

5. Darstellung der Ergebnisse

5.1 Stichprobenvariablen und Auswahl der verwendeten Ereignisse

Insgesamt nahmen 67 Schüler erster Klassen an der Untersuchung teil. Die kindlichen Probanden waren zu Beginn der Untersuchung 6;5 bis 8;6 Jahre alt; das Durchschnittsalter der Kinder lag bei 7;5 Jahren ($SD=0.4$). Es nahmen 31 Mädchen (46.3 %) und 36 Jungen (53.7 %) an der Untersuchung teil. Ein statistisch signifikanter Unterschied zwischen Jungen und Mädchen zeigte sich zwar hinsichtlich des Alters ($p=0.16$; $t=2.469$); jedoch war dieser insofern geringfügig, als die Mädchen durchschnittlich nur drei Monate jünger waren (7;4 Jahre; $SD=0.4$) als die Jungen (7;7 Jahre; $SD=0.3$).

Zur Bestimmung der kognitiven Leistungsfähigkeit der Kinder wurden die Ergebnisse der Kinder in der CPM-Form des Progressiven Matrizentests von Raven (1962) ausgewertet, indem die Rohwerte bestimmt und anhand der aktuell vorliegenden Normentabellen - unter Berücksichtigung des Alters der einzelnen Kinder - IQ-Werte ermittelt wurden.

Von 36 möglichen Rohwerten erzielten die Kinder durchschnittlich 25 Punkte bei einem Range von 16 bis 35 ($SD=4.6$). Dies entspricht einem durchschnittlichen IQ von 112 bei einem Range von 89 bis 135 ($SD=12$). Signifikante Zusammenhänge zwischen den Variablen Geschlecht und kognitive Leistungsfähigkeit ergaben sich nicht.

Wie in Abschnitt 4.2 dargestellt wurden für jedes Kind individuell anhand der von den Eltern ausgefüllten Fragebögen ein reales und ein fiktives Ereignis festgelegt, zu denen die Kinder bei den sechs Interviewterminen befragt werden sollten. Eine Aufstellung der verwendeten Ereignisse findet sich im Anhang.

5.2 Suggestionenübernahme und Suggestionenresistenz

5.2.1 Beeinflussungsphase

5.2.1.1 Reaktionen der Kinder bei den ersten vier Befragungszeitpunkten

Um festzustellen, wie sich die suggestiven Einflüsse im Zeitverlauf auf die Antworten der Kinder ausgewirkt haben, wurden die Schilderungen aller Kinder zu den ersten vier Erhebungszeitpunkten dahin gehend analysiert, ob die Kinder das fiktive Ereignis verneint haben, ob sie unsicher waren, ob sie es zunächst verneint, aber im Verlauf des Termins bejaht haben oder ob sie es spontan bejaht haben.

Bei der Analyse der Schilderungen bzw. der Reaktionen der Kinder wurde folgendes Kategoriensystem zugrunde gelegt:

Kategorie 1: verneinend

Eine Verneinung des fiktiven Ereignisses wurde immer dann angenommen, wenn die Kinder das fiktive Ereignis verneinten bzw. angaben, es sei nicht passiert oder sie könnten sich nicht erinnern. Auch Antworten, die Erklärungen des Kindes beinhalteten, warum das Ereignis gar nicht habe passiert sein können (z.B. Angabe, die Mutter gefragt zu haben, noch zu klein gewesen o.ä.), wurden unter diese Kategorie subsumiert. Eine Verneinung des fiktiven Ereignisses wurde auch dann angenommen, wenn das Kind bei Anwendung der Als-ob-Technik zwar entsprechende Angaben lieferte, jedoch

deutlich erkennbar war, daß das Kind damit lediglich der Aufforderung der Interviewerin nachkam, ohne von der eigenen Schilderung subjektiv überzeugt zu sein.

Kategorie 2: unsicher, ambivalent

Die Antworten der Kinder wurden als unsicher bzw. ambivalent kategorisiert, wenn sich die Kinder hinsichtlich des Realitätsgehalts des fiktiven Ereignisses unsicher waren, beispielsweise äußerten, nicht sicher zu wissen, ob das Ereignis passiert sei oder nicht, oder wenn sie längere Zeit überlegten. Auch wenn die Kinder Details nannten, diese dann aber widerriefen, wenn also ein Wechsel zwischen Zustimmung und Ablehnung bzw. Rücknahme deutlich wurde, erfolgte die Zuordnung zu dieser Kategorie.

Kategorie 3: zögerlich bejahend

Verneinten die Kinder das fiktive Ereignis zunächst, stimmten sie ihm dann aber nach suggestiver Einflußnahme im Verlauf desselben Interviews doch noch zu, wurde die Antwort - unabhängig davon, ob sie Details enthielt oder nicht - der Kategorie zögerlich bejahend zugeordnet.

Kategorie 4: spontan bejahend

Als spontan bejahend wurden schließlich die Antworten eingeordnet, die auf offene Frage ein spontanes Bejahen des fiktiven Ereignisses beinhalteten - unabhängig davon, ob die Kinder Details zum fiktiven Ereignis lieferten oder nicht.

Tabelle 1: Reaktionen der Kinder bei den Zeitpunkten T1 bis T4

	T1 (N= 67)	T2 (N= 64)	T3 (N= 64)	T4 (N= 63)
verneinend	40.3 % (n=27)	29.7 % (n=19)	28.1 % (n=18)	19.0 % (n=12)
unsicher, ambivalent	3.0 % (n= 2)	7.8 % (n= 5)	4.7 % (n= 3)	4.8 % (n= 3)
zögerlich bejahend	28.4 % (n=19)	10.9 % (n= 7)	4.7 % (n= 3)	3.2 % (n= 2)
spontan bejahend	28.4 % (n=19)	51.6 % (n=33)	62.5 % (n=40)	73.0 % (n=46)

Die Ergebnisse zeigen, daß beim ersten Befragungszeitpunkt 40 % aller Kinder das fiktive Ereignis verneint haben. 28 % haben das fiktive Ereignis spontan bejaht, weitere 28 % haben zögerlich zugestimmt, was bedeutet, daß ihre erste Reaktion in einer Verneinung bestand, sie aber noch im Verlauf des ersten Interviews - vermutlich infolge der wiederholten Nachfrage bzw. suggestiven Einflußnahme - zustimmten.

Im Vergleich dazu nahmen beim zweiten und dritten Befragungszeitpunkt der Anteil der Kinder, die das fiktive Ereignis verneinten, sowie der Anteil der Kinder, die zögerlich zustimmten, weiter ab; die Anzahl der Kinder, die das fiktive Ereignis spontan bejahten, verdoppelte sich. Beim vierten Erhebungszeitpunkt verneinten dann nur noch 19 % der befragten Kinder das fiktive Ereignis, während 73 % spontan zustimmten.

Verneinten die Kinder beim zweiten bis vierten Befragungstermin das fiktive Ereignis, so kam bei der Mehrzahl dieser Kinder (beim dritten Termin bei allen verneinenden Kindern) die Als-ob-Technik zur Anwendung. Daraufhin schilderten beim zweiten Termin 9 Kinder, beim dritten Termin 15 Kinder und beim vierten Termin 8 Kinder - wenn auch teilweise nur sehr wenige - Details zum fiktiven Ereignis, wobei aber deutlich war (häufig durch die Verwendung des Konjunktivs), daß die Kinder vom Realitätsgehalt ihrer Darstellungen subjektiv nicht überzeugt waren. Die übrigen sechs Kin-

der konnten nicht dazu motiviert werden, sich vorzustellen, wie sich das fiktive Ereignis zugetragen haben könnte, und/oder dieses zu schildern.

Tabelle 2: Anwendung und Effekte der Als-ob-Technik bei Verneinung des fiktiven Ereignisses

	T2	T3	T4
Als-ob-Technik angewandt	n=11	n=18	n= 9
auf Als-ob-Technik eingegangen	n= 9	n=15	n= 8
nicht auf Als-ob-Technik eingegangen	n= 2	n= 3	n= 1

In einem weiteren Analyseschritt wurde das oben dargestellte und verwendete Kategoriensystem insoweit modifiziert, als bei den Kategorien 2 bis 4 zusätzlich differenziert wurde, ob die Kinder Details zum fiktiven Ereignis berichtet haben oder nicht. Ein Berichten von Details wurde dann bejaht, wenn das Kind mindestens ein Detail zum fiktiven Ereignis schilderte, das über unmittelbar zuvor von der Interviewerin vorgegebene Inhalte hinausging.

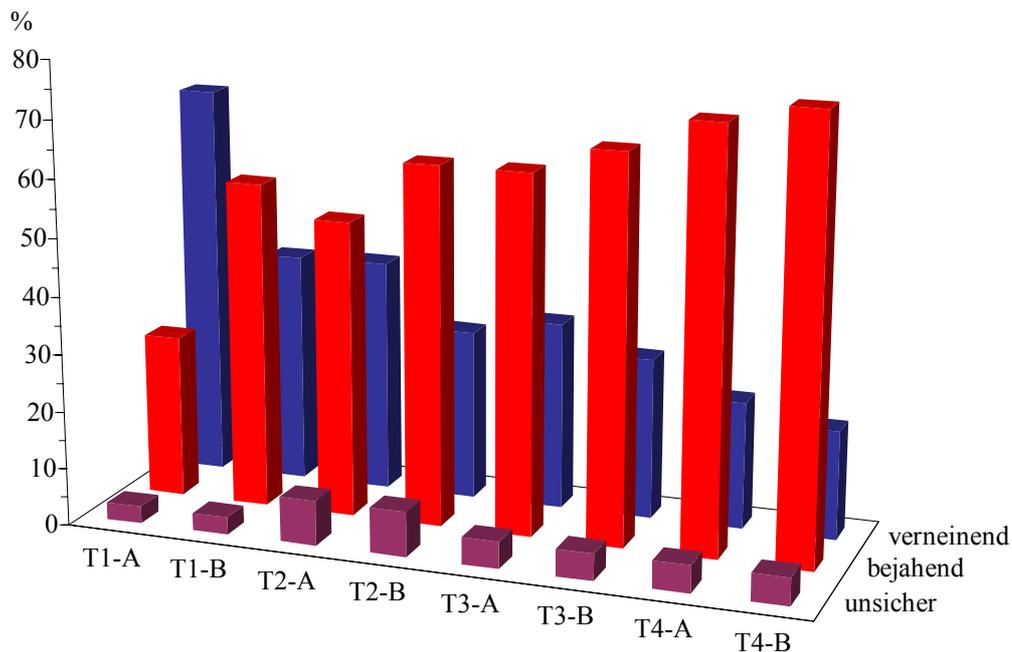
Beim ersten Befragungstermin schilderten die beiden Kinder, die sich hinsichtlich des Realitätsgehalts des fiktiven Ereignisses unsicher waren, dennoch Details. Nur zwei der Kinder, die dem fiktiven Ereignis zögerlich zustimmten, und nur ein Kind, das das Ereignis spontan bejahte, nahmen keine Präzisierungen zum fiktiven Ereignis vor. Auch bei den Interviewzeitpunkten zwei bis vier lieferten fast alle Kinder, die das fiktive Ereignis nicht explizit verneinten oder die nicht nur auf die Als-ob-Technik eingingen, Details zum fiktiven Ereignis. Ausnahmen ergaben sich beim dritten und vierten Termin lediglich in bezug auf jeweils ein Kind, das sich unsicher über den Realitätsgehalt war und auch keine Details schilderte.

Somit läßt sich zusammenfassend festhalten, daß die meisten Kinder, die das fiktive Ereignis bei den einzelnen Befragungsterminen nicht verneinten, den Realitätsgehalt nicht nur bejahten, sondern auch präzisierende Angaben lieferten.

Nachfolgend wird das Antwortverhalten der Kinder im Zeitverlauf zwischen erstem und viertem Befragungszeitpunkt betrachtet und darüber hinaus das Antwortverhalten jeweils zu Beginn eines Termins (vor suggestiver Einflußnahme; Tx-A) sowie am Ende der jeweiligen Befragung (nach suggestiver Einflußnahme; Tx-B) berücksichtigt, indem die Gruppe der zögerlich bejahenden Kinder für den Beginn des Termins der Gruppe der verneinenden Kinder zugeordnet werden, für das Ende der Befragung hingegen der Gruppe der bejahenden Kinder.

Die so ermittelten Ergebnisse zeigen eine sprunghafte Veränderung des Antwortverhaltens in Form einer Zunahme an Zustimmungen und einer Abnahme an Verneinungen des fiktiven Ereignisses während der ersten Befragung. Deutliche Suggestionseffekte traten auch innerhalb des zweiten Termins auf. Ab dem dritten Termin trat insofern eine allmähliche Stabilisierung ein, als die Zustimmungsraten zwar noch kontinuierlich zunahmen und die Verneinungen stetig abnehmen. Allerdings waren diese Veränderungen zwischen den verschiedenen Zeitpunkten nur noch vergleichsweise geringfügig (Veränderungsrate zwischen den verschiedenen Zeitpunkten bei etwa 4 %).

Abbildung 2: Antwortverhalten der Kinder zu den Zeitpunkten T1 bis T4



Um zu überprüfen, ob das Aussageverhalten der einzelnen Kinder über die verschiedenen Befragungszeitpunkte hinweg konstant blieb, wurden die Korrelationen zwischen den einzelnen Befragungszeitpunkten hinsichtlich der Antwortkategorien ermittelt.⁷

Tabelle 3: Stabilität des Antwortverhaltens im Zeitverlauf (Rangkorrelationskoeffizienten nach Spearman)⁸

	T2	T3	T4	T6
T1	.591*** (N= 64)	.370*** (N= 64)	.289** (N= 63)	.320*** (N= 64)
T2		.694*** (N= 62)	.653*** (N= 60)	.615*** (N= 61)
T3			.813*** (N= 61)	.611*** (N= 62)
T4				.666*** (N= 60)

Während zwischen dem Antwortverhalten der Kinder beim ersten Interviewtermin und den späteren Befragungen zwar signifikante, aber eher niedrige Korrelationen bestehen, ergaben sich mittlere bis hohe Korrelationen zwischen dem zweiten, dritten und vierten Interviewtermin. Die Ergebnisse lassen sich dahin gehend interpretieren, daß im Zeitverlauf eine zunehmende Stabilisierung im Antwortverhalten der Kinder eingetreten ist,

⁷ Bei der Berechnung der Korrelationskoeffizienten nach Spearman wurde hinsichtlich der Variable Antwortverhalten von Ordinalskalenniveau ausgegangen (1= verneinend, 2= unsicher, ambivalent, 3= zögerlich zustimmend, 4= spontan zustimmend).

⁸ In dieser und allen folgenden Ergebnisdarstellungen gelten folgende Signifikanzniveaus: *** p≤ .01; ** p≤ .05; * p≤ .10; n.s. p> .10 (jeweils zweiseitig).

d.h. daß sich das zukünftige Antwortverhalten der Kinder mit wiederholter Befragung immer besser vorhersagen läßt.

5.2.1.2 Typische Aussageentwicklungen in der Beeinflussungsphase

Des weiteren wurde geprüft, ob sich zwischen dem ersten und vierten Befragungszeitpunkt typische Aussageverläufe aufzeigen lassen. So wurden die Daten dahin gehend ausgewertet, ob es Kinder gibt, die das fiktive Ereignis beispielsweise durchgehend verneint haben, die zwischen Zustimmung und Ablehnung gewechselt haben, die anfänglich verneint und später zugestimmt haben oder die das fiktive Ereignis durchgehend bejaht haben.

Um die Kinder aufgrund der Entwicklungen ihrer Aussagen unterschiedlichen Gruppen zuzuordnen, wurde folgendes Kategoriensystem entwickelt:

Kategorie 1: durchgehend verneinend

Dieser Kategorie wurden diejenigen Kinder zugeordnet, die das fiktive Ereignis über die ersten vier Befragungstermine hinweg durchgehend verneint haben und sich höchstens bei einem Termin hinsichtlich des Realitätsgehalts unsicher waren.

Kategorie 2: unsicher, zwischen Zustimmung und Verneinung wechselnd

Kinder dieser Gruppe waren sich bei Betrachtung der ersten vier Befragungstermine hinsichtlich des Realitätsgehalts des fiktiven Ereignisses offenbar unsicher, was sich darin zeigte, daß sie im Zeitverlauf zwischen Zustimmung und Verneinung des fiktiven Ereignisses wechselten oder anfangs zustimmten, ihre Zustimmungen aber später widerriefen.

Kategorie 3: anfänglich verneinend oder unsicher, später eindeutig zustimmend

Kinder, die dieser Kategorie zugeordnet wurden, waren sich bei den ersten Befragungsterminen unsicher über den Realitätsgehalt des fiktiven Ereignisses, oder sie verneinten ihn. Im Verlauf der wiederholten Befragungen und suggestiven Einflußnahmen stimmten sie dem fiktiven Ereignis aber eindeutig zu und lieferten entsprechende Schilderungen.

Kategorie 4: durchgehend zögerlich oder spontan bejahend

Diese Kategorie umfaßt die Kinder, die das fiktive Ereignis durchgehend bejaht haben, die es jedoch nicht bei allen vier Terminen spontan bejaht haben, sondern die bei mindestens einem Termin zunächst gezögert bzw. das Ereignis verneint haben, bevor sie dann - nach suggestiver Einflußnahme - noch beim selben Termin zugestimmt haben.

Kategorie 5: durchgehend spontan bejahend

Im Unterschied zu Kategorie 4 wurden Kategorie 5 diejenigen Kinder zugeordnet, die das fiktive Ereignis - ohne zu zögern oder zunächst zu verneinen - bei allen vier Terminen spontan bejaht haben.

Die Ergebnisse zeigen, daß knapp 20 % der Kinder das fiktive Ereignis während der ersten vier Befragungen durchgehend verneint haben, d.h. daß die suggestiven Einflußnahmen keinerlei Wirkung auf diese Kinder hatten. Weitere 12 % der Kinder waren sich

hinsichtlich des Realitätsgehalts des fiktiven Ereignisses unsicher oder wechselten zwischen Zustimmung und Ablehnung. Demgegenüber haben 21 % der Kinder das fiktive Ereignis bereits beim ersten Interviewtermin spontan bejaht und dieses Antwortverhalten auch bei den drei nachfolgenden Befragungen beibehalten. Weitere 48 % der Kinder haben der suggestiven Einflußnahme im Verlauf der Befragungen nachgegeben. Bei 25 % der Kinder geschah dies bereits im Verlauf des ersten Interviews; bei weiteren 21 % der Kinder erfolgte die Übernahme erst in den nachfolgenden Befragungen (vgl. Tabelle 4).

Bei einem Kind, dessen Aussageverlauf der Kategorie 2 zugeordnet wurde, ergab sich eine besonders auffällende Aussageentwicklung, wenn man neben den ersten vier Befragungen auch die fünfte sowie die Nachbefragung berücksichtigt. So bejahte dieser Junge das fiktive Ereignis (Kugel in die Nase gesteckt bekommen) bei den ersten beiden Terminen spontan, verneinte es beim dritten Termin mit der Begründung, er habe mit seiner Mutter gesprochen, die gesagt habe, das Ereignis stimme nicht, lieferte dann aber doch einige Angaben dazu. Beim vierten Termin gab der Junge zunächst an, sich nicht erinnern zu können, lieferte dann aber wiederum eine der vorherigen Darstellung ähnliche Schilderung. Bei der fünften suggestionsfreien Befragung verneinte der Junge (als einziges Kind im Rahmen der Befragung durch die uninformierten Experten), das fiktive Ereignis erlebt zu haben. Im Rahmen der Nachbefragung bejahte er es wiederum und lieferte eine entsprechende Schilderung.

Zur Klärung der Frage, ob sich die Kinder, bei denen sich unterschiedliche Aussageverläufe gezeigt haben, hinsichtlich personaler Voraussetzungen wie Geschlecht, Alter und kognitive Leistungsfähigkeit voneinander unterscheiden, wurden entsprechende Analysen durchgeführt.

Das Ergebnis eines Chi-Quadrat-Tests mit den Variablen Aussageverlauf und Geschlecht zeigte, daß sich Jungen und Mädchen annähernd gleichmäßig auf die fünf Gruppen verteilten, d.h. sich hinsichtlich der Verläufe ihrer Aussagen nicht signifikant voneinander unterscheiden.

Eine univariate Varianzanalyse mit dem Aussageverlauf als Faktorvariable und dem Alter als abhängige Variable brachte kein signifikantes Ergebnis, was bedeutet, daß die unterschiedlichen Aussageverläufe unabhängig vom Alter der Kinder waren.

Die Durchführung von zwei weiteren univariaten Varianzanalysen zeigte schwach signifikante Mittelwertsunterschiede zwischen den fünf Gruppen hinsichtlich Rohwert ($p = .090$, $F = 2.115$) und IQ-Wert ($p = .059$; $F = 2.410$) im CPM. T-Tests für unabhängige Stichproben zeigten, daß die Kinder der ersten Gruppe, die sogenannten *hoch suggestionsresistenten* Kinder, gegenüber den anderen Kindern signifikant bessere Ergebnisse im CPM erzielten (RW: 27.5 versus 24.4; IQ: 117.6 versus 110.5). Bezüglich der Rohwerte war dieses Ergebnis auf dem 5 %-Niveau signifikant ($p = .028$; $t = 2.243$), bezüglich der IQ-Werte verfehlte das Ergebnis knapp das 5 %-Niveau ($p = .055$; $t = 1.953$). Signifikante Unterschiede hinsichtlich der Ergebnisse im CPM ergaben sich auch zwischen den Kindern der Gruppe 1 und denen aus den Gruppen 3 bis 5 (alle Kinder, die die Suggestionen übernommen haben; RW: 27.5 versus 24.5; IQ: 117.6 versus 110.9; RW: $p = .029$; $t = 2.241$; IQ: $p = .070$; $t = 1.848$) sowie zwischen den Kindern aus Gruppe 1 und denen aus Gruppe 5 (RW: $p = .007$, $t = 2.940$; IQ: $p = .022$, $t = 2.451$).

Tabelle 4: Typische Aussageentwicklungen in der Beeinflussungsphase sowie Ergebnisse im CPM in Abhängigkeit vom Aussageverlauf

	N	CPM-Rohwert		CPM-IQ-Wert	
		M	SD	M	SD
durchgehend verneinend	13 19.4 %	27.46	3.62	117.62	11.05
unsicher, zwischen Zustimmung und Verneinung wechselnd	8 11.9 %	23.75	6.16	108.25	13.83
anfänglich verneinend, später zustimmend	15 22.4 %	23.87	4.02	108.33	12.32
durchgehend zögerlich oder spontan bejahend	17 25.4 %	26.00	4.95	116.00	10.92
durchgehend spontan bejahend	14 20.9 %	23.36	3.63	107.43	10.54

5.2.2 Befragung der Kinder durch uninformierte Experten

Im Rahmen der Expertenbefragung wurden nur die Kinder zu fiktiven Ereignissen befragt, die im Vorfeld zusammenhängende sowie plausible Schilderungen über das fiktive Ereignis geliefert hatten und bei denen davon auszugehen war, daß sie dieses Aussageverhalten auch bei einer suggestionsfreien Befragung beibehalten würden. Für die Expertenbefragung zu mindestens einem fiktiven Ereignis wurden dementsprechend 20 Mädchen und 19 Jungen (insgesamt 58.2 % der Kinder aus der Gesamtstichprobe) ausgewählt. Hinsichtlich keiner der abhängigen Variablen Geschlecht, Alter sowie Rohwert und IQ-Wert im CPM ergaben sich signifikante Unterschiede zwischen den Kindern, die von den Experten befragt wurden, und denen, die nicht befragt wurden.

Von allen fünf Interviewerinnen, die an der Beeinflussungsphase beteiligt waren, wurden Kinder, die sie an den ersten drei Terminen befragt hatten, einbezogen. Der Anteil der Kinder, die von den Experten zu mindestens einem fiktiven Ereignis befragt wurden, lag - bezogen auf die einzelnen Suggestorinnen - zwischen 41.2 % und 68.2 %, durchschnittlich bei 58.8 %.

Alle bis auf eines der von den Experten zum fiktiven Ereignis befragten 39 Kinder bejahten, das fiktive Ereignis erlebt zu haben, und lieferten entsprechende Schilderungen (56.7 % der Gesamtstichprobe); ein Junge gab hingegen an, das fiktive Ereignis sei nicht passiert (dieser Junge gab allerdings bei der Nachbefragung wieder an, das fiktive Ereignis erlebt zu haben; vgl. Abschnitt 5.2.1.2).

Betrachtet man die Angaben der Kinder, die beim fünften Termin nicht von den Experten zum fiktiven Ereignis befragt worden waren (n= 28; 41.8 % der Gesamtstichprobe), so ist davon auszugehen, daß auch einige dieser Kinder gegenüber den blinden Interviewern Schilderungen zum fiktiven Ereignis geliefert hätten, wengleich die Angaben möglicherweise nicht immer plausibel gewesen wären.⁹

⁹ So wurden zwei der von den Experten nicht befragten Kinder beim fünften Termin von einer der früheren (informierten) Interviewerinnen neutral zum fiktiven Ereignis befragt. Eines dieser Kinder stimmte dem fiktiven Ereignis zu und schilderte es, das andere Kind war sich hinsichtlich des Realitätsgehalts unsicher.

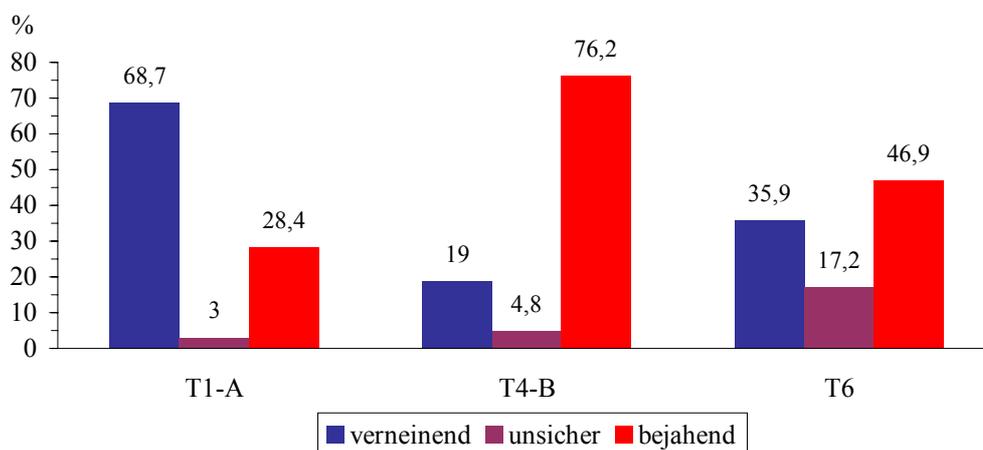
5.2.3 Teilaufklärung und Nachbefragung der Kinder

Nachdem die Kinder zu Beginn des sechsten Befragungstermins von einer neuen Befragungsperson darauf hingewiesen worden waren, daß sie früher möglicherweise zu einem Ereignis befragt worden seien, das sie gar nicht erlebt hätten, und nachdem sie gebeten worden waren, noch einmal gut zu überlegen, ob sie die vorgegebenen Ereignisse tatsächlich erlebt hätten oder nicht, verneinten 35,9 % (N= 23) der Kinder, das fiktive Ereignis erlebt zu haben; 17,2 % (N= 11) waren sich hinsichtlich des Realitätsgehaltes unsicher, aber immer noch 46,9 % (N= 30) der Kinder gaben an, das fiktive Ereignis tatsächlich erlebt zu haben. Davon hatten 24 Kinder (36 % der Gesamtstichprobe) auch gegenüber den blinden Experten beim fünften Termin Angaben zum fiktiven Ereignis geliefert. Drei Kinder konnten aufgrund von Umzügen bzw. Krankheit nicht in die Nachbefragung einbezogen werden.

Keines der befragten Kinder verneinte, das reale Ereignis erlebt zu haben, oder zeigte sich hinsichtlich des Realitätsgehaltes unsicher. Allerdings gaben einige Kinder an, sich nicht mehr gut an das reale Ereignis (N= 9) oder sogar schlechter als an das fiktive Ereignis (N= 2) erinnern zu können.

Zusammenfassend ergibt sich bezüglich des Antwortverhaltens der Kinder beim ersten Befragungszeitpunkt (vor suggestiver Einflußnahme; T1-A), beim vierten Termin (nach suggestiver Einflußnahme; T4-B) und beim sechsten Termin (nach Teilaufklärung; T6) folgendes Bild:

Abbildung 3: Antwortverhalten der Kinder zu den Zeitpunkten T1-A, T4-B und T6



5.2.4 Nachbefragung der Eltern

Im Rahmen der Nachbefragung der Eltern - zum Zweck der Validierung des Realitätsgehaltes der Schilderungen ihrer Kinder - erhielten die Eltern derjenigen Kinder, die beim fünften Befragungszeitpunkt zu einem realen und einem fiktiven Ereignis befragt worden waren und die das fiktive Ereignis bei dieser Befragung auch bejaht haben, nach der sechsten Befragung der Kinder einen Fragebogen. Da von den 38 Kindern, die zu je einem realen und einem fiktiven Ereignis befragt worden waren, ein Kind bei der fünf-

ten Befragung verneint hatte, das fiktive Ereignis erlebt zu haben, wurden somit für die Eltern von 37 Kindern individuelle, d.h. auf die jeweiligen Angaben der Kinder abgestimmte Fragebögen entwickelt. Da zwei Kinder aufgrund von Umzügen in der Nachbefragungsphase nicht mehr erreichbar waren, wurden nur an die Eltern von 35 Kindern Fragebögen verteilt. Davon wurden 25 Fragebögen ausgefüllt zurückgesandt; dies entspricht einer Rücklaufquote von 71.4 %.

Der Realitätsgehalt der realen Ereignisse wurde in 24 Fällen im wesentlichen bejaht. In einem Fall war den Angaben zu entnehmen, daß es zu Vermischungen zwischen erlebten und nicht erlebten Elementen gekommen ist, was mit den eigenen Analysen der Schilderungen dieses Kindes zum realen Ereignis in Einklang steht. So waren die Schilderungen im Zeitverlauf - gerade auch in bezug auf zentrale Elemente wie gravierende Verletzungsfolgen - in hohem Maße inkonstant.

Wider Erwarten wurde nur für acht Kinder (32 %) der Realitätsgehalt des fiktiven Ereignisses verneint. In neun Fällen (36 %) hatten die Eltern angegeben, das fiktive Ereignis habe sich so zugetragen wie im Fragebogen geschildert, was insofern verwunderlich ist, als die Eltern im Rahmen der mittels eines Fragebogens vorgenommenen Vorbefragung verneint hatten, daß ihr Kind das fiktive Ereignis erlebt habe. Sechs Antworten (24 %) bestanden in der Aussage, daß das Kind das Ereignis grundsätzlich erlebt habe; allerdings wurden die Folgen von den Eltern als weit weniger gravierend angegeben. In einem Fall wurde angegeben, daß einige Aspekte des Ereignisses realitätsbegründet seien, andere nicht. Die Eltern eines weiteren Kindes beantworteten zwar die Fragen zum realen, nicht jedoch die zum fiktiven Ereignis.

Auf die offene Frage zum realen Ereignis lieferten sieben Eltern entsprechende Angaben, die vorrangig in ergänzenden, präzisierenden oder korrigierenden Anmerkungen bestanden. Auf die offene Frage zum fiktiven Ereignis wurde in 14 Fällen geantwortet. Davon beinhalteten sechs Antworten Überlegungen dazu, auf welchen Erfahrungen oder Beobachtungen die Schilderungen des Kindes beruhen könnten. In sieben Fällen wurden Präzisierungen, Ergänzungen oder Korrekturen der Angaben des Kindes vorgenommen, unter anderem wurden die negativen Konsequenzen des fiktiven Ereignisses als weniger gravierend dargestellt. Eine Mutter gab an, daß sie sich zwar nicht an das fiktive Ereignis erinnern könne, daß dies aber nicht notwendigerweise bedeute, daß das Ereignis nicht stattgefunden habe.

Vergleicht man die im Rahmen der Nachbefragung von den Kindern erhobenen Angaben mit den Antworten der Eltern, so stimmen die Angaben zu den realen Ereignissen in allen 25 Fällen überein.

Hinsichtlich der fiktiven Ereignisse fanden sich in zehn Fällen Übereinstimmungen zwischen den Angaben der Eltern und denen der Kinder: So gaben drei Kinder sowie deren Eltern an, das fiktive Ereignis habe nicht stattgefunden, sieben Kinder sowie deren Eltern bejahten den Realitätsgehalt des Ereignisses hingegen.

Eindeutige Diskrepanzen zwischen den Angaben von Kindern und Eltern waren in fünf Fällen zu verzeichnen. So bejahten drei Kinder, das fiktive Ereignis erlebt zu haben, während die Eltern dies verneinten. In zwei Fällen verneinten die Kinder den Realitätsgehalt des fiktiven Ereignisses, während die Eltern ihn bejahten.

Bei zwei Kindern, die sich hinsichtlich des Realitätsgehalts des fiktiven Ereignisses unsicher waren, verneinten die Eltern den Realitätsgehalt. Bei einem weiteren „unsiche-

ren“ Kind gaben die Eltern an, daß die Folgen des Ereignisses weniger gravierend gewesen seien.

Die Eltern von sechs Kindern stimmten dem Realitätsgehalt des fiktiven Ereignisses in dem vom Kind geschilderten Ausmaß der Folgen nicht zu, während fünf dieser Kinder den Realitätsgehalt des fiktiven Ereignisses bejahten, ein Kind ihn verneinte.

5.2.5 Spezielle Aussagecharakteristika

5.2.5.1 Merkmale der Suggestionresistenz

In einem weiteren Auswertungsschritt wurde geprüft, ob sich typische Antworten oder Antwortmuster nachweisen lassen, derer die Kinder sich bedienen, um die suggestiven Vorgaben zurückzuweisen bzw. um das fiktive Ereignis begründet zu verneinen.

Zunächst lassen sich vier Typen von Verneinungen unterscheiden:

Kategorie 1: Verneinung ohne weitere Erläuterung

In diese Kategorie wurden Antworten eingeordnet, die aus einem bloßen Verneinen des fiktiven Ereignisses bestanden, ohne daß erkennbar war, ob das Kind der Auffassung war, das Ereignis habe nicht stattgefunden, oder ob es das Ereignis für möglich hielt, sich aber nicht erinnern konnte. Des weiteren wurden hier Antworten subsumiert, die keiner der nachfolgenden vier Kategorien eindeutig zugeordnet werden konnten.

Kategorie 2: Angabe, das Ereignis sei nicht passiert

Aus Antworten, die in diese Kategorie eingeordnet wurden, geht eindeutig hervor, daß das Kind davon überzeugt war, das Ereignis habe nicht stattgefunden (z.B. „Das stimmt nicht“, „Das ist mir nicht passiert“ u.ä.).

Kategorie 3: Angabe, sich nicht erinnern zu können

Dieser Kategorie wurden Antworten zugeordnet, in denen das Kind auf Erinnerungs- und/oder Wissenslücken verwies, ohne explizit zu sagen, daß das Ereignis nicht stattgefunden habe (z.B. „Ich weiß nicht“, „Ich kann mich nicht erinnern“, „Dazu fällt mir nichts ein“ u.ä.).

Kategorie 4: Verneinung und Anführen von Gründen

Ähnlich wie bei den Antworten aus Kategorie 2 beinhalten auch die Antworten dieser Kategorie eine ausdrückliche Verneinung des fiktiven Ereignisses. Darüber hinaus wurden bei Antworten dieser Kategorie von den Kindern aber Gründe angeführt, warum das Ereignis gar nicht passiert sein kann (z.B. „Wenn man klein ist, macht man keine Schneeballschlacht“), die Kinder verwiesen auf dritte Personen, die entweder den fehlenden Realitätsgehalt bestätigt oder die (im Elternfragebogen) etwas Falsches aufgeschrieben hätten, die Kinder erklärten, daß ein anderes Kind so etwas erlebt habe, oder sie berichteten über ein anderes Ereignis, das - wenn auch wenige - Parallelen zu dem vorgegebenen aufwies. Antworten dieser Kategorie zeigen somit ein Bemühen, den aus Sicht des Kindes vorliegenden Irrtum bzw. das Mißverständnis aufzuklären.

Tabelle 5 zeigt die Verteilung der Antworten auf die verschiedenen Arten von Verneinungen.

Tabelle 5: Arten von Verneinungen bei den Zeitpunkten T1 bis T4

	T1 (N= 27)	T2 (N= 19)	T3 (N= 18)	T4 (N= 12)
bloÙe Verneinung	11.1 % (n= 3)	5.3 % (n= 1)	33.3 % (n= 6)	16.7 % (n= 2)
Ereignis nicht passiert	18.5 % (n= 5)	10.5 % (n= 2)	0	25.0 % (n= 3)
keine Erinnerung	48.1 % (n=13)	15.8 % (n= 3)	22.2 % (n= 4)	8.3 % (n= 1)
Verneinung mit Begründung	22.2 % (n= 6)	68.4 % (n=13)	44.4 % (n= 8)	50.0 % (n= 6)

Unterzieht man die Antworten der Kinder, die das fiktive Ereignis nicht nur verneinten oder Erinnerungsverluste anführten, sondern darüber hinausgehende Erklärungen lieferten, einer weiteren Analyse, so lassen sich drei Typen von Einwänden nachweisen:

- Hinweise, die Mutter/Eltern gefragt zu haben
- Verweise auf ein anderes Ereignis, ein anderes Kind oder auf Irrtümer der Eltern
- Einwände gegen die Plausibilität des Ereignisses

Hinweise, die Mutter bzw. Eltern gefragt zu haben:

Obwohl sowohl die Eltern im Vorfeld als auch die Kinder nach jedem Befragungstermin gebeten worden waren, nicht miteinander über die in der Untersuchung erörterten Ereignisse zu sprechen, gaben zwölf der Kinder (17.9 %) bei mindestens einem der sechs Interviewtermine (acht Kinder beim zweiten, je zwei beim dritten und sechsten) an, daß auch die Mutter bzw. die Eltern gesagt hätten, das fiktive Ereignis habe nicht stattgefunden, und bekräftigten ihre Verneinungen damit. Zwei weitere Kinder gaben zwar an, mit ihrer Mutter über das fiktive Ereignis gesprochen zu haben, sie äußerten sich jedoch nicht dazu, was die Mutter dazu gesagt habe. Vier der Kinder, die beim zweiten Termin angaben, ihre Mutter/Eltern hätten das fiktive Ereignis verneint, übernahmen die Suggestion im weiteren Verlauf dennoch. Hingegen nahmen zwei Kinder ihre früheren Zustimmungen und Angaben zum fiktiven Ereignis nach dem Gespräch mit den Eltern zurück und erwiesen sich im weiteren Verlauf als suggestionsresistent. Von den beiden Kindern, die erstmals im Rahmen der Nachbefragung angegeben hatten, die Mutter habe gesagt, das Ereignis habe nicht stattgefunden, hatte ein Kind den Realitätsgehalt des fiktiven Ereignisses zuvor durchgehend bejaht, das andere hatte ihn durchgehend verneint.

Verweise auf ein anderes Ereignis, ein anderes Kind oder auf Irrtümer der Eltern:

Häufig waren in den Antworten der Kinder Verweise auf ein anderes Ereignis oder ein anderes Kind enthalten, d.h. die Kinder gaben an, daß das von der Interviewerin vorgegebene Ereignis nicht ihnen, sondern (möglicherweise) einem anderen Kind passiert sei, oder sie schilderten ein anderes Ereignis, das sie selber erlebt hatten und das gewisse Parallelen zu dem von der Interviewerin vorgegebenen fiktiven Ereignis aufwies. Darüber hinaus äußerten einige Kinder, daß die Mutter bzw. die Eltern dann wohl etwas Falsches aufgeschrieben hätten. Bei Reaktionen dieser Art entstand bei einigen Kindern der Eindruck, daß sie davon ausgingen, die Interviewerin unterliege einem Irrtum, und daß sie bemüht waren, dieses Mißverständnis aufzuklären. Bei anderen Kindern schien der Verweis auf ein anderes Ereignis oder ein anderes Kind vorrangig dem Ziel zu dienen, die (suggestive) Befragung zum fiktiven Ereignis zu beenden und sich dem Befragungsdruck zu entziehen.

Ein Junge gab auf die Frage, ob er sich daran erinnern könne, seinen Finger in eine Flasche gesteckt und nicht mehr herausbekommen zu haben, zunächst an, sich daran nicht mehr erinnern zu können. „Ich war schon mal schon irgendwo eingesteckt, aber in der Flasche noch nie. ... Mit dem Gummistiefel in so'n Balken reingegangen, bin reingekommen, aber nicht mehr wieder rausgekommen. ... Da mußte mein Opa mich ausgraben.“

Ein Mädchen, das beim ersten Termin der suggestiven Vorgabe, in der Bustür eingeklemmt worden zu sein, zögernd zugestimmt hatte, äußerte beim zweiten Termin, nachdem sie mit ihrer Mutter gesprochen hatte: „Das war ich gar nicht.“ Beim vierten Termin sagte sie: „Das war eigentlich nicht so. ... Das war gar nicht mir eingeklemmt. ... Da hab' ich mich nicht eingeklemmt, sondern das war bestimmt 'n anderes Kind.“

Ein Junge, der beim ersten Termin zugestimmt hatte, sich einen Splitter zugezogen zu haben, der von einem Arzt habe entfernt werden müssen, äußerte beim darauffolgenden Termin: „Da hab' ich nochmal nachgedacht. ... Nee, das war nicht.“ Auf Hinweis auf seine frühere Schilderung: „Ja, das war glaub' ich was anderes.“ Beim dritten Termin: „Also, da hab' ich nachgedacht, weil's wegen meine, wegen meine Polypen waren. Nämlich, nämlich das wußt' ich nicht mehr, ob ich mir da 'n Splitter eingehauen hab' oder ... oder Polypen, ja.“

Ein Junge, dem suggeriert werden sollte, daß ihm ein anderes Kind eine Erbse in die Nase gesteckt habe, erwiderte beim ersten Termin: „Nee, das war keine Erbse. ... Das war, das war so'n Tuch, weil ich im Judo war, und da hatt' ich Nasenbluten. Das sah aus wie 'ne Erbse, aber das war keine“, beim zweiten Termin wiederum, „Das war keine Erbse. Mama wußte das nicht!“, beim dritten Termin, „Meine Mutter hat ... Dann hat mein Vater geschrieben. ... Meine Mutter weiß davon nichts.“ Beim vierten Termin gab der Junge an, „Aber da hat Mama es gedacht, daß es bestimmt 'ne Erbse war.“

Auf den Versuch, einem Mädchen einen Wespenstich zu suggerieren, erwiderte dieses: „Nein, aber ich hab' 'n Mückenstich. Hier bei Papa, bei Papa an See, da sind ja ganz viele Mücken.“

Ein Junge, dem vorgegeben wurde, er habe sich seinen Fuß verstaucht, äußerte: „Nee, mein' Fuß hab' ich noch nie verstaucht. ... Dann hat meine Mutti das nämlich falsch aufgeschrieben.“

Entsprechende Angaben wurden von insgesamt 22 Kindern (32.8 %) geliefert. Davon stellten 12 Kinder mehrfach, d.h. an verschiedenen Terminen derartige Überlegungen an. 8 Kinder lieferten im weiteren Verlauf der Interviews dennoch Schilderungen zum fiktiven Ereignis, während sich die übrigen 14 (20.9 %) als suggestionsresistent erwiesen. Damit haben offenbar alle Kinder, die sich in der Beeinflussungsphase durchgehend als suggestionsresistent erwiesen haben (N= 13; 19.4 %; vgl. Abschnitt 5.2.1.2), verbale Abgrenzungen der vorgenannten Art gegen die suggestiven Einflußnahmen vorgenommen.

Einwände gegen die Plausibilität des Ereignisses:

Als Einwände gegen die Plausibilität des Ereignisses wurden Angaben der Kinder aufgefaßt, die Gründe beinhalten, warum das Ereignis in der von der Interviewerin vorgegebenen Form gar nicht stattgefunden haben könne. Dabei waren in den Antworten der Kinder häufig Angaben oder Überlegungen zum möglichen Kontext vorhanden. Entweder wurde aber schon begründet, warum es den Kontext gar nicht gegeben haben könne, oder aber das vorgegebene Ereignis wurde als nicht kompatibel mit dem möglichen Kontext angesehen.

So äußerte ein Junge, auf die Bitte zu erzählen, wie es gewesen sei, als er barfuß in eine Scherbe getreten sei, daß dies nicht stimme; er sei nie da barfuß gelaufen, wo Scherben gelegen hätten.

Ein Junge, der gebeten wurde, zu schildern, wie es gewesen sei, als ihn im Streichelzoo eine Ente gebissen habe, antwortete: Selbst wenn es passiert wäre, wäre es „nicht so schlimm, ... weil Enten haben ja keine Zähne.“

Ein Mädchen erwiderte auf die Vorgabe, im Streichelzoo von einer Ziege gebissen worden zu sein: „Wenn sie's gemacht hätte, hätte ich es glaube ich gemerkt.“ Beim nächsten Termin gebeten, es sich

vorzustellen, „Ich glaub’ ich weiß, wie es ist, wenn man, wenn ‘ne Ziege mich beißen würde. ... Genau so, wie als mein Zwergkaninchen mich gebissen hat.“

Auf die Vorgabe, er sei auf einer Bootsfahrt, als es geschaukelt habe, seekrank geworden, antwortete ein Junge, „Also wenn es schaukelt, dann schlaf’ ich meistens ein.“

Ein Mädchen, dem vorgegeben wurde, es sei mit dem Fahrrad in einen Nadelbaum gefahren und habe sich an den Nadeln gekratzt, erwiderte: „Eigentlich war’n ja immer meine Eltern dabei, wenn ich Fahrrad gefahren bin. ... Und auf dem Hof, da war’n ja sowieso keine Nadelbäume.“ Beim nächsten Termin: „Das war ja nur auf’m Sand. Da waren ja keine Tannen, da hat Papa mich auch immer gehalten. ... Wenn im Wald, aber da war ich noch nicht mit’m Fahrrad.“

Einem Mädchen sollte suggeriert werden, es habe bei einer Schneeballschlacht einen Schneeball an die Lippe bekommen, so daß diese geblutet habe. Das Mädchen antwortete dazu beim ersten Termin: „Nö, der Schnee war zwar ganz hart, aber ich hab’ mich umgedreht und hab’ das auf den Rücken gekriegt.“ Beim zweiten Termin äußerte sie: „Nö, hab’ ich noch nie. ... Nein, sonst hätte Mutti ja schon mal gesagt, also hiervor. Oder da wär’ ‘n Photo von dagewesen.“ Auf den Hinweis der Interviewerin, daß das Ereignis möglicherweise schon sehr lange zurückliege, erwiderte das Mädchen: „Naja, aber als kleineres Kind werd’ ich ja noch keine Schneeballschlacht machen.“ Beim vierten Termin: „Aber wenn ich den Schneeball sehe, dann dreh’ ich mich ganz schnell um, dann krieg’ ich den nur an die Jacke.“

Auf Vorgabe, er sei im Streichelzoo von einem Esel gebissen worden, erwiderte ein Junge, daß er zwar in zwei verschiedenen Zooparks gewesen sei; in dem einen habe es aber gar keine Streichelwiese gegeben, in dem anderen seien keine Esel auf der Streichelwiese gewesen.

Ein Junge, dem suggeriert werden sollte, daß ihm ein anderes Kind eine Erbse in die Nase gesteckt habe, erwiderte, wenn es passiert wäre, „dann hätt’ ich gewußt, daß mir ‘n Kind ... Im Kindergarten hatt’ ich ‘n Freund, der heißt Jonny. Der hat’s nicht sowas gemacht.“

Bei Versuchen der Interviewerin, einem Mädchen einen möglichen Kontext vorzugeben, in welchem sie von einer Wespe gestochen worden sei, äußerte das Mädchen auf die Vorgabe, daß es möglicherweise bei der Mutter im Garten passiert sei, „Mama hat keinen Garten“, auf die Frage, ob es passiert sei, wenn sie mit der Mutter in den Park gegangen sei, „Ich bin immer alleine rausgegangen.“ Beim vierten Termin äußerte das Mädchen „Kann nicht, kann nirgendwo passiert sein. Und das ist gar nicht passiert. ... Sonst könnte ich mich ja erinnern.“

Derartige Plausibilitätsüberlegungen wurden von insgesamt neun Kindern (13.4 %) vorgenommen, wobei einige dieser Kinder entsprechende Argumente an mehreren Terminen anführten. Acht dieser Kinder (11.9 %) haben sich als durchgängig suggestionsresistent erwiesen; nur ein Kind hat die Suggestion übernommen und auch bei der Nachbefragung angegeben, daß das fiktive Ereignis tatsächlich stattgefunden habe.

Offenbar ging die Fähigkeit der Kinder, sich verbal gegen die suggestiven Einflußnahmen abzugrenzen, indem eigenständige Plausibilitätsüberlegungen angestellt oder Klärungsversuche unternommen wurden (anderes Ereignis schildern, auf anderes Kind verweisen, Irrtümer der Eltern anführen), mit einer erhöhten Suggestionsresistenz einher, d.h. wenn entsprechende Äußerungen vorgebracht wurden, war die Wahrscheinlichkeit geringer, daß die suggestiven Vorgaben übernommen wurden. Hingegen hatten Gegeninstruktionen der Eltern einen weniger eindeutigen Einfluß auf das Antwortverhalten der Kinder, da die Gegeninstruktionen durch erneute suggestive Einflußnahmen bei einigen Kindern wieder aufgehoben werden konnten.

Bei der Analyse der kindlichen Schilderungen hinsichtlich Plausibilitätsüberlegungen fiel allerdings auch auf, daß einige Kinder (vermeintliche) Außenkriterien anführten, die den Realitätsgehalt des fiktiven Ereignisses positiv belegen sollten.

So äußerte ein Junge, der beim ersten Termin anfänglich verneint hatte, von einer Feuerqualle berührt worden zu sein, noch während desselben Interviews: „Den Hautausschlag hab’ ick immer noch.“

Ein Mädchen, dem (erfolgreich) suggeriert worden war, vom Pferd gefallen zu sein, gab im Rahmen der Nachbefragung an, wegen dieses Ereignisses immer noch Angst vor dem Reiten zu haben.

Ein Mädchen bekräftigte das Zutreffen des fiktiven Ereignisses (Loch im Kopf durch Kippen) bei der Nachbefragung damit, daß ein entsprechendes Photo existiere.

Ein anderes Mädchen äußerte wiederholt, sich eigentlich nicht mehr an das fiktive Ereignis (in der Bustür eingeklemmt) erinnern zu können, führte aber an, daß ihre Cousine, die dabei gewesen sei, sich noch erinnern könne. Auf Befragen, ob sie jetzt immer schön vorsichtig sei, äußerte sie: „Ja, auch im Fahrstuhl und bei so ‘ne Einkaufstüren, ... bin ich ganz vorsichtig immer.“

Drei Kinder führten an, ihre Mutter habe ihnen gesagt, das fiktive Ereignis habe tatsächlich stattgefunden.

Entsprechende Belege für den Realitätsgehalt wurden von den Kindern aber auch in bezug auf die realen Ereignisse angeführt, wobei die Angaben der Kinder nicht immer plausibel waren:

Ein Junge zeigte der Interviewerin beim ersten Termin seinen verletzten Zeh, auf den er eine Holzplatte habe fallen lassen.

Zwei Kinder wiesen die Interviewerin auf Narben, die infolge der realen Verletzungen entstanden seien, hin.

Ein Mädchen äußerte auf die Frage, wieso sie sich an die (mindestens ein Jahr zurückliegende) Situation, als sie in die Brennesseln gefallen sei, noch so gut erinnern könne, „Weil es mir immer noch ein bißchen weh tut.“

5.2.5.2 Reflexionen über den eigenen Erinnerungsprozeß

Die Schilderungen der Kinder wurden dahin gehend betrachtet, inwiefern sie sogenannte Metakognitionen, d.h. Reflexionen, Gedanken oder Überlegungen zum eigenen Erinnerungsprozeß enthalten. Insbesondere sollte festgestellt werden, inwiefern die Angaben der Kinder darauf hindeuten, daß die Kinder sich gegebenenfalls über die suggestiven Prozesse bewußt waren und inwiefern sie ihre eigenen Bekundungen auf deren Quellen (tatsächliches Erlebnis versus Vorstellung bzw. Befragung) zurückführen konnten.

Zunächst erfolgte eine Analyse der Aussagen aller Kinder zu den fiktiven Ereignissen zu allen sechs Befragungszeitpunkten. Dabei wurden nur solche Angaben berücksichtigt, die über die Äußerung, sich nicht mehr (gut) erinnern zu können, es sei schon zu lange her, das Kind sei zum Zeitpunkt des Ereignisses noch zu klein gewesen, um sich jetzt zu erinnern, es könne sich an das reale Ereignis besser erinnern o.ä., hinausgehen. Entsprechende Äußerungen der Kinder, die in den unterschiedlichen Erhebungsphasen geliefert wurden, werden nachfolgend dargestellt.

Spontane Bekundungen der Kinder während der Beeinflussungsphase:

Nachdem ein Junge das fiktive Ereignis (Feuerqualle) zunächst verneint hatte, äußerte er dann: „Muß gewesen sein, als ich noch klein war.“ (Wie es gewesen sein könnte?) „Na als klein kann ich mich nicht mehr erinnern.“

Beim zweiten Termin auf das fiktive Ereignis (von Ente gebissen) angesprochen, das der Junge beim ersten Termin bejaht hatte, äußerte er, sich beim ersten Termin „ein bißchen verfehlt“ zu haben. Er sei von seinem Hund und nicht von einer Ente gebissen worden. „Habe ich mich ein bißchen vertäuscht.“

Nachdem ein Junge das fiktive Ereignis (Lausbefall) bei den ersten beiden Terminen verneint hatte und ihm inhaltliche Vorgaben von der Interviewerin gemacht worden waren, schilderte er beim dritten Termin

einige Details. Auf die Frage der Interviewerin, woher er das wisse, äußerte er: „Na, du hast es mir ja gesagt, wenn man Läuse hat und man sich dann kratzt, dann brennt es.“

Ein Mädchen, die das fiktive Ereignis (Erbsen in die Ohren gesteckt bekommen) bei den ersten beiden Terminen verneint hatte, beim zweiten Termin aber bei Anwendung der „Als-ob-Technik“ eine plastische Geschichte dazu erzählt hatte, lieferte (ohne subjektive Überzeugung) auch beim dritten Termin eine entsprechende Darstellung und äußerte zwischendurch: „Das hab’ ich nur mal mir so ausgedacht. Weil er (beschuldigter Junge) der Größte ist.“

Nach anfänglichem Zögern lieferte ein Junge beim ersten Termin eine kurze Darstellung des fiktiven Ereignisses (Tür an den Kopf bekommen) und fügte hinzu: „Aber ich denk’ mal, das war ein Traum, aber ich denk’ mal, der Traum war wahr.“ Beim zweiten Termin sagte der Junge zur Interviewerin, als diese das fiktive Ereignis etwas modifizieren wollte: „Du hattest doch letztes Mal gesagt, ich sollte mich überlegen, weil die Tür gegen mein Kopf geknallt ist.“ Nach kurzer Schilderung: „Das war mal passiert, das war aber passiert, und dann hatte ich das aber noch mal als Traum.“

Ein Mädchen äußerte beim vierten Termin, als sie von einer neuen Interviewerin zum fiktiven Ereignis befragt wurde: „Aber, na, die andere Tante (frühere Interviewerin), die hat gesagt, ich soll sie nur als Geschichte so aufsagen.“

Ein Junge revidierte seine Schilderung des fiktiven Ereignisses vom ersten Termin bei der zweiten Befragung, indem er äußerte: „Also, da hab’ ich eben mal falsch überlegt. Also, das erste Mal hab’ ich falsch überlegt, weil da war ich nicht.“ Auf entsprechende Aufforderung lieferte er dann bei diesem und beim nächsten Termin eine Schilderung im Konjunktiv. Beim vierten Termin von einer neuen Interviewerin auf das fiktive Ereignis angesprochen äußerte er: „Also, das war in echt jar nicht passiert, aber ich soll’s mir ausdenken, so, wie’s passieren könnte.“

Ein Mädchen äußerte beim zweiten Termin - vor Beginn des eigentlichen Interviews - von sich aus, sich an das fiktive Ereignis nicht mehr erinnern zu können. Sie fügte später auf nochmalige Nachfrage hinzu: „Ich werd’ mich auch nie erinnern können!“ Beim dritten Termin gefragt, ob es passiert sein könne, als sie noch klein gewesen sei, erwiderte das Mädchen: „Wenn man klein ist, dann kann man sich ja auch nicht mehr erinnern!“

Ein Junge äußerte beim vierten Termin auf die Frage nach dem fiktiven Ereignis, zu dem er bei den früheren Terminen bereits recht inkonsistente Angaben geliefert hatte, „Das weiß ich nicht so richtig. Aber daß meine Mutter das gesagt hat ... Ich denk’ ‘mal, das weiß ich auch nicht, aber ich denk’ mal ...“ (es folgte eine Schilderung des Ereignisses). Beim fünften Termin äußerte er wiederum zunächst, „Ach so, das weiß ich nicht mehr weiter. Die andere, meine andere (gemeint war die frühere Interviewerin) ... Aber wenn Mama das nicht meinte, dann denk’ ich ‘mal, daß die (frühere Interviewerin) was falsch gemacht hatte. ... Ich weiß nicht, ob das meine Mutter meinte. ... Ich denk’ mal, meine Mutter hat was anderes gedacht ...“ (es folgten wiederum unzusammenhängende Äußerungen zum Ereignis).

Bekundungen der Kinder im Rahmen des Experteninterviews auf Befragen:

Auf die Frage, ob sie sich noch gut an das fiktive Ereignis - in der Bustür eingeklemmt - erinnern könne, antwortete ein Mädchen: „Ja, das hab’ ich mir einfach so’n bißchen den Kopf eingebildet, und dann ging das schon immer wieder. ... Dann konnt’ ich immer wieder erzählen.“ (Ob sie es schon häufiger erzählt habe, beispielsweise Verwandten oder Bekannten?) „Nee, denen hab’ ich das nicht erzählt. Das hab’ ich mir dann erstmal als geheim in meinen Kopf getan.“

Auf Befragen, ob er das fiktive Ereignis (Sturz vom Klettergerüst mit Fußverletzung) schon einmal vergessen gehabt habe, äußerte ein Junge: „Das hatt’ ich immer in meinen Gedanken drin.“

Zunächst gab ein Mädchen auf Befragen zum realen Ereignis an, sich daran noch gut erinnern zu können. (Ob sie das immer gewußt habe, oder ob ihr das mal irgendwann wieder eingefallen sei?) „Das hab’ ich immer gewußt.“ (Und das fiktive Ereignis - Sturz von der Treppe - auch?) „Nein, das habe ich irgendwann mal vergessen.“ (Ob es ihr wieder eingefallen sei, als sie jemand danach gefragt habe?) „Nein, hat mich keiner nach gefragt.“

Bekundungen der Kinder im Rahmen der Nachbefragung auf Befragen:

Zum fiktiven Ereignis (mit Schere in den Finger geschnitten) befragt gab ein Mädchen, das im Verlauf der Befragungen dem Ereignis zugestimmt hatte, an, sich unsicher zu sein, ob es passiert sei: „Ich weiß

nicht, ob das wahr war. Ich weiß nicht, ich muß da noch mal meine Mutti fragen.“ Auf die Frage, woran sie sich noch erinnern könne: „Eigentlich, wo das erste Mal gefragt wurde, da konnt’ mich nur noch dran erinnern, daß ich mich vielleicht, vielleicht war’s auch wie ich größer war, mal, ich weiß ja nicht das Alter, und jemand anders war das, der als Kleines sich in’ Finger geschnitten hat.“

Zum fiktiven Ereignis gab ein Junge, der die Suggestion bei den früheren Terminen übernommen hatte, an, das stimme gar nicht, da habe er sich „verredet“.

Ein Mädchen äußerte zum fiktiven Ereignis (Hand in der Tür geklemmt), sich daran nicht erinnern, es sich aber „vorstellen“ zu können. Dieses Mädchen hatte die Suggestion im Verlauf der früheren Befragungen übernommen gehabt.

Auf das fiktive Ereignis (seekrank geworden) angesprochen, bejahte der Junge zunächst, sich erinnern zu können. Er sei sich jedoch nicht sicher, ob es passiert sei. Vielmehr hätte die andere Frau (frühere Interviewerin) gesagt, es sei ihm passiert. Eigentlich könne er sich gar nicht mehr daran erinnern und habe es auch damals nicht gekannt, als sie ihn gefragt habe.

Auf Befragen zum fiktiven Ereignis (in den Bach gefallen) äußerte ein Junge, der die Suggestion nie übernommen hatte, die anderen Interviewerinnen hätten gesagt, es sei passiert, aber er könne sich nicht daran erinnern.

Ein Mädchen verneinte, daß das fiktive Ereignis (mit Fahrrad in Nadelbaum gefahren) passiert sei und fügte hinzu, daß sie das auch nie erzählt habe, was ihrem tatsächlichen Aussageverhalten entspricht.

Ebenso äußerte ein Junge, der das fiktive Ereignis bei den vorangegangenen Befragungen vehement verneint hatte, daß das fiktive Ereignis nicht stimme und daß er das auch schon immer gesagt habe.

Noch während die Interviewerin die Instruktion für diesen Termin gab und über die früheren Befragungen durch andere Interviewerinnen sprach, äußerte ein Mädchen spontan: „Da wußte ich gar nicht mehr, daß ich mich da so an einem Wespenstich gestochen hatte.“ Sie könne sich an das fiktive Ereignis, dem sie nie zugestimmt hatte, nicht erinnern.

Ein Junge, der die suggestive Vorgabe, er sei im Streichelzoo von einem Esel gebissen worden, bis dahin nicht übernommen hatte, äußerte noch während der Erläuterung der Interviewerin, daß sie herausfinden wolle, ob die früheren Interviewerinnen bei seiner Befragung einen Fehler gemacht hätten, spontan, „wie das mit dem Esel.“

Ein Mädchen, die das fiktive Ereignis bei den ersten beiden Terminen bejaht hatte, dann aber von ihrer Mutter erfahren hatte, daß sich das Ereignis tatsächlich nicht ereignet hat, äußerte im Rahmen der Nachbefragung, daß sie „durcheinander“ gekommen sei; das habe sie zu Hause selber gemerkt.

Die Interviews, die von den uninformierten Experten beim fünften Befragungszeitpunkt geführt worden waren, wurden zusätzlich einer genauen inhaltlichen Analyse hinsichtlich des Vorhandenseins von Metakognitionen und Merkmalen der Wirklichkeitskontrolle unterzogen. Dabei wurden im Gegensatz zu der zuvor dargestellten Auswertung sowohl die Schilderungen zu den fiktiven als auch die zu den realen Ereignissen einbezogen. Ferner wurden nicht nur besonders prägnante Äußerungen der Kinder berücksichtigt, sondern auch bloße Angaben dazu, ob bzw. wie gut die Kinder sich an das jeweilige Ereignis erinnern könnten, ob sie sicher seien, daß es passiert sei, an welches der beiden relevanten Ereignisse sie sich besser erinnern könnten, ob sie sich schon immer daran hätten erinnern können u.ä. Es wurde differenziert zwischen Angaben, die für den Realitätsgehalt der Schilderung sprechen, und solchen, die dagegen sprechen. Des weiteren wurde unterschieden zwischen Angaben der Kinder, die spontan erfolgten, und Angaben, die auf konkrete Frage geliefert wurden, so daß sich insgesamt vier Kategorien ergaben. Schließlich wurde für jede Schilderung auf einer dreistufigen Rating-skala vermerkt, ob gar keine Angaben vorhanden waren oder ob sie gegebenenfalls eher schwach oder eher stark ausgeprägt waren. War dabei mehr als eine relevante Angabe pro Kategorie vorhanden, wurde die am stärksten ausgeprägte kodiert.

Die Ergebnisse zeigen, daß insgesamt nur drei spontane Bekundungen von den Kindern vorgebracht wurden, die als Metakognitionen bzw. Wirklichkeitskontrollkriterien zu klassifizieren waren. Ebenso ergab sich nur eine geringe Anzahl an entsprechenden Kriterien, die als eher stark ausgeprägt klassifiziert wurden (n= 8). Betrachtet man spontane und erfragte Angