorhanden bis stark ausgeprägt) eingeschätzt. Signifikante Unterschiede zwischen den Gruppen der wahren und unwahren Aussagen zeigten sich bei den Kriterien Logische Konsistenz, Quantitativer Detailreichtum, Raum-zeitliche Verknüpfungen, Schilderung von Komplikationen im Handlungsverlauf, Schilderung ausgefallener Einzelheiten, Schilderung nebensächlicher Einzelheiten, Phänomengemäße Schilderungen unverstandener Handlungselemente, Indirekt handlungsbezogene Schilderungen und Schilderungen eigener psychischer Vorgänge. Die Klassifizierung der Aussagen in die Gruppen der wahren und unwahren Aussagen erfolgte in 76,7% (wahr) und 67 % (unwahr) der Fälle richtig, wenn sie von geschulten Ratern vorgenommen wurde. Kriterienunkundige Rater ordneten die wahren Aussagen zu 58,44% und die unwahren Berichte zu 49,52% korrekt zu. Bemerkenswert an dieser Studie sind neben der hohen methodischen Qualität die Überlegungen zur Übertragbarkeit der Ergebnisse auf die Realität, die sich darin widerspiegeln, dass psychologische Variablen verwandt wurden, die sowohl die Situation bei einem sexuellen Missbrauch kennzeichnen als auch in Situationen vorkommen, die in einer experimentellen Untersuchung ohne ethische Bedenken simuliert werden können.

In einer Simulationsstudie mit Studenten fanden Landry und Brigham (1992), dass die Realkennzeichen Quantitativer Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Wiedergabe von Gesprächen, Schilderungen ausgefallener Einzelheiten, Schilderungen nebensächlicher Einzelheiten, Schilderung eigener psychischer Vorgänge, Spontane Verbesserungen der eigenen Aussage, Eingeständnis von Erinnerungslücken und Einwände gegen die Richtigkeit der eigenen Aussage signifikant häufiger in wahren Schilderungen auftraten als in erdachten Darstellungen. Die Kriterien Logische Konsistenz und Schilderung psychischer Vorgänge des Beschuldigten traten dagegen signifikant häufiger in unwahren Schilderungen auf.

Fünfunddreißig Versuchspersonen schilderten jeweils ein wahres und ein erdachtes Erlebnis mit traumatischem Inhalt. Aus den 70 Geschichten wurden sechs wahre und sechs erfundene ausgewählt. Als weitere Variablen gingen die Art des Basismaterials (Video-

aufzeichnung vs. Transkript) und die Kenntnis der Rater von der kriterienorientierten Aussagenanalyse (geschult vs. nicht geschult) in die Auswertung ein. Die Auswertung ergab, dass beide Variablen signifikante Effekte bei der Klassifizierung der Darstellungen zeigten. Trainierte Rater identifizierten 75,4% der wahren Schilderungen richtig, untrainierte nur 58,6%. Beide Gruppen ordneten nur 35% der unwahren Geschichten richtig ein. Videoaufzeichnungen als Beurteilungsgrundlage erhöhten die Trefferquoten sowohl in der geschulten als auch in der ungeschulten Gruppe. Rater, die eine kriterienorientierte Aussagenanalyse auf der Basis von Videoaufzeichnungen vornahmen, erzielten die höchsten Trefferquoten.

Köhnken, Schimossek, Aschermann und Höfer (1995) fanden in einer experimentellen Studie mit 59 erwachsenen Versuchspersonen im Alter von 19 bis 39 Jahren signifikante Unterschiede hinsichtlich des Auftretens von Realkennzeichen in wahren und phantasiegeleiteten Aussagen. Erlebnishintergrund der wahren Aussagen war ein Film über eine Blutspende. Den nicht erlebnisbegründeten Aussagen lagen lediglich verbale Informationen über den Filminhalt zugrunde. Weitere Untersuchungsvariablen waren die Befragungsstrategie (kognitives Interview vs. standardisiertes Interview) und die Einführung von sechs zusätzlichen Kriterien, die über den Merkmalskatalog von Steller und Köhnken (1989) hinausgingen. Diese Kriterien waren Reporting style, Display of insecurities, Providing reasons for lack of memory, Clichés, Repetitions and Comments and interpretations regarding the event. In die Auswertung gingen dreizehn der neunzehn Realkennzeichen nach Steller und Köhnken ein. Die Kriterien Raum-zeitliche Verknüpfungen, Indirekte handlungsbezogene Schilderungen, Schilderung eigener psychischer Vorgänge, Selbstbelastung und Deliktspezifische Aussagelemente wurden nicht einbezogen, da sie keine Relevanz für die Schilderungen über den Filminhalt hatten. Die Kriterien Wiedergabe von Gesprächen und Interaktionsschilderungen wurden zu einem Kriterium zusammengefasst, da Voruntersuchungen zeigten, dass die Rater zwischen diesen Kriterien nicht ausreichend reliabel unterscheiden konnten. Von den Aussagen konnten 85% auf der Basis der dreizehn Kriterien korrekt klassifiziert werden. Die Interviewstrategie hatte dabei keinen signifikanten Effekt auf die Trefferquote. Die diskriminanzanalytische Auswertung ergab, dass die Versuchsgruppen zwar mittels der Kriterien überwiegend richtig

klassifiziert werden konnten, allerdings fanden sich nur für die Kriterien Ungeordnete Darstellung und Detailreichtum signifikante Effekte zwischen den Gruppen. Dagegen differenzierten fünf der sechs hinzugenommen Merkmale (Reporting style, Display of insecurities, Providing reasons for lack of memory, Clichés, Comments and interpretations regarding the event) signifikant zwischen wahren und falschen Aussagen.

Zur externen Validität dieser Untersuchung ist kritisch anzumerken, dass, wie auch in der Untersuchung von Köhnken und Wegener (1982), die elementare Variable des Erlebens durch Beobachten ersetzt wurde. Nur acht Probanden hatten in dieser Untersuchung eigene Erfahrungen mit dem Blutspenden. Das Ansehen eines Films wurde dabei als Erlebnishintergrund operationalisiert.

Unterstützt wird diese Kritik durch die Ergebnisse einer Untersuchung von Akehurst, Köhnken und Höfer (2001) in der die Wahrnehmungsvariable über Erleben eines Geschehens, Beobachten des Geschehens per Video und Verbale Informationen über das Geschehen variiert wurde. Hinsichtlich des Auftretens inhaltlicher Kriterien und Trefferquoten ergaben sich signifikante Unterschiede zwischen den beiden Erlebnisgruppen. Das bedeutet, dass das Beobachtungsparadigma, meist realisiert durch das Betrachten eines Filmes, nicht oder nur unzureichend das Erleben eines Geschehens repräsentiert.

Wolf und Steller (1997) überprüften die Validität der Realkennzeichen in einer Studie, in der erwachsene Frauen über das Erleben einer Geburt berichteten. Wie in der Studie von Wellerhaus und Wolf (1989) sollten durch die Wahl des Basismaterials die psychologischen Grundvariablen Eigenbeteiligung am Geschehen, Kontrollverlust und starke emotionale Tönung des Erlebten vorhanden sein, und damit eine hohe ökologische Validität erreicht werden. Die Auswertung der Geburtsberichte von 15 Müttern und 15 Nichtmüttern durch drei geschulte Rater zeigte, dass die Kriterien Unstrukturierte Darstellung, Quantitativer Detailreichtum, Raum-zeitliche Verknüpfungen, Interaktionsschilderungen, Wiedergabe von Gesprächen, Komplikationen im Handlungsverlauf, Ausgefallene Einzelheiten, Nebensächliche Einzelheiten, Eigenpsychische Vorgänge, Schilderung psychischer Vorgänge Beteiligter und Spontane Verbesserungen der eigenen Aussage in wahren Berichten signifikant häufiger als in unwahren Schilderungen zu finden waren. Die

diskriminanzanalytische Verrechnung der Daten zeigte, dass mittels der Realkennzeichen 100% der Fälle richtig zugeordnet werden konnten.

Sporer (1997) fand in einer Studie, in der Schüler, Auszubildende und Studenten je 100 selbst erlebte und erfundene Geschichten schriftlich verfassten, keine signifikanten Unterschiede in der Verteilung der Realkennzeichen zwischen den beiden Gruppen. Zweiundsechzig Prozent der erfundenen Geschichten und 64 % der selbst erlebten Geschichten wurden von zwei trainierten Ratern richtig zugeordnet. Allerdings ist das Paradigma des Schreibens von Erlebnissen nicht auf den forensischen Alltag zu übertragen, da hier der mündliche Bericht eines Erlebnisses Mittelpunkt der Aussagebeurteilung steht, zumal insbesondere Realkennzeichen aus dem motivbezogenen Bereich, z.B. das Zugeben von Erinnerungslücken, gerade auf die Erzählweise abzielen. In seinem Fazit schränkt der Autor ein, dass nicht die Überprüfung der Validität der Glaubwürdigkeitskriterien im Mittelpunkt des Forschungsinteresses gestanden habe, sondern es um eine theoretische Fundierung der kriterienorientierten Unterscheidung von Wahrheit und Lüge gegangen sei.

Akehurst, Köhnken und Höfer (2001) führten eine experimentelle Studie mit Kindern der Altersgruppen sieben bis acht und zehn bis elf Jahren und mit Erwachsenen u.a. mit dem Ziel durch, die Brauchbarkeit der kriterienorientierten Aussagenanalyse im direkten Vergleich der beiden Altersgruppen zu überprüfen. Als Untersuchungsparadigma wurde das Erleben während einer Fotoaufnahme gewählt. Während von einem Teil der Versuchspersonen tatsächlich ein Foto aufgenommen wurde, sah die zweite Gruppe der Versuchspersonen lediglich ein Video von einer Fotoaufnahme. Eine dritte Gruppe sollte auf der Grundlage spezifischer verbaler Informationen (z.B. „Es wurden drei Fotos gemacht“, „Der Fotograph war ein junger Mann“, „Bei der ersten Aufnahme hast du nach rechts geguckt“ usw.) realistisch berichten, tatsächlich fotografiert worden zu sein. Die Personen der beiden Altersgruppen wurden gleichmäßig auf die experimentellen Bedingungen verteilt. Im Interview erfolgte nach einem freien Bericht der Versuchspersonen ein Befragungsteil, in dem konkrete Fragen nach Einzelheiten der Photoaufnahme gestellt wurden. In einer dritten Interviewphase wurden die Versuchspersonen erneut aufgefor-

dert, alles zu berichten, was sie bisher geschildert hatten und ihre Schilderung gegebenfalls zu ergänzen.

Die Auswertung erbrachte keinen Alterseffekt. Es ergaben sich aber hinsichtlich des Auftretens der Realkennzeichen signifikante Unterschiede zwischen den erlebnisbegründeten Schilderungen (tatsächliche Fotoaufnahme) und den auf verbalen Informationen beruhenden Berichten. Außerdem gab es einen signifikanten Unterschied zwischen erlebnisbegründeten und nur auf Beobachtung beruhenden Aussagen. In der Gruppe der wahren Aussagen wurden die Realkennzeichen häufiger signiert. Keine Unterschiede ergaben sich bezüglich des Vorhandenseins der Realkennzeichen zwischen der Beobachtungsgruppe und der Fantasiegruppe. Die diskriminanzanalytische Auswertung der Ergebnisse zeigte, dass auf Basis der Merkmalskombinationen 69,7% der wahren Aussagen und 84,8% der ausgedachten Berichte richtig zugeordnet werden konnten. Bei den Aussagen der Kinder unterschieden die Kriterien Ungeordnete Darstellung, Detailreichtum, Nebensächliche Einzelheiten, Einwände gegen die Richtigkeit der eigenen Aussage zwischen den Gruppen der wahren und unwahren Berichte, während bei den Erwachsenen Logische Konsistenz und Eigenpsychische Schilderungen, die trennschärfsten Kriterien waren.

2.4.4 Überlegungen zu den bisherigen experimentellen Untersuchungen

Bei Betrachtung der experimentellen Paradigmen in den Untersuchungen zur Validität der kriterienorientierten Aussageanalyse fällt auf, dass nur wenige Studien eine ausreichende ökologische Validität aufweisen können. Nach Steller (1988) bedarf es jedoch zur experimentellen Überprüfung des qualitativen Unterschieds zwischen erlebnisbegründeten und erdachten Darstellungen keineswegs einer ereignistreuen Abbildung solcher Geschehnisse wie sexueller Missbrauch, Vergewaltigung und Kindesmisshandlung, bei denen in der forensischen Praxis psychologische Sachverständigengutachten zu Glaubhaftigkeitsbegutachtung angefordert werden. Es gehe vielmehr darum, die Vergleichbarkeit von Experiment und Realität in Bezug auf die

für die Aussage wesentlichen psychologischen Bedingungen zu gewährleisten. Diesem Gedanken folgend realisierten Wellershaus und Wolf (1989) in ihrer Untersuchung die für eine Missbrauchssituation relevanten Grundvariablen Eigenbeteiligung am Geschehen, negative emotionale Tönung des Geschehens und Kontrollverlust in der Situation, indem sie medizinische Eingriffe und körperliche Angriffe als Grundlage für erlebnisbegründete Berichte und Fantasiegeschichten verwendeten. Während in den oben beschriebenen Feldstudien Daten von sexuell missbrauchten Kindern die Erleben-Kategorie repräsentieren, gibt es in den experimentellen Untersuchungen nur wenige Paradigmen, in denen die genannten Grundvariablen enthalten sind. Das Betrachten von Filmen und das Schildern von Alltagserlebnissen bilden in den meisten Studien den Erlebnishintergrund (Köhnken & Wegener, 1982; Köhnken et al., 1995; Akehurst et al., 2001, Yuille, 1988; Rüth- Bemelmans, 1984). Lediglich in den Untersuchungen von Landry und Brigham (1992) sowie Wolf und Steller (1997) in denen Ereignisse mit traumatischem Inhalt oder das Geburtserleben von Frauen als tatsächliches Erleben geschildert wurden, waren die oben angeführten Variablen vorhanden. Möglicherweise hängt die größere Bandbreite gefundener Kriterien in den letztgenannten Studien mit der Operationalisierung dieser Variablen zusammen.

Eine genauere Betrachtung der signierten Realkennzeichen ergibt, dass ihre Auftretenshäufigkeit über die dargestellten Feld- und Simulationsstudien sehr unterschiedlich ist. Sie variiert zwischen zwei (z.B. Köhnken et al., 1995) und achtzehn Kriterien (Espkin et al., 1988). Einerseits ist das Vorhandensein kriterienerfüllender Textstellen vom Paradigma der Untersuchung abhängig. So sind in einigen Erhebungen motivbezogene Kriterien durch die Wahl des Paradigmas, z.B. das Betrachten eines Filmes, von vornherein ausgeschlossen. Andererseits ist es möglich, dass bestimmte Kriterien grundsätzlich eine größere Trennkraft aufweisen als andere. Anhaltspunkte hierfür ergeben sich aus der Verteilung von Kriterien in den bisherigen Arbeiten. In der Mehrzahl der Studien wurden die Kriterien Detailreichtum, Eigenpsychische Vorgänge, Ungeordnete Darstellung, Raum-zeitliche Verknüpfungen, Ausgefallene Einzelheiten und die Wiedergabe von Gesprächen signifikant häufiger in erlebnisbegründeten Aussagen als in nicht erlebnisbegründeten Schilderungen gefunden. Im Hinblick auf die bisher fehlende theoretische Be-

gründung der inhaltsorientierten Kriterienanalyse könnte das Auffinden spezifischer Kriterienmuster oder besonders trennscharfer Kriterien hilfreich sein.

In mehreren Untersuchungen zur Validität der Glaubwürdigkeitskriterien wurde in Bezug auf das Auftreten von Kriterien ein Alterseinfluss deutlich. Die Feldstudie von Boychuk (1991) zeigte, dass einige Kriterien signifikant häufiger in den Schilderungen älterer Kinder auftraten. Ähnliche Ergebnisse erhielt Lamers-Winkelmann (1992a, b) in einer Feldstudie. In Aussagen älterer sexuell missbrauchter Kindern fand sie mehr Kriterien und höhere Gesamtscores als bei jüngeren Kindern, bei denen der sexuelle Missbrauch als belegt galt. Einzelne Kriterien traten bei acht- bis elfjährigen Kindern signifikant häufiger auf als bei vier- bis fünfjährigen Versuchspersonen. Diese Ergebnisse werden durch die Arbeit von Anson et al. (1993) gestützt, die in ihrer Untersuchung ebenfalls eine positive Korrelation zwischen Alter und dem Gesamtscore sowie einzelner Kriterien fanden. Vrij, Akehuerst, Soukara und Bull (2002) konnten in ihrer Studie ebenfalls einen Alterseffekt für das Auftreten der Realkennzeichen zeigen. Sie verglichen wahre und unwahre Aussagen (bewusst gelogen) in einer experimentellen Studie und verglichen fünf- bis sechsjährige, zehn- bis elfjährige, vierzehn- bis fünfzehnjährige und Studierende, deren mittleres Alter bei 22,37 Jahren lag. In einer varianzanalytischen Auswertung konnten sie zeigen, dass sich die Anzahl der Realkennzeichen in den verschiedenen Altersgruppen signifikant unterschied. Dabei waren in der jüngsten Altersgruppe im Vergleich zu den drei anderen am wenigsten Realkennzeichen in den Berichten aufzufinden. Bei zehn- bis elfjährigen Probanden fanden sich signifikant weniger Realkennzeichen, als in der Gruppe der Studierenden. Die weiteren Gruppenvergleiche führten zu keinen signifikanten Ergebnissen.

Eine Erklärung für das seltener Auftreten von Glaubwürdigkeitskriterien in den Aussagen jüngerer Kinder lässt sich möglicherweise aus den weitgehend konsistenten Ergebnissen der Forschung zu Zeugenaussagen von Kindern ableiten. Danach ist der Aussageumfang im *freien Bericht* altersabhängig. Unter einem freien Bericht wird die zusammenhängende Darstellung ohne Unterbrechung durch Fragen verstanden. Die Berichte jüngerer Kinder sind meist weniger ausführlich und vollständig als die älterer Kinder oder Er-

wachsener, während sich Genauigkeit und Richtigkeit ihrer Angaben im freien Bericht nicht unterscheiden. Sowohl Marin, Holmes, Guth und Kovac (1979) als auch Goodman und Reed (1986) berichten von einem nahezu linearen Anstieg der Anzahl geschilderter Details mit zunehmendem Alter der Versuchspersonen. In allen Untersuchungen mit Kindern unter vier Jahren(z.B. Cassel, Roebers und Bjorklund, 1996) zeigte sich, dass diese nicht oder nur sehr wenig frei berichten und direkte Fragen als Aussageanstoß benötigen. Für die kriterienorientierte Aussagenanalyse, deren Grundlage idealiter der freie Bericht ist, bedeuten diese Ergebnisse, dass die Grundvoraussetzung für die Durchführung der Inhaltsanalyse - die ausreichende Detaillierung der Schilderung - von jungen Kindern möglicherweise nicht immer erfüllt werden kann und somit keine oder nur wenige kriterienerfüllende Textstellen gefunden werden können. Die Schwierigkeiten bei der Anwendung der inhaltlichen Kriterien bei jüngeren Kindern können entwicklungspsychologisch erklärt werden. So können - wie bereits angeführt- die altertypischen Einschränkungen in der kognitiven und sprachlichen Kompetenz dazu führen, dass Schilderungen jüngerer Kinder weniger elaboriert sind. Die Entwicklungstypischen Besonderheiten bei Kindergarten- und Vorschulkindern, nämlich die noch nicht voll ausgebildete Fähigkeit zum Perspektivenwechsel, zur Bildung metakognitiver Strategien und die noch nicht abgeschlossene moralische Entwicklung, können zudem bewirken, dass spezifische Kriterien wie fremdpsychische Schilderungen, dass Zugeben von Erinnerungsunsicherheiten oder die Entlastung des Angeschuldigten bei jüngeren Kindern eine geringere Auftretenswahrscheinlichkeit haben und die Bandbreite der Kriterien dadurch eingeschränkt wird.

Für die aussagepsychologische Begutachtungspraxis bedeuten diese Ergebnisse, dass der Einsatz der kriterienorientierten Inhaltsanalyse bei sehr jungen Kindern schon aufgrund des unzureichenden Aussageumfangs Schwierigkeiten bereiten, bzw. nur wenig zur Beantwortung der Frage nach dem Realitätsgehalt der zu beurteilenden Aussage beitragen kann. Andererseits zeigen die Mängel der angeführten Studien, insbesondere bei der Wahl des Untersuchungs-Paradigmas, dass es voreilig wäre, die Ergebnisse dahingehend zu interpretieren, dass die Validität der kriterienorientierten Aussagenanalyse bei jüngeren Kindern grundsätzlich in Frage zu stellen ist. Vielmehr wird die Notwendigkeit deutlich,

weitere experimentelle Studien mit jungen Kindern durchzuführen, in denen deren entwicklungsspezifischen Besonderheiten Rechnung getragen wird. Dabei sollte auch geprüft werden, in welcher Weise äußere Einflüsse Auswirkungen auf den Aussageinhalt und damit auf die Richtigkeit der Klassifizierung der Aussage haben können. So liegt in der aussagepsychologischen Begutachtungspraxis neben der Erhebung der Aussage über ein mutmaßliches Erlebnis ein weiterer Schwerpunkt in der Rekonstruktion der Aussageentstehungs- und Entwicklungsgeschichte. Damit sollen mögliche suggestive Einflussnahmen aufgedeckt werden, die einen bedeutsamen Einfluss auf die Qualität der Aussage und damit auf die Auftretenswahrscheinlichkeit von Realkennzeichen haben können.

3. Suggestion und Suggestibilität

Bei der Suche nach einer Definition der Begriffes Suggestion und Suggestibilität findet sich in der Literatur ein weites Spektrum, dessen Darstellung den Rahmen dieser Arbeit sprengen würde (vgl. dazu Burger, 1971; Pieters, 1994). Daher werden im Folgenden beispielhaft einige Definitionen angeführt. Gudjonsson (1986) beschreibt Suggestibilität als das Ausmaß, in dem Personen nachträgliche Informationen akzeptieren und mit den bereits bestehenden Gedächtnisinhalten verknüpfen. Er formulierte damit eine Definition, unter der die bisherigen Forschungsaktivitäten in diesem Bereich stimmig eingeordnet werden können.

In der aussagepsychologisch Literatur wird vielfach eine Definition von Ceci und Bruck (1993) zitiert, in der eine *breitere* von einer *engen* eher traditionellen Definition von Suggestibilität unterschieden wird: „According to its broadest definition, suggestibility concerns the degree by which...encoding, storage, retrieval and reporting of events can be influenced by a range of social and psychological factors. This broad definition contrasts with the narrower and more traditional definition, which asserts that it is the extent to which individuals come to accept and subsequently incorporate postevent information into their memory recollections“ (S. 404). Die Autoren vertreten die Meinung, die Anwendung der breiteren Definition biete die Möglichkeit, mehr Fehlerquellen erfassen zu

können, z.B. suggestive Einflüsse, die zeitlich vor dem kritischen Ereignis liegen. Auch ist nicht auszuschließen, dass bestehende Voreinstellungen und Vorurteile des Befragten den Inhalt der Wiedergabe beeinflussen könnten. In der Definition von Ceci und Bruck wurden auch intentionale Falschaussagen subsumiert. Dies erschwert nach Ansicht von Steller, Volbert und Wellershaus (1993) das Verständnis der Problematik von Suggestionseffekten. Sie halten es für fragwürdig, von Suggestionseffekten zu sprechen, wenn sich der Aussagende einer objektiv unwahren Aussage bewusst ist und in Verabredung mit dem Instruierenden handelt. Nach Steller (1998) handelt es sich „bei suggestionsbedingten Aussageverfälschungen um eine Form des Irrtums: Suggestionen können zu der subjektiv gegebenen Überzeugung führen, etwas erlebt zu haben, was in Wirklichkeit nicht der Fall war. Wenn keine Verfälschungsintention bei dem Aussagenden vorliegt, ist die Bezeichnung Lüge für suggestionsbedingte Falschaussagen nicht sinnvoll, selbst dann nicht, wenn z.B. zu Beginn suggestiver Einflussnahme noch ein (Rest-) Bewusstsein bei dem Aussagenden über die Falschheit des Gesagten besteht und der Aussagende zunächst der Suggestion wegen des hohen Aufforderungsgehaltes der spezifischen Kommunikationsbedingung wider besseres eigenes Wissen nachgibt. Bei nicht erlebnisbegündeten Kinderaussagen stellen absichtliche Falschaussagen nur eine denkbare Kategorie dar, nach allen vorliegenden Befunden die statistisch am wenigsten bedeutsame.“ (S.13)

Volbert (2003) führt eine Unterscheidung der Begriffe Suggestion und Suggestibilität an. Danach ist unter Suggestion eine Form der Kommunikation, oder Technik zu verstehen, durch die eine Beeinflussung geschehen kann, während Suggestibilität zum einen die allgemeine Tendenz, zum anderen die individuelle Bereitschaft auf eine Suggestion zu reagieren beschreibt. Eine suggestive Aufforderung ist nach Volbert einseitig in ihrer Richtung. Dabei kann diese, muss allerdings nicht, zu einer erfolgreichen Beeinflussung führen. Das erfolgreiche Wirken einer suggestiven Aufforderung beschreibt sie als abhängig von einer affektiven, kognitiven oder strukturellen Mangelsituation, in der durch die Übernahme suggerierter Inhalte dieser Mangel ausgeglichen wird. Ein Suggestionseffekt wird nach Volbert erst durch das Bestehen einer suggestiven Aufforderung und die durch die Mangelsituation bestimmte Empfänglichkeit wirksam.

In der gegenwärtigen forensisch-psychologischen Praxis ist die Überprüfung möglicher suggestiver Einflüsse ein unverzichtbares Element der Glaubhaftigkeitsbegutachtung geworden. Aussageentstehung und Aussageentwicklung werden einer eingehenden Prüfung unterzogen, um Hinweise auf suggestive Einflüsse zu erkennen. Schon Arntzen und Michaelis (1970) berücksichtigten in der praktischen Begutachtung das Problem der Suggestibilität bei der Analyse kindlicher Zeugenaussagen. Sie wiesen darauf hin, dass es bei vier bis sechsjährigen Kindern durch Suggestionen zu Aussageverfälschungen kommen kann. Insbesondere wurden *beinhaltete* Fragen als Ursache von Erinnerungsverfälschungen angeführt, da sie in die Erinnerungen eingriffen und diese veränderten. Neben dieser Art der Beeinflussung wurden von Arntzen und Michaelis (1970) weitere potentiell suggestive Faktoren erkannt, wie die emotionale Bindung des Befragten an den Befragenden, Gruppeneinflüsse, Alter der Zeugen, Erinnerungsunsicherheiten, Erwartungen des Befragenden und Wiederholungsbefragungen. Diese möglichen Einflüsse wurden in den achtziger und neunziger Jahren in der stark auflebenden experimentellen Suggestionforschung ebenfalls als wesentliche potentielle Einflussfaktoren untersucht (vgl. Goodman & Clark-Stewart, 1991; Ceci & Bruck, 1993; Volbert & Pieters, 1996). Das starke Interesse an der Suggestionforschung in den achtziger Jahren wurde durch mehrere Aufsehen erregende Strafprozesse in den USA (z.B. der McMartin Fall) mit angeblich sexuell missbrauchten Kindern ausgelöst. Die belastenden Aussagen dieser Kinder mussten nach offensichtlich schweren Fehlern bei der Befragung als unsicher und zum Teil als sicher falsch beurteilt werden (Wegener, 1997) und lösten damit eine erneute Skepsis hinsichtlich der Glaubhaftigkeit von Kinderaussagen aus.

Die früheren Zweifel an der Validität von Aussagen kindlicher Opferzeugen hatten ihren Ursprung in den Ergebnissen experimenteller Untersuchungen zu Beginn des letzten Jahrhunderts, die zu der negativen Bewertung der Aussageleistungen, insbesondere von kindlichen Zeugen, geführt hatten. Stern (1902) folgerte aus den Ergebnissen seiner Experimente, die er in Bezug zum forensischen Kontext stellte, dass die Aussagegüte mit dem Alter der Versuchspersonen ansteige. Die oberflächlichen Interpretationen seiner viel beachteten Versuche führten zu einer ungünstigen Beurteilung von Zeugenleistun-

gen, die in den folgenden Jahrzehnten auch in den Gerichtssälen vorherrschte und zu einer Unterschätzung insbesondere der kindlichen Aussagen führte (Wegener, 1997).

Das Interesse an der Suggestibilität von Zeugen wurde zum Zeitpunkt der vorherigen Jahrhundertwende durch die Arbeit von Binet (1900) unterstrichen. Binet führte als erster eine Untersuchung über Aussagesuggestibilität mit Kindern durch. Er stellt mit zunehmendem Ausmaß der Suggestion in der Befragung ein erhebliches Ansteigen von Falschantworten fest. Die folgenschweren Auswirkungen von fehlerhaften Befragungen auf die Qualität der Aussagen von Zeugen im Strafprozess wurden durch Varendonck (1911) bestätigt. In einer Untersuchung mit achtjährigen Schülern fand er heraus, dass die Schüler trotz fehlender Erlebnisgrundlage nach suggestiver Befragung detaillierte Angaben über das Zusammentreffen mit einer Person machten. Lipmann (1908, zit. nach Endres, Scholz & Summa, 1997) stellte in seinen Untersuchungen Zusammenhänge der Suggestibilität mit Alter, Intelligenz, sozialer Schicht und Geschlecht fest. Schon Stern (1904) fand einen Alterseffekt, der sich darin zeigte, dass die Qualität der Aussage mit zunehmendem Alter der Befragten anstieg. Außerdem konstatierte er, dass pubertierende Mädchen eine erhöhte Suggestibilität zeigten, die Stern mit hormonellen Veränderungen erklärte.

Ergebnisse der Forschung in der zweiten Hälfte des letzten Jahrhunderts führten zu einer Korrektur der Auffassung, die Kindern wegen ihrer hohen Anfälligkeit für Suggestionen generell eine geringere Zeugentüchtigkeit zuschrieb. Die Abkehr von experimentellen Studien und die Auswertung von Erfahrungen und Kenntnissen aus praktischen Gerichtsfällen in den fünfziger Jahren führten zu einer Revision der generell negativen Einstellung gegenüber kindlichen Zeugen (Undeutsch, 1967). Diese Neubetrachtung manifestierte sich in einem BGH-Grundsatzurteil aus dem Jahre 1955. In diesem Urteil wurde konstatiert, dass Kinderaussagen nicht häufiger unglaublich seien als die Aussagen von Erwachsenen und dass Kinder häufig sogar bessere Zeugen seien (BGHSt 1955, 7, 82).

Bis zum erneuten Aufleben von Zweifeln an kindlichen Zeugenaussagen in den achtziger Jahren war die Frage der Suggestibilität von kindlichen Zeugen in den Hintergrund der

Forschung getreten. Im Vordergrund des Interesses stand in Forschung und Praxis vielmehr die Qualität der Zeugenaussagen, abgebildet durch die Realitäts- bzw. Glaubwürdigkeitskriterien (s. Kap. 2). Die Suggestibilität von Zeugen wurde in der praktischen Begutachtung lediglich als *eine* von mehreren möglichen Fehlerquellen berücksichtigt.

3.1 Empirische Ergebnisse zur Suggestibilität

In einer umfassenden Analyse zahlreicher Untersuchungen zum Thema Suggestibilität bei Kindern unterscheiden Ceci und Bruck (1993) zwischen, kognitiven, sozialen und biologischen Faktoren, die als übergeordnete Klassen von Faktoren die Anfälligkeit von Kindern für Suggestionen beeinflussen. Als kognitive Faktoren werden Gedächtnis-theoretische Modelle, linguistische Kompetenzen, semantisches-, stereotypes- und Skriptwissen und Probleme der Quellendiskriminierung bezeichnet. Die Stärke der Gedächtnisspur wird in verschiedenen Studien in ursächlichem Zusammenhang mit den Annahmen über die Wirkungsweise von Suggestionen diskutiert:

Der *Hypothese des destruktiven Aktualisierens* liegt die Annahme zugrunde, dass für die Erinnerung an ein Ereignis Gedächtnisspuren abgerufen und dem Bewusstsein zugänglich gemacht werden. Nachträgliche Informationen überschreiben die ursprüngliche Gedächtnisspur und zerstören dabei gleichzeitig die ursprüngliche Erinnerung (Loftus, Miller & Burns, 1978; Loftus & Loftus, 1980). Eine korrekte Erinnerung an das ursprüngliche Ereignis ist nach diesem Ansatz nach einer suggestiven Beeinflussung nicht mehr möglich.

In der *Koexistenzhypothese* wird dagegen davon ausgegangen, dass die nachträgliche Information zusätzlich zum ursprünglichen Ereignis im Gedächtnis gespeichert wird. Beide Inhalte sind also gleichzeitig im Gedächtnis vorhanden. Die ursprüngliche Gedächtnisspur ist zwar schwerer zugänglich, kann aber möglicherweise unter bestimmten Umständen wieder abgerufen werden. (Bekerian & Bowers, 1983).

Nach der *Hypothese des Lückenfüllens* führen Vergessensprozesse zu Lücken in der Erinnerung (McCloskey & Zaragoza, 1985). Diese Lücken können dann durch nachträgliche Informationen ausgefüllt werden und damit als Teil der Erinnerung in das Ur-

sprungereignis eingebunden werden. Aufgrund der Annahme, die Stärke der Gedächtnisspur stehe mit dem Ausmaß der Empfänglichkeit für Suggestionen in kausalem Zusammenhang, wird diskutiert, dass Kinder generell schwächere Gedächtnisspuren enkodieren und daher anfälliger für nachträgliche Fehlinformationen sein könnten (Toglia, 1991; Loftus & Hoffman, 1989). Die Übernahme von nachträglichen Informationen geschieht insbesondere dann, wenn die Gedächtnisspur für das suggerierte Ereignis deutlich stärker ist als die Spur für das tatsächliche Ereignis (Ceci, Ross & Toglia 1987). In verschiedenen Untersuchungen (z. B. Ceci & Bruck, 1993) konnte gezeigt werden, dass suggestive Einflussnahmen bei den Kindern am erfolgreichsten waren, die nur schwache Erinnerungen an das ursprüngliche Ereignis hatten. Geht man von einem Zusammenhang zwischen der Stärke der Gedächtnisspur und der Empfänglichkeit für Suggestion aus, so lässt sich die höhere Suggestibilität von Kindern im Vergleich zu Erwachsenen damit erklären, dass eher einzelne Aspekte eines Ereignisses wahrgenommen werden, die Informationen weniger effektiv verarbeitet werden und daher weniger zusammenhängende Erinnerungen vorhanden sind als bei Erwachsenen. Das wiederum führt zu einer schwächeren Ausprägung der Gedächtnisspur.

Als weitere mögliche Einflussgröße auf die Suggestibilität führen Ceci und Bruck (1993) linguistische Kompetenzen an. Komplizierte Formulierungen von suggestiven Fragen können dazu führen, dass Kinder aufgrund ihres eingeschränkten Sprachverständnisses zu falschen Rekognitionen kommen. Andererseits können Äußerungen der Kinder aber auch wegen ihrer eingeschränkten verbalen Fähigkeiten fehlinterpretiert werden.

Ergänzend zu den oben angegebenen gedächtnistheoretischen Überlegungen werden zudem Probleme der Quellenidentifikation als mögliche Ursache für die Empfänglichkeit für Suggestionen angeführt.

Eine wichtige Funktion des Gedächtnisses besteht darin, neben der Erinnerung an Informationen auch die Erinnerung über deren Ursprung bzw. die Quelle bereitzustellen. Diese Funktion wird als *Quellenidentifikation* bezeichnet.

Man unterscheidet drei Grundformen der Quellenidentifikation. So können Erinnerungsquellen in der Person selbst lokalisiert (internal source monitoring) oder externen Ursprungs (external source monitoring) sein. Bei ersterem geht es z.B. um die Frage, ob

man etwas getan hat oder es sich nur vorgestellt hat, während die zweite Form sich auf die Unterscheidung mehrerer externer Quellen bezieht. Die dritte Form (reality monitoring) bezieht sich auf die Unterscheidung zwischen selbst erzeugten Erinnerungen und solchen, die auf selbst Erlebtem beruhen. Die zentrale Annahme des Modells ist, dass die Quellenidentifikation sowohl von den Erinnerungscharakteristika als auch von Beurteilungsprozessen beeinflusst wird. Erinnerungen können sich hinsichtlich ihrer Detailliertheit, räumlicher und zeitlicher Merkmale, semantischer Informationen und emotionaler Tönungen unterscheiden. Derartige Charakteristika werden zusammen mit dem Ereignis gespeichert. Die Identifizierung von Quellen beruht auf den Unterschieden zwischen verschiedenen Quellen. Die Erinnerungen an erlebte Ereignisse zeichnen sich durch eine stärkere Detailliertheit v.a. hinsichtlich perzeptueller Details und Umgebungsinformationen aus. Erinnerungen an erdachte Ereignisse enthalten dagegen mehr Informationen über kognitive Operationen. Eine Erinnerung, die viele räumliche und visuelle Informationen enthält, aber nur wenig Erinnerungen an kognitive Operationen, wird daher eher einer externen Quelle (wirklich erlebt) als einer internen Quelle (erdacht) zugeschrieben (Johnson, Hastroudi & Lindsay, 1993). Vor dem Hintergrund dieser Überlegungen kann Suggestibilität als Quellenverwechslung interpretiert werden. Sowohl die Erinnerung an das ursprüngliche Ereignis als auch die nachträglichen Informationen sind im Gedächtnis gespeichert. Erinnerungsfehler werden dann dadurch verursacht, dass die Informationen in der Abrufsituation nicht mehr den korrekten Quellen zugeordnet werden können, so dass es zu einer Quellenverwechslung kommt und die nachträgliche Information als Originalinformation erinnert wird. Im forensischen Kontext muss ein Zeuge zwischen tatsächlicher Erinnerung an ein Geschehen und späteren intensiven Gesprächen über das Erlebnis unterscheiden.

Besondere Bedeutung scheint auch der Person zuzukommen, von der die suggestive Beeinflussung ausgeht. Eine dem Kind vertraute Person, wie Mutter oder Vater, kann die Wirkung der Suggestion und damit die Leistung bei der Quellenzuordnung erheblich beeinflussen. Insbesondere bei einer schwachen Gedächtnisspur kann durch die subjektiv wahrgenommenen Autorität bzw. Kompetenz von Erwachsenen die richtige Identifizierung der "Ur-Information" erschwert werden. Die Häufigkeit der Präsentation falscher Infor-

mationen erhöht möglicherweise auch die Wahrscheinlichkeit falscher Quellenzuordnung. Eine Erklärung kann sein, dass Kinder intern realistische Bilder auch von solchen Informationen generieren, die sie nur gehört haben (Ackil & Zaragoza, 1995). Die falsche Quellenzuordnung kann zudem durch unterschiedlich lange zeitliche Abläufe verstärkt werden. Je größer das Zeitintervall zwischen Originalereignis und Abruf ist, desto schwerer fällt Kindern möglicherweise die richtige Quellenzuordnung, da mit verblassender Erinnerung die Erinnerungsbilder von tatsächlich erlebten und nur geschilderten Ereignissen zunehmend ähnlicher werden. Auch die zeitnahe Präsentation von Fehlinformationen zum Originalereignis könnte eine korrekte Quellenzuordnung erschweren, wenn die Informationen aus ursprünglicher und nachträglicher Quelle ähnlich sind.

Ackil & Zaragoza (1995) zeigten Erwachsenen und Kindern im Grundschulalter einen kurzen Videofilm und präsentierten anschließend verbale Falschinformationen, die sich auf den Inhalt des Films bezogen. In einer späteren Befragung sollten die Versuchspersonen angeben, ob sie eine Information im Film gesehen oder nur gehört hatten. Es zeigte sich, dass sowohl Erwachsene als auch Kinder verbale Informationen fälschlich als gesehen klassifizierten. Kinder machten dabei mehr Fehler bei der Quellenzuordnung als Erwachsene.

Poole und Lindsay (1995) setzen in einem Experiment mit drei- bis siebenjährigen Kindern die Eltern als Suggestoren ein und verstärkten damit die suggestiven Effekte. Die Kinder führten unter Anleitung eines Erwachsenen namens „Mr. Science“ mehrere einfache Versuche durch: z.B. Falten von Papierflugzeugen und Herstellen eines Telefons aus einer Klebetube und zwei Trichtern. Im Anschluss wurden die Kinder über das Erlebte befragt. Eine zweite Befragung fand drei Monate später statt. In der Zwischenzeit erhielten die Eltern ein Buch mit Beschreibungen von zwei Experimenten, die die Kinder tatsächlich erlebt hatten sowie mit der Darstellung zweier fiktiver Versuchsabläufe. Außerdem endete das Buch mit der Behauptung, dass „Mr. Science“ dem Kind etwas Ekeliges in den Mund getan habe. Dieses Buch sollte den Kindern von den Eltern dreimal vorgelesen werden. In der Befragung nach den Erlebnissen mit „Mr. Science“ zeigte sich, dass die Kinder nur gehörte Geschehnisse häufig als real erlebte Vorgänge berichteten. Bei

wiederholten Befragungen nahm die Anzahl der falschen Quellenzuordnungen zu. Einundsiebzig Prozent der Kinder gaben auf die Frage, ob Mr. Science ihnen tatsächlich etwas Ekeliges in den Mund getan habe oder ihnen dieses von der Mutter nur vorgelesen wurde, an, dass ihnen dieses tatsächlich passiert sei. Achtundsiebzig Prozent der Kinder sagten aber zudem auch, dass dieses in dem vorgelesenen Buch gestanden habe. Die Autoren konnten daher nicht ausschließen, dass Missverständnisse über die Aufgabenstellung in der Befragung die Antworten der Kinder beeinflusst haben können.

Besondere Bedeutung für die forensische Praxis haben empirische Ergebnisse, nach denen Kinder unter bestimmten Bedingungen überzeugt werden konnten, dass sie ein tatsächlich nicht geschehenes Ereignis erlebt hatten. Ceci, Huffman, Smith & Loftus (1994) zeigten, dass es möglich ist, Vorschulkinder dahingehend zu beeinflussen, dass sie Ereignisse als selbst erlebt berichteten, die sie sich lediglich bildlich vorgestellt hatten oder über die sie nachgedacht hatten. An der Untersuchung nahmen 96 Vorschulkinder im Alter von drei bis vier sowie von fünf und sechs Jahren teil. Die Kinder wurden sieben bis zehnmal zu zwei real erlebten Ereignissen aus den vergangenen 12 Monaten und zu zwei tatsächlich nicht erlebten Vorkommnissen befragt. Letztere waren für alle Kinder gleich: eines der fiktiven Ereignisse war das Fahren mit den Klassenkameraden in einem Heißluftballon, das andere war das Stecken eines Fingers in eine Mausfalle mit dem anschließenden Entfernen der Mausfalle im Krankenhaus. Den Kindern wurde zu Beginn des ersten Interviews mitgeteilt, dass ihnen eine Liste mit Ereignissen vorgelesen werden, die sie möglicherweise erlebt hätten. Sie wurden dann aufgefordert, zu überlegen ob sie einzelne Ereignisse tatsächlich erlebt hätten. Außerdem wurden sie darüber informiert, dass die Liste mithilfe der Eltern zusammengestellt worden sei, aber die Kinder nicht alle Ereignisse erlebt hätten. Die Auswertung zeigte, dass die wahren Ereignisse von fast allen Kindern zutreffend bejaht wurden. Von den fiktiven Ereignissen wurden 34% als erlebt bestätigt, unabhängig davon ob die Kinder beim ersten oder beim letzten Mal danach gefragt wurden. Betrachtet man die beiden Altersgruppen getrennt, zeigte sich, dass die jüngeren Kinder bei der ersten Befragung 44% der fiktiven Ereignisse als erlebt angaben, während sie bei der letzten Befragung 36% als reales Geschehen einordneten. Bei den älteren Kindern ergab sich eine Zunahme der Zustimmungsrate von der ersten Sitzung

zur letzten von 25% auf 32%. Nur wenige Kinder wechselten zwischen Zustimmung und Ablehnung im Verlauf der verschiedenen Interviews. Es zeigte sich außerdem, dass eine frühe Bestätigung der fiktiven Vorkommnisse die Wahrscheinlichkeit erhöhte, dass die Kinder in späteren Befragungen bei dieser Einschätzung blieben. Die Sicherheit, mit der die Kinder die fiktiven Geschehnisse bestätigten, nahm im Verlauf der Befragungen zu. In einer Nachbefragung mit Kindern, die den fiktiven Ereignissen durchgehend zugesimmt hatten, blieb ein Teil dieser Kinder trotz Aufklärung über den fehlenden Realitätsgehalt der Erlebnisse dabei, die Geschehnisse tatsächlich erlebt zu haben.

Huffman, Crossman und Ceci, (1997) untersuchten die Stabilität von Erinnerungen an fiktive Ereignisse. Zweiundzwanzig Kindern wurden acht fiktive und zwei realen Erlebnisse vorgegeben. Die Kinder wurden gefragt, ob sie diese selbst erlebt hatten und bei Bestätigung aufgefordert, darüber zu berichten. In einer zweiten Befragung zwei Jahre nach der ersten zeigte sich eine signifikant niedrigere Zustimmungsrate zu den fiktiven Ereignissen (13% vs. 22% zwei Jahre zuvor). Die Hälfte der Kinder stimmte mindestens einem fiktiven Ereignis zu. Mindestens eines der realen Ereignisse erinnerten 90% der Kinder. Insgesamt wurde den wahren Ereignissen häufiger zugestimmt als den fiktiven. Weiter zeigte sich, dass die Kinder die realen Ereignisse überwiegend durchgängig bestätigten (75% Übereinstimmung). Dagegen wurden die meisten Zustimmungen bei fiktiven Ereignissen in der zweiten Befragung nach zwei Jahren zurückgezogen (75%). In der Nachbefragung wurden 7% der fiktiven Ereignisse erstmalig als real bestätigt. Die Autoren begründen die geringe Zustimmungsrate zu fiktiven Erinnerungen damit, dass diese Ereignisse in den Familien wahrscheinlich nicht mehr besprochen wurden während die realen bedeutsamen Geschehnisse weiterhin in Unterhaltungen präsent waren. Außerdem argumentieren sie mit unterschiedlichen Gedächtnisrepräsentationen auf neuronaler Ebene von real wahrgenommenen Geschehnissen und nur vorgestellten Ereignissen.

In einer Untersuchung von Ceci, Loftus, Leichtman & Bruck (1994) wurden Kinder (jeweils 20 drei- bis vierjährige und fünf- bis sechsjährige) zu acht Ereignissen, davon vier fiktiven befragt. Letztere wurden ihnen mit dem Hinweis vorgegeben, dass sie diese tatsächlich erlebt und ihre Eltern dies bestätigt hätten. Die Kinder wurden aufgefordert,

über die Ereignisse intensiv nachzudenken und sich diese bildhaft vorzustellen. Nach einer offenen Befragung zu den Ereignissen wurden sie gefragt, ob sie sich daran erinnern könnten oder nicht. Die Kinder wurden insgesamt zwölfmal gefragt. Das letzte Interview wurde durch eine andere Interviewerin durchgeführt als die vorherigen. Diese erklärte den Kindern, dass sie in den vorherigen Befragungen irrtümlich zu Ereignissen gefragt worden seien, die sie nicht erlebt hätten. Die Kinder wurden dann aufgefordert, Angaben zu den Ereignissen zu machen, die sie tatsächlich erlebt hätten. Die Auswertung zeigte, dass 91 bis 100% der wahren Ereignisse von den Kindern richtig erinnert wurden. Den fiktiven Ereignissen stimmten die jüngeren Kinder in der ersten Befragung zu 34% zu, die älteren Kinder zu 25%. In der elften Befragung steigerten sich die Zustimmungsraten auf 54% bei den jüngeren und auf 49% bei den älteren Kindern. Nach Aufklärung am letzten Termin bestätigten die Kinder 45%, bzw. 40% der fiktiven Ereignisse als real erlebt. Je früher die Kinder den fiktiven Ereignissen im Verlauf der Befragungen zustimmten, desto eher stimmten sie auch in den späteren Interviews zu, wechselten also nicht in ihrer Einschätzung. Drei Kinder bestätigten in der elften Sitzung erstmalig ein fiktives Erlebnis als real erlebt. Die Autoren interpretieren ihre Ergebnisse dahingehend, dass die mehrmalige Zustimmung zu einem fiktiven Ereignis darauf hinweist, dass diese Kinder tatsächlich davon ausgingen, das Ereignis erlebt zu haben. Kinder die wechselnd antworteten und nach Aufklärung in der letzten Sitzung ihre Zustimmung zum Erleben des Ereignisses widerriefen, sind sich nach Ansicht der Autoren entweder unsicher hinsichtlich der Realität des Ereignisses oder gehen davon aus, dass sie das Ereignis tatsächlich nicht erlebt haben.

Aus den angeführten Studien kann nicht abgeleitet werden, ob falsche Quellenzuordnungen ursächlich für die Bestätigung fiktiver Ereignissen sind oder andere Erklärungen wie Ja-Sage-Tendenz, bewusstes Lügen, Geltungsstreben, Gruppendruck oder Tendenzen im Sinne der sozialen Erwünschtheit zu antworten, dass gezeigte Verhalten erklären.

McBrien und Dagenbach (1998) bezogen in ihren Experimenten diese Faktoren mit ein. Darüber hinaus gingen sie der Frage nach, ob die Plausibilität eines Ereignisses einen Einfluss auf die Entstehung von nicht erlebnisgebundenen Erinnerungen hat. Kindern im Vorschulalter (15 drei- bis vierjährige, 30 fünf- bis sechsjährige) und im Grundschulalter

(30 sieben- bis achtjährige) wurde je fünf plausible und fünf unplausible Ereignisse als Bilder vorgegeben. Ein unplausible Ereignis war zum Beispiel die Geschichte mit einer Rakete zum Mond gefahren zu sein. In Einzel- und Gruppensituationen wurden sie mehrfach aufgefordert, darüber nachzudenken, ob sie das Ereignis erlebt hatten. Vorgelegt wurden jedem Kind zunächst zwei fiktive Ereignisse. In einer letzten Befragung wurden sie außerdem zu vier neuen fiktiven Ereignissen gefragt, die bisher nicht weiter thematisiert worden waren. Die Ergebnisse zeigen, dass die älteren Kinder (fünf- bis achtjährige) weder den mehrfach nachgefragten fiktiven Ereignissen noch den nur einmal thematisierten Ereignissen zustimmten. Dagegen stimmte etwa die Hälfte der jungen Kinder sowohl den mehrfach angesprochenen als auch nur den einmal vorgegebenen fiktiven Ereignissen zu. Die Plausibilität des Ereignisses und die Variable Einzel- bzw. Gruppenbefragung hatte keinen Einfluss auf die Zustimmungsrate. Die Autoren schließen aus ihren Resultaten, dass die falschen Zustimmungen zu den fiktiven Ereignissen sowohl durch fehlerhafte Quellenzuordnungen als auch durch andere Faktoren, z.B. "Ja-Sage-Tendenz" bestimmt werden.

In einem zweiten Experiment von McBrien und Dagenbach (1998) wurden drei bis vierjährigen Kinder vier plausible und vier unplausible fiktive Ereignisse bildhaft mehrfach mit der Aufforderung vorgegeben, über jedes Ereignis intensiv nachzudenken. Die Kinder wurden dann zu diesen Ereignissen und zu acht neuen fiktiven Ereignissen befragt. Dabei wurden sowohl positiv als auch negativ formulierte Fragen gestellt. Es zeigte sich, dass den mehrfach vorgelegten fiktiven Ereignissen häufiger (27%) zugestimmt wurde als den einmalig thematisierten Geschehnissen (19%). Achtunddreißig Prozent der Kinder stimmten mindestens einem neuen Ereignis zu. Weiter zeigte sich, dass sich bei wiederholt vorgegebenen Ereignissen höhere Zustimmungsraten ergaben, wenn diese plausibel waren.

Verschiedene Studien (z.B. Ceci & Bruck, 1993; Volbert & Pieters, 1996) konnten zeigen, dass unterschiedliche soziale und motivationale Faktoren die Wirksamkeit von Suggestion beeinflussen können. Hierzu gehören u.a. die Induktion von Stereotypen, Gruppendruck durch Gleichaltrige, Voreingenommenheit der Interviewers, der emotionale

Ton des Interviews, das Machtgefälle zwischen fragendem Erwachsenen und antwortendem Kind, die Wiederholung von Fragen und Mehrfachbefragungen mit falschen Informationen. Das Bemühen eines Kindes einer Autoritätsperson gerecht zu werden, kann in einer Befragung dazu führen, dass Fehlinformationen übernommen werden, auch wenn Erinnerungen an ein tatsächliches Ereignis vorhanden sind. So fanden Ceci et al. (1987) mehr Suggestionseffekte bei Vorschulkindern, wenn die nachträglichen Fehlinformationen durch einen Erwachsenen präsentiert wurden als wenn diese von einem siebenjährigen Kind vorgegeben wurden. Kinder schätzen erwachsene Gesprächspartner möglicherweise als kompetenter und glaubwürdiger ein als andere Kinder und werten daher die eigenen Erinnerungen als weniger zuverlässig. Memon, Cronin, Eaves, Bull und Küpper (1992) beschreiben den *Omniscient Adult Effect*, der besagt, dass Kinder ihr Wissen durch Frage-Antwort-Spiele von Erwachsenen erwerben und folglich davon ausgehen, dass Erwachsene die Antworten auf ihre eigenen Fragen kennen, also „allwissend“ sind. Somit lassen sich Kinder eher von Erwachsenen als von anderen Kindern suggestiv beeinflussen, da sie die „Glaubwürdigkeit“ einer Informationsquelle zunächst abhängig vom Alter derselben einschätzen (Ceci et al. 1987). Diese Tendenz wird gemindert, wenn Kinder vorher darüber informiert werden, dass die gestellten Fragen möglicherweise irreführend sein können, oder wenn der Erwachsene, der die irreführende Information gibt, vorher den Kindern gegenüber diskreditiert wird (Warren, Hulse-Trotter & Tubbs, 1991; Lampinen & Smith, 1995).

Unsicherheiten hinsichtlich der eigenen Erinnerung zeigen Kinder auch dann, wenn ihnen inhaltsgleiche Fragen mehrfach gestellt werden. Verschiedene Untersuchungen haben gezeigt, dass Kinder dazu neigen, ihre Antworten zu ändern, wenn ihnen in kurzen zeitlichen Abständen dieselben Fragen gestellt werden. Dieser Effekt kann durch die Verwendung spezieller Fragetypen verstärkt werden. So fanden Poole und White (1991), dass vierjährige Kinder ihre Antworten änderten, wenn ihnen „Ja-“ oder „Nein“-Fragen gestellt wurden. Wiederholungen von offenen Fragen hatten dagegen keinen Effekt auf die Antworten der Kinder. Sie können sogar dazu beitragen, Vergessensprozesse zu reduzieren (Poole & White, 1995; Warren & Lane, 1995). Möglicherweise liegt dies daran, dass offene Fragen von den Kindern als Aufforderung mehr Informationen zu geben, interpre-

tiert werden. Demgegenüber signalisiert die Wiederholung von "Ja-" oder "Nein-" Fragen dem Kind, dass seine erste Antwort falsch gewesen sein muss und eine korrigierte Antwort erwartet wird.

Andere Forschungsergebnisse zeigen, dass versteckte oder offene Beschuldigungen Kinder in einer Befragung derart verunsichern können, dass sie Einzelheiten über ein Geschehen berichten, an das sie sich tatsächlich nicht erinnern können.

Goodman, Wilson, Hazan & Reed, 1989 (zitiert nach Ceci & Bruck, 1993) befragten sieben- bis elfjährige Kinder über ein experimentell simuliertes Ereignis, welches vier Jahre zuvor stattgefunden hatte. Durch das Schaffen einer die Kinder anschuldigenden Atmosphäre erreichten sie, dass ein Drittel der Kinder falsche Informationen der Interviewer über das frühere Geschehen als korrekt bestätigten.

Bereits eine eher negative Atmosphäre ohne explizite Anschuldigungen kann in der Befragung zu mehr falschen Antworten führen als eine emotional positiv geführte Befragung. Goodman, Rudy, Bottoms & Aman (1990) ließen drei- bis siebenjährige Kinder durch freundliche und durch neutrale Interviewer zu einer zwei Wochen vorher stattgefundenen Impfung befragen. Es zeigte sich, dass die jüngeren Kinder gegenüber den freundlichen Personen weniger falsche Antworten gaben und weniger auf Fehlinformationen eingingen als bei der Befragung durch neutrale Interviewer.

Leichtman & Ceci (1995) überprüften die Wirkung von Vorurteilen auf die Empfänglichkeit für Suggestionen. Sie konnten zeigen, dass die Vorgaben eines Personenstereotyps, (z.B. Tolpatschigkeit), vor dem Erleben eines Geschehens dazu führte, dass Kindergartenkinder in ihren Aussagen nicht nur auf Falschinformationen reagierten, sondern ihre Schilderungen auch mit Darstellungen anreicherten, die dem zuvor vorgegebenen Stereotyp entsprachen.

Weitere Variablen, die in Zusammenhang mit Suggestionseffekten diskutiert werden, sind *biologischen Faktoren*. Dazu gehört auch das Alter der Kinder. Ceci und Bruck (1993) folgerten aus einer Analyse experimenteller Untersuchungen, dass Kinder unter sechs Jahren anfälliger für suggestive Einflüsse sind als ältere Kinder. Dabei waren drei

bis vierjährige Kinder besonders suggestibel. Allerdings waren die Alterseffekte nicht in allen Studien vorhanden, sondern traten in Abhängigkeit von den Suggestionsbedingungen auf. Während eine Vielzahl von Studien, z.B. Cohen und Harnick (1980), King und Yuille (1987), Ceci, Ross und Toglia (1987), Ornstein, Gordon und Larus (1992) und Lepore & Sesco (1994), McBriar und Dagenbach (1998), zeigte, dass Kinder deutlich anfälliger für Suggestivfragen waren als Erwachsene, wurden in wenigen anderen Studien keine altersabhängigen Unterschiede festgestellt (Goodman & Clarke-Stewart, 1991), oder nur bei bestimmten Suggestivfragen, z.B. zu Fragen nach peripheren Details, beobachtet (Goodman & Reed, 1986; Rudy & Goodman, 1991). Dagegen zeigten sich in einer Studie von Vrij und Bush (2000) Altersunterschiede in der Suggestibilität: fünf- bis sechsjährige Kinder ließen sich durch suggestive Fragen signifikant stärker beeinflussen, als zehn- bis elfjährige.

Auch Stress und emotionale Erregung werden als biologische Faktoren in Verbindung mit Suggestibilität angeführt (Ceci & Bruck, 1993). Kinder, die Ereignisse unter stark belastenden Umständen erlebt hatten, werden durch die Befragung erneut einer erheblichen emotionalen Belastung ausgesetzt. Die Ergebnisse experimenteller Untersuchungen lassen bisher aber keine eindeutigen Zusammenhänge zwischen Stress und Gedächtnis erkennen.

Ceci und Bruck (1993) fanden in 15 Studien zu diesem Thema nur in zwei Untersuchungen eine positive Korrelation zwischen Stress und Gedächtnis. In fünf Untersuchungen konnte kein Zusammenhang festgestellt werden. In sechs Studien wurde ein negativer Zusammenhang gefunden wurde, was darauf hinweist, dass ein erhöhtes Erregungsniveau beim Erleben und Reproduzieren des Erlebten negative Auswirkungen auf die Erinnerungsleistungen hatte. Auch die Analyse der Studien, die suggestive Beeinflussungen einbezogen, ergab nach Ceci und Bruck (1993) kein einheitliches Bild. In den meisten Untersuchungen wurde kein Zusammenhang zwischen Stress und suggestiver Beeinflussung festgestellt. In einigen Studien wurden zwar Korrelationen gefunden, diese hatten aber unterschiedliche Richtungen. Goodman, Hirschmann, Hepps & Rudy, (1991; zitiert nach Ceci & Bruck, 1993) folgerten aus ihren Ergebnissen mit Vorschulkindern, dass

eine erhöhte emotionale Belastung die Genauigkeit und Richtigkeit der Erinnerungen verbesserte und die Anfälligkeit für suggestive Einwirkungen verringerte. Ornstein et al. (1992) fanden in Abhängigkeit vom Stresserleben Unterschiede in der Anzahl berichteter Details über eine kinderärztliche Untersuchung. Je größer die emotionale Belastung der Kinder durch den Arztbesuch war, desto geringer war die Anzahl der geschilderten Einzelheiten. Peters (1991, zitiert nach Ceci & Bruck, 1993) schloss aus seinen Experimenten, in denen er verschiedene Stress erzeugende Ereignisse und physiologische Maße einsetzte, dass ein erhöhter Stresszustand der Kinder deren Erinnerungsleistung verbesserte. In keiner Untersuchung reduzierte ein erhöhtes Erregungsniveau die Gedächtnisleistungen. Bei Einbezug suggestiver Beeinflussung stellte er fest, dass Kinder, die in einer der Versuchsbedingungen einem starkem Stresserlebnis ausgesetzt wurden, sich als empfänglicher für suggestive Falschinformationen in der Befragung zu dem Ereignis erwiesen als Kinder, die moderatem Stress ausgesetzt wurden.

Möglicherweise sind die unterschiedlichen Ergebnisse auf die verschiedenen Operationalisierungen der untersuchten Variablen zurückzuführen. So kann zum Beispiel Stress, durch Selbsteinschätzung, Verhaltensbeobachtung, Fremdeinschätzung oder physiologische Maße wie Herzrate und Blutdruck erfasst werden. Aber auch die unterschiedlichen Stress-Level in Situationen, ungleiche Verknüpfungen zwischen Stress und Ereignis, der Transfer des Stress-Erlebens auf die Abrufsituation könnten die heterogenen Befunde beeinflusst haben.

Die dargestellten Faktoren, die die Empfänglichkeit von Kindern für Suggestionen beeinflussen, führen zu der Überlegung, dass diese Bedingungen die Beurteilung von Aussagen, z.B. im forensischen Kontext, erheblich beeinträchtigen können. So stellt sich z. B. die Frage, ob Aussagen, die unter suggestiven Bedingungen zustande gekommen sind mittels inhaltlicher Merkmale richtig eingeordnet werden können.

3.2 Suggestibilität im Kontext aussagepsychologischer Beurteilung

In den 80er und 90er Jahren des letzten Jahrhunderts wurden durch spektakuläre Prozesse ("McMartin-Fall", "Kelly-Michaels-Fall" in den USA, "Montessori-Prozess", "Wormser Kinderschänderprozess" in Deutschland, (vgl. z.B. Friedrichsen, 1995a, b, Steller, 1998) zahlreiche Forschungsaktivitäten ausgelöst. In diesen Prozessen wurden gravierende, immer bizarner werdende Anschuldigungen des sexuellen Missbrauchs gegen Kindergärtnerinnen oder Familienangehörige erhoben. Im Vordergrund der empirischen Untersuchungen standen demzufolge suggestive Befragungsformen und deren Auswirkungen auf die Schilderungen von Ereignissen. Vergleichsweise weniger Untersuchungen gibt es, in denen die Bedeutung suggestiver Einflüsse im unmittelbaren forensischen Kontext überprüft werden. So nahmen suggestive Einflüsse auch dadurch zu, dass als Folge der verstärkten öffentlichen Sensibilität für Sexualdelikte gegen Kinder generell eine erhöhte Aufdeckungsbereitschaft im Bereich des sexuellen Missbrauchs besteht und Anleitungen zur Aufdeckung tatsächlichen oder vermeintlichen sexuellen Missbrauchs große Verbreitung gefunden haben (z.B. Furniss, 1993).

Der praktisch tätige Sachverständige wird mit dem Problem konfrontiert, suggestive Einflüsse in der Aussagegeschichte zu identifizieren und ihre möglichen Auswirkungen auf die Aussage festzustellen und zu bewerten. Die Relevanz dieser möglichen Fehlerquelle für die Beurteilung der in Frage stehenden Aussage ergibt sich aus den vorangegangenen Darstellungen der Ergebnisse aus der Suggestionsforschung, nach denen suggestive Einflussnahme erhebliche Auswirkungen auf den Inhalt kindlicher Schilderungen haben kann. Für die praktische Begutachtung ergibt sich daraus die Frage, ob die kriterienorientierte aussagepsychologische Analyse ein valides Vorgehen zur Unterscheidung von erlebnisbegründeten, partiell oder vollständig suggerierten Aussagen sein kann.

Studien über die Brauchbarkeit der kriterienorientierten Inhaltsanalyse bei Aussagen, die unter suggestiven Bedingungen zustande gekommen sind, weisen auf Schwierigkeiten bei der Anwendung der kriterienorientierten Aussagenanalyse hin. Es zeigt sich nämlich, dass die Realkennzeichen weniger gut zwischen erlebnisbegründeten und suggerierten Aussagen, bzw. Aussageteilen trennen können, als bisherige experimentelle Studien zeig-

ten, in denen mit Hilfe der Realkennzeichen mit gutem Erfolg zwischen erlebnisbegründeten und phantasiegeleiteten Aussagen unterschieden werden konnte.

In einer experimentellen Untersuchung von Lehr (1995, vgl. für eine ausführliche Beschreibung des experimentellen Paradigmas Pieters, 1994) wurde erstmalig suggestive Beeinflussung als unabhängige Variable in den Versuchsplan aufgenommen. Ausgangspunkt der Überlegungen war die Beobachtung, dass in der aussagepsychologischen Begeutachtungspraxis mit Kindern weniger die vollständig falsche Aussage zu finden ist, sondern eher eine Kombination aus erlebten und nicht erlebten Erfahrungen. In der Studie wurden daher Beeinflussungen in ein tatsächlich erlebtes Geschehen eingebettet. Drei Kindergruppen im Alter zwischen vier und sechs Jahren erlebten getrennt eine Clownvorführung, die aus mehreren Teilepisoden bestand (z.B. Clown macht Seifenblasen, Clown macht Zaubertee, Clown singt ein Lied). Eine Gruppe erlebte das vollständige Programm, das in Basisereignisse und kritische Ereignisse unterteilt worden war. Die Suggestionsgruppe und die Instruktionsgruppe sahen dagegen nur die Basisereignisse. Während der Instruktionsgruppe die kritischen Ereignisse verbal mit der Aufforderung mitgeteilt wurden, diese in der späteren Befragung als tatsächlich erlebt zu schildern, wurden die Kinder der Suggestionsgruppe nur beiläufig über die kritischen Ereignisse informiert. 65 vollständige Interviews wurden in die Auswertung einbezogen. In diesen schilderten die Kinder 227 Teilepisoden. In die Auswertung wurden nur die Aussagen von Interviews aufgenommen, in denen die Kinder tatsächlich Schilderungen zu allen Episoden machten. Die Ergebnisanalyse zeigte, dass das Kriterium Wiedergabe von Gesprächen in wahren Aussagen signifikant häufiger auftrat als in instruierten und suggerierten Schilderungen. Auch Spontane Verbesserungen der eigenen Aussage fanden sich im Vergleich von wahren und suggerierten Aussagen signifikant häufiger in der Gruppe der wahren Berichte.

Der Gesamtscore der Realkennzeichen konnte nicht zwischen den drei Untersuchungsgruppen trennen. Kritisch ist zu dieser Untersuchung anzumerken, dass das Basismaterial für eine kriterienorientierte Aussagenanalyse keine ausreichende Grundlage bietet, da im Rahmen der fünf Teilepisoden nur Kurzaussagen zustande kommen konnten. Von der Autorin wird dazu angemerkt, dass nur wenige kriterienerfüllende Textstellen in den

Schilderungen der Episoden gefunden werden konnten. Auch das Fehlen einer Kontrollgruppe ohne Basisereignis schränkt die Interpretationsmöglichkeiten und Generalisierbarkeit der Studie ein (das Untersuchungsparadigma wurde für eine andere Studie gewählt, in der es unabhängig von der Validität der Realkennzeichen um Suggestionseinflüsse ging, vgl. Pieters, 1994).

Weitere Untersuchungsergebnisse zeigten, dass kriterienkundige Rater signifikant höhere Trefferquoten bei unwahren Aussagen (gebildet aus instruierten und suggerierten Berichten) erzielten als kriterienunkundige Beurteiler. Dagegen erzielten die kriterienunkundigen Rater bei wahren Aussagen tendenziell höhere Trefferquoten. Dieses Ergebnis könnte hinsichtlich der praktischen Aussagebegutachtung dahingehend interpretiert werden, dass die Anwendung der inhaltlichen Aussagenanalyse gerade dort besonders hilfreich ist, wo die meisten Zuordnungsfehler gemacht werden, nämlich bei der Identifizierung falscher Aussagen. Die Ergebnisse der Studie, insbesondere die nicht zwischen den Gruppen differenzierenden Gesamtscores, weisen auf die Probleme hin, die bei der Anwendung der kriterienorientierten Analyse bei beeinflussten Aussagen offensichtlich zu erwarten sind.

Leichtmann und Ceci (1995) zeigten in ihrer Untersuchung die Schwierigkeit auf, suggestiv beeinflusste Aussagen von unbeeinflussten erlebnisbegründeten Aussagen zu unterscheiden. Insgesamt 119 Praktiker und Wissenschaftler, die Erfahrungen in der Diagnose des sexuellen Missbrauchs bei Kindern hatten, sahen Videoaufnahmen von Schilderungen dreier Kinder über den zweiminütigen Besuch einer Person namens Sam Stone im Kindergarten. Während nur ein Kind eine vollständig richtige Darstellung dieses Ereignisses gab, enthielten die Berichte der beiden anderen Kinder in unterschiedlichem Ausmaß suggerierte und ausgedachte Details. Die Experten lagen bei der Klassifizierung der Berichte als erlebt oder nicht erlebt aufgrund ihres klinischen Urteils nicht über dem Zufallsniveau.

In einer Untersuchung von Bruck, Hembrooke und Ceci (1997) wurde u.a. eine Analyse der inhaltlichen Qualität von suggerierten Schilderungen vorgenommen. Sechzehn Vor-

schulkinder wurden fünfmal zu zwei wahren und zwei fiktiven Erlebnissen befragt. Beim ersten Termin wurden die Kinder gefragt, ob die Erlebnisse stattgefunden hätten, in drei weiteren Terminen wurde unter Einsatz suggestiver Techniken zu den Ereignissen befragt, u.a. wurden sie aufgefordert so sich so zu verhalten, als ob die fiktiven Ereignisse stattgefunden hätten. In einer letzten Befragung wurden sie suggestionsfrei zu den Ergebnissen exploriert. Es zeigte sich, dass die Kinder die fiktiven Ereignisse im Verlauf der Befragung je nach emotionaler Tönung des Ereignisses (positiv/negativ) zu 80% bzw. 70% als erlebt bestätigten. Die inhaltliche Qualität der Aussagen wurde durch Detailmenge, Anzahl spontaner Äußerungen, Kohärenz und Elaboriertheit bestimmt. Kohärenz wurde durch zeitliche Angaben und die Schilderung von Dialogen operationalisiert. In die Variable Elaboriertheit gingen die Schilderungen von emotionalen Inhalten und die Verwendung von Adjektiven und Adverbien ein. In der ersten Befragung waren die Schilderungen über reale, positiv getönte Erlebnisse detailreicher als Darstellungen über fiktive und reale negativ getönte Ereignisse. Im Verlauf der Befragungen berichteten die Kinder mehr Details über die fiktiven Ereignisse als über die realen Geschehnisse. Hinsichtlich der Kohärenz und der Elaboriertheit der Darstellungen ergaben sich kaum Unterschiede zwischen wahren und fiktiven Ereignissen. Tendenziell waren die Schilderungen über fiktive Ereignisse kohärenter und elaborierter als die Angaben zu den realen Erlebnissen. Zwar wurden in dieser Untersuchung keine expliziten Klassifizierungen der Aussagen nach dem Erlebnishintergrund vorgenommen, anzunehmen ist aber, dass bei einer Beurteilung des Erlebnisgehaltes auf der Grundlage der beschriebenen inhaltlichen Qualitäten, Schilderungen über fiktive Ereignisse häufig als reale Erlebnisse diagnostiziert worden wären.

Erdmann, Volbert und Böhm (2004) untersuchte die Induktion von Pseudoerinnerungen bei Kindern sowie deren Niederschlag in der Qualität von Schilderungen. In einer experimentellen Untersuchung wurden 67 Erstklässler jeweils viermal zu einem realen und fiktiven Ereignis befragt. Die Befragung zu fiktiven Ereignissen erfolgte unter Einsatz suggestiver Techniken. An einem fünften Termin erfolgte eine Befragung eines Teils der Kinder durch für die Versuchsbedingung blinde Experten wiederum zu beiden Geschehnissen. Nach einem mehrwöchigen Abstand wurden die Kinder von anderen Interviewern

erneut befragt. Sie wurden darüber informiert, dass die früheren Interviewer Fehler gemacht hätten und sie zu Ereignissen befragt hätten, die sie nicht erlebt hätten. Die Kinder wurden nun aufgefordert zu überdenken, welches der Ereignisse sie erlebt hätte und welches nicht.

Die Ergebnisse zeigten, dass vor der suggestiven Einflussnahme 28% der Kinder das fiktive Ereignis bejahten, 69% es verneinten. Beim vierten Termin zeigte sich ein umgekehrtes Verhältnis, d.h. die Mehrzahl der Kinder stimmte im Verlauf suggestiver Einflussnahmen dem fiktiven Ereignis zu. Nach einer mehrwöchigen Pause bejahten immer noch 47% der Kinder den Realitätsgehalt des fiktiven Ereignisses, 17% waren sich unsicher, 36% verneinten ihn. In nachfolgenden Befragungen durch informierte Interviewer, bei suggestionsfreier Befragung durch blinde Experten sowie nach Teilaufklärung bejahten ein Drittel der Kinder weiterhin den Realitätsgehalt des fiktiven Ereignisses. Die Autorin wertete dies als deutlichen Hinweis auf die Entstehung von Pseudoerinnerungen. Weiterhin wurden in einer fünften Befragung suggestionsfrei erhobene Schilderungen einer inhaltlichen Qualitätsanalyse unterzogen. Suggerierte und erlebnisbegründete Aussagen zeigten kaum qualitative Unterschiede, während beim ersten Befragungszeitpunkt deutlich mehr Realkennzeichen in den erlebnisbegründeten Schilderungen als in der suggerierten Darstellungen vorhanden waren. Die Autorin zog daraus den Schluss, dass die merkmalsorientierte Inhaltsanalyse in der praktischen Glaubhaftigkeitsbeurteilung keine ausreichende Validität aufweist, wenn in der Entstehungs- und Entwicklungsgeschichte der Aussage Hinweise auf intensive suggestive Einflussfaktoren vorhanden sind.

Zu einem ähnlichen Ergebnis kam Niehaus (2001). Sie untersuchte die Validität der inhaltsbezogenen Kriterien, wobei sie nicht lediglich zwischen wahren und frei erfundenen Aussagen bei der Operationalisierung der Versuchbedingungen unterschied, sondern zwei weitere Gruppen untersuchte, die sich im Grad der Wahrnehmungsgrundlagen, die den Kindern bei ihrer Schilderung eines nicht erlebten Ereignisses zur Verfügung standen. Die Kinder wurden zu einem traumatischen Verkehrsunfallerlebnis befragt. Dabei bestanden die zwei zusätzlichen Gruppen aus einer, die eine Erzählung eines anderen Kindes über ein solches Erlebnis gehört hatte und einer weiteren, die ebenfalls diese Geschichte gehört hatte und zudem auf eine ähnliche Erfahrung in ihrer eigenen Vergangen-

heit zurückgreifen konnten. Dabei zeigte sich, dass die verwendete Kriteriologie sehr gut zwischen wahren und frei erfundenen Aussagen trennen konnte. Auch die Unterschiede zwischen der Gruppe mit Erlebnishintergrund zu den Gruppen, die auf eine Wahrnehmungsgrundlage zurückgreifen konnten zeigten sich signifikant. Es zeigte sich zudem, dass in den beiden Gruppen, in denen die Kinder auf einen gewissen Wahrnehmungshintergrund zurückgreifen konnten signifikant mehr Realkennzeichen zu aufzufinden waren, als in der Gruppe der frei erfundenen Erzählungen. Die Autorin zieht aus diesen Ergebnissen und aus den geringen Trefferquoten für die beiden Gruppen mit Wahrnehmungsgrundlage (43,75% und 55%) den Schluss, dass "das Merkmalssystem sehr nützlich für die Identifizierung frei erfundener Aussagen ist, die Anwendbarkeit jedoch stark eingeschränkt ist, wenn Aussagende auf einschlägige Erfahrung oder angeeignetes Wissen zurückgreifen können." (S. 305). Zwar ist das Hören einer Erzählung über das Erlebnis und ein bestehender Erfahrungshintergrund über ein ähnliches Ergebnis nicht gleichzusetzen mit suggestiver Beeinflussung, jedoch zeigt sich in den Ergebnissen der Studie ein Trend, der für den Bereich der Suggestion von Bedeutung sein kann: Die Zunahme der Realkennzeichen von rein fantasiegeleiteten Aussagen hin zu solchen, denen ein gewisser Gedächtnisinhalt für das Erlebnis (wenn auch lediglich als Erzählung gehört oder in Form der Erinnerung an ein ähnliches Erlebnis) zu Grunde liegt.

Ähnlich führte schon Köhnken (1999) aus, dass die kriterienorientierte Inhaltsanalyse lediglich zur Trennung von absichtlichen Falschaussagen brauchbar ist, nicht aber zur Aufdeckung suggestionsbedingter Falschaussagen.

4. Fragestellung und Hypothesen

4.1 Ableitung der Fragestellung

Die vorangegangenen Ergebnisse experimenteller Studie zeigen, dass bisher unsicher ist, ob eine kriterienorientierte Inhaltsanalyse anhand festgelegter Realkennzeichen auch im Falle einer möglichen Suggestion eine angemessene Differenzierung zwischen erlebnisbegründeten und suggerierten Berichten ermöglicht. Zudem zeigen die in Kapitel 2 beschriebenen Feldstudien in der Mehrzahl betätigenden Ergebnisse für die Undeutsch-Hypothese, allerdings zeigten die Studien große Varianz in der Validität der kriterienbezogenen Inhaltsanalyse. So konnte Boychuk (1991) in ihrer Untersuchung wahre und nicht wahre Aussagen mittels der Kriterien gut differenzieren, während in anderen Studien (Lehr, 1995; Lamb et al., 1997) weniger deutliche Unterschiede hinsichtlich des Vorkommens von Realkennzeichen in den Experimentalgruppen vorhanden waren. Dabei waren die Studien stark unterschiedlich in Bezug auf die Größe der Stichprobe, die Größe der Gruppen in den Bedingungen und die untersuchten Altersbereiche. Verschiedene oben genannte Studien konnten einen Alterseffekt für Suggestibilität aufzeigen mit einer höheren Suggestibilität für die jüngeren Altersgruppen (z.B Cohen & Harnick, 1980, King & Yuille, 1987, Ceci et al., 1987, Ornstein et al., 1992, Lepore & Sesco, 1994, McBriar & Dagenbach, 1998 und Vrij & Bush, 2000). Aus diesem Grund ist für die Frage nach der Anwendbarkeit kriterienorientierter Inhaltsanalysen wie der Realkennzeichenanalyse bei möglichen suggerierten Inhalten die Gruppe der Vorschulkinder von besonderem Interesse, da hier das höchste Ausmaß an suggestiven Einflüssen zu erwarten ist. Zwar wurden in mehrere Feldstudien Kinder ab einem Alter von vier Jahren miteinbezogen - in eine Studie (Lamers-Winkelmann, 1992a) sogar Kinder ab zwei Jahren - allerdings lassen sich unter den experimentellen Studien nur die Untersuchungen von Honts et al. (1992) und Lehr (1995) finden, die auch Vorschulkinder miteinbezogen.

Bei Honts et al. (1992) deutet allerdings der relativ weit gespannte Altersbereich von vier bis zehn Jahren bei einer Stichprobengröße von 17 Versuchspersonen darauf hin, dass die

Ergebnisse dieser Untersuchung auf die Altersgruppe der Vorschulkinder aufgrund der kleinen Stichprobe nicht zu generalisieren sind.

Die Studie von Lehr (1995) mit vier bis sechsjährigen Kindern war ursprünglich nicht für die Überprüfung der Validität der Kriterien konzipiert, sondern verfolgte die Überprüfung von Suggestionen auf Kinderaussagen (vgl. Pieters, 1994). Daher kann angenommen werden, dass die Aussagekraft dieser Studie hinsichtlich der Validität der Kriterien eingeschränkt sein dürfte. Alle anderen experimentellen Untersuchungen überprüfen die Brauchbarkeit der kriterienbezogenen Aussagenanalyse bei Erwachsenen und Schulkindern. Erdmann (2001) begründete ihre Wahl sechs bis achtjähriger Versuchspersonen damit, dass für eine inhaltliche Qualitätsanalyse möglichst umfangreiche Schilderungen vorliegen müssen, die von den Kindern auf offene Fragen und Erzählanstöße produziert werden sollten. Diese Voraussetzungen liegen nach ihrer Einschätzung eher bei Kindern ab dem Grundschulalter vor. Obgleich diese Überlegungen nachvollziehbar sind, zeigt die forensische Praxis, dass Gutachtenaufträge auch gerade dann erteilt werden, wenn sehr junge Kinder betroffen sind.

Die Erfahrungen aus der Gutachtenpraxis zeigen, dass von Seiten der Auftraggeber häufig Verunsicherungen bezüglich der entwicklungsbedingten Besonderheiten von Vorschulkindern und deren Auswirkungen auf die zu erstattende Aussage als Auslöser für die aussagepsychologische Begutachtung angegeben werden. Unterstützt werden diese Bedenken, dass bei der Beurteilung von Aussagen junger Kinder möglicherweise andere Maßstäbe anzulegen sind als bei Einordnung von Schilderungen älterer und Erwachsener auch durch die Befunde zur kognitiven Entwicklung und zum Spracherwerb (vgl. Oerter & Montada, 2002). Diese weisen daraufhin, dass in der Altersgruppe der Vorschulkinder noch bedeutsame Entwicklungsschritte vollzogen werden müssen, bevor die kognitiven und kommunikativen Fähigkeiten und Fertigkeiten dann mit etwa acht Jahren weitgehend ausgebildet sind. Bezogen auf die inhaltliche Analyse von Aussagen könnten diese entwicklungsbedingten Grenzen oder Einschränkungen dazu führen, dass in Schilderungen von Vorschulkindern weniger Kriterien gefunden werden und damit die Beurteilung der Glaubhaftigkeit der Aussagen eine geringere Validität aufweist. Ursächlich können mög-

licherweise fehlende intellektuelle Kompetenzen angenommen werden, die mit einem geringeren Aussageumfang einhergehen. So erfordern Kriterien wie Entlastung des Angeschuldigten, Beschreibung eigen- und fremdpsychischer Vorgänge und Interaktions- schilderungen kognitive Operationen, die in dieser Alterstufe erst noch ausgebildet werden. Unterstützt werden diese Überlegungen durch verschiedene Untersuchungen (Anson et al., 1993; Lamb et al., 1997; Craig et al., 1999; Buck et al., 2002, Vrij et al. 2002, Blandon- Gitlin, Pezdek, Rogers & Brodie, 2005), die Alterseffekte bezüglich des Auftretens der Realkennzeichen zeigten. Allgemein ließen sich in den Schilderungen von älteren Kindern und Erwachsenen mehr Realkennzeichen finden als in denen jüngerer Kinder. Bei Vrij et al. (2002) zeigte sich der Alterseffekt besonders deutlich in der jüngsten Versuchsgruppe der fünf- bis sechsjährigen im Vergleich zu allen anderen Versuchsgruppen.

Diese Studien weisen aber häufig wie oben bereits erwähnt sehr weite Altersbereiche auf. Darüber hinaus ist der Anteil der jüngeren Kinder zumeist sehr gering, so dass mit den bisher vorliegenden Studien keine sicheren Schlussfolgerungen über die Brauchbarkeit der merkmalsorientierten Inhaltsanalyse bei Vorschulkindern getroffen werden können. In der vorliegenden Untersuchung wurde dieser Altersbereich daher als besonders forschungsbedürftig betrachtet. Erreicht werden sollte mit der Beschränkung auf Vorschulkinder, dass bei isolierter Betrachtung gerade dieser Altersgruppe Alterseffekte innerhalb der Untersuchungsstichprobe von vornherein ausgenommen werden konnten.

Vor dem Hintergrund der genannten entwicklungsbedingten Besonderheiten interessierte im Rahmen der vorliegenden Untersuchung auch die Erfassung möglicher Einflüsse von intellektuellen und sprachlichen Kompetenzen der Vorschulkinder auf die inhaltliche Qualität von Schilderungen über einen Sachverhalt. Anzunehmen ist, dass eine höhere intellektuelle und sprachliche Kompetenz mit dem stärkeren Auftreten von Realkennzeichen- der Prüfgröße der Qualität - einhergeht. In den referierten Ergebnissen zur Validität der Realkennzeichen wurde deutlich, dass in den meisten Studien jeweils nur einige Kriterien aus dem Kriterienkatalog von Steller & Köhnken (1989) ausreichend trennscharf waren und diese je nach Untersuchung unterschiedlich waren. In den meistens

Untersuchungen – sowohl bei Erwachsenen als auch bei Kindern- fanden sich die Kriterien Quantitativer Detailreichtum und Ungeordnete Darstellung, darüber hinaus war kein durchgehendes, systematisches Auftreten einzelner Kriterien zu beobachten. Eine Untersuchung von Niehaus (2003) weist daraufhin, dass die Unterscheidung von nicht-motivationalen Merkmalen und motivationsbezogenen Kriterien hinsichtlich der Validität der Kriterien bedeutsam sein könnte. In der Studie erwiesen sich motivationsbezogene Merkmale auch bei teilweise erlebnisbasierten Falschaussagen als diskriminationsrelevant. Es zeigte sich auch, dass die Trennschärfe der nicht-motivationalen Merkmale durch Wissensvermittlung über den zu berichtenden Sachverhalt negativ beeinträchtigt wurde. In der vorliegenden Untersuchung soll jedoch das Hauptaugenmerk auf die Gesamtheit der Realkennzeichen gelegt werden. Der Mangel an Forschungsergebnissen zur kriterienorientierten Inhaltsanalyse bei Vorschulkindern liefert keinen ausreichenden Hintergrund zur systematischen Validierung einzelner Realkennzeichen. Vielmehr steht zunächst eine methodisch ausreichende Validierung des gesamten Kriterienkatalogs im Mittelpunkt des Interesses.

Es soll vor dem Hintergrund der referierten Ergebnisse auch das Augenmerk auf die Quantität der Aussage, also deren Umfang, gelegt werden. Da das Auffinden von Realkennzeichen naturgemäß von der Anzahl gesprochener Worte abhängig ist, stellt sich die Frage, ob hier ein Zusammenhang besteht, der Auswirkungen auf die Einschätzung der Validität der Inhaltsanalyse haben könnte. So wäre denkbar, dass umfangreiche Schilderungen schon unter quantitativen Aspekten als erlebnisbegründet eingeordnet werden und die in den Realkennzeichen abgebildeten inhaltlichen Qualitäten weniger bedeutsam sind.

Ferner ist im Zusammenhang mit der Validität der Realkennzeichenanalyse nicht nur das grundsätzliche Auftreten von Realkennzeichen zu betrachten sondern auch die Kernfrage der aussagepsychologischen Praxis, ob mittels der gefundenen Kriterien erlebnisbegründete Aussagen von Darstellungen unterschieden werden können, die keine unmittelbare Erlebnisbasis haben. Auch diesem Aspekt soll in der vorliegenden Arbeit nachgegangen werden, indem der Erlebnishintergrund der erhobenen Aussagen beurteilt wurde.

Die im Kapitel 3 referierten Ergebnisse aus der Suggestionsforschung zeigen, dass eine Vielzahl von Faktoren Schilderungen über Geschehnisse beeinflussen, also verfälschend wirksam sein können. Auf der kognitiven Ebene sind z.B. Komplexität des Geschehens, Vorwissen, Einstellungen (Ceci & Bruck, 1993), Stärke der Erinnerungsspur (Volbert, 1997), eigene Beteiligung, sowie das Verstehen und Behalten von Informationen Variablen, die die Suggestibilität von Kindern und/oder auch Erwachsenen beeinflussen können. Dabei - und das zeigt die Notwendigkeit einer experimentellen Studien mit Vorschulkindern - ist wiederum ein deutlicher Alterstrend vorhanden: Kinder sind suggestibler als Erwachsene, jüngere Kinder sind suggestibler als ältere.

Zwar ist aus den Ergebnissen der angeführten Studien nicht abzuleiten, dass Vorschulkinder generell nicht geeignet sind, zurückliegende Erlebnisse realitätsgerecht zu beschreiben, die Ergebnisse sind aber dahingehend konsistent, dass verschiedene Faktoren die Zuverlässigkeit von Aussagen über erlebte oder aus anderer Quelle aufgenommene Erlebnisse beeinträchtigen können. Besonders eindrucksvoll sind dabei die unter 3.1. beschriebenen Ergebnisse aus mehreren experimentellen Studien (z.B. Ceci, Loftus, Leichtman & Bruck, 1994), die zeigen, dass es möglich ist, Kindern Erinnerungen über Ereignisse zu implantieren, die sie tatsächlich nicht erlebt haben. Auch hier zeigen sich wiederum Alterseffekte: In einer Untersuchung von McBrien und Dagenbach (1998) zur Übernahme fiktiver Ereignissen als selbst erlebte Erlebnisse zeigten sich diese erfolgreiche Implantation besonders innerhalb der Gruppe jüngerer Kinder. Während fünf- bis sechsjährige und sieben- bis achtjährige Kinder mehrere vorgegebene fiktive Ereignisse nicht als eigenes Erleben übernahmen, bestätigten 64% der drei- bis vierjährigen Kinder, das fiktive Ereignis selbst erlebt zu haben. Als besonders problematisch wird bei derartigen Prozessen gesehen, dass Kinder, die suggestiven Einflüssen ausgesetzt wurden, teilweise subjektiv von der Richtigkeit ihrer "falschen" Angaben überzeugt waren, somit die Motivation sich als glaubwürdiger Zeuge darzustellen entfällt und damit gleichzeitig den motivbezogenen Realkennzeichen die theoretische Grundlage entzogen wird (Köhnken, 1997). Der Schluss, dass bei derartigen Bedingungen die Grenzen der kriterienorientierten Inhaltsanalyse erreicht sind, liegt nahe und sollte daher für die hier relevante Altersgruppe auf experimenteller Basis gezogen werden.

In vielen Studien wurden den Kindern mehrere fiktive Erlebnisse vorgegeben, die bei Bestätigung durch weitere, meist suggestive Fragen von den Kindern selbst ausgestaltet wurden und erst im Verlauf der Befragungen in mehr oder weniger umfangreichen Aussagen mündeten. In der vorliegenden Untersuchung soll im Gegensatz dazu überprüft werden, ob es gelingen kann, Kindern einen umfangreichen Sachverhalt, der keine Erlebnisbasis hat, so nahe zu bringen, dass sie in der Lage sind eine ausführliche Schilderung zu geben, und Idealerweise auch von dem eigenen Erleben überzeugt sind. Untersuchungen zur Quellenidentifikation zeigen, dass jüngere Kinder größere Schwierigkeiten haben als ältere oder Erwachsene zu unterscheiden, ob ein Ereignis tatsächlich erlebt wurde oder nur darüber gesprochen wurde bzw. sie davon gehört haben (s. Kap. 3.1). Für die Übernahme von Informationen spielt auch die subjektiv wahrgenommene Autorität oder Kompetenz des Suggestors eine Rolle. So ergaben sich größere Suggestionseffekte wenn Erwachsene als Suggestoren fungierten als wenn andere Kinder die Befragung vornahmen. Anzunehmen ist, dass vertraute Erwachsene diesen Effekt vergrößern, da sie für das Kind eine "geprüfte" Autorität darstellen. In der vorliegenden Untersuchung werden daher vertraute Autoritätspersonen zur Übermittlung des suggerierten Sachverhaltes eingesetzt. Dies soll die Wahrscheinlichkeit der Aufnahme der umfangreichen verbalen Darbietung fördern mit dem Ziel, daraus eine den Grundvoraussetzungen der Aussagenanalyse genügende ausreichend detaillierte Darstellung zu erlangen. Gleichzeitig sollte damit auch das Zustandekommen einer fehlerhaften Quellenzuordnung (selbst erlebt vs. vorgelesen bekommen) begünstigt werden. Darüber hinaus sollte auch die Einbettung des experimentellen Settings (Kindergarten, Familie) in alltägliche Abläufe der Kinder dazu beitragen, dass die vorgegebenen Inhalte, bzw. Suggestionen eher aufgenommen werden als unter alltagsfernen Bedingungen. Das gewählte experimentelle Paradigma zeigt zudem eine große Praxisnähe, da im Falle der Begutachtung von Kinderaussagen bei sexuellem Missbrauch in der Regel mögliche suggestive Einflüsse vor allem im alltäglichen Umfeld des Kindes durch bekannte Personen zu erwarten sind. Dieses könnte zu einer höheren ökologischen Validität beitragen und die Generalisierbarkeit der Ergebnisse begünstigen.

Aus der empirisch gestützten Überzeugung, dass Kinder suggerierte Inhalte aufnehmen und als eigenes Erleben bewusst oder unbewusst ausgeben können, ergibt sich die Frage nach der Qualität von Schilderungen mit diesem Entstehungshintergrund. In der aussagepsychologischen Forschung sind die Realkennzeichen Prüfgröße für die Aussagequalität. Weiteres zentrales Anliegen dieser Untersuchung ist somit die Überprüfung der Eignung der inhaltlichen Qualitätsanalyse bei Vorschulkindern unter suggestiven Bedingungen. Bisher vorliegende Untersuchungen bieten kein einheitliches Bild im Hinblick auf qualitative Unterschiede zwischen Aussagen über erlebnisbasierte oder nicht erlebnisbasierte Geschehnisse. Es ist anzunehmen, dass sich die heterogenen Ergebnisse weitgehend mit dem unterschiedlichen Vorgehen in den Untersuchungen erklären lassen.

In den Untersuchungen, in denen sich Schwierigkeiten zeigten, mittels der Realkennzeichen zwischen erlebnisbegründeten und suggerierten Schilderungen zu unterscheiden, wurden suggerierte Schilderungen wegen der guten inhaltlichen Qualität häufiger als erlebnisbegründet eingeschätzt, als dass erlebnisbegründete Darstellungen als suggeriert, bzw. nicht erlebnisbegründet klassifiziert werden. In der praktischen Begutachtung hätte dieser Fehler weit reichende Konsequenzen für den Ausgang eines Verfahrens, da mit der Beurteilung einer Anschuldigung als erlebnisbegründet die Weichen für eine falsche Verurteilung gestellt werden. Anliegen der vorliegenden Untersuchung ist es daher, festzustellen, ob die genannte Problematik bei jüngeren Kindern möglicherweise in noch stärkerer Weise zu Tage tritt. Untersucht werden sollen dabei nicht nur die Qualitätsunterschiede in den Aussagen, sondern es soll auch geprüft werden, ob die Kernaufgabe der aussagepsychologischen Begutachtung, - die Klassifizierung einer Aussage als erlebnisbasiert oder nicht erlebnisbasiert- auf der Grundlage der Qualitätseinschätzung erfolgreich gelöst werden kann. Die Schwierigkeiten bei der Einschätzung von Aussagen bilden sich dabei möglicherweise nicht nur in den Trefferquoten ab, sondern auch in der Sicherheit mit der diese Einschätzungen vorgenommen werden. Aus der praktischen Begutachtung ist bekannt, dass die Klassifizierung einer Aussage als glaubhaft durch die positive Feststellung von Qualitäten subjektiv eine stabilere Basis hat als die gegenteilige Einordnung. Verknüpft mit den Befunden, nach denen suggerierte Aussagen eher als erlebnisbegründet eingestuft werden, ergibt sich die Überlegung, ob sich dies möglicherweise auch in der Sicherheit, mit der ein Urteil abgegeben wird, widerspiegelt. So könnte

ein Urteiler zwar zu der Einschätzung kommen, dass eine suggerierte Aussage erlebnisbegründet ist, dabei aber im Vergleich zum selben Urteil bei einer tatsächlich erlebnisbegründeten Aussage weniger sicher bezüglich dieser Einschätzung sein. Möglicherweise könnte dies mit der Ausprägung der Realkennzeichen zusammenhängen. Dieser Aspekt soll hier ebenfalls betrachtet werden.

In der vorliegenden Untersuchung soll zunächst geprüft werden, ob die kriterienorientierte Aussagenanalyse bei Vorschulkindern eine geeignete Methode zur Feststellung des Realitätsgehaltes ihrer Aussagen ist. Es wird untersucht, ob Schilderungen über ein reales Ereignis mit Hilfe der kriterienorientierten Aussagenanalyse von Schilderungen unterschieden werden können, die keinen faktischen Erlebnishintergrund haben. Weiterhin wird erforscht, ob Schilderungen, die ohne Erlebnishintergrund auf Suggestionen basieren, von Schilderungen ohne Erlebnishintergrund unterschieden werden können. Dieser Vergleich dient der Überprüfung der Frage, ob Suggestionen so erfolgreich implantiert werden können, dass sich Unterschiede zum Nicht-Erleben ergeben und so einen Vergleich mit erlebnisbegründeten Aussagen sinnvoll machen. Geprüft wird zudem, ob Schilderungen, die auf einem realen Erlebnis basieren von solchen, die auf intensiven verbalen Informationen ohne Erlebnishintergrund beruhen, unterschieden werden können.

Zusammenfassend sind die Annahmen dieser Arbeit, dass die Anwendung der kriterienorientierten Inhaltanalyse bei Vorschulkindern valide Ergebnisse erbringt, dass Vorschulkinder suggerierte Informationen derart aufnehmen, dass Realkennzeichen in den entsprechenden Schilderungen zu finden sind und dass erlebnisbegründete Aussagen und suggestionsbedingte Aussagen mit Hilfe der Realkennzeichen nicht zu trennen sind.

Darüber hinaus soll der Einfluss der intellektuellen und verbalen Kompetenz auf die Aussagequalität geprüft werden, so wie die Bedeutung des Aussageumfangs für die Beurteilung einer Aussage hinsichtlich ihres Erlebnishintergrundes.

4.2. Hypothesen der Untersuchung

Für die vorliegende Arbeit lassen sich folgende Hypothesen formulieren:

Die erste Annahme dieser Untersuchung ist, dass die Anwendung der kriterienorientierten Inhaltsanalyse bei Vorschulkindern valide Ergebnisse bringt. Daraus ergibt sich:

Hypothese1: In erlebnisbegründeten Aussagen sind Realkennzeichen in stärkerer Ausprägung vorhanden als in nicht erlebnisbegründeten Aussagen.

Des Weiteren wird angenommen, dass Vorschulkinder suggerierte Informationen derart aufnehmen, dass Realkennzeichen in den entsprechenden Schilderungen zu finden sind. Daraus ergibt sich:

Hypothese2: In nicht erlebnisbegründeten Aussagen, die auf wiederholten verbalen Informationen (Suggestion) basieren, sind Realkennzeichen in stärkerer Ausprägung vorhanden als in nicht erlebnisbegründeten Aussagen.

Außerdem wird die Annahme untersucht, dass erlebnisbegründete Aussagen und suggestionsbedingte Aussagen mit Hilfe der Realkennzeichen nicht zu trennen sind. Daraus ergibt sich:

Hypothese3: In erlebnisbegründeten Aussagen und in Aussagen, die auf wiederholten verbalen Informationen (Suggestion) basieren, sind Realkennzeichen in ähnlich starker Ausprägung vorhanden.

5. Methodik

5.1 Operationalisierung der Variablen

5.1.1 Unabhängige Variable

Die unabhängige Variable *Basis der Aussage* basiert auf drei unterschiedlichen Darbietungen eines Ereignisses, mit dem Vorschulkinder in einem Feldexperiment konfrontiert wurden. Die Darbietungsgruppen werden mit den Buchstaben A, B und C bezeichnet.

In der Darbietungssituation A haben die Kinder die Aufführung eines Kaspertheaters erlebt. Die Vorschulkinder der Gruppe B haben keine Aufführung gesehen. Stattdessen wurde ihnen der Inhalt der aufgeführten Geschichte mehrfach vorgelesen. Durch dieses Vorgehen wurde die Stufe *Suggestive Beeinflussung* operationalisiert. Es erfolgte keine suggestive Befragung, als suggestiver Reiz wurde lediglich die Eingangsfrage bei den Interviews ("Du hast doch den Verkehrskasper gesehen.... vgl.6.5.2) eingesetzt und die Abschlussfrage, ob das Kind den Verkehrskasper tatsächlich gesehen habe.

Der Begriff Suggestion wurde in dieser Untersuchung in der weit gefassten Definition nach Ceci und Bruck (1993) verwendet, da keine Verknüpfung mit schon vorhandenen Gedächtnisspuren intendiert war. Durch das mehrfache Vorlesen sollte Einfluss auf Wahrnehmung, Speicherung, Abruf und Reproduktion genommen werden. Verstärkt wurde dieser Prozess dadurch, dass das Vorlesen, d.h. die Vermittlung der relevanten Informationen, durch vertraute Personen (Eltern und Erzieherinnen) vorgenommen wurde (vgl. Poole & Lindsay, 1995). Instruktionen dazu finden sich in Anhang A.

Den Kindern der Kontrollgruppe (Gruppe C), die das Kaspertheater ebenfalls nicht gesehen hatten, wurde der Inhalt der Aufführung einmal vorgelesen. Da nicht davon ausgegangen werden konnte, dass alle Kinder außerhalb des experimentellen Settings eine Kasperaufführung gesehen hatten, sollten damit Basisinformationen gegeben werden, die eine Befragung überhaupt ermöglichen.

5.1.2 Abhängige Variable

Die abhängige Variable stellen die Realkennzeichen dar, die in den Aussagen der Vorschulkinder der drei Gruppen festgestellt wurden.

Zu Grunde gelegt wurde der in Tabelle 1 dargestellte Kriterienkatalog von Steller und Köhnken (1989). Die Kriterien 17. Selbstbelastung des Aussagenden, 18. Entlastung des Angeschuldigten und 19. deliktspezifische Aussageelemente wurden nicht berücksichtigt, da sie sich auf forensische Sachverhalte beziehen. Das Vorhandensein der folgenden Kriterien wurde überprüft:

1. Logische Konsistenz und Widerspruchsfreiheit
2. Widerspruchsfrei ungeordnete Darstellung der Aussage
3. Quantitativer Detailreichtum
4. Kontextuelle Einbettung
5. Beschreibung von nonverbalen Interaktionen
6. Wiedergabe von Gesprächen
7. Komplikationen im Handlungsverlauf
8. Ausgefallene Einzelheiten
9. Nebensächliche Einzelheiten
10. Phänomengemäße Schilderung unverstandener Handlungselemente
11. Inhaltliche Verschachtelungen
12. Schilderung fremdpsychischer Vorgänge
13. Schilderung eigenpsychischer Vorgänge
14. Spontane Verbesserungen der eigenen Aussage
15. Zugeben von Erinnerungslücken
16. Unsicherheiten bei der Darstellung

Auf einer sechsstufigen Skala, die von nicht vorhanden (=0) bis sehr stark vorhanden (=5) reichte, wurde die Ausprägung der gefundenen Realkennzeichen eingeschätzt. Die 16 Realkennzeichen wurden für jede Versuchsperson durch Addieren der Skalenwerte

für die einzelnen Realkennzeichen zur abhängigen Variable *Summenscore* zusammengefasst.

5.1.3 Kontrollvariablen

Die Ausprägungsgrade der Realkennzeichen können durch verschiedene Faktoren beeinflusst werden: So können Realkennzeichen ausgeprägter erscheinen, wenn ein Kind quantitativ mehr berichtet, es sich besser ausdrücken kann, bzw. Sachverhalte besser umschreiben kann und/oder flüssiger spricht. Das Sprachvermögen von Kindern und ihre Merkfähigkeit nehmen in der Regel mit zunehmendem Alter zu. Ebenso hat die Intelligenz eines Kindes einen Einfluss auf das Ausdrucksvermögen und das Gedächtnis. Zur Kontrolle dieser Einflüsse wurde die Intelligenz mit dem Mann-Zeichen-Quotient (Ziler, 1975) und die Verbalität über die Untertests Wortverständnis und Gegenstände beschrieben aus dem Psycholinguistischen Entwicklungstest (PET, Angermeier, 1974) erhoben.

5.1.4 Zusätzliche Variablen

Ebenso wurde die *Wortanzahl* pro Aussage jedes Kindes erfasst. Die Wortanzahl kann durch die kognitiven, verbalen Fähigkeiten beeinflusst werden, außerdem ist auch eine Beeinflussung der Wortanzahl durch das Treatment anzunehmen. So könnte eine Aussage, die auf einem real erlebten Geschehen basiert, möglicherweise umfangreicher sein, als eine Aussage, die nicht auf real erlebten Geschehnissen basiert. Dies kann zum Beispiel mit einer größeren Anzahl an Sinneseindrücken bei einem real erlebten Ereignis zusammenhängen. Ebenso wurden als zusätzliche Variablen das *Urteil* über den Wahrheitsgehalt der Erzählungen erhoben, sowie die *Sicherheit über das gefällte Urteil*. Explorativ wird daher in einer zusätzlichen Auswertung untersucht, welchen Einfluss diese Variablen haben.

5.2 Experimentelles Design

Die Untersuchung wird mit einem einfaktoriellen Versuchsdesign durchgeführt, in dem die unabhängige Variable in drei Stufen variiert wird.

Tabelle 2:

Einfaktorielles Versuchsdesign

UV:	AV:
Basis der Aussage	Ausprägung der Realkennzeichen
A	
B	
C	

- A: Basis der Aussage: Erlebnis ohne suggestive Beeinflussung
B: Basis der Aussage: Kein Erlebnis, mehrmalige Vorgabe
verbaler Informationen (Suggestive Beeinflussung)
C: Basis der Aussage: Kein Erlebnis, einmalige Vorgabe
verbaler Informationen

Das Design enthält die unabhängige Variable *Basis der Aussage* sowie die abhängige Variable *Ausprägung der Realkennzeichen*. Die unabhängige Variable wurde in drei Stufen variiert. Diese drei Stufen unterscheiden sich bezüglich der Art der Darbietung und Wahrnehmung eines Ereignisses. Die abhängige Variable ist die Ausprägung der Realkennzeichen in Aussagen über das verschieden dargebotene Erlebnis.

6. Durchführung der Untersuchung

6.1. Kaspertheater als Stimulusmaterial

Die Polizeidirektionen Neumünster und Flensburg unterhalten Lehreinrichtungen für elementare Verhaltensregeln im Verkehr, die insbesondere Verkehrsanfänger im Vor- und Grundschulalter erreichen. Über das Medium Kaspertheater werden den Kindern alters-

entsprechend Verkehrsregeln vermittelt. Den Verkehrskasperbühnen (Steinbrink, Schrader & Caspers, 1993) liegt das Konzept zugrunde, den Kindern durch eine Kaspergeschichte sicheres Verhalten im Verkehr beizubringen. Die Verknüpfung der Elemente Theater und pädagogische Inhalte soll bewirken, dass die Aufnahme und Speicherung der Verhaltensregeln im Verkehr optimiert werden. Die starke emotionale Beteiligung von Kindern bei Kasperaufführungen soll die Aufnahme und Verankerung der zu lernenden Verkehrsregeln in besonderem Maße fördern.

Die Polizei-Puppenspielbühne Flensburg spielte im Gemeindekindergarten Gettorf das Stück „Luzies Weg durch das Entental“. Ort des Geschehens ist Bauer Ottos Garten, der an einem großen Teich im Entental liegt. Hier ist das Zuhause vieler Tiere. Ente Luzie ist Mutter von sechs Entenküken. Der freche Hase Eberhard macht sich über den Kohl in Bauer Ottos Garten her und stößt dabei auf den Wurm Fridolin, der in einem schönen, grünen Kohlkopf wohnt. Er verschwindet schnell, als Kasper kommt, um im Teich zu angeln. Leider wurde der schöne Teich von rücksichtslosen Zeitgenossen mit Unrat verschmutzt, und Kasper zog statt der erhofften Riesenfische nur Müll aus dem Wasser. Anschließend hilft er der Ente Luzie bei der Suche nach ihren Entenküken, die verschwunden sind. Sie kommen während ihrer Suche an belebten Straßen vorbei. Dort entgeht der Junge Rudi, der unbekümmert auf der Straße Ball spielt, nur knapp einem Zusammenstoß mit dem Trecker des Bauern Otto. Dieser fährt, um Rudi auszuweichen, in die Fensterscheibe des Bäckers. Der Hase Eberhard trifft das Küken Zippel-Zappel, das sich von ihm zu verkehrswidrigem Verhalten auf der Straße anstiften lässt. Das erzählen die Kindergartenkinder Kasper, der daraufhin beschließt, den Hasen zu fangen. Kasper stellt dem Hasen mit einer Möhre eine Falle und bringt ihn zur Polizei. Schließlich finden Kasper und Luzie mit Hilfe der Kindergartenkinder die Küken beim fröhlichen Bad im Teich.

6.2 Versuchspersonen

6.2.1 Rekrutierung der Versuchspersonen

An der Untersuchung für die vorliegende Studie nahmen 78 Kinder im Alter zwischen drei und sieben Jahren teil. Die Kinder besuchten den Gemeindekindergarten in Gettorf, einem Vorort der Schleswig-Holsteinischen Landeshauptstadt Kiel. Der Kindergarten stellt 114 Plätze, aufgeteilt in vier Vormittagsgruppen und zwei Nachmittagsgruppen.

In der Planungsphase der Untersuchung fand ein intensiver Informationsaustausch zwischen den beiden Versuchsleitern, den Mitarbeiterinnen des Kindergartens und den Eltern der Kinder statt. Das Erzieherteam wurde über das Untersuchungsvorhaben informiert und war geschlossen bereit, die Untersuchung aktiv zu unterstützen und die üblichen Gruppenaktivitäten während der ca. einmonatigen Zeit der Durchführung, zugunsten der Untersuchungsaktivitäten (Erlebnis, Suggestion, Befragung) einzuschränken. Das Untersuchungsvorhaben wurde als Untersuchung von Gedächtnisleistungen mit Relevanz auch für den forensischen Kontext definiert.

Auf Elternabenden wurde den Eltern der Ablauf der Untersuchung von den beiden Versuchsleitern vorgestellt. Den Eltern wurde ein Schreiben mit einer kurzen Beschreibung des Untersuchungsablaufes und der Bitte, ihr Kind an der Untersuchung teilnehmen zu lassen, übergeben. Erteilten die Eltern ihr Einverständnis, konnten sie dann entscheiden, ob sie ihr Kind nur grundsätzlich teilnehmen lassen wollten oder ob sie zusätzlich bereit wären, aktiv an der Untersuchung teilzunehmen, indem sie ihrem Kind eine Geschichte vorlesen würden.

6.2.2 Zuordnung der Versuchspersonen zu den Versuchsgruppen

In die vorliegende Arbeit wurden 78 Versuchspersonen einbezogen, die auf drei Versuchsgruppen aufgeteilt wurden. Insgesamt nahmen an der Untersuchung 109 Kinder teil. Die hier nicht einbezogenen Versuchspersonen gehören zu einer anderen experimentellen Arbeit (Schelle, 1996).

Die Aufteilung auf die Versuchsgruppen erfolgte in zwei Teilschritten. Zuerst wurden jene Kinder, deren Eltern nicht aktiv an den suggestiven Maßnahmen (mehrfaches Vorlesen) teilnehmen wollten, zufällig auf die Gruppen verteilt, in denen die Beteiligung der Eltern nicht vorgesehen war (Gruppe A und Gruppe C). Anschließend erfolgte die zufällige Zuordnung der anderen Versuchspersonen auf alle Gruppen bis zum Erreichen der vorgesehenen Gruppengrößen. Die Festlegung der Gruppengrößen zeigt Tabelle 3.

Tabelle 3:

Aufteilung der Versuchspersonen auf die drei Versuchsgruppen

Versuchsgruppe	Anzahl Vpn	Quelle der Aussage
Gruppe A	25	Kaspertheater erleben
Gruppe B	33	Kein Kaspertheater erleben, Inhalt des Stückes mehrmals vorgelesen bekommen
Gruppe C	20	Kein Kaspertheater erleben, Inhalt einmal vorgelesen bekommen

Die Gruppe B wurde stärker besetzt, um einer höheren Wahrscheinlichkeit von Ausfällen begegnen zu können. Diese Ausfälle hätten aufgrund der häufigeren Untersuchungstermine und der damit höheren Auftretenswahrscheinlichkeit von Abwesenheiten entstehen können.

Die Kinder der Gruppe A nahmen an der Kasperaufführung teil. Die Gruppen B und C hatten kein real erlebtes Ereignis für ihre Erinnerungen, sondern griffen im Interview zur

Kasperaufführung auf Erinnerungen an die ein- oder mehrfach vorgelesene Geschichte zurück.

6.3 Interviewer

Die Interviews wurden von drei Studentinnen im Hauptstudium des Instituts für Psychologie der Universität Kiel durchgeführt. Diese hatten vorher in einem von dieser Arbeit unabhängigen Kurs am Institut für Psychologie die Durchführung von Interviews mit Kindern erlernt und geübt. Um einen guten Zugang zu den Kindern zu erhalten und mögliche Ängste vor unbekannten Personen nicht erst entstehen zu lassen, gingen die Interviewerinnen einige Tage vor dem Interview während der regulären Spielzeiten in die verschiedenen Gruppen und beteiligten sich an den Spielen. Zur Optimierung der Aussagequalität wurden die Interviewerinnen vor der Durchführung der Interviews ausdrücklich angewiesen, die folgenden Regeln zu berücksichtigen:

- Herstellung einer freundlichen Arbeitsatmosphäre,
- den Kindern Zeit lassen,
- die Kinder nachdenken und ausreden lassen,
- die Kinder nicht drängen,
- Berücksichtigung der kindlichen Ausdrucksmöglichkeiten,
- offene Fragen stellen,
- Suggestivfragen vermeiden,
- das Gesagte wiederholen,
- verbale und nonverbale Verstärker benutzen,
- die Kinder für ihre Bemühung loben.
-

Die Interviewerinnen waren weder über den gesamten Versuchsablauf noch über den Zweck der Untersuchung informiert. Für die Interviews waren sie lediglich darauf hingewiesen worden, dass die Kinder dieser Gruppen eine Verkehrskasperaufführung gesehen hätten und ihnen Geschichten vorgelesen worden waren. Die Interviewerinnen hatten

dann in den Interviews die Aufgabe, die Kinder aufzufordern, ihnen vom Verkehrskasper zu berichten.

6. 4 Rater

Die Auswertungen der Interviewprotokolle nach dem Vorhandensein inhaltlicher Realkennzeichen und die Einschätzung Aufführung gesehen oder Aufführung nicht gesehen wurden durch geschulte, für die Versuchsbedingung blinde Rater durchgeführt. Da die Beurteilung der Aussage ausschließlich aufgrund ihres Inhalts vorgenommen wurde und so eine Konfundierung von Aussageinhalt und nonverbalen Merkmalen ausgeschlossen werden sollte, war eine Trennung der Funktionen Interviewer und Rater notwendig. Als Raterinnen standen zwei Psychologiestudentinnen aus dem Hauptstudium des Psychologischen Instituts der Christian-Albrechts-Universität zu Kiel zur Verfügung. Ihre Rater-Qualifikation hatten sie durch ein sechswöchiges institutsinternes Praktikum erworben. Während des Praktikums waren sie für andere Untersuchungen in der Anwendung der kriterienbezogenen Aussagenanalyse mit den von Steller und Köhnken (1989) aufgeführten neunzehn Realkennzeichen trainiert worden. Außerdem wurden ihnen Kenntnisse über andere Kriteriologien sowie theoretische und praktische Hintergründe vermittelt (vgl. Höfer 1995 b, 1996). Diese Ausbildung beinhaltete jedoch weder die Auswertung von Interviews von Kindern im Vorschulalter noch die Bearbeitung von Transkripten in Dialogform. Daher wurde vor Beginn der aussageanalytischen Tätigkeit für die vorliegende Untersuchung eine Schulung der Raterinnen durch die Versuchsleiter durchgeführt. Im Rahmen der Studie beurteilte jede Raterin insgesamt 43 Interviews. Zwölf Interviews wurden von beiden Raterinnen zur Bestimmung der Interrater- Reliabilität bearbeitet. Die Raterinnen wurden für ihre Tätigkeit bezahlt. Sie hatten keine Kenntnis über das Untersuchungsziel.

6.5 Ablauf der Untersuchung

6.5.1 Untersuchungsphase 1 : Erleben des Kaspertheaters

Es fanden zwei Vorführungen des Verkehrskaspers jeweils am Vormittag im Kindergarten statt, die eine Woche auseinander lagen. Diese Staffelung erfolgte, um die Zeitspanne von der Vorführung bis zum jeweiligen Interview für alle Versuchspersonen möglichst gleich zu halten. Sie betrug acht bis zehn Tage. Die Kinder der Versuchsgruppe A erlebten die Vorführung während eines normalen Kindergartenvormittags, sie wurde dafür aus den Spielgruppen geholt und in den Vorführraum gebracht. Teilnehmende Nachmittagskinder wurden von ihren Müttern zur Aufführung gebracht. Der Inhalt des Stückes entspricht der Vorlesegeschichte für die Gruppen B und C (Anhang B).

Beide Aufführungen wurden auf Video aufgezeichnet, um mögliche Abweichungen in den Aufführungen bei der Interviewauswertung berücksichtigen zu können. Da sich beim Vergleich der Aufzeichnungen keine Unterschiede zwischen den Aufführungen ergaben, konnte von der Inhaltsgleichheit der Aufführung für alle Kinder ausgegangen werden. Die Dauer der Vorführung betrug ca. 35 Minuten. Während der gesamten Aufführung wurden die Kinder von den Versuchsleitern beobachtet, um eventuelle Aufmerksamkeitsausfälle, die das spätere Erinnern des Erlebnisses verhindern könnten, vermerken zu können. Alle Kinder schienen konzentriert und beteiligt zu sein. Die Kinder waren aktiv in das Geschehen eingebunden. Sie führten Dialoge mit den dargestellten Figuren, in einer Szene wurden sie als Teil der Aufführung nass gespritzt. Sie zeigten emotionale Beteiligung durch lautes Rufen und gespanntes, stilles Verhalten in entsprechenden Situationen sowie durch spontanen Szenenapplaus. Das gewählte experimentelle Setting zeigt zudem eine große Praxisnähe, da im Falle der Begutachtung von Kinderaussagen bei sexuellem Missbrauch in der Regel mögliche suggestive Einflüsse vor allem im alltäglichen Umfeld des Kindes durch bekannte Personen zu erwarten sind

6.5.2 Untersuchungsphase 2: Suggestive Einflüsse

Den Versuchsgruppen B und C wurden nach der Aufführung in unterschiedlichem Ausmaß verbale Informationen geboten, die die suggestive Beeinflussung darstellten.

Den Kindern der Versuchsgruppe B wurde in einem Zeitraum von drei Wochen insgesamt zehnmal im Kindergarten und Zuhause der Inhalt der Kasperaufführung als Vorlesegeschichte vorgegeben. Die Vorlesegeschichte war so konstruiert, dass ein direkter Bezug zum Kindergarten-Alltag enthalten war. Ziel war, den Kindern eine Identifikation mit der Situation im Kindergarten zu ermöglichen. So wurde zu Beginn der Geschichte beschrieben, dass ein Kaspertheater im Kindergarten aufgebaut wurde. Diese Beobachtung konnten auch alle Kinder machen, da sie bei der Ankunft in den Kindergarten das aufgebaute Kaspertheater sahen. Im Gegensatz zu den Kindern aus der Gruppe A nahmen die Kinder der Gruppe B und C aber nicht an der Aufführung teil. Durch diese Verknüpfung der vorgelesenen Geschichte mit dem tatsächlichem Erleben durch die Beobachtung des Aufbaus des Theaters in den vertrauten Räumen sollte gefördert werden, dass die Kinder nach der Phase des intensiven Vorlesens annehmen, das Kaspertheaterstück tatsächlich gesehen zu haben.

Der Gruppe C wurde die Geschichte nur einmal vorgelesen. Das Vorlesen der Geschichte in dieser Kontrollgruppe sollte ein Basiswissen für die Befragung erzeugen. Schließlich konnte nicht davon ausgegangen werden, dass jedes Kind schon einmal ein Kaspertheater gesehen hat und in der Befragung auf eine solche Erinnerung zurückgreifen könnte. Als mögliche Fehlerquelle für die Erinnerungsleistungen der Kinder wurde die Verarbeitung des Erlebens der Aufführung bzw. des Hörens der Geschichte in gemeinsamen Rollenspielen oder Gesprächen angesehen, bei denen sich die Kinder gegenseitig beeinflussen könnten. Eine Trennung der Kinder aus gleichen Kindergartengruppen während der Suggestionsphase war organisatorisch nicht möglich. Die Erzieherinnen der Gruppen wurden deshalb gebeten, das Auftreten entsprechender Reaktionen der Kinder zu beobachten und den Versuchsleitern zu berichten. Die Erzieher beobachteten kaum derartige

Reaktionen, wenn doch, beschränkten sich diese auf kurze und inhaltsarme Dialoge zwischen einzelnen Kindern, wie „Wie war der Kasper?“, „Gut!“.

Eine inhaltliche Vertiefung der Geschichte vom Verkehrskasper durch die Erzieher und die Einbindung in ihre pädagogische Arbeit fand erst nach Ablauf der Untersuchung statt. Auch die Eltern waren in den Instruktionen angehalten, ihre Kinder zwar so ausführlich über den Verkehrskasper erzählen zu lassen, wie diese wollten, aber nicht nachzufragen.

6.5.3 Untersuchungsphase 3: Erhebung der Aussage

Die Befragung der Kinder erfolgte durch zwei Interviewerinnen. Ihnen standen zwei Räume des Kindergartens zur Verfügung, in denen die Befragung ungestört erfolgen konnte. Für die Interviews wurden die Kinder einzeln in nicht festgelegter Reihenfolge aus ihren Gruppenräumen geholt und später wieder zurückgebracht. Interviewerin und Kind saßen dabei über Eck an einem Tisch.

Nach der Begrüßung wurde zunächst der Mann-Zeichen-Test von Ziler (1975) durchgeführt, um dem Kind in dieser ungewohnten Situation als Aufwärmung eine gut lösbarer Aufgabe zu geben. Gleichzeitig wurde damit ein Befund für die Einschätzung des intellektuellen Entwicklungsstandes erhoben. Anschließend erfolgte die Befragung zum Kaspertheater. Abschließend wurden zwei Untertests aus dem Psycholinguistischen Entwicklungstest (PET) von Angermaier (1974), WV (Wortverständnis) und GB (Gegenstände beschreiben), zur Überprüfung des Sprachverständnisses und der Ausdrucksfähigkeit vorgegeben. Am Ende der Befragung bekam jedes Kind als Belohnung einen Tierstempel auf den Handrücken und einen Aufkleber vom Verkehrskasper. Die Datenerhebung betrug abhängig von der Erzählfreudigkeit der Kinder insgesamt eine halbe bis dreiviertel Stunde für jedes Kind. Die Befragung zum Verkehrskasper dauerte zehn bis dreißig Minuten. In einer Verhaltensbeobachtung wurde besonders auf Aufmerksamkeitsschwankungen und Konzentrationsmängel geachtet. Diese wurden protokolliert, da derartige Einbrüche Einfluss auf die Qualität der Aussagen haben und damit als Fehlerquelle wirk-

sam sein könnten. Eines der Kinder war während des gesamten Interviews müde und lustlos, andere Kinder waren leicht abgelenkt durch Geräusche auf dem Flur oder durch die Erwartung der abschließenden Belohnung. Zwei der Kinder schwiegen in der Interviewsituation.

Angelehnt an das Vorgehen bei der praktischen Begutachtung gliederte sich die Befragung zur Kasperaufführung in einen Freien Bericht und in einen standardisierten Frage teil. Für den Freien Bericht hatten die Interviewerinnen die Instruktion erhalten, das Kind zunächst aufzufordern, alles zu erzählen, was es vom Kaspertheater noch erinnere. Es wurde dabei ausdrücklich auf die erlebte Aufführung hingewiesen. Dazu wurde folgende Formulierung benutzt: „Du hast doch den Verkehrskasper gesehen. Erzähl mir bitte alles, was du davon noch weißt!“ Die Kinder äußerten in keinem Interview Einwände gegen diese Vorgabe. Im weiteren Verlauf des Freien Berichts konnte die Interviewerin unspezifische Gesprächsverstärker benutzen z.B. nicken oder „Mmh!“ sagen, spiegeln (das vom Kind Gesagte wörtlich wiederholen) und zum Weiterreden auffordern („Erzähl mir noch mehr darüber“, „Was weißt du noch vom Verkehrskasper?“).

Erst wenn die Interviewerinnen den Eindruck hatten, dass das Kind von sich aus nichts mehr erzählen würde, wurde zum standardisierten Frageteil (Anhang D) übergegangen. Dieser bestand aus zwanzig Fragen, die sich auf das Theater, die Figuren und die Handlungen bezogen. Die Fragen wurden im Sinne von Erinnerungshilfen zunehmend konkreter und inhaltsbezogener. Die Kinder durften jederzeit zusätzliche Dinge erzählen, die ihnen zum Kaspertheater einfielen. Am Ende des Fragenkatalogs wurden die Kinder erneut offen gefragt, ob ihnen noch etwas vom Kaspertheater einfiele. Außerdem wurden sie gefragt, ob sie das Kaspertheater tatsächlich gesehen haben.

6.5.4 Auswertung der Daten

Alle Interviews wurden mit Tonband aufgezeichnet. Aus Sicherheitsgründen geschah dies mit zwei getrennten Bandgeräten, so dass Ausfälle durch technische oder bedienungsbedingte Fehler vollständig vermieden werden konnten.

In die Transkripte der Interviews wurden sämtliche verbale und paraverbale Äußerungen (z.B. Stöhnen, Seufzen, „mmh“, „ääh“) der Kinder und der Interviewerinnen einbezogen. Das Transkript begann mit der o.a. Eingangsfrage und endete mit dem Beginn der Durchführung des Psycholinguistischen Entwicklungstests. Lücken in den Interviews ergaben sich dort, wo trotz der Kontrollabhörung durch die zweite Tonbandaufnahme bestimmte Begriffe oder Satzteile nicht verstanden werden konnten. Derartige Ausfälle waren gering. Ebenfalls ließen sich auf diese Weise Hörfehler aus dem ersten Transkriptionsvorgang entdecken und korrigieren. Unverständlich gebliebene Aussageteile wurden durch Punkte gekennzeichnet: „...“. Wenn ein Ausdruck zwar wörtlich verstanden schien, aber dennoch keine vollständige Sicherheit über die Richtigkeit erlangt werden konnte, wurde ein Fragezeichen in Klammern hinten angestellt: „(?)“. Diese Unsicherheiten ergaben sich häufiger durch das Auftreten von individuellen kindlichen Sprachschöpfungen und Ausdrucksweisen. Diese wurden wörtlich niedergeschrieben, um Fehldeutung durch eine interpretierende Übersetzung in die Erwachsenensprache zu verhindern.

Beispiel:

Frage: Wie sah denn das Kaspertheater aus?

Antwort: „Hm, es war so tell (?), tell (?) und da noch so, so'n kleines Loch.“

Gesprächspausen wurden durch waagerechte Striche markiert, lange Pausen durch mehrere Striche, z.B.: „-“, „---“. Da häufig Bejahungen und Verneinungen nicht mit Worten, sondern durch zustimmendes oder ablehnendes Brummen signalisiert wurden, erschienen in den Transkripten als Bejahung: „Mmh“ und als Verneinung: „Mh-mh“. Nachdenkliches "Brummen" wurde durch „Hm“ aufgeschrieben.

In vielen Interviews fanden sich unvollständige Sätze ohne den geregelten Aufbau der geschriebenen Sprache. Auch diese Eigentümlichkeit wurde in den Transkripten berücksichtigt.

Beispiel:

Antwort: „Und - denn wollt er einen Wurm ha, haben und denn fragte er Fri, Fri, do, ähm, Kasper fragte Fridolin und denn hat er gesagt, ähm, hier: Das sind doch ähm, hier, äh, alte Schuhe!“

Von allen durchgeführten Interviews wurden zunächst Urfassungen der Transkripte erstellt. Urfassungen enthalten den vollständigen Aussagedialog mit Fragen und Antworten, den Namen des Kindes, seine Versuchspersonennummer und Gruppenzugehörigkeit. Diese Urfassungen wurden für die inhaltliche Analyse durch die Raterinnen in einem weiteren Arbeitsgang aufbereitet. Die Urfassungen wurden anonymisiert, um keine Informationen über Alter und Geschlecht zu geben, indem entsprechende Passagen aus dem Text entfernt wurden. Einige Kinder gaben während des Interviews unwissentlich Hinweise auf ihre Gruppenzugehörigkeit. Sie erwähnten zum Beispiel, dass ihnen zu Hause oder von einer Erzieherin eine Kaspergeschichte vorgelesen worden sei. Sämtliche Äußerungen, die bei der folgenden Analyse Rückschlüsse auf die Gruppenzugehörigkeit zugelassen hätten, wurden von den Versuchsleitern entfernt. Dies geschah in etwa einem Drittel der Fälle. Entsprechende Passagen wurden durch Punkte ersetzt: „...“.

Die Eingangsfrage, mit der die Interviewerinnen das Gespräch einleiteten, wurde für alle Aussagen vereinheitlicht. Sie lautete für alle Transkripte: „Ich möchte jetzt von dir etwas über den Verkehrskasper hören. Was weißt du denn davon?“

6.5.5 Einschätzung der Realkennzeichen

Für die inhaltsanalytische Auswertung der Schilderungen wurde ein Fragebogen mit sechzehn Kriterien entwickelt (Anhang C). Für jedes Kriterium wurde eine 6-stufige Skala mit den Endpunkten sehr stark vorhanden und nicht vorhanden vorgegeben. Zu den Kriterien wurden auf dem Auswertebogen erläuternde Hinweise gegeben, um eine korrekte und einheitliche Anwendung des Kriteriums zu erleichtern. Abbildung 1 zeigt ein Beispiel aus dem verwendeten Fragebogen.

A: Allgemeine Merkmale		Bitte ankreuzen:					
<u>1. Logische Konsistenz und Widerspruchsfreiheit</u>		nicht vorhanden 0	schwach vorhanden 1	eher weniger vorhanden 2	eher mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
1.a Ist die Aussage in sich stimmig und folgerichtig?		nicht vorhanden 0	schwach vorhanden 1	eher weniger vorhanden 2	eher mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
1. b Enthält die Aussage Widersprüche?		nicht vorhanden 0	schwach vorhanden 1	eher weniger vorhanden 2	eher mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
1.c Wie bedeutsam sind die Widersprüche für das Geschehen?		nicht vorhanden 0	schwach vorhanden 1	eher weniger vorhanden 2	eher mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5

Abbildung 1: Beispiel aus dem Fragebogen für die Rater

Der Fragebogen enthält weitere Items mit gleichen Skalen, die speziellere Aspekte der Kriterien zur Beurteilung vorgeben. Diese zusätzlichen Items wurden nicht in die Auswertung einbezogen. Sie wurden für eine andere Untersuchung bzw. Fragestellung erhoben.

Um Einflüsse durch die Reihenfolge der zu bearbeitenden Transkripte zu vermeiden, wurde den Raterinnen eine feste Bearbeitungssequenz vorgegeben. Diese wurde so ausgewählt, dass aus derselben Gruppe stammende Interviews nicht unmittelbar aufeinander folgten, sondern Aussagen aus allen drei Gruppen ständig wechselten.

Im Anschluss an die Beurteilung der Kriterien sollten die Raterinnen auf der Grundlage der Ausprägungen der Realkennzeichen entscheiden, ob die jeweilige Aussage von einem Kind gemacht wurde, das die Aufführung des Kaspertheater wirklich gesehen hatte oder nicht (Urteil). Diese Entscheidung sollten die Raterinnen hinsichtlich ihrer subjektiven Sicherheit, auf einer fünfstufigen Skala bewerten (Urteilssicherheit).

7. Darstellung der Ergebnisse

7.1 Kennwerte der Stichprobe

In die Auswertung wurden die Daten von 74 der anfangs 78 Versuchspersonen dieser Untersuchung aufgenommen. Im Verlauf der Studie waren vier Kinder durch Krankheit (2) und Verweigerung des Interviews (2) ausgefallen.

Das Alter der Kinder lag zwischen 3; 4 und 7; 3 Jahren, das Durchschnittsalter betrug 5; 4 Jahre. Es nahmen 43 Jungen und 31 Mädchen an der Untersuchung teil. Das Durchschnittsalter der Jungen betrug 65,4 Monate ($S = 11,53$), die Mädchen waren im Durchschnitt 63,1 ($S = 10,3$) Monate alt. Die Verteilung der Kinder nach Alter und Geschlecht in den drei Versuchsgruppen ist Tabelle 4 zu entnehmen.

Tabelle 4:

Altersverteilung in den drei Versuchsgruppen

	Anzahl	Mittelwert Alter in Monaten	Standardabweichung Alter in Monaten
Gruppe A	21	66.9	11.0
Gruppe B	33	63.2	11.3
Gruppe C	20	64.0	10.6

7.2 Interrater-Reliabilität

Um festzustellen, ob eine ausreichende Übereinstimmung der zwei Raterinnen bei der Beurteilung der Aussagen hinsichtlich der Realkennzeichen vorhanden war, wurde die Interrater- Reliabilität berechnet. Es wurden die ersten zwölf Interviews der festgelegten Bearbeitungssequenz, die Interviews aus den verschiedenen Gruppen beinhalteten, in die Berechnung aufgenommen. Diese Aussagen bearbeiteten die Raterinnen getrennt von einander, so dass für jedes dieser 12 Interviews zwei Beurteilungen vorlagen. Es wurden Korrelationen zwischen den Ausprägungen der Einschätzungen durch die beiden Rate-

rinnen (r_1) über alle Kriterien berechnet. Dieses Vorgehen wurde nach einer Schulung der Rater wiederholt (r_2). Es zeigte sich, dass die Reliabilitäten bei einigen Kriterien niedrig waren. Leere Zellen in den Spalten bedeuten, dass dieses Realkennzeichen in keiner der zwölf Aussagen gefunden wurde. Insgesamt zeigen die Berechnungen, dass nach der Nachschulung eine ausreichende Übereinstimmung der Raterinnen vorliegt. Die Spalte r_1 zeigt niedrige Übereinstimmungen bei den Kriterien 2. Ungeordnete Darstellung (.48), 4. Kontextuelle Einbettung (.47), 11. Inhaltliche Verschachtelungen (-.28) und 14. Spontane Verbesserungen (.38).

Tabelle 5:

Interrater- Reliabilitäten über 12 Aussagen im ersten Durchgang r_1 und nach erneuter Schulung r_2

Kriterium	r_1 (vor Schulung)	r_2 (nach Schulung)
1. Logische Konsistenz und Widerspruchsfreiheit	.64	.59
2. Widerspruchsfreie ungeordnete Darstellung der Aussage	.48	.87
3. Quantitativer Detailreichtum	.81	.81
4. Kontextuelle Einbettung	.47	.87
5. Beschreibung von nonverbalen Interaktionen	.72	.72
6. Wiedergabe von Gesprächen	.89	.89
7. Komplikationen im Handlungsverlauf	.80	.86
8. Ausgefallene Einzelheiten	.79	.79
9. Nebensächliche Einzelheiten	---	.63
10. Phänomengemäße Schilderung unverstanden. Handlungselemente	---	---
11. Inhaltliche Verschachtelungen	-.28	1.0
12. Schilderung fremdpsychischer Vorgänge	1.0	1.0
13. Schilderung eigenpsychischer Vorgänge	---	---
14. Spontane Verbesserungen der eigenen Aussage	.38	.86
15. Zugeben von Erinnerungslücken	.66	.66
16. Unsicherheiten bei der Darstellung	1.0	1.0

Die unterschiedlichen Beurteilungen des Auftretens dieser Kriterien ergaben sich überwiegend durch unklare Vorstellungen der Raterinnen über die Operationalisierung einzelner Kriterien und konnten im Rahmen der Nachschulung geklärt werden. Es wurde eine erneute Bearbeitung der jeweils zwölf Interviews durch die Raterinnen vorgenommen, nach der erneut die Korrelationen (r_2) berechnet wurden. Es ergaben sich deutliche Verbesserungen gegenüber r_1 .

7.3 Einfluss der Kontrollvariablen auf den Summenscore der Realkennzeichen

Um zu überprüfen, ob die Variable Alter Einflüsse auf die abhängige Variable Summenscore ausübt, wurde eine Korrelation nach Pearson zwischen diesen beiden Variablen berechnet. Diese wurde mit .42 auf dem 1%-Niveau signifikant. Je älter die Kinder waren, desto höher war der Summenscore. Um zu überprüfen, ob der Einfluss der Variable Alter durch eine unterschiedliche Verteilung auf die Versuchsgruppen zu Unterschieden in den Summenscores der Versuchgruppen geführt haben kann, wurde eine einfaktorielle Varianzanalyse mit der abhängigen Variable Alter und dem Faktor Versuchsgruppe berechnet. Das Ergebnis wurde mit Werten von $F= 0,725$ und $p= 0,488$ auf dem 5 %-Niveau nicht signifikant. Es bestehen demnach keine systematischen Altersunterschiede in den Versuchsgruppen, weshalb anzunehmen ist, dass die Variable Alter mögliche Unterschiede in den Summenscores der Realkennzeichen in den drei Versuchsgruppen nicht beeinflusst hat.

Mit dem Mannzeichentest (MZT) wurde der intellektuelle Entwicklungsstand der Kinder erhoben. Der Mittelwert der Gesamtgruppe ($N=74$) betrug 12,4 ($S= 4,9$). Die Rohwerte in diesem Verfahren wiesen zwar auf dem 1 %-Niveau mit $r= .425$ signifikante Korrelationen mit der Variable Summenscore auf, es zeigten sich jedoch in einer einfaktoriellen Varianzanalyse mit den MZT- Rohwerten als abhängiger Variable und dem Faktor Gruppe mit Werten von $F= 0,694$ und $p= 0,503$ auf dem 5 %-Niveau keine signifikanten Unterschiede zwischen den Versuchsgruppen. Es bestehen demnach keine systemati-

schen Unterschiede in den Versuchsgruppen, weshalb anzunehmen ist, dass die Variable Rohwerte im Mannzeichentest mögliche Unterschiede in den Summenscores der Realkennzeichen nicht beeinflusst hat.

Tabelle 6:

Rohwerte im „Mannzeichentest“ in den Versuchsgruppen

	Anzahl (N)	M	S
Gruppe A	21	13.4	5.6
Gruppe B	33	12.4	4.9
Gruppe C	20	11.6	4.0

Zur Überprüfung der verbalen Fähigkeiten wurde der Untertest Gegenstände beschrieben aus dem Psycholinguistischen Entwicklungstest (PET) durchgeführt. Der Mittelwert für die Gesamtgruppe (N=74) betrug 18,2 (S= 6,8). Die Streubreite umfasste Werte von 4 bis 36.

Tabelle 7:

Rohwerte im Untertest „Gegenstände beschreiben“ aus dem Psycholinguistischen Entwicklungstest (PET)

	Anzahl (N)	M	S
Gruppe A	21	17.3	5.6
Gruppe B	33	19.8	7.0
Gruppe C	20	16.4	4.9

Als weiterer Test zur Überprüfung der verbalen Fähigkeiten wurde das Wortverständnis erfasst. Der Mittelwert für die Gesamtgruppe beträgt $M= 15,5$ ($S= 6,6$). Die Streubreite umfasste Werte von 3 bis 34.

Um zu überprüfen, ob die Variable Verbale Fähigkeit einen Einfluss auf die abhängige Variable Summenscore hat, wurde zunächst eine Korrelation nach Pearson zwischen den Variablen Summenscore und den Ergebnissen aus dem Untertest Gegenstände beschrei-

ben, berechnet. Diese wurde mit $r = .507$ auf dem 1 %-Niveau signifikant. Höhere Werte im Untertest gingen mit höheren Summenscores einher. Um zu überprüfen, ob der Einfluss dieser Variable zu Gruppenunterschieden beigetragen hat, wurde eine einfaktorielle Varianzanalyse mit den Ergebnissen aus dem Untertest Gegenstände beschreiben als abhängiger Variable und dem Faktor Gruppe gerechnet. Das Ergebnis wurde mit Werten von $F = 1,904$ und $p = 0,159$ auf dem 5 %-Niveau nicht signifikant.

Weiterhin wurde eine Korrelation nach Pearson zwischen der Variable Summenscore und den Ergebnissen aus dem Untertest Wortverständnis, gerechnet. Diese Korrelation wurde mit $r = .588$ auf dem 1 %-Niveau signifikant. Eine einfaktorielle Varianzanalyse mit den Ergebnissen aus dem Untertest Wortverständnis als abhängige Variable und dem Faktor Gruppe wurde mit Werten von $F = 1,483$ und $p = 0,234$ auf dem 5 %-Niveau nicht signifikant.

Tabelle 8:

Rohwerte im Untertest „Wortverständnis“ aus dem Psycholinguistischen Entwicklungs-test (PET).

	Anzahl (N)	Mittelwert (M)	Standardabw. (S)
Gruppe A	21	16.8	8.3
Gruppe B	33	16.0	5.9
Gruppe C	20	13.5	5.5

Insgesamt ergibt sich aus den Berechnungen, dass die Variable Verbale Fähigkeit nicht systematisch zwischen den Gruppen variierte und somit anzunehmen ist, dass mögliche Unterschiede nicht signifikant zu Unterschieden im Summenscore der Realkennzeichen zwischen den Gruppen beigetragen haben.

7.4 Einfluss der Wortanzahl auf den Summenscore der Realkennzeichen

Es wurden die von den Kindern gesprochenen Worte im freien Bericht und in der anschließenden Befragung gezählt. In der Auszählung wurden diese beiden Abschnitte des Interviews nicht getrennt. Der Gesamt-Mittelwert der Worte über alle Versuchsgruppen betrug 698 (S= 411).

Tabelle 9:

Wortanzahl in den Versuchsgruppen

	Anzahl Vpn	M: Worte	S: Worte
Gruppe A	21	846	478
Gruppe B	33	712	361
Gruppe C	20	519	364

Um zu überprüfen, ob die Variable Wortanzahl einen Einfluss auf die abhängige Variable Summenscore hat, wurde eine Korrelation nach Pearson zwischen diesen beiden Variablen gerechnet. Diese wurde mit $r= .80$ auf dem 1 %-Niveau signifikant. Je mehr Worte verwendet wurden, desto höher war der Summenscore.

Um zu überprüfen, ob der Einfluss der Variable Worte zu Unterschieden zwischen Gruppen beigetragen haben könnte, wurde eine einfaktorielle Varianzanalyse mit der abhängigen Variable Wortanzahl und dem Faktor Gruppe gerechnet. Das Ergebnis wurde mit Werten von $F= 3,505$ und $p= 0,035$ auf dem 5 %- Niveau signifikant. Die Variable Worte variierte systematisch zwischen den Gruppen und somit ist es möglich ist, dass diese Unterschiede signifikant zu Unterschieden im Summenscore der Realkennzeichen zwischen den Gruppen beigetragen haben. Dieses Ergebnis wird in einer zusätzlichen Auswertung aufgegriffen (s. 7.8.3)

7. 5 Auftreten der Realkennzeichen in den Aussagen

Zur Feststellung der grundsätzlichen Eignung des Erlebnishintergrundes Kaspertheater für eine kriterienorientierte Inhaltsanalyse wurde das Vorhandensein der Kriterien in den Aussagen geprüft. Ab einer Bewertung von 1 auf der Skala zum Auftreten der Kennzeichen wurde ein Realkennzeichen als vorhanden signiert.

Tabelle 10:

Absolute und relative Häufigkeiten der Realkennzeichen in den Aussagen

Kriterium	abs. Häufigkeit	rel. Häufigkeit
1. Logische Konsistenz und Widerspruchsfreiheit	74	100 %
2. Widerspruchsfreie ungeordnete Darstellung der Aussage	57	77,0 %
3. Quantitativer Detailreichtum	72	97,3 %
4. Kontextuelle Einbettung	64	86,5 %
5. Beschreibung von nonverbalen Interaktionen	7	9,5 %
6. Wiedergabe von Gesprächen	64	86,5 %
7. Komplikationen im Handlungsverlauf	58	78,4 %
8. Ausgefallene Einzelheiten	70	94,6 %
9. Nebensächliche Einzelheiten	45	60,8 %
10. Phänomengemäße Schilderung	3	4,1 %
11. Inhaltliche Verschachtelungen	4	5,4 %
12. Schilderung fremdpsychischer Vorgänge	38	51,4 %
13. Schilderung eigenpsychischer Vorgänge	5	6,8 %
14. Spontane Verbesserungen der eigenen Aussage	56	75,7 %
15. Zugeben von Erinnerungslücken	69	93,2 %
16. Unsicherheiten bei der Darstellung	11	14,9 %

Aus Tabelle 10 ist zu ersehen, dass sechzehn Kriterien mit unterschiedlichen Auftretenshäufigkeiten in den Aussagen vorhanden waren. In allen Aussagen wurde das Kriterium der Logischen Konsistenz signiert. Quantitativer Detailreichtum und Ausgefallene Ein-

zelheiten wurden jeweils in mindestens 70 der Aussagen gefunden. Die Kriterien Phänomen gemäß Schilderung, Inhaltliche Verschachtelungen und Schilderung eigenpsychischer Vorgänge traten jeweils in höchstens 5 Aussagen auf.

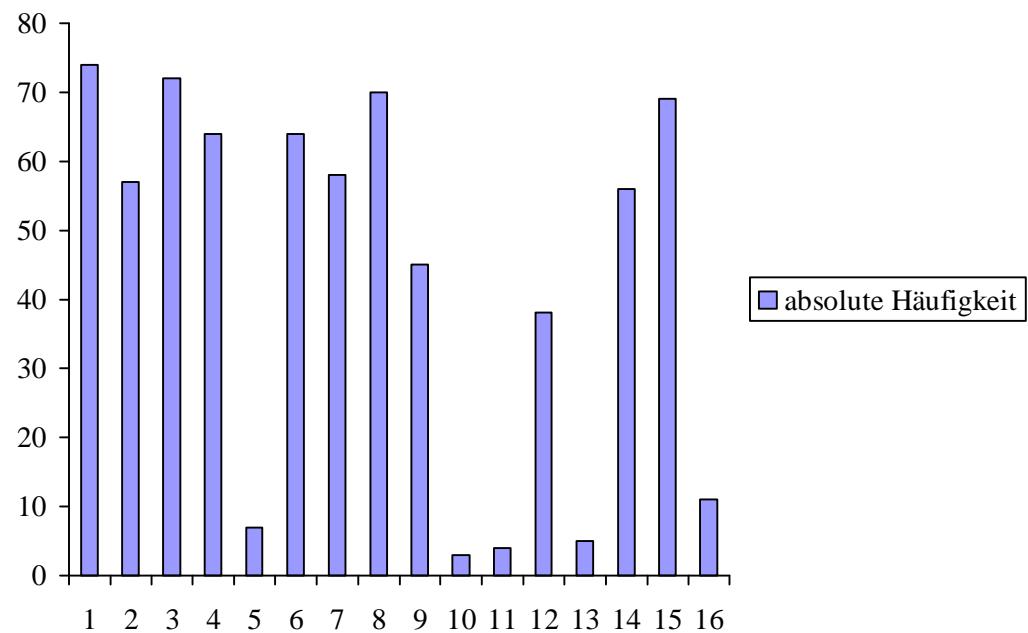


Abbildung 2: Absolute Häufigkeit der Realkennzeichen (1-16)

Das Auftreten der Kriterien in den drei Versuchsgruppen ergibt sich aus Tabelle 11.

Tabelle 11:

Anzahl der Aussagen, in denen inhaltliche Realkennzeichen gefunden wurden, getrennt nach Versuchsgruppen

Kriterium	Anzahl in Gruppe A (in %)	Anzahl in Gruppe B (in %)	Anzahl in Gruppe C (in %)
1. Logische Konsistenz und Widerspruchsfreiheit	21 (100 %)	33 (100 %)	20 (100 %)
2. Widerspruchsfreie ungeordnete Darstellung der Aussage	18 (85,7 %)	26 (78,8 %)	13 (65 %)
3. Quantitativer Detailreichtum	20 (95,2 %)	25 (75,8 %)	19 (95 %)
4. Kontextuelle Einbettung	19 (90,5 %)	31 (93,9 %)	14 (70 %)
5. Beschreibung von nonverbalen Interaktionen	2 (9,5 %)	4 (12,1 %)	1 (5 %)
6. Wiedergabe von Gesprächen	19 90,5 %	30 90,9 %	15 75 %
7. Komplikationen im Handlungsverlauf	17 (81,0 %)	28 (84,8 %)	13 (65 %)
8. Ausgefallene Einzelheiten	90,5 % (19)	97,0 % (32)	95 % (19)
9. Nebensächliche Einzelheiten	57,1 % (12)	72,7 % (24)	45 % (9)
10. Phänomengemäße Schilderung	9,5 % (2)	0 % (0)	5 % (1)
11. Inhaltliche Verschachtelungen	4,8 % (1)	9,1 % (3)	0 % (0)
12.a Schilderung fremd-psychischer Vorgänge	38,1 % (8)	69,7 % (23)	35 % (7)
12.b Schilderung eigen-psychischer Vorgänge	19,0 % (4)	3,0 % (1)	0 % (0)
13. Spontane Verbesserungen der eigenen Aussage	85,7 % (18)	69,7 % (23)	75 % (15)
14. Zugeben von Erinnerungslücken	95,2 % (20)	90,9 % (30)	95 % (19)
15. Unsicherheiten bei der Darstellung	9,5 % (2)	9,1 % (3)	30 % (6)

Aus Tabelle 12 ergibt sich, dass deskriptiv hinsichtlich des Auftretens der Realkennzeichen einige Unterschiede zwischen den Gruppen bestehen, z.B. wurden phänomenengemäße Schilderungen nur in den Gruppen A und C signiert und das Kriterium Detailreichtum wurde in 95,2 % Aussagen der Gruppe A, aber nur in 75,8% Aussagen der Gruppe B gefunden.

Tabelle 12:

Mittelwerte (M) und Standardabweichungen (S) der Ausprägung der Kriterien

Realkennzeichen	Gruppe A		Gruppe B		Gruppe C	
	M	S	M	S	M	S
Logische Konsistenz	2.76 <i>1.09</i>		2.97 <i>1.07</i>		2.55 <i>0.89</i>	
Inkontinenz	1.43 <i>0.98</i>		1.27 <i>1.04</i>		0.85 <i>0.81</i>	
Detailreichtum	2.95 <i>1.28</i>		2.85 <i>1.39</i>		2.10 <i>0.79</i>	
Kontextuelle Einbettung	1.86 <i>1.24</i>		1.67 <i>1.08</i>		1.05 <i>0.89</i>	
Nonverbale Interaktion	0.24 <i>0.77</i>		0.18 <i>0.53</i>		0.05 <i>0.22</i>	
Wiedergabe von Gesprächen	3.05 <i>1.47</i>		2.42 <i>1.37</i>		1.55 <i>1.23</i>	
Komplikationen im Handlungsverlauf	1.95 <i>1.16</i>		2.15 <i>1.44</i>		1.05 <i>1.05</i>	
Ausgefallene Einzelheiten	2.67 <i>1.06</i>		2.42 <i>1.06</i>		1.95 <i>0.69</i>	
Nebensächliche Einzelheiten	0.57 <i>0.51</i>		0.97 <i>0.85</i>		0.45 <i>0.51</i>	
Phänomengemäße Schilderung	0.14 <i>0.48</i>		0.00 <i>0.00</i>		0.05 <i>0.22</i>	
Inhaltliche Verschachtelungen	0.05 <i>0.22</i>		0.15 <i>0.51</i>		0.00 <i>0.00</i>	
Fremdpsychische Vorgänge	0.48 <i>0.68</i>		1.33 <i>1.24</i>		0.50 <i>0.83</i>	
Eigenpsychische Vorgänge	0.24 <i>0.54</i>		0.09 <i>0.52</i>		0.00 <i>0.00</i>	
Spontane Verbesserungen	1.95 <i>1.20</i>		1.27 <i>1.26</i>		1.05 <i>0.85</i>	
Zugeben von Erinnerungslücken	2.00 <i>1.05</i>		1.76 <i>1.15</i>		2.00 <i>1.26</i>	
Unsicherheiten bei der Darstellung	0.10 <i>0.30</i>		0.12 <i>0.42</i>		0.55 <i>1.10</i>	

Tabelle 13 zeigt die mittlere Ausprägung der Realkennzeichen in den einzelnen Versuchsgruppen. Es zeigen sich hier deskriptiv Unterschiede in den Variationen der Mittelwerte über die Gruppen für die einzelnen Kriterien. Zudem zeigen sich analog zur Häufigkeit des Auftretens Schwankungen innerhalb der Versuchgruppen für die Intensität der einzelnen Realkennzeichen.

7.7 Ergebnisse zur Trennung der Versuchsgruppen mittels Realkennzeichen

Zur Überprüfung der Hypothesen beurteilten die beiden Raterinnen für jeweils die Hälfte der Schilderungen der Versuchspersonen die Ausprägung der Realkennzeichen. Die 16 Realkennzeichen wurden zu der neuen abhängigen Variable Summenscore zusammengefasst. Dieses Vorgehen basierte auf Berechnungen der Interkorrelationen der Realkennzeichen, die zeigten, dass von den 120 möglichen Interkorrelation 59 statistisch signifikant wurden. Dass bedeutet, dass die 16 Realkennzeichen nicht als von einander unabhängig anzusehen sind. Weiterhin ergab eine Hauptkomponentenanalyse eine Einfaktorstruktur mit einer Kommunalität von 5.17 und einer Varianzaufklärung von 32.3% für die erste Hauptkomponente. Cronbachs Alpha für die 16 Realkennzeichen lag bei .83 (standardisiert .79). Für jede Versuchsperson wurde ein Summenscore über die festgestellten Ausprägungen der Realkennzeichen gebildet. Anschließend wurde ein Mittelwert der Summenscores in der jeweiligen Gruppe erstellt.

Tabelle 13:

Mittelwerte (M) und Standardabweichungen (S) der Ausprägung der Kriterien

Basis der Aussage	Summenscore	
	M	S
Gruppe A	22.43	7.94
Gruppe B	21.63	8.98
Gruppe C	15.75	6.13

Gruppe A: Kaspertheater gesehen

Gruppe B: Kaspertheater nicht gesehen, Inhalt mehrfach vorgelesen

Gruppe C: Kaspertheater nicht gesehen, Inhalt einmal vorgelesen

Die statistische Auswertung der Hypothesen wurde aufgrund des varianzanalytischen Designs mit Hilfe von Kontrasten (Einzelvergleichen) vorgenommen. Als Vortest wurde dabei eine einfaktorielle Varianzanalyse gerechnet, die mit Werten von $F = 4,426$ und $p = 0,015$ auf dem 5 %-Niveau signifikant wurde.

Hypothese 1

Nach der ersten Hypothese werden in erlebnisbegründeten Aussagen Realkennzeichen in stärkerer Ausprägung erwartet als in nicht erlebnisbegründeten Aussagen.

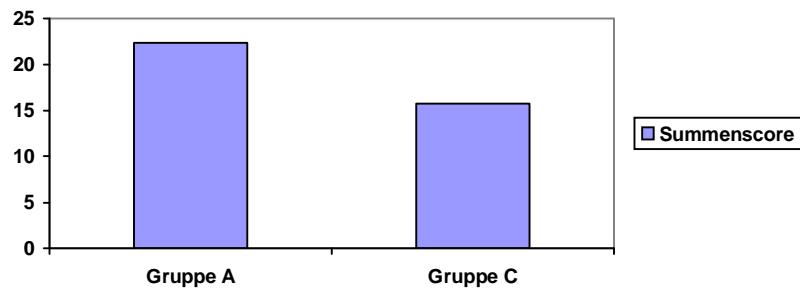


Abbildung 3: Diagramm der Summenscores der Realkennzeichen der Gruppen A und C

Der Mittelwert des Summenscores in Gruppe A betrug 22,43 ($S = 7,94$). In Gruppe C betrug der Mittelwert des Summenscores 15,75 ($S = 6,13$). Der Einzelvergleich der beiden Summenscores ergab mit Werten von $t = 2,668$ und $p = 0,009 / 2$ auf dem 5 %-Niveau einen signifikanten Unterschied zwischen den beiden Gruppen.

Somit kann die Nullhypothese, die davon ausgeht, dass kein Unterschied in den Mittelwerten der Summenscores in den beiden Gruppen besteht, verworfen werden. Das Ergebnis stützt die Hypothese, dass in erlebnisbegründeten Aussagen Realkennzeichen in höherer Ausprägung vorhanden sind als in nicht erlebnisbegründeten Aussagen.

Hypothese 2

Nach der zweiten Hypothese werden in nicht erlebnisbegründeten Aussagen, die auf wiederholten verbalen Informationen (Suggestion) basieren, Realkennzeichen in stärkerer Ausprägung erwartet, als in nicht erlebnisbegründeten Aussagen.

Der Mittelwert des Summenscores in Gruppe B betrug $M = 21,63$ ($S = 8,98$). In Gruppe C betrug der Mittelwert des Summenscores $15,75$ ($S = 6,13$). Der Einzelvergleich der beiden Summenscores ergab mit Werten von $t = 2,593$ und $p = 0,012 / 2$ auf dem 5 %-Niveau einen signifikanten Unterschied zwischen den beiden Gruppen. Somit kann die Nullhypothese, die davon ausgeht, dass kein Unterschied in den Mittelwerten der Summenscores in den beiden Gruppen besteht, verworfen werden.

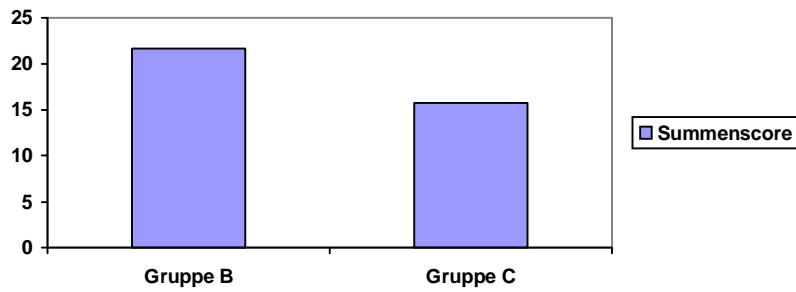


Abbildung 4: Diagramm der Summenscores der Realkennzeichen der Gruppen B und C

Das Ergebnis stützt die Hypothese, dass in nicht erlebnisbegründeten Aussagen, die auf wiederholten verbalen Informationen basieren Realkennzeichen in höherer Ausprägung vorhanden sind, als in nicht erlebnisbegründeten Aussagen.

In der Versuchsgruppe B (Suggestionsgruppe) wurde jedes Kind im Anschluss an das Interview zu dem Kaspertheater gefragt, ob es das Kaspertheater tatsächlich gesehen hat. Von den 33 Kindern der Gruppe B bestätigten dies acht Kinder. Um auszuschließen, dass Unterschiede hinsichtlich der Ausprägung der Realkennzeichen zwischen den Kindern bestehen, die möglicherweise aufgrund der Suggestion glauben, das Kaspertheater

real erlebt zu haben und denen, die wissen, dass sie es tatsächlich nicht gesehen haben, wurde die Gruppe in zwei Untergruppen (8 Vpn und 25 Vpn) aufgeteilt.

Ein t-Tests mit der in zwei Gruppen geteilten Gruppe B als unabhängige Variable und der Variable Summenscore als abhängige Variable ergab, dass sich unter der Annahme ungleicher Streuung mit $t= 0,447$ und $p= 0,663$ auf dem 5 %-Niveau keine signifikanten Unterschiede in der Ausprägung der Realkennzeichen zeigten. Aufgrund dieses Ergebnisses wurde auf eine Trennung oder Reduktion der Gruppe B bei der weiteren Auswertung verzichtet.

Hypothese 3

Nach der dritten Hypothese werden in erlebnisbegründeten Aussagen und in Aussagen, die auf wiederholten verbalen Informationen (Suggestion) basieren, Realkennzeichen in ähnlich starker Ausprägung erwartet.

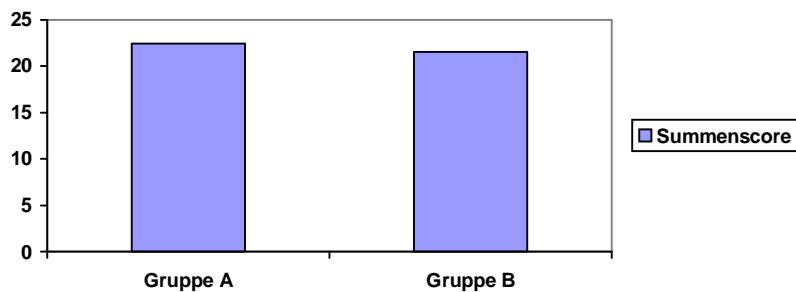


Abbildung 5: Diagramm des Summenscores der Realkennzeichen der Gruppen A und B

Der Mittelwert des Summenscores in Gruppe A betrug 22,43 ($S= 7,94$). In Gruppe B betrug der Mittelwert des Summenscores 21,63 ($S= 8,98$). Der Einzelvergleich (Anpassungstest) der beiden Summenscores ergab mit Werten von $t= 0,354$ und $p= 0,724$ keinen signifikanten Unterschied auf dem 20 %-Niveau zwischen den beiden Gruppen.

Somit kann die Hypothese, die davon ausgeht, dass kein Unterschied in den Mittelwerten der Summenscores in den beiden Gruppen besteht, nicht verworfen werden. Das bedeu-

tet, dass in erlebnisbegründeten Aussagen und in Aussagen, die auf wiederholten verbalen Informationen basieren, kein Unterschied in der Ausprägung der Realkennzeichen besteht.

7.8 Zusätzliche statistische Auswertungen

Wie oben beschrieben ergeben sich in Zusammenhang mit den untersuchten Hypothesen einige darüber hinausgehende Überlegungen, die in den folgenden zusätzlichen statistischen Analysen berücksichtigt werden sollen.

7.8.1 Trefferquoten

Analog zu anderen experimentellen Untersuchungen (Rüth-Bemelmanns, 1984; Yuille, 1988; Wellershauß & Wolf, 1989; Landry & Brigham, 1992; Köhnken et al., 1995; Wolf & Steller, 1997; Akehurst et al., 2001) hatten die Raterinnen die Aufgabe, nach Durchführung der kriterienbezogenen Aussageanalyse einzuschätzen, ob die jeweilige Aussage erlebnisbegründet war oder keinen realen Erlebnishintergrund hatte. Die sich ergebenden Trefferquoten sind in Tabelle 15 dargestellt.

² Mit einem χ^2 -Test soll geprüft werden, ob die Zuordnung der Schilderungen (Urteil) auf die beiden Erlebniskategorien erlebnisbegründet (Gruppe A) und nicht erlebnisbegründet (Gruppe C) per Zufall erfolgte. Das Ergebnis ist mit einem χ^2 -Wert von 5,707 auf dem 5 %-Niveau signifikant. Die Nullhypothese, dass kein Zusammenhang der beiden Dimensionen "tatsächliches Ereignis" und "Urteil der Rater" besteht, kann also verworfen werden. Die tatsächlich erlebnisbegründeten Aussagen der Kinder werden zu 81% von den Ratern in die richtige Kategorie eingeordnet. Die tatsächlich nicht erlebnisbegründeten Aussagen der Kinder wurden zu 55% richtig zugeordnet. Daraus ergibt sich insgesamt eine durchschnittliche Trefferquote von 68 %. Die Wahrscheinlichkeit, tatsächliche "Nicht-Erlebnisse" der Kinder auch als solche zu identifizieren, liegt mit 55% sehr eng an der Zufallswahrscheinlichkeit von 50%. Die richtige Identifizierung gelingt offensichtlich

erfolgreicher bei der Zuordnung erlebnisbegründeter Aussagen als bei der Identifizierung von Aussagen ohne Erlebnisbasis.

Tabelle 14:

Trefferquoten bei der Zuordnung der Schilderungen zu den Kategorien erlebnisbegründet/ nicht erlebnisbegründet

	<u>Urteil:</u> "Nicht Erlebnisbegründet"	<u>Urteil:</u> "Erlebnisbegründet"	Gesamt
<u>Gruppe A</u>	4 (19%)	17 (81%)	21 (100%)
<u>Gruppe B</u>	13 (39%)	20 (61%)	33 (100%)
<u>Gruppe C</u>	11 (55%)	9 (45%)	20 (100%)
Gesamt	28	46	74

Gruppe A: Kaspertheater gesehen

Gruppe B: Kaspertheater nicht gesehen, Inhalt mehrfach vorgelesen
Gruppe C: Kaspertheater nicht gesehen, Inhalt einmal vorgelesen

Einundsechzig Prozent der nicht erlebnisbegründeten Aussagen (Gruppe B) wurden falsch zugeordnet, 39 % richtig klassifiziert. Dagegen wurden 81 % der tatsächlich erlebnisbegründeten Aussagen richtig zugeordnet, 19% falsch zugewiesen. Die durchschnittliche Trefferquote eines richtigen Urteils beläuft sich demnach auf 60 %. Um zu überprüfen, ob die Zuordnung der Schilderungen (Urteil) auf die beiden Erlebniskategorien (Gruppe A und B) per Zufall erfolgte, wurde ein χ^2 -Test durchgeführt. Das Ergebnis ist mit einem χ^2 -Wert von 2,463 nicht signifikant. Die Nullhypothese, dass kein Zusammenhang besteht kann also nicht verworfen werden und es muss Unabhängigkeit

zwischen Urteil und Erlebnishintergrund angenommen werden. Die beobachteten Tref-ferquoten bei den Gruppen A und B könnten demnach zufällig zustande gekommen sein. Bezuglich des Summenscores der Realkennzeichen ergab ein Einzelvergleich der beiden Urteilsgruppen erlebnisbegründet und nicht erlebnisbegründet mit den Werten $t = 5,904$ und $p = 0,000 / 2$ einen hochsignifikanten Unterschied. In der Urteilsgruppe Erlebnisbe- gründet betrug der Mittelwert 23,98 ($S=7,25$), in der Gruppe Nicht erlebnisbegründet war der Mittelwert 14,18 ($S=6,35$).

7.8.2 Urteilssicherheit

In Verbindung mit der Signierung der Realkennzeichen und der Einschätzung des Erlebnishintergrundes der Aussagen wurde auch die Urteilssicherheit der Raterinnen erhoben. Die Raterinnen hatten auf einer Fünf-Punkte-Skala (1= unsicher; 5= sehr sicher) einzuschätzen, wie sicher sie sich bezüglich ihres Urteils über die Erlebnisgrundlage der je- weiligen Aussage waren. In der folgenden Tabelle wird die Urteilssicherheit bei der Zu- ordnung zu den Kategorien Erlebnisbegründet oder Nicht erlebnisbegründet dargestellt.

Tabelle 15:

Urteilssicherheit				<u>Urteilssi- cherheit und Urteil</u>
Gruppe	N	M	S	
A	21	3,76	1,30	
B	33	3,06	1,14	
C	20	2,55	1,28	

Der t-Test ergab auf dem 1 %- Niveau einen signifikanten Unterschied in der Urteilssi- cherheit bezogen auf die Art des Urteils (erlebnisbegründet / nicht erlebnisbegründet).

Die Raterinnen waren unabhängig von der tatsächlichen Richtigkeit ihres Urteils sicherer, wenn sie Schilderungen als erlebnisbegründet klassifizierten. Bezieht man die drei Versuchgruppen mit ein, so ergibt sich das in Tabelle 17 dargestellte Bild.

Tabelle16:

Urteilssicherheiten in den drei Versuchsgruppen

Urteil	N	M	S	t	sign.
erlebnis-begründet	46	3,57	1,31		
Nicht erlebnisbegründet	28	2,39	0,88	- 4.609	P< .001

Die größte Urteilssicherheit bestand mit einem Mittelwert von 3,76 bei den Aussagen, denen das tatsächlich gesehene Kasperstück zugrunde lag (Gruppe A). Deutlich geringer war die Urteilssicherheit mit einem Mittelwert von 2,55 bei der Zuordnung der Aussagen aus der Kontrollgruppe (Gruppe C) in die Kategorien erlebnisbegründet/nicht erlebnisbegründet. Schilderungen, die nur auf intensiven verbalen Informationen basierten (Gruppe B) wurden mit einer höheren Urteilssicherheit klassifiziert ($M= 3,06$) als die Darstellungen der Kontrollgruppe (Gruppe C), aber geringer als bei Gruppe A. Eine einfaktorielle Varianzanalyse zeigt, dass sich die Mittelwerte der Urteilssicherheit lediglich in den Gruppen A und C signifikant unterscheiden. ($F = 5,082$, $p= .009$).

Um festzustellen ob ein Zusammenhang zwischen Urteilssicherheit und Richtigkeit des Urteils besteht, wurde die Korrelation der Variablen Urteil und Urteilssicherheit berechnet. Diese ergab mit einer Korrelation von 0,44 auf dem 1 %-Niveau einen hochsignifikanten Zusammenhang zwischen der Art des Urteils (erlebnisbegründet / nicht erlebnisbegründet) und der Urteilssicherheit.

Eine zweifaktorielle Varianzanalyse mit den Faktoren Gruppe und Urteil (erlebnisbegründet / nicht erlebnisbegründet) und der abhängigen Variable Urteilssicherheit wurde nur auf dem Faktor Urteil mit $F= 10,99$ und $p= 0,001$ auf dem 5 %-Niveau signifikant.

Es bestehen demnach zwar keine signifikanten Unterschiede zwischen den Gruppen in Bezug auf die Urteilssicherheit, jedoch kann angenommen werden, dass die Art des Urteils einen Einfluss darauf hat, wie sicher sich die Rater in ihrer Beurteilung waren, auch wenn diese Beurteilung nicht dem tatsächlichen Erlebnishintergrund entsprach. Offensichtlich wurde die relativ häufig falsche Zuordnung als erlebnisbegründet in der Gruppe B, die tatsächlich nur verbale Informationen erhielt, mit einer relativ hohen Sicherheit bezüglich der Richtigkeit der Einschätzung, abgegeben.

7.8.3 Variable Wortanzahl

In Kap. 7.4 wurde gezeigt, dass die Wortanzahl einen signifikanten Zusammenhang mit den Summenscores hat und sich signifikant in den drei Versuchsgruppen unterscheidet. Folglich stellt sich die Frage, welche Variable besser geeignet ist, die Gruppenzugehörigkeit vorherzusagen. Sollte die Wortanzahl eine ebenso gute oder sogar überlegene Vorhersageleistung erbringen, wäre die Durchführung einer zeitaufwändigen inhaltlichen Analyse zumindest bei experimentellen Untersuchungen aus ökonomischen Gründen in Frage zu stellen. Mittels einer logistischen Regression wurde daher geprüft, ob die Realkennzeichen über die Wortanzahl hinaus zur Vorhersage der Gruppenzugehörigkeit beitragen. Die Ergebnisse werden in der folgenden Tabelle dargestellt:

Tabelle 17:

Vorhersage der Gruppenzugehörigkeit durch Wortanzahl und Realkennzeichen

Versuchsgruppen	β -Koeffizient	Sig.
A vs C	-,001	,473
	-,105	,110
B vs C	,000	,771
	-,109	,106
A vs B	-,001	,248
	,024	,601

Es zeigte sich, dass die Variablen Worte und Realkennzeichen nicht zur Vorhersage der Gruppenzugehörigkeit beitragen konnten. Die β -Koeffizienten sind zwar durchgängig bei

den Realkennzeichen höher als bei der Wortanzahl, aber nicht signifikant. Da der Summenscore lediglich die Ausprägungsstärke der Realkennzeichen abbildet, aber möglicherweise nicht die diagnostische Valenz, wurde die subjektive Gewichtung der Realkennzeichen folgendermaßen berücksichtigt: Durch die Multiplikation des Summenscores mit den Werten der Urteilsicherheit wurde eine neue Variable gebildet. Diese nimmt nur bei hohem Summenscore und großer Sicherheit hohe Werte an.

Tabelle 18:

Vorhersage der Gruppenzugehörigkeit durch Wortanzahl und Realkennzeichen x Urteilssicherheit

Versuchsgruppen	ß-Koeffizient	Sig.
A vs C	-,001	,319
	-,024	,025 *
B vs C	-,001	,636
	-,015	,149
A vs B	-,001	,443
	,005	,421

Während die Wortanzahl keinen Beitrag zur Vorhersage der Gruppenzugehörigkeit leisten konnte, zeigte sich, dass die kombinierte Variable aus Realkennzeichen und Urteilssicherheit signifikant zwischen erlebnisbegründeten Aussagen und Schilderungen, die keinen Erlebnishintergrund haben, differenzieren konnte.

8. Diskussion der Ergebnisse

Mit dieser Arbeit sollte die Validität der Realkennzeichen bei der Anwendung von Aussagen mit unterschiedlichem Erlebnishintergrund bei Vorschulkindern untersucht werden.

Bei den Aussagen, die in dieser Arbeit untersucht wurden, handelt es sich um Berichte von Kindern im Vorschulalter zu einem Ereignis, dass die Kinder in unterschiedlicher Form wahrgenommen haben. Die Aussagen basierten entweder auf einem tatsächlich erlebten Ereignis oder hatten eine vorgelesenen Geschichte als Grundlage, wobei die Intensität des Vorlesens stark variiert wurde, in dem die Erzählung einmal oder zehnmal vorgegeben wurde. Die vorgelesenen Geschichten entsprachen im Inhalt dem tatsächlichen Erlebnis.

In der ersten Fragestellung sollte untersucht werden, ob Aussagen von Kindern im Vorschulalter zu einem tatsächlich erlebten Ereignis hinsichtlich des Ausprägungsgrades inhaltlicher Realkennzeichen von Schilderungen unterschieden werden können, die keinen Erlebnishintergrund haben. Die erste Hypothese postuliert, dass in Aussagen, die sich auf ein tatsächlich erlebtes Ereignis beziehen, Realkennzeichen in einem durchschnittlich höheren Ausprägungsgrad vorhanden sind als in Aussagen, die lediglich auf einmalig verbal vorgegebenen Informationen beruhen.

Des Weiteren sollte untersucht werden, ob sich Aussagen von Kindern im Vorschulalter zu einer mehrfach verbal vorgetragenen Geschichte von Aussagen unterscheiden, die nahezu keinen Erlebnishintergrund haben. In der Hypothese zu dieser Fragestellung wird postuliert, dass der Ausprägungsgrad der inhaltlichen Realkennzeichen in den Berichten zu den intensiv verbal vorgegebenen Inhalten höher ist als in den Schilderungen, die lediglich nahezu keinen Erlebnishintergrund haben.

In der dritten Fragestellung sollte überprüft werden, ob sich Berichte von Kindern im Vorschulalter zu Ereignissen mit realem Erlebnisgrund von Schilderungen unterscheiden, die auf intensiven verbalen Informationen beruhen. Dazu wurde die Hypothese aufgestellt, dass sich hinsichtlich der Ausprägung der inhaltlichen Realkennzeichen keine Unterschiede ergeben.

8.1 Allgemeine Feststellungen zur Untersuchung

Zu den Ergebnissen der vorliegenden Arbeit lässt sich zunächst feststellen, dass alle Vorschulkinder in der Lage waren, ein selbst erlebtes, komplexes Ereignis detailreich, weitgehend logisch konsistent und weitgehend fehlerfrei wiederzugeben. Dieses Ergebnis entspricht im Wesentlichen den Befunden der Forschung zu kindlichen Zeugenaussagen. Diese weisen darauf hin, dass verbale Darstellungen jüngerer Kinder im freien Bericht zwar weniger ausführlich und vollständig sind als die Schilderungen älterer Kinder, dass aber hinsichtlich der Korrektheit der Schilderungen keine Unterschiede gefunden wurden. Mit zunehmendem Alter steigt zwar die Anzahl der berichteten Details, gleichzeitig nimmt aber auch die Fehlerrate zu (Memon, Cronin, Eaves, Bull & Küpper, 1992, Marin, Holmes, Guth & Kovac, 1979). In anderen Studien (Goodman & Reed, 1986; Goodman, Aman & Hirschman, 1987) mit sehr jungen Kindern wurde festgestellt, dass Kinder unter vier Jahren im freien Bericht sehr wenig oder nichts berichten und fast ausschließlich nur auf konkrete Fragen antworteten. Abweichend dazu reagierten in der vorliegenden Untersuchung die Kinder im Vorschulalter nicht nur auf Fragen, sondern schilderten ihre Erlebnisse auch in längeren Zusammenhängen. Dies zeigt, dass die verwendeten Quellen der jeweiligen Aussagen, nämlich das Kaspertheater bzw. eine Geschichte über das Stück, den kindlichen Verständnis- und Erfahrungshorizont derart ansprechen konnten, dass es den Kindern möglich war daraus zusammenhängende, detailreiche und inhaltlich plausible Schilderungen zu produzieren. Zudem bestätigt es die Wahl des Untersuchungsparadigmas für diese Altersgruppe und spricht für die ökologische Validität der vorliegenden Untersuchung.

Die Analyse früherer experimenteller Studien ergab, dass nur wenige Arbeiten eine ausreichende ökologische Validität nach Brunswick (1955) aufweisen und eine Generalisierbarkeit der Ergebnisse zulassen. Daraus ergab sich die Überlegung, für diese Arbeit ein Untersuchungsparadigma zu entwickeln, in dem wichtige Variablen forensisch relevanter Sachverhalte berücksichtigt werden sollten, ohne dabei ethische Grenzen zu überschreiten. Als relevante Grundvariablen für sexuelle Missbrauchssituationen - dieses Delikt nimmt in der psychologischen Begutachtung den größten Raum ein – nennen Steller,

Wellershaus und Wolf (1992): Eigenbeteiligung am Geschehen, negative emotionale Tönung des Geschehens und Kontrollverlust in der Situation.

Diese Variablen konnten in der vorliegenden Untersuchung nur annähernd realisiert werden. Das ergab sich daraus, dass das von den Kindern zu schildernde Geschehen in das Untersuchungsdesign einbezogen werden sollte, um ein inhaltlich vergleichbares Basisergebnis zu haben und um die Richtigkeit der späteren Schilderungen überprüfen zu können. Während man bei der Schilderung zurückliegender Lebensereignisse auch auf Geschehnisse zurückgreifen kann, die eine negative emotionale Tönung haben, z.B. ein Zahnarztbesuch oder das Erleben eines Verkehrsunfalls, sind bei der Konstruktion eines Basisereignisses für Kinder schnell ethische Grenzen erreicht, die das Erzeugen eines unangenehmen Erlebnisses als nicht geboten erscheinen lassen. Daher wurde in der vorliegenden Untersuchung eine *negative* emotionale Tönung durch eine *positive* emotionale Tönung ersetzt und das Paradigma Kaspertheater ausgewählt. Die Bedeutsamkeit der Variablen emotionale Tönung des Geschehens ergibt sich aus der Annahme, dass emotional getönte Erlebnisse intensiver wahrgenommen und stabiler eingeprägt werden, als Alltagserlebnisse ohne besondere Prägnanz. Die Verhaltensbeobachtung der Kinder während der Vorführung deutete auf eine hohe emotionale Beteiligung der Kinder am Geschehen hin. Daher kann davon ausgegangen werden, dass das dargebotene Erlebnis von den Kindern intensiv wahrgenommen und gespeichert werden konnte. Dies bildete sich in den überwiegend detailreichen Schilderungen ab.

Durch das für ein Kaspertheater typische Einbeziehen der Kinder in die Handlung - fallsweise Dialoge mit den Kindern - wurde eine Eigenbeteiligung der Kinder am Geschehen erreicht. Auch wenn ein Kind nicht unmittelbar selbst Kommunikator war, so ist anzunehmen, dass das Beobachten der Interaktion eines vertrauten Spielgefährten mit einem Darsteller des Kaspertheaters eine höhere Erlebnisqualität hat als das Betrachten einer ähnlichen Handlung mit fremden Akteuren, z.B. in einer Videosequenz. Damit einher geht eine intensivere Wahrnehmung und Speicherung des Geschehens, da die Eigenbeteiligung höher ist, auch wenn diese wahrscheinlich deutlich geringer ist als z. B. beim Erleben eines Unfalls.

Die oben genannte Variable Kontrollverlust in der Situation konnte in der Untersuchung nur in eingeschränktem Maß verwirklicht werden. Sie beschränkte sich auf eine Teilnahmepflicht der Kinder an der Vorstellung sowie einem Verbot, das Kaspertheater während der Vorführung zu verlassen. Anzunehmen ist, dass ein Kontrollverlust in einer negativ getönten Situation erheblich gravierender erlebt wird als in einer positiven. Der Wunsch, sich dem Geschehen zu entziehen, hatte im Rahmen des gewählten Paradigmas sicherlich eine weitaus geringere Auftretenswahrscheinlichkeit als z.B. in Situationen mit körperlichen Übergriffen.

Insgesamt kann davon ausgegangen werden, dass sich die Kinder mit dem Erleben des Kaspertheaters in einer emotional und kognitiv nicht alltäglichen Situation befanden. Durch dieses positiv getönte Erlebnis wurden die Prozesse der Wahrnehmung, des Einspeicherns und des Erinnerns in besonderer Weise aktiviert, wenn auch möglicherweise nicht in der gleichen Qualität wie beim Erleben realer negativ getönter Ereignisse. Dabei wurden nicht alle Sinneskanäle beansprucht. In erster Linie waren die visuellen und auditiven Kanäle eingebunden. Allein bei Kindern, die im Laufe des Erlebnisses nass gespritzt worden waren, wurden auch taktile Kanäle beansprucht. Nach Wegener (1997) ist jedoch die physische Identität zwischen der zu untersuchenden Lebensbedingung und der experimentellen Bedingung für das Ergebnis einer Untersuchung weniger ausschlaggebend als die Vergleichbarkeit der Prozesse, die durch beide Situationen ausgelöst werden.

Bei der Anlage des Experiments ging es zunächst darum, eine experimentelle Situation zu schaffen, die hinreichend auswertbares Basismaterial für die nachfolgende inhaltliche Analyse liefern sollte. Es sollte ein Ereignis gefunden werden, dass den Kindern ausreichend umfangreiche Schilderungen ermöglichte, um überhaupt eine inhaltliche Analyse vornehmen zu können. Mit der Vorführung des Kaspertheaters wurden die Kinder mit einem sehr komplexen Sachverhalt konfrontiert, der durch verschiedene parallel laufende und inhaltlich miteinander verwobene Handlungsstränge gekennzeichnet war. Dadurch erhielten die Kinder ein sehr umfangreiches Wahrnehmungsangebot und somit eine gute Grundlage für ihre späteren Schilderungen. Der Wortumfang der Schilderungen, im

Durchschnitt sprachen die Kinder 674 (S=375) Worte, kann als Beleg für die treffende Wahl des Untersuchungsparadigmas gewertet werden.

Der Umstand, dass es in der vorliegenden Untersuchung zudem gelang, umfangreiche Schilderungen von den Kindern zu erhalten, denen das Geschehen lediglich häufig vorgelesen wurde, kann mit dem Einsatz vertrauter Autoritätspersonen aus ihrem Alltag als „Suggestoren“ zusammenhängen. Verschiedene Untersuchungen konnten zeigen, dass Kinder Informationen von Erwachsenen eher übernehmen als von Kindern, da sie die Kompetenz von Erwachsenen höher einschätzen. Allerdings liegen zum Status von Suggestoren bislang keine weiteren Befunde vor. Lepore und Sesco (1994) stellten zwar fest, dass Kinder im freien Bericht weniger Fehler produzierten, wenn sie zu von einem vertrauten Erwachsenen gefragt wurden. Es wurden aber keine Unterschiede zwischen der Befragung durch vertraute und fremde Erwachsene hinsichtlich der Beantwortung suggerierter Inhalte gefunden. Allerdings waren in der Untersuchung vertraute Personen dadurch definiert, dass die Kinder der Versuchgruppe den jeweiligen Erwachsenen lediglich drei Mal für ein halbe Stunde erlebt hatten. Möglicherweise hat in dieser Untersuchung der höhere Vertrautheitsgrad von Eltern und Erziehern im Vergleich zu Fremden einen verstärkenden Einfluss auf die Aufnahmebereitschaft der Kinder für die vorgelesenen Geschichten und das Ausmaß der Speicherung und Reproduktion gehabt. Zum Einfluss dieser Variablen sind weitere Studien angebracht, da in der Praxis Beeinflussungen von Kindern häufig von vertrauten Personen vorgenommen werden und hier möglicherweise erheblich verstärkende Mechanismen wirksam sein könnten.

Die Überlegung, der relativ große Schilderungsumfang stelle zwar eine ausreichende Quantität dar, enthalte aber möglicherweise nur wenige kriterienerfüllte Textstellen, bestätigte sich nicht. Es zeigte sich, dass in den Schilderungen der Kinder alle der sechzehn in die Untersuchung aufgenommenen Realkennzeichen mit unterschiedlichen, aber hinreichenden Auftretenshäufigkeiten vorhanden waren. Daher konnte davon ausgegangen werden, dass das ausgewählte Untersuchungsparadigma geeignet war, um eine Inhaltsanalyse durchzuführen.

In dieser Arbeit sollte u. a. untersucht werden, ob die Inhaltsanalyse bei der Beurteilung von Aussagen angewendet werden kann, die nicht auf eigenem Erleben beruhen, sondern auf intensiven mündlichen verbalen Informationen, also suggestiven Einflüssen. Das Ergebnis anderer Untersuchungen, in denen verbale Informationen von den Versuchspersonen (Leichtman & Ceci, 1995) fälschlich als reale Erlebnisse wiedergegeben wurden, konnte in dieser Arbeit nicht in demselben Maße reproduziert werden. Nur 8 von 33 Kindern glaubten, dass sie die mehrfach erzählte Geschichte tatsächlich erlebt hatten. Da sich diese bezüglich der Ausprägung der Realkennzeichen nicht von den übrigen Kindern unterschieden, wurden die Daten zusammen ausgewertet. In der praktischen Begutachtung könnte der Teil der Gruppe, dem bewusst war, dass er eine gehörte Geschichte vortrug, den Kindern entsprechen, die intensiven Gesprächen mit Erwachsenen ausgesetzt waren und dann in einer Befragung vorgeben, das Berichtete selbst erlebt zu haben, obgleich sie wissen, dass dieses nicht der Fall ist.

8.2 Validität der kriterienorientierten Aussageanalyse bei Kindern im Vorschulalter

Mit der vorliegenden Untersuchung konnte gezeigt werden, dass die kriterienorientierte Aussageanalyse auch bei Kindern im Vorschulalter eine geeignete Methode ist, den Realitätsbezug ihrer Aussagen zu überprüfen.

Der festgestellte durchschnittliche Ausprägungsgrad der inhaltlichen Realkennzeichen in den Schilderungen der Kinder im Vorschulalter zu einem erlebten Ereignis und in den Berichten zu einem inhaltlich vergleichbaren Sachverhalt, der den Kindern nur verbal vorgetragen worden war, stützt die erste Hypothese. In den Aussagen der Kinder zu einem tatsächlich erlebten Ereignis sind Realkennzeichen in einem Ausprägungsgrad festgestellt worden, der signifikant über dem der Berichte über den verbal transportierten Sachverhalt liegt. Damit kann die Undeutsch-Hypothese, die einen qualitativen Unterschied zwischen Aussagen über selbst erlebte Ereignisse und Aussagen über nicht selbst

erlebte Begebenheiten postuliert, auch bei Kindern im Vorschulalter als grundsätzlich zutreffend angesehen werden.

Für die psychologische Begutachtung von Kindern im Vorschulalter, ein nicht geringer Teil der psychologischen Begutachtung befasst sich mit der Glaubwürdigkeitsuntersuchung von Kinderaussagen in mutmaßlichen Fällen sexuellen Missbrauchs, ist das Ergebnis dieser Untersuchung bedeutsam. Damit kann dem wiederholt geäußerten Einwand von Juristen, aber auch von Psychologen argumentativ begegnet werden, Kinderaussagen seien zur Feststellung eines Erlebnishintergrundes nicht geeignet. Als Argument wird häufig die noch nicht abgeschlossene Sprachentwicklung von jungen Kindern angeführt. Kinder seien demnach aufgrund ihres eingeschränkten Wortumfangs und ihren begrenzten Schilderungsfähigkeiten noch nicht befähigt, eine umfassende Darstellung eines erlebten Sachverhaltes zu geben. In der vorliegenden Untersuchung konnte jedoch gezeigt werden, dass drei- bis sechsjährige Kinder in beeindruckender Weise in der Lage waren, einen komplexen Sachverhalt mit unterschiedlichen Handlungssträngen erlebnisgerecht darzustellen. Zwar variierten der Umfang der Schilderungen und die Anzahl berichteter Details erheblich, es gab allerdings keine Schilderung, in der keine Details berichtet wurden.

Als weiteres Argument gegen die Anwendung der kriterienbezogenen Aussageanalyse bei sehr jungen Kindern wird angeführt, dass sie aufgrund ihres eingeschränkten Wahrnehmungsvermögens zwar möglicherweise quantitativ ausreichende Darstellungen eines Sachverhalts geben könnten, dass ihre sprachliche Ausdrucksfähigkeit aber noch nicht ausreichend ausgebildet seien, um jene qualitativen Merkmale in einem derartigen Ausprägungsgrad hervorzubringen, die für die Bewertung des Erlebnisgehaltes ihrer Aussage erforderlich seien. Die Ergebnisse dieser Untersuchung weisen darauf hin, dass auch Vorschulkinder in der Lage sind, ein Ereignis so zu schildern, dass die gesamte Bandbreite des Kriterienkatalogs zum Einsatz kommen kann. In der vorliegenden Untersuchung wurden nicht nur vereinzelte Realkennzeichen signiert, sondern sämtliche in diese Untersuchung einbezogenen Realkennzeichen vorgefunden. Grundsätzlich hängen Anzahl, Ausprägung und die Verschiedenheit der Realkennzeichen vom Ablauf und der Komple-

xität des zu schildernden Ereignisses ab. In der Darstellung eines einfachen, wenig komplexen Geschehens wird man weniger Realkennzeichen finden als in Berichten über komplexe Sachverhalte.

Zwar bestätigen die Ergebnisse der Untersuchung die grundsätzliche Eignung der kriterienorientierten Aussageanalyse bei Vorschulkindern, deutlich wurde aber auch, dass keine validen Ergebnisse zu erwarten sind, wenn suggestive Einflüsse im Vorfeld einer Aussage vorhanden sind. Es konnten keine Qualitätsunterschiede zwischen erlebnisbegründeten und durch intensive verbale Vorgabe entstandenen Schilderungen festgestellt werden. Bezogen auf die gutachterliche Praxis in Fällen des sexuellen Missbrauchs bedeutet dieses, dass wiederholte, konkrete Vorwürfe thematisierende Gespräche mit angenommenen Opfern dazu führen können, dass Inhalte aufgenommen werden und in späteren Anhörungen als eigenes Erleben berichtet werden. Das Argument, dass junge Kinder nicht in der Lage sind, detailreiche Vorgaben aufzunehmen, zu speichern und später wiederzugeben, wird durch die Ergebnisse dieser Arbeit eindrucksvoll widerlegt. Komplizierte Sachverhalte, die ausschließlich verbal dargeboten wurden, konnten von den Kindern überwiegend widerspruchsfrei und detailliert berichtet werden. Alter, Intelligenz und verbale Fähigkeiten hatten zwar einen allgemeinen Einfluss auf die Ausprägung der Realkennzeichen, nicht aber bezüglich der verschiedenen Untersuchungsbedingungen. Das Argument, dass nur Kinder mit besonderen intellektuellen und verbalen Fähigkeit in der Lage seien, verbale Informationen aufzugreifen und später zu reproduzieren, wird für die vorliegenden Ergebnisse widerlegt, da kein Gruppeneffekt bezüglich dieser personalen Variablen festzustellen war.

Anzunehmen ist, dass die offensichtlich erfolgreiche Aufnahme der verbalen Informationen u.a. durch die wiederholte Vorgabe und möglicherweise auch durch den Einsatz vertrauter Personen als "Suggestoren" zustande gekommen ist. Dabei wurde im Untersuchungskontext auf spezifische suggestive Techniken, wie suggestive Frageformen, inhaltliche Vorgaben verzichtet. Lediglich die Eingangsfrage unterstellte, dass das Geschehen erlebt wurde. Die Suggestoren hatten auch keine Anweisungen, den Kindern Inhalte nahe zulegen oder auf deren Erlebnishintergrund zu bestehen. Das bedeutet, dass schon bei

moderaten suggestiven Bedingungen qualitativ hochwertige Aussagen generiert werden konnten. Für die gutachterliche Praxis bedeutet dies, dass nicht nur bei dem Verdacht von gezielter "Aufdeckungsarbeit" die Entstehungsgeschichte der Aussage umfassend recherchiert werden muss, sondern auch in Fällen, in denen keine gezielten suggestiven Einflüsse angenommen werden, nach Quellen und Bedingungen, die die Qualität der Aussage beeinflusst haben könnten, geforscht werden muss. Die in vielen Studien festgestellte erhöhte Suggestionsanfälligkeit von jungen Kindern (z.B. Vrij und Bush, 2000, Bruck, Hembroooke und Ceci, 1997) unterstützt diese Forderung.

Für die rechtspychologische Praxis bedeutsam ist, ob aus den Ergebnissen der Aussageanalyse auf den realen Erlebnishintergrund einer Aussage geschlossen werden kann. In dieser Untersuchung wurden 81% der erlebnisbegründeten Aussagen richtig klassifiziert. Bei den nicht erlebnisbegründeten Aussagen waren es lediglich 55%, also ein Ergebnis, das eher im Zufallsbereich anzusiedeln ist. Vergleichbare Unterschiede, wenn auch nicht so ausgeprägt, wurden in anderen Untersuchungen festgestellt. Auch dort wurde deutlich, dass die richtige Klassifizierung erlebnisbegründeter Aussagen besser gelang als die Klassifizierung nicht erlebnisbegründeter Darstellungen. Yuille (1988) wies Trefferquoten von 90,9% für die Einordnung wahrer Geschichten und 74,4% für die Identifizierung ausgedachter Berichte nach. In der Untersuchung von Wellershaus und Wolf (1989) wurden 76,7% der erlebnisbegründeten Schilderungen als wahr und immerhin 67% der nicht erlebnisbegründeten Berichte als unwahr eingestuft. Bei Landry und Brigham (1992) identifizierten trainierte Rater 75,4% der erlebnisbegründeten Aussagen als wahr, aber nur 35% der nicht erlebnisbegründeten als unwahr. Santilla, Roppola, Runtti und Niemi, (2000) fanden Trefferquoten von 61,2% für die richtige Einschätzung der Aussagen als erlebnisbegründet. Generell ist somit festzustellen, dass Aussagen von erlebten Ereignissen von geschulten Ratern mit einer höheren Trefferquote eingeordnet werden als Aussagen zu nicht erlebten Ereignissen.

Von der Undeutsch-Hypothese ausgehend erfolgt die Klassifizierung einer Aussage auf der Grundlage der gefundenen Realkennzeichen. Je ausgeprägter die Qualität der Merkmale desto größer müsste die Wahrscheinlichkeit sein, dass eine Aussage als erlebnisbe-

gründet eingeschätzt wird. Da aber auch in nicht erlebnisbegründeten Aussagen Realkennzeichen vorhanden, sind auf Seiten des Urteilers ein Entscheidungsprozess stattfinden, bei dem abzuwägen ist, ob *wegen* des Vorhandenseins von Realkennzeichen ein tatsächlich erlebtes Ereignis wahrscheinlich ist, oder ob *trotz* des Vorhandenseins von Realkennzeichen der Erlebnishintergrund der Aussage fraglich ist.

In der vorliegenden Untersuchung wurde versucht, diesen Abwägungsprozesses anhand der Urteilssicherheit der Rater, mit der sie die Identifizierung und Klassifizierung von Aussagen vornehmen, darzustellen. Unabhängig von der Richtigkeit ihres Urteils waren die Rater sicherer, wenn sie Schilderungen als erlebnisbegründet klassifizierten. Es zeigte sich zudem ein signifikanter Unterschied in der Urteilssicherheit zwischen der Gruppe, die das Kaspertheater gesehen hatte und der Kontrollgruppe zugunsten der Erlebnisgruppe. Anzunehmen ist, dass die geringere Ausprägung der Realkennzeichen in der Kontrollgruppe dazu führte, dass die Aussagen weniger sicher beurteilt werden konnten. Dagegen bestand zwischen den Kindern, die das Kaspertheater gesehen hatten und den Kindern, die es intensiv vorgelesen bekamen, kein Unterschied hinsichtlich der Urteilssicherheit. Offensichtlich waren die Rater bei der falschen Einschätzung, dass die zweite Gruppe auch erlebnisbegründet berichtete, sehr sicher. Analog zu der oben angeführten Überlegung kann man davon ausgehen, dass die nahezu gleiche Ausprägung der Summenscores in diesen beiden Gruppen das Resultat begründen kann. Aus diesen Ergebnissen sind mehrere Folgerungen abzuleiten: Zum einen wird die Untersuchungshypothese nach der suggerierte Aussagen mittels Realkennzeichen nicht von erlebnisbegründeten Aussagen zu unterscheiden sind, bestätigt. Hierbei ist aber zwischen experimenteller Situation und praktischer Begutachtung zu unterscheiden. Während im experimentellen Setting der vorliegenden Arbeit und auch vieler anderer Studien keine weiteren Informationen zur Einschätzung des Wahrheitsgehaltes der Aussage zur Verfügung standen, werden in der gutachterlichen Praxis im Rahmen des hypothesenleiteten Vorgehens Befunde aus anderen Bereichen erhoben und in den Beurteilungsprozess einbezogen. Wie eingangs in Kapitel 2.1 beschrieben, ist die Durchführung der kriterienorientierte Inhaltsanalyse lediglich *ein* Element im Begutachtungsprozess. So werden in der Praxis Persönlichkeitsdaten des Aussagenden, dessen mögliche Motive, die Aussagegeschichte

und Aussageentwicklung in den Entscheidungsprozess eingebunden. Dabei zeigt die praktische Erfahrung, dass gerade die Befunde außerhalb der kriterienorientierten Aussageanalyse, insbesondere die Aussagegeschichte und Fehlerquellenanalyse für die psychologische Begutachtung von großer Bedeutung sind, auch wenn deren Validität experimentell bisher kaum bestätigt wurde. Lediglich Landry und Brigham (1992) konnten zeigen, dass sich die Trefferquoten von Ratern signifikant verbesserten, wenn die Rater die Klassifizierung von Aussagen nicht nur auf der Grundlage von Interviewprotokollen vornahmen, sondern Videoaufzeichnungen der Aussagen nutzen konnten. Ob nonverbale Verhaltensweisen der Aussagenden in den Urteilsprozess einbezogen wurden oder andere Faktoren zu den treffsichereren Urteilen beitrugen, wurde nicht diskutiert. In diesem Zusammenhang ist eine Metaanalyse von Sporer (2007) anzuführen, die zeigte, dass nonverbale Indikatoren bei der Entdeckung von Täuschung keine bedeutsame Rolle spielen.

Weiterhin festzustellen ist, dass die geringen Trefferquoten der Rater bei hoher Urteilssicherheit wenig überraschend sind. Die Rater wurden in dem Training lediglich in der Anwendung der kriterienorientierten Aussageanalyse intensiv geschult. Sie erhielten keine Schulung bezüglich des diagnostischen Prozesses, der der Einschätzung des Realitätsgehaltes einer Aussage zugrunde liegt. Betrachtet man die Summenscores der Realkennzeichen in den beiden Urteilsgruppen Erlebnisbegründet und Nicht erlebnisbegründet ist festzustellen, dass sich diese hochsignifikant unterscheiden. In der Gruppe der erlebnisbegründet eingeschätzten Aussagen sind diese deutlich höher als in der Gruppe der nicht erlebnisbegründeten Schilderungen. Das heißt, dass die Rater aus methodischer Sicht richtig bei der Zuordnung zu den beiden Urteilsgruppen die qualitative Ausprägung der Realkennzeichen zu Grunde gelegt haben, und dementsprechend qualitativ hochwertige Aussagen als erlebnisbegründet beurteilt haben. Bezogen auf die praktische Begutachtung wird hierdurch anschaulich eine negative Folge abgebildet, die sich aus der Verbreitung der Realkennzeichen in den Kreisen der Justiz ergibt- nämlich die Anwendung der Realkennzeichen im Sinne einer Merkmalskatalogs, der dann zur Untermauerung von Anklagen genutzt wird. Das damit einhergehende Unverständnis für den umfassenderen spezifische psychologische Kenntnisse erfordern den diagnostischen Prozess bei der

Glaubhaftigkeitsbeurteilung, bedarf der Aufklärung in den entsprechenden Berufsbereichen.

Die grundsätzliche Validität der merkmalsbezogenen Aussageanalyse konnte in dieser Untersuchung nicht allein durch das Auftreten von Realkennzeichen belegt werden, es konnte auch die Überlegenheit dieses Vorgehens gegenüber einem rein quantitativen Ansatz, nämlich der Erhebung des Umfanges einer Aussage via Wortanzahl belegt werden. Zwar zeigten die Ergebnisse, dass in den Versuchsgruppen Unterschiede in der Wortanzahl auftraten und ein signifikanter Zusammenhang mit dem Summenscore der Realkennzeichen bestand. Durch Einbeziehung der Urteilssicherheit als subjektives Gewicht im Urteilsprozess wurde deutlich, dass mittels der Realkennzeichen im Verbund mit dieser Variable erlebnisbegründete und nicht erlebnisbegründete Schilderungen identifiziert werden konnten. Aufgrund der Wortanzahl allein konnte dagegen in dieser Analyse keine Klassifizierung vorgenommen werden.

Zusammenfassend ergibt sich aus der vorliegenden Studie, dass eine Anwendung der Realkennzeichenanalyse nach Steller und Köhnken (1989) auch bei Kindern im Vorschulalter eine ausreichende Validität aufweist, wenn es um die Unterscheidung erlebnisbegründeter und nicht erlebnisbegründeter Aussagen geht. Unter suggestiven Einflüssen allerdings zeigte sich die Validität stark eingeschränkt. Das bisherige Fehlen von Untersuchungen zur Validität der Realkennzeichen bei Vorschulkindern, insbesondere unter Berücksichtigung suggestiver Bedingungen, wirft im Zusammenhang mit den Ergebnissen dieser Studie weiteren Forschungsbedarf auf. Anregungen dazu ergeben sich z.B. aus einer Untersuchung von Niehaus (2001), in der die einzelnen Realkennzeichen unter verschiedenen Wahrnehmungsbedingungen betrachtet wurden. Allerdings wurde auch in dieser Studie nicht der Altersbereich der Vorschulkinder berücksichtigt. Eine ähnlich differenzierte Untersuchung der einzelnen Kriterien in der Gruppe der Kindergarten- und Vorschulkinder unter der Berücksichtigung suggestiver Einflüsse könnte für die weitere Forschung von Interesse sein.

Ein weiterer Schluss, der aus den Ergebnissen der vorliegenden Studie gezogen werden kann, ist von Bedeutung für die gutachterliche Praxis:

Die fehlende Trennschärfe der Realkennzeichen unter suggestiven Bedingungen zeigt für die Glaubhaftigkeitsbeurteilung die zwingende Notwendigkeit, Informationen zur Aussageentstehung zu erheben, um mögliche suggestive Einflüsse aufdecken zu können. Gerade für die Gruppe der jungen Kinder, in der am ehesten suggestive Effekte zu erwarten sind, sind Informationen über die Aussagegenese in der Regel nur von Bezugspersonen der Kinder zu erhalten. Bestrebungen, den praktisch tätigen, psychologischen Sachverständigen bei der Aufdeckung der Entstehungsgeschichte zu beschränken (Nagler, 2006), indem die informatorische Befragung von Personen, die darüber Auskunft geben könnten, zur Besorgnis der Befangenheit bei Sachverständigen führt, sind daher kontraproduktiv.

Literatur

- Ackil, J.K. & Zaragoza, M.S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of Experimental Child Psychology*, 60, 57 - 83.
- Akehurst, L., Köhnken, G. & Höfer, E. (2001) Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology*, 6, 65-83.
- Angermaier, M. (1974). *Psycholinguistischer Entwicklungstest, Manual*. Weinheim: Beltz.
- Anson, D.A., Golding, S.L. & Gully, K.J. (1993). Child sexual abuse allegation: Reliability of criteria-based content analysis. *Law and Human Behavior*, 17 (3), 331-341.
- Arntzen, F. (1982) (2. Auflage). *Die Psychologie der Zeugenaussage*. Göttingen: Hogrefe.
- Arntzen, F. (1983). Psychologische Beurteilung der Glaubwürdigkeit von Zeugenaussagen. In F. Lösel (Hrsg.), *Kriminalpsychologie*. Weinheim: Beltz.
- Arntzen, F. & Michaelis, E. (1970). Psychologie der Kindervernehmung. Wiesbaden: Bundeskriminalamt.
- Bekerian, D. A. & Bowers, J. M. (1983). Eyewitness Testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 139-145.
- Binet, A. (1897). Description d'un object. *Annee Psychologique*, 3, 296-332
- Binet, A. (1900). *La Suggestibilité*. Paris: Schleicher Frères.
- Bortz, J. (1989). *Statistik für Sozialwissenschaftler* (3. Aufl.) Berlin: Springer.
- Boychuk, T.D. (1991). *Criteria-based content analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished Dissertation. Arizona State University, Tempe (AZ).
- Brickenkamp, R. (1985). *Handbuch psychologischer und pädagogischer Test*. Göttingen: Hogrefe.
- Brosius, G. (1989). *SPSS, PC+ advanced statistics und tables*. Hamburg: McGraw-Hill.

- Brosius, G. & Brosius, F. (1995). *SPSS Base System und Professional Statistics*. Bonn: International Thomson Publishing.
- Bruck, M. Hembrooke, H. & Ceci, S. (1997). Children's reports of pleasant and unpleasant events. In J.D. Read & D.S. Lindsay (Eds.) *Recollections of Trauma. Scientific Evidence and Clinical Practice* (NATO ASI Series. Series A: Life Sciences Vol. 29; pp. 199-213) New York: Plenum
- Buck, J. A., Warren, A. R., Betman, S., & Brigham, J. C. (2002). Age differences in Criteria- Based Content Analysis scores in typical child sexual abuse interviews. *Journal of Applied Developmental Psychology*, 23, 267-283.
- Burger, H. (1971). *Die suggestive Beeinflussbarkeit von Aussagen über Beobachtungen*. Unveröffentlichte Dissertation, Universität Freiburg.
- Cassel, W.S., Roebers, C.E.M. & Bjorklund, D.F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology*, 61, 116-133
- Cattell, J.M. (1895). Measurements of the accuracy of recollection. *Science*, 2, 761-766
- Ceci, S. J. (1993). *Cognitive and social factors in children's testimony*. American Psychology Association: Master Lecture.
- Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, Vol. 113(3), 403-439.
- Ceci, S. J., Huffman, M.L.C., Smith, E. & Loftus, E.W. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. *Conscious and Cognition*, 3, 388- 407.
- Ceci, S.J.; Loftus, E.; Leichtman, M.D. & Bruck, M. (1994) The possible role of source misattributions in the creation of false beliefs among preschoolers. *International Journal of Clinical and Experimental Hypnosis*, 42, 304-320.
- Ceci, S. J., Ross, D. F., Toglia, M. P. (1987). Age differences in suggestibility: narrowing the uncertainties. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's Eyewitness Memory* (79-91). New York.
- Claparède, E. (1906). La psychologie judiacaire. *Année psychologique*, 12, 295-302
- Cohen, R.L. & Harnick, M. A. (1980). The susceptibility of child witnesses to suggestion. *Law and human behaviour*, 4, 201.210.
- Craig, R. A., Scheibe, R. Raskin, D. C. Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Journal of Applied Developmental Science*, 3, 77-85.

- Eggers, J. (2002) *Glaubwürdigkeit von Zeugenaussagen – Eine Evaluation des Kieler Trainingsprogramms zur Beurteilung der Glaubwürdigkeit von Zeugenaussagen (KTBG)*. Diplomarbeit im Fach Psychologie. Universität Kiel.
- Endres, J., Scholz, O. B. & Summa, D. (1997). *Aussagesuggestibilität bei Kindern. Vorstellung eines neuen diagnostischen Verfahrens und erste Ergebnisse*. In L. Greuel, T. Fabian & M. Stadler (Hrsg.), *Psychologie der Zeugenaussage* (S. 189-204). Weinheim: Psychologie Verlags Union.
- Erdmann, K. (2001). *Induktion von Pseudoerinnerungen bei Kindern*. Regensburg: Roderer Verlag.
- Erdmann, K., Volbert, R. & Böhm, C. (2004). Children report suggested events even when interviewed in a non- suggestive manner: what are its implications for credibility assessment? *Applied Cognitive Psychology*, 18, 589- 611
- Esplin, P., Boychuk, T. & Raskin, D. (1988). *Application of statement validity analysis*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.
- Friedrichsen, G., Mauz, G. (1995a). Mißbrauch der Strafjustiz (Zum Freispruch im Montessori- Prozeß). *Der Spiegel*, Nr. 21, 76-83.
- Friedrichsen, G., Mauz, G. (1995b). Blind die Blinden angeführt (Über die Lehren aus dem Montessori Prozeß). *Der Spiegel*, Nr. 22, 82-90.
- Fürniss, T. (1993). Kinder und Familien im traumaorganisierten System von Sexringen. *Familiendynamik*, 18 (3), 264-286.
- Goodman, G. S., & Reed, R. S. (1986). Age differences in eyewitness testimony. *Law and Human Behaviour*, 10, 317-332.
- Goodman, G. S., Aman, C. & Hirschman, J. (1987). Child sexual and physical abuse: children's testimony. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's Eyewitness Memory* (1-23). New York.
- Goodman, G. S., Wilson, M. E., Hazan, C., & Reed, R. S. (1989). *Children's testimony nearly four years after an event*. Paper presented at the annual meeting of the eastern Psychological Association. Boston, MA.
- Goodman, G. S., Rudy, L., Bottoms, B. L. & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush & J. A. Hudson (Eds.), *Knowing and remembering in young children* (249- 284). New York: Cambridge University Press.
- Goodman, G.S., Hirschman, J.E., Hepps, D. & Rudy, L. (1991). Childrens memory for stressful events. *Merrill Palmer Quarterly*, 37, 109-158.

- Goodman, G.S. & Clarke-Stewart, A. (1991). Suggestibility in children's testimony. Implications for child sexual abuse investigations. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp.92-105). Washington, D.C: American Psychological Association.
- Greuel, L., (1997). Suggestibilität und Aussagezuverlässigkeit: ein (neues) Problem in der forensisch-psychologischen Praxis? In: L. Greuel, T. Fabian & M. Stadler (Eds.), *Psychologie der Zeugenaussage*. (pp 121-130). Weinheim: Beltz.
- 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: Beltz.
- Gudjonsson, G. (1986). The relationship between interrogative suggestibility and acquiescence: empirical findings and theoretical implications. *Personality and Individual Differences*, 7, 195-199.
- Höfer, E. (1995a). *Glaubwürdigkeitsdiagnostik unter differentiellen Beanspruchungsbedingungen. Eine experimentelle Studie zur Überprüfung der Sensitivität und Spezifität distaler Glaubwürdigkeitskriterien*. Kiel: Dissertationsschrift.
- Höfer, E. (1995b). *Operationalisierung der inhaltlichen Glaubwürdigkeitskriterien*. Unveröffentlichtes Manuskript, Christian-Albrecht-Universität, Kiel.
- Höfer, E. (1996). *Ablauf des Trainings in der Anwendung des CBCA*. Unveröffentlichtes Manuskript, Christian-Albrecht-Universität, Kiel.
- Honts, C.R., Peters, D.P., Devitt, M.K. & Amato, S.L. (1992). *Detecting children's lies with statement validity assessment: A pilot study of a laboratory paradigm*. Paper presented at the NATO Advanced Study Institute „The child witness in context: Cognitive, social and legal perspectives“, Il Ciocco, Italy.
- Horowitz, S. (1991). Empirical Support for statement validity assessment. *Behavioral Assessment*, 13, 293-313.
- Horowitz, S., Lamb, M., 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.
- Huffmann, M.L., Crossman, A.M. & Ceci, S.J. (1997). "Are false memories permanent?": An investigation of the long-term effects of source misattributions. *Consciousness and Cognition*, 6, 482-490
- Joffe, R. & Yuille, J.C. (1992) *Criteria-based Content Analysis: An experimental investigation*. Paper presented at the NATO Advanced Study Institute in the childwitness in legal context: Cognitive, social and legal perspectives, Lucca, Italy.

- Johnson, M.K., Hashtroudi, S. & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28
- King, M. & Yuille, J. (1987). Suggestibility and the child witness. In S. Ceci, M. Bruck & D. Ross (Eds.), *Children's eyewitness memory* (pp. 24-35). New York: Springer.
- Köhnken, G. (1987). Nachträgliche Informationen und die Erinnerung komplexer Sachverhalte - Empirische Befunde und theoretische Kontroversen. *Psychologische Rundschau*, 38, 190-203.
- Köhnken, G. (1990). *Glaubwürdigkeit: Untersuchungen zu einem psychologischen Konstrukt*. München: Psychologie Verlags Union.
- Köhnken, G. (1997) Suggestive Prozesse in Zeugenbefragungen: Formen und theoretische Erklärungsansätze. *Monatsschrift für Kriminologie und Strafrechtsreform*, 80, 290-299.
- Köhnken, G. (1999). Suggestion und Suggestibilität. In R. Lempp, G. Schütze & G. Köhnken (Hrsg), *Forensische Psychiatrie und Psychologie des Kindes- und Jugendalters* (S. 342- 362). Darmstadt: Steinkopff.
- Köhnken, G. & Wegener, H. (1982). Zur Glaubwürdigkeit von Zeugenaussagen. Experimentelle Überprüfung ausgewählter Glaubwürdigkeitskriterien. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 29, 92-111.
- Köhnken, G. & Wegener, H. (1985). Zum Stellenwert des Experiments in der Forensischen Aussagepsychologie. *Zeitschrift für experimentelle und angewandte Psychologie*, 32, 104- 119.
- Köhnken, G., Schimossek, E., Aschermann, E. & Höfer; E. (1995). The cognitive Interview and the assessment of the credibility of adult's statements. *Journal of Applied Psychology*, 80, 671-684.
- Krahé, B. & Kundrotas, S. (1992). Glaubwürdigkeitsbeurteilung bei Vergewaltigungsanzeigen: Ein aussagepsychologisches Feldexperiment. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 39(4), 598-620.
- Lamb, M. E., Sternberg, K. J., Esplin, P. W. (1994). Factors influencing the reliability and validity of statements made by young victims of sexual maltreatment. *Journal of Applied Developmental Psychology*, 15, 225-280
- 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 and Neglect*, 21, 3, 255- 264.

- Lamers-Winkelmann, F. (1995). Kinder und Justiz oder warum es nicht so einfach ist. In: Landesregierung Schleswig Holstein (Hrsg.), *Neue Wege zum Opferschutz nach sexueller Mißhandlung* (pp. 12-19). Kiel: A.C.Ehlers.
- Lamers-Winkelmann, F., Buffing, F., Van der Zanden, A.P. (1992a). *What children can or will tell about sexual abuse: preliminary results*. Poster presented at the NATO Advanced Study Institute: The child witness in context: cognitive, social and legal perspectives, Lucca, Italy.
- Lamers-Winkelmann, F., Buffing; F., Van der Zanden, A.P. (1992b). *Statement validity analysis in child sexual abuse cases: a field study*. Poster presented at the Third European Conference of Law and Psychology, Oxford, United Kingdom.
- Lampinen, J.M. & Smith, V. (1995). The Incredible (am Sometimes Incredulous) Child Witness: Child Eyewitnesses`Sensitivity to Source Credibility Cues. *Journal of Applied Psychology*, 80, 621-627.
- 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.
- Lehr, S. (1995) *Qualitativer Vergleich von instruierten, suggerierten und erlebnisbegründeten Aussagen*. Unveröffentlichte Diplomarbeit an der Freien Universität Berlin.
- Lepore, S. J. & Sesco, B. (1994). Distorting children`s reports and interpretations of events through suggestion. *Journal of Applied Psychology*, 79, 108-120.
- Leichtman, M. D. & Ceci, S.J. (1995) The effects of stereotypes and suggestions on preschoolers` reports. *Developmental Psychology*, 31, 568-578.
- Lipmann, O. (1908). *Die Wirkung von Suggestivfragen*. Leipzig: J.A. Barth.
- Littmann, E. (1992). Psychische Schäden nach sexuellem Kindesmißbrauch. *psychomed*, 4, 13-19.
- Littmann, E. & Szewczyk, H. (1983). Zu einigen Kriterien und Ergebnissen forensisch-psychologischer Glaubwürdigkeitsbegutachtung von sexuell mißbrauchten Kindern und Jugendlichen. *Forensia*, 4, 55-72.
- Loftus, E.F. & Hoffman, H.G. (1989) Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, 118, 100-104
- Loftus, E. F. & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, 35, 409-420.

- Loftus, E. F., Miller, D. G. & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, Vol 4(1), 19-31.
- Marin, B.V., Holmes, D. L., Guth, M. & Kovac, P. (1979). The potential of children as eyewitnesses. *Law & Human Behavior*, 3, 295-305.
- Mc Brien, C.M. & Dagenbach, D. (1998). The contributions of source misattribution, acquiescence, and response bias to children's false memories. *American Journal of Psychology*, 111, 509-528
- McCloskey, M. & Zaragoza, M. S. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, Vol. 114(1), 1-16.
- Memon, A., Cronin, O., Eaves, R., Bull, R. & Küpper, B. (1992). The cognitive Interview & Child Witness. Cognitive, social and legal perspectives. *Paper presented at the NATO Advanced Study Institute*. Lucca, Italy.
- Michaelis-Arntzen, E. (1987). Unglaubwürdige Zeugenaussagen. *Forensia*, 8(2), 73-80.
- Michaelis-Arntzen, E. (1997). Zur Suggestibilität von Kleinkindern. In L. Greuel, T. Faßbian & M. Stadler (Hrsg.) *Psychologie der Zeugenaussage* (S. 205-210). Weinheim: Psychologie Verlags Union.
- Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. New York: Clak Boardman.
- Nagel, A. (2006). Besorgnis der Befangenheit einer Sachverständigen wegen "Informatorischer Befragung" von Zeugen, §§74 Abs. 1, 80 Abs. 1. *Praxis der Rechtspsychologie*, 16 (1/2), 307- 308
- Niehaus, S. (2001). *Zur Anwendbarkeit inhaltlicher Glaubhaftigkeitsmerkmale bei Zeugenaussagen unterschiedlichen Wahrheitsgehaltes*. Frankfurt: Peter Lang
- Niehaus, S. (2003). Diskriminationsfähigkeit der merkmalsorientierten Inhaltsanalyse bei teilweise erlebnisbasierten Falschaussagen. *Praxis der Rechtspsychologie*, 13 (2), 309- 328
- Oerter, R., & Montada, L., (2002). *Entwicklungspsychologie*. Weinheim: .Psychologie Verlags Union.
- Ornstein, P. A., Gordon, B. N. & Larus, D. M. (1992). Children's memories for an personally experienced event: Implications für testimony. *Applied Cognitive Psychology*, 6, 49-60

- Peters, D. (1987). The impact of naturally occurring stress on children's memory. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (122-141). New York: Springer.
- Peters, D. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.) *The suggestibility of children's recollections* (pp. 60-76). Washington, DC: American Psychology Association.
- Peters, K. (1972). *Fehlerquellen im Strafprozess*, Band 2. Karlsruhe: Müller.
- Petersen, R. (1997). *Konzeption und Evaluation der Reliabilität des Kieler Trainingsprogrammes zur Beurteilung der Glaubwürdigkeit von Zeugenaussagen (KTBG)*. Diplomarbeit im Fach Psychologie. Universität Kiel.
- Pieters, V. (1994). *Beeinflussungen von Kinderaussagen durch Suggestionen und Interaktionen*. Unveröffentlichte Diplomarbeit an der Freien Universität Berlin.
- Poole, D.A. & Lindsay, D.S. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching, and leading questions on reports of nonexperienced events. *Journal of Experimental Child Psychology*, 60, 129-154.
- Poole, D. A. & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, 27, 975-986.
- Poole, D.A. & White, L.T. (1995). Tell me again and again: Stability and change in the repeated testimonies of children and adults. In M.S. Zaragoza, J.R. Graham, G.C.N. Hall, R. Hirschman & Y.S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp.24-43). Thousand Oaks: Sage.
- Raskin, D.C. & Esplin, P. W. (1991). Assessment of children's statements of sexual abuse. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 153-164). Washington, D.C.: American Psychological Association.
- Rudy, L. & Goodman, G.S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology*, 27, 527-538.
- Rüth-Bemelmans, E. (1984) Experimentelle Erprobung der Kriterien zur Aussageanalyse. Unveröffentlichte Diplomarbeit, Universität Köln.
- Santtila, P., Roppola, H., Runtti, M. & Niemi, P. (2000). Assessment of child witness statements using criteria-based content analysis (CBCA): The effect of age, verbal ability, and interviewers emotional style. *Psychology, Crime and Law*, 6, 159-179.
- Sarris, V. (1992). *Methodologische Grundlagen der Experimentalpsychologie, Bd. 2: Versuchsplanung und Studien*. München: E. Reinhardt.

- Schelle, H. (1996). Nachträgliche irreführende Informationen in Kinderaussagen und ihre Wirkung auf inhaltliche Realkennzeichen. Unveröffentlichte Diplomarbeit. Universität Kiel.
- Sporer, S.L. (1997). Realitätsüberwachungskriterien und forensische Glaubwürdigkeitskriterien im Vergleich: Validitätsüberprüfungen anhand selbsterlebter und erfundener Geschichten. In L. Greuel, T. Fabian & M. Stadler (Hrsg.), *Psychologie der Zeugenaussage*. (S. 71-85). Weinheim: Beltz. Psychologie Verlags Union.
- Sporer, S.L. & Küpper, B. (1995). Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie. *Zeitschrift für Sozialpsychologie*, 26, 173-193.
- Steinbrink, H., Schrader, H. & Caspers, W. (1993). *35 Jahre Polizeipuppenspiel: Polizeipuppentage 4. bis 8. Oktober 1993 (Festschrift)*. Polizeidirektion Schleswig - Holstein Nord (Hrsg.), Flensburg.
- Steller, M. (1988). Die vierte Phase der Aussagepsychologie. *Forensia*, 9, 23-28.
- Steller, M. (1989). Recent developments in statement analysis. In J.C. Yuille (Ed.), *Credibility Assessment* (pp. 135-154) London: Kluwer Academic Publishers.
- Steller, M. (1998). Aussagepsychologie vor Gericht – Methodik und Probleme von Glaubwürdigkeitsgutachten mit Hinweisen auf die Wormser Missbrauchsproze se. *Recht & Psychiatrie*, 16 (1), 11-18.
- Steller, M. & Köhnken, G. (1989). Statement analysis: credibility assessment of children's testimonies in sexual abuse cases. In D.C. Raskin (Eds.), *Psychological methods in criminal investigation and evidence* (217-245). New York: Springer.
- Steller, M. & Volbert, R. (1997). Glaubwürdigkeitsbegutachtung. In M. Steller & R. Volbert (Hrsg.), *Psychologie im Strafverfahren* (S. 12-239). Bern: Huber.
- Steller, M., Volbert, R. & Wellershaus, P. (1993). Zur Beurteilung von Zeugenaussagen: Aussagepsychologische Konstrukte und methodische Strategien. In Montada, L. (Hrsg.), *Bericht über den 38. Kongreß der Deutschen Gesellschaft für Psychologie in Trier 1992*, Göttingen.
- Steller, M., Wellershaus, P. & Wolf, T. (1992). Realkennzeichen in Kinderaussagen: Empirische Grundlagen der Kriterienorientierten Aussagenanalyse. *Zeitschrift für experimentelle und angewandte Psychologie*, 39(1), 151 - 170.
- Stern, W. (1902). Zur Psychologie der Aussage. *Zeitschrift für die gesamte Strafrechswissenschaft*, 22, 315-370.
- Stern, W. (1904). Die Aussage als geistige Leistung und als Verhörsprodukt. *Beiträge zur Psychologie der Aussage*, 1. (S.269-326). Leipzig: Barth Verlag.

- Szewczyk, H. (1973). Kriterien der Beurteilung kindlicher Zeugenaussagen. *Probleme und Ergebnisse der Psychologie*, 46, 47-66.
- Toglia, M. (1991). Commentary: memory impairment – it is more common than you think. In J. L. Doris (Hrg.) *The Suggestibility of Children's Recollections*. (s. 40-46). Washington, DC.: American Psychological Association.
- Trankell, A. (1971). *Der Realitätsgehalt von Zeugenaussagen*. Göttingen: Hogrefe.
- Tully, B. (1999). Statement validation. In D. Canter & L. Alison (Hrsg.). *Interviewing and deception* (S. 83-104). Aldershot: Darmouth.
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In: U. Undeutsch (Hrsg.), *Handbuch der Psychologie Vol.11: Forensische Psychologie*. Göttingen: Hogrefe.
- Undeutsch, U. (1993). Die aussagepsychologische Realitätsprüfung bei Behauptung sexuellen Mißbrauchs. In S. Kraheck-Brägelmann (Hrsg.) *Die Anhörung von Kindern als Opfer sexuellen Mißbrauchs* (S. 69-162). Rostock : Hanseatischer Fachverlag für Wirtschaft.
- Varendonck, J. (1911). Les temoigagnes d'enfants dans un proces retentissant. *Archives de Psychologie*, 11, 129-171.
- Volbert, R. (1992). Sexueller Mißbrauch von Kindern - Empirische Befunde und psychosoziale Trends. *Psychomed*, 4, 8-12.
- Volbert, R. (1997) Suggestibilität kindlicher Zeugen. In: M. Steller & R. Volbert (Eds), *Psychologie im Strafverfahren*. (pp. 40-62) Bern: Huber.
- Volbert, R. (2003). Suggestion. In Kubinger K.,D. & Jäger R.S. (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S.395-398). Weinheim: Beltz.
- Volbert, R. (2004) Beurteilung von Aussagen über Traumata. Bern: Huber.
- Volbert, R. & Pieters, V. (1996). Suggestive Beeinflussungen von Kinderaussagen. *Psychologische Rundschau*, 47, 183-198.
- Volbert, R. & Steller, M. (1998). Aussagefähigkeit von Kindern. Entwicklungspsychologische Aspekte der forensischen Aussagepsychologie. In H.-L. Kröber & K.-P. Dahle (Hrsg.), *Sexualstraftaten und Gewaltdelinquenz. Verlauf – Behandlung – Opferschutz* (S. 235-357). Heidelberg: Kriminalistik. Wissenschaft und Praxis.
- Vrij, A. (2002). Deception in children: A literature review and implications for children's testimony. In H. Westcott, G. Davies & R. Bull (Eds), *Children's testimony: A handbook for psychological research and forensic practice* (pp. 175-194). Chichester, Great Britain: Wiley.

- 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. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and human behaviour*, 26 (3), 261-283
- Vrij, A. & Bush, N. (2000). Differences in suggestibility between 5-6 and 10-11 year olds: The relation with self- confidence. *Psychology, Crime and Law*, 6, 127- 138
- Warren, A.R., Hulse-Trotter, K. & Tubss, E. (1991) Inducing Resistance to Suggestibility in Children. *Law and Human Behavior*, 15, 273-285.
- Warren, A.R. & Lane, P. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M.S. Zaragoza, J.R. Graham, G.C.N. Hall; R.Hirschman & Y.S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp.44-60). Thousand Oaks: Sage.
- Wegener, H. (1981). *Einführung in die forensische Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Wegener, H. (1992). Psychologische Aspekte der Einvernahme. *Aktuelle Juristische Praxis*. 9/92, 1115-1119.
- Wegener, H. (1997). Die Entwicklung der experimentellen Aussagepsychologie. In: L. Greuel, T. Fabian & M. Stadler (Eds.), *Psychologie der Zeugenaussage*. (pp 13-21). Weinheim: Beltz.
- Wellershaus, P. & Wolf, T. (1989). *Kriterienorientierte Aussagenanalyse: Empirische Überprüfung der diagnostischen Leistungsfähigkeit inhaltlicher Realkennzeichen zur Trennung wahrer von unwahren Kinderaussagen*. Unveröffentlichte Diplomarbeit Institut für Psychologie an der Christian-Albrecht-Universität, Kiel.
- Wells, G. L. & Loftus, E. F. (1991). Commentary: Is this child fabricating? Reactions to a new assessment technique. In J. Doris (ed.), *The suggestibility of children's recollections* (pp 168-171). Washington, DC: American Psychological Association.
- 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 (Eds.), *Psychologie der Zeugenaussage*. (S. 121-130). Weinheim: Beltz Psychologie Verlags Union.
- Wertheimer, M. (1906) Experimentelle Untersuchungen zur Tatbestandsdiagnostik. *Archiv für die gesamt Psychologie* 6, 59-131.
- Yuille, J.C. (1988) The systematic assessment of childrens` s testimony. *Canadian Psychology*, 29,247-262

Zaragoza, M.S. (1987). Memory, suggestibility, and eyewitness testimony in children and adults. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's Eyewitness Memory* (53-78). New York.

Ziler, H. (1975). *Der Mann-Zeichen-Test in detail-statistischer Auswertung* (5. Aufl.). Münster: Aschendorff.

Anhang

Anhang A: Vorleseinstruktionen an die Eltern

Liebe Eltern,

Vielen Dank, dass Sie unsere Untersuchung durch Ihre Mithilfe unterstützen. Mit diesem Brief bekommen Sie die Vorlesegeschichte für Ihr Kind.

Bitte lesen Sie Ihrem Kind die Geschichte an den nächsten drei Wochenenden jeweils einmal vor (9./10., 16./17. und 23./24. März). Sie brauchen für das Vorlesen einer Geschichte etwa 15 Minuten. Sie können die Geschichte aber auch in zwei oder mehr Teilen vorlesen, falls es Ihnen oder Ihrem Kind zu lange dauert. Wichtig ist nur, dass die Geschichte an einem Wochenende einmal ganz vorgelesen wird. Sie können sie als Gute-Nacht-Geschichte vorlesen oder zu einer anderen Tageszeit.

Ihr Kind hat diese Geschichte bereits im Kindergarten von Christel vorgelesen bekommen, kennt also den Inhalt schon etwas und wird ihn im Laufe der Zeit immer besser kennen lernen. Lassen sie sich davon nicht irritieren. Lesen sie bitte trotzdem vor. Während der Woche werden die Kinder die Geschichte auch mehrmals im Kindergarten hören. Von Montag dem 25.3. bis Mittwoch, den 27.3. werden die Kinder dann ins „Kinderbüro“ eingeladen und befragt.

Die Geschichte ist zum Teil so geschrieben, als ob Ihr Kind das Kaspertheater schon gesehen hätte. Sollte Ihr Kind an diesen Stellen Einwände haben, reagieren Sie "cool", indem sie nur kurz darauf eingehen und z.B. sagen: "Ja, ich weiß, aber das ist eben eine Geschichte." Sprechen Sie möglichst nicht darüber, dass Ihr Kind tatsächlich bald das Kaspertheater sehen wird. Lesen Sie einfach nur vor.

Wir haben für Sie ein Protokollblatt vorbereitet. Hier sollen Sie jeweils nach dem Vorlesen eintragen, wann Sie dem Kind vorgelesen haben, welchen Teil Sie vorgelesen haben und, falls Ihnen bei Ihrem Kind etwas aufgefallen ist, wie es reagiert hat (z.B.: hat aufmerksam zugehört, ist unruhig gewesen, hat sich gelangweilt, ist eingeschlafen, wollte nicht mehr zuhören, usw.). Bitte geben Sie den Protokollbogen nach dem dritten Wochenende im Kindergarten ab. Damit alle Kinder die gleichen Voraussetzungen für die Befragung haben, ist es wichtig, dass Sie die Geschichte an den drei Wochenenden jeweils einmal vorlesen. Sollte Ihnen dieses nicht möglich sein, bitte wir Sie, es im Protokoll zu notieren, damit wir es in der späteren Auswertung berücksichtigen können.

Sollten sich Fragen ergeben, können Sie uns gerne anrufen, auch am Wochenende.

Wir wünschen Ihnen viel Spaß beim Vorlesen.

Anhang B: Vorlesegeschichte für die Gruppen B und C

Luzies Weg durch's Entental

Es war einmal ein Kindergarten. Und in diesen Kindergarten sollte ein Kaspertheater kommen. Alle Kinder waren schon sehr aufgeregt und eines Morgens war es dann so weit. Im Kindergarten stand ein großes Kaspertheater mit roten, blauen und gelben Vorhängen und davor durften sich die Kinder setzen. Als alle Kinder auf ihren Plätzen saßen und mucksmäuschenstill waren, kam ein Polizist und begrüßte die Kinder. Plötzlich tauchte auch eine große Ente auf und sprach mit dem Polizisten. Dann gingen beide weg und das Kaspertheater beginnt nun endlich.

Die Kinder sind ganz gespannt und sie sehen, wie das Licht angeht und ein Hase angehoppelt kommt. Das ist der Hase Eberhard. Der mag schrecklich gerne Kohl, und am liebsten mag er den Kohl auf dem Kohlfeld von Bauer Otto. Er schleicht sich an einen großen grünen Kohlkopf heran und beißt hinein. „Aua, Aua!“ schreit es plötzlich mitten aus dem Kohlkopf, und es erscheint ein langer, orange-roter Wurm: „Du hast was von meinem Schwanz abgebissen!“ jammert er. „Oh, das tut mir aber leid!“ sagt der Hase zu dem Wurm, der Fridolin heißt. Dann erzählt er Fridolin, um ihn von seinen Schmerzen abzulenken, dass der Bauer ihn gestern beim Kohlklauen fast erwischt hätte, aber er, der schnelle Hase Eberhard, war einfach blitzschnell über die Straße gelaufen. Der Hase will gerade noch mehr von dem schönen Kohl in Bauer Ottos Garten erzählen, als sich jemand mit einer langen, roten Zipfelmütze auf dem Kopf nähert. Die Kinder im Kindergarten wissen sofort, wer das ist und rufen laut: „Kasper, Kasper!“ Davon bekommt der Hase Eberhard einen großen Schrecken und er hoppelt schnell davon und versteckt sich. Der Kasper mit der roten Zipfelmütze und dem roten Mantel trägt eine Angel. „Guten Tag, liebe Kinder! Guten Tag, Christel und! Ich will gleich angeln gehen!“ sagt er zu den Kindern im Kindergarten.

Plötzlich spritzt Wasser hoch und Kasper und die Kinder bekommen einige Spritzer ab. Sie wissen aber nicht, woher das Wasser kommt.

Kasper sucht einen Wurm für seine Angel, denn wenn man einen Fisch angeln will, braucht man einen Wurm. Und Kasper fragt Fridolin, der gerade auftaucht, ob er wohl an seiner Angel hängen wolle. „Aber nein!“ sagt Fridolin und zuckt zurück. „Auf keinen

Fall, das tut ja weh, an einem Haken aufgespießt zu werden!“ So nimmt Kasper ein Stück altes Brot, befestigt es an der Angel und hängt die Angel ins Wasser.

Plötzlich ruckt und zuckt es an der Angel. Kasper zieht aufgeregt die Schnur hoch und was hängt dran? Kein Fisch, sondern eine Cola Dose und Müll. Das ärgert Kasper sehr und er schimpft auf die Leute, die ihren Müll überall hinwerfen. „Genau, genau!“ quäkt da Lucy, die Ente, die die Kinder am Anfang schon einmal gesehen haben. „In unserem Teich liegen sogar alte Schuhe!“ Kasper fragt die Kindergartenkinder, wo sie denn ihren Müll lassen. Und was antworten diese? Richtig: „Der Müll kommt in den Mülleimer!“

Aber Lucy ist aus einem anderen Grund sehr aufgeregt: ihre Entenküken sind verschwunden. Nach dem Frühstück sollten die Küken in den Kindergarten gehen. Dort sind sie aber nicht angekommen und Lucy weiß nun nicht, wo sie geblieben sind. Sie macht sich große Sorgen. „Wir suchen sie gemeinsam, wir finden deine Küken!“ verspricht Kasper der Ente. Lucy ist sehr froh, dass Kasper ihr helfen will und beide machen sich auf den Weg. Da hören die Kinder plötzlich Bremsen quietschen, ein lautes Krachen und Bauer Otto schimpft laut: „Rudi, du verflixter Bengel!“ Rudi hat auf der Straße vor dem Bäckerladen Ball gespielt und Bauer Otto konnte in seinem Trecker gerade noch das Steuer herumdrehen, sonst hätte er Rudi überfahren. Aber leider ist er dafür in das Fenster vom Bäckerladen gefahren und das ist nun zerbrochen und auch der Trecker fährt nicht mehr. Nun sehen die Kindergartenkinder den Hasen Eberhard und ein Küken mit einem roten Schal, das heißt Zippel-Zappel. Eberhard erzählt dem Küken, das Mutter Lucy und Kasper es suchen. „Ich war mit meinen Geschwistern im Teich, da haben wir geplantscht und getobt!“ erzählt das Küken. „Und dann ist mir langweilig geworden und ich bin zur Straße gegangen! Dort habe ich eine grüne Kiste gesehen, mit drei bunten Lichtern drin, was das wohl ist?“ Der Hase Eberhard ist nun ganz stolz, denn er weiß, was Zippel-Zappel gesehen hat: Eine Ampel! Und er ruft dem Küken und den Kindergartenkindern zu: „Ich kenne ein schönes Ampelgedicht: Bei Rot kann man gehn, bei Grün bleibt man!“ Da schreien alle Kindergartenkinder ganz laut, denn sie wissen, dass das nicht richtig ist. Denn es heißt natürlich: „Bei Rot bleibt man stehn, bei Grün kann man gehn!“ Zippel-Zappel will nun wieder baden gehen, Eberhard will in die Sauna gehen und beide gehen weg. Da sehen die Kindergartenkinder Lucy und Kasper wiederkommen. Sie haben die Kükenkinder noch nicht gefunden und sind ganz traurig. Aber die

Kindergartenkinder rufen, dass sie das Küken Zippel-Zappel gesehen haben und dass es mit dem Hasen weggegangen ist. Und die Kinder erzählen Kasper und Lucy auch, dass der Hase was Falsches über die Ampel gesagt hat. „Das ist ja gemein“, ruft Lucy. „Und gefährlich“, sagt Kasper. „Wir müssen den Hasen fangen und ihm sagen, dass er den Kindern nichts Falsches erzählen darf.“ Die Kindergartenkinder haben eine gute Idee. Sie erzählen Kasper, dass Hasen gerne Möhren mögen und Kasper ihn damit heranlocken kann. Lucy holt schnell zwei große, dicke Möhren, sie legen sie hin und Kasper und Lucy verstecken sich. Das ist schwierig, denn Lucy ist ganz schön groß und hat kaum Platz hinter dem Vorhang. Aber die Kinder helfen ihr, bis keine Schnabelspitze mehr zu sehen ist. Es wird dunkel im Kaspertheater und die Kindergartenkinder sind wieder mucksmäuschenstill. Da erscheint schon der Hase Eberhard. „Mmh!“ schnuppert er, „riecht das gut!“ Er sieht die großen, roten Möhren, beißt hinein und im Nu hat er eine verspeist. Schon will er die zweite anknabbern, da tauchen plötzlich Lucy und Kasper auf und halten ihn ganz fest. Die Kinder klatschen und freuen sich, dass die Falle gelungen ist. Kasper sagt dem Hasen sehr ernst, dass er den Kindern nichts Falsches über Ampeln und Straßen erzählen darf und dass Lucy ihn zur Polizei bringt, damit ihm die Polizisten einmal genau erzählen, worauf man im Straßenverkehr achten muss. Plötzlich sehen die Kinder fünf Küken im Teich auf und nieder tanzen und Kasper ruft laut nach Lucy. Lucy kommt schnell angerannt und sieht, dass es ihre Küken sind. Aber eines, nämlich Zippel-Zappel mit dem roten Schal, fehlt und Lucy ist ganz erschrocken. Doch plötzlich taucht auch Zippel-Zappel im Teich auf und tanzt mit den anderen. Er war schon wieder an der Straße, obwohl es verboten ist. Aber Mutter Lucy ist froh, dass er wieder da ist. Sie hatte große Angst gehabt, dass ihren Küken etwas im Straßenverkehr passiert war. Die Kindergartenkinder sind auch froh, dass alles gut ausgegangen ist und klatschen fröhlich in die Hände. Dann kommt noch einmal der Polizist und erzählt ihnen, dass sie im Straßenverkehr immer gut aufpassen müssen. Und dann ist das Kaspertheater zu Ende und die Kinder gehen nach Hause und erzählen, was sie an diesem Tag im Kindergarten erlebt haben.

Anhang C: Fragebogen für die Rater

Vpn.-Nr.: Rater: Bearbeitungs-Reihenfolge-Nr.: .

Die Merkmale 1-3 (Teil A) beziehen sich auf die Aussage in ihrer Gesamtheit, die Merkmale 4-15 (Teil B) beziehen sich auf einzelne Inhalte.

A: Allgemeine Merkmale

Bitte ankreuzen:

1. Logische Konsistenz und Widerspruchsfreiheit	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

1.a Ist die Aussage in sich stimmig und folgerichtig?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

1.b Enthält die Aussage Widersprüche?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---------------------------------------	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

1.c Wie bedeutsam sind die Widersprüche für das Geschehen?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2. Widerspruchsfreie ungeordnete Darstellung der Aussage	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Berücksichtige die sprachlichen Fähigkeiten der aussagenden Kinder. Undeutliche Ausdrucksweise ist nicht mit Inkontinenz gleichzusetzen.

2.a Wird der Handlungsverlauf ungeordnet und sprunghaft dargestellt?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2.b Sind Bruchstücke eines Sachverhalts über die gesamte Aussage verstreut?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2.c Werden bei der Schilderung des Handlungsverlaufs häufiger zeitliche Rückgriffe gemacht?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2.d Wird der Handlungsablauf aus verschiedenen Blickwinkeln geschildert?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2.e Werden Teile des Geschehens zunächst nur gestreift, später aber präzisiert?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

2.f Bleibt die Darstellung lange Zeit unverständlich?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

3. Quantitativer Detailreichtum	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---------------------------------	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Achte bitte auf Wiederholungen! Nur neue Details sollen gewertet werden.

3.a Werden zahlreiche Einzelheiten geschildert?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

3.b Sind Ortsangaben vorhanden?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

3.c Erfolgen Personenbeschreibungen?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

3.d Werden Handlungen Schritt für Schritt wiedergegeben?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

3.e Wie präzise werden die einzelnen Handlungsteile beschrieben?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

3.f Werden Gespräche reproduziert?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

3.g Werden nebensächliche Einzelheiten erwähnt?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

B: Spezielle Merkmale

4. Kontextuelle Einbettung

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

4.a Werden Verknüpfungen mit persönlichen Lebensumständen vorgenommen?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

4.b Wird das Geschehen (innerhalb/außerhalb der Geschichte) in einem räumlichen Kontext geschildert?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

4.c Wird das Geschehen in einem zeitlichen Kontext geschildert?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

5. Beschreibung von nonverbalen Interaktionen

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

Bitte berücksichtige bei deiner Bewertung, daß eine nonverbale Interaktion nur dann vorliegt, wenn erkennbar wird, welche Handlungen von welcher Person ausgeführt werden.

5.a Wie häufig werden nonverbale Interaktionen geschildert?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

5.b Wie gut wird erkennbar, welche Handlung von welcher Person ausgeführt wird?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

5.c Werden regelrechte Handlungsketten (Aktion, Reaktion, Rereaktion, etc.) geschildert?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

5.d Wie nachvollziehbar (gut vorstellbar) werden diese Handlungen beschrieben?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

5.e Wird die räumliche Position der Personen untereinander erkennbar?

nicht vorhanden	schwach vorhanden	ehler weniger vorhanden	ehler mehr vorhanden	stark vorhanden	sehr stark vorhanden
0	1	2	3	4	5

<u>6. Wiedergabe von Gesprächen</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-------------------------------------	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Bitte denke daran, daß eine Schilderung innerhalb dieser Kategorie nur dann gerechtfertigt ist, wenn erkennbar ist, welche Inhalte von welcher Person geäußert werden. Die Schilderung von Dialogen kann sowohl in direkter als auch in indirekter Rede erfolgen.

6.a Wie häufig werden Gespräche wiedergegeben?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

6.b Werden die Gespräche wortwörtlich zitiert bzw. Gesprächsinhalte in direkter Rede wiedergegeben?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

6.c Wie ausführlich und präzise werden die Gesprächsinhalte wiedergegeben?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

6.d Verwendet die aussagende Person für sie unangemessenes Vokabular?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

6.e Werden regelrechte Dialogketten wiedergegeben? (A sagt, dann sagt B, worauf A sagt, etc.)	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

<u>7. Komplikationen im Handlungsverlauf</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Bitte denke daran, daß ein Vorgang an sich noch keine Komplikation im Handlungsverlauf darstellt, selbst dann, wenn der Beginn des Geschehens so beschrieben wird, als sei er plötzlich und überraschend.

7.a Werden gestörte und/oder abgebrochene Handlungsabläufe geschildert?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

7.b Berichtet die Person von plötzlich auftauchenden Hindernissen?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

7.c Werden unerwartete Vorgänge (bspw. das unerwartete Auftauchen einer dritten Person) geschildert?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

7.d Werden Äußerungen vorgebracht, die nicht den erwünschten Erfolg zeigen?	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

<u>8. Ausgefallene Einzelheiten</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-------------------------------------	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Werden außergewöhnliche, originelle Details geschildert?

<u>9. Nebensächliche Einzelheiten</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---------------------------------------	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Kommen Details vor, die für den Handlungsgang selbst nicht zwingend notwendig sind?

<u>10. Phänomengemäße Schilderung unverstandener Handlungselemente</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Werden von der aussagenden Person Geschehensabläufe oder Gegebenheiten geschildert, die sie selber offensichtlich nicht einordnen kann oder nicht verstanden hat, die ihr unklar sind, von deren Wahrnehmung sie aber überzeugt ist?

<u>11. Inhaltliche Verschachtelungen</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Werden verschiedene Handlungen berichtet, die ineinander verschachtelt sind?

<u>12. Schilderung fremdpsychischer Vorgänge</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

12.a Werden fremdpsychische Vorgänge berichtet?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

12.b Wie konkret werden diese fremdpsychischen Vorgänge geschildert?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

12.c Wird von Veränderungen in der Gefühlslage berichtet?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

12.d Wird auch von ungewöhnlichen Gefühlsregungen berichtet?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

<u>13. Schilderung eigenpsychischer Vorgänge</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

<u>14. Spontane Verbesserungen der eigenen Aussage</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Denke bitte daran, daß sich spontane Verbesserungen der Aussage nicht auf grammatischale, sondern inhaltliche Aspekte der Aussage beziehen. Spontan meint hier nicht, daß die Verbesserungen augenblicklich erfolgen müssen! Spontane Verbesserungen sind auch dann zu registrieren, wenn sie aus eigener Initiative im späteren Verlauf der Aussage gemacht werden. Verbesserung ist nicht mit korrigieren gleichzusetzen. Präzisierungen, d.h. die „Anreicherung“ bereits erwähnter Sachverhalte sollten daher auch als „spontane Verbesserungen“ registriert werden!

<u>15. Zugeben von Erinnerungslücken</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
--	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

15.a Werden Erinnerungslücken in Bezug auf einzelne Details zugegeben?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

15.b Werden Erinnerungslücken in Bezug auf ganze Handlungen zugegeben?

nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

<u>16. Unsicherheiten bei der Darstellung</u>	nicht vorhanden 0	schwach vorhanden 1	ehler weniger vorhanden 2	ehler mehr vorhanden 3	stark vorhanden 4	sehr stark vorhanden 5
---	-----------------------------	-------------------------------	-------------------------------------	----------------------------------	-----------------------------	----------------------------------

Es soll beurteilt werden, ob die aussagende Person sich hinsichtlich der von ihr geschilderten Handlung unsicher zeigt, ob sie wirklich so stattgefunden hat.

Bitte beantworte abschließend folgende Fragen:

(Bitte ankreuzen)

1. Das Kind hat das Kasperstück:

2. Wie sicher bist du in deinem Urteil?

3. Falls du meinst, das Kind habe das Kasperstück nicht gesehen, versuche bitte kurz zu begründen, warum.

.....

.....

.....

.....

.....

.....

.....

.....

4. Fiel etwas am Inhalt auf?

Anhang D: Liste der Interviewfragen

Fragenliste:

1. Wie sah das Kaspertheater aus?
2. War da nur ein Kasper oder auch andere?
3. War da ein Hase?
4. Was hat der gemacht?
5. War da ein Wurm?
6. Was war mit dem Wurm?
7. Was hat der Kasper gemacht?
8. War was mit einer Angel?
9. Was hing an der Angel?
10. War was mit naßspritzen?
11. War da eine Ente?
12. Waren da Entenküken?
13. Was haben die gemacht?
14. War jemand verschwunden/weg?
15. Sind die wiedergekommen?
16. Hat jemand gesucht?
17. Wer hat sie gefunden?
18. War was mit einem Trecker?
19. Hat der Hase was falsch gemacht?
20. Haben Kasper und Luzy was mit dem Hasen gemacht?
21. Fällt Dir noch irgendetwas ein vom Kaspertheater?

Lebenslauf: Petra Hänert, geb. Hackmann

Persönliche Daten

22.01.58 Geburt in Kiel
Eltern: Helmuth und Hildegard Hackmann, geb. Baedt
Staatsangehörigkeit: deutsch
Familienstand: verheiratet seit 1982 mit Dr. med. Wolf Hänert
drei Kinder: 1983, 1992, 1994
Wohnort: Kiel

Schulischer Werdegang / Ausbildung / Berufstätigkeit

1965 - 1969	Grundschule, Matthias-Claudius-Schule, Kiel
1966 - 1969	Umzug und Besuch der Schule am Sonderburger Platz, Kiel
1969 - 1976	Ernst-Barlach-Gymnasium, Kiel
12.05.76	Allgemeine Hochschulreife
1.08.76 - 1.08.77	Ausbildung in der Kommunalverwaltung Kiel mit ausbildungsbegleitendem Besuch der Kaufmännischen Berufsschule Kiel
1.08.77 - 30.11.78	Ernennung zur Stadtinspektorenanwärterin, Übernahme in das Beamtenverhältnis auf Widerruf und Ausbildung an der Verwaltungsfachhochschule Kiel-Altenholz
30.11.78	Abbruch der Ausbildung und Beendigung des Beamtenverhältnisses
10.11.78 - 26.06.85	Studium Psychologie
26.06.85	Diplom-Hauptprüfung für Psychologie Thema der Diplomarbeit: „Der Einfluß der Variablen Extraversion und Elektrodermale Labilität auf die Validität des Tatwissentests“
15.10.85 - 15.3.86	Wissenschaftliche Zeitangestellte am Institut für Psychologie der Universität Kiel
1986	Beginn der freiberuflichen forensischen Tätigkeit als Sachverständige unter Supervision von Prof. Dr. Dr. Wegener
bis gegenwärtig	und eigenständige hauptberufliche Tätigkeit in diesem Bereich
1988 bis gegenwärtig	Lehraufträge am Institut für Psychologie der Universität Kiel in den Bereichen Entwicklungspsychologie, Persönlichkeitspsychologie, Pädagogische Psychologie und Rechtspsychologie
27.03.01	Fachpsychologin für Rechtspsychologie BDP /DGPs

Mitgliedschaften in berufsbezogenen Verbänden

1988	Deutsche Gesellschaft für Psychologie e.V. (DGPs)
1992	European Association of Psychology an Law (EAPL)
1993	Berufsverband deutscher Psychologen e.V. (BDP)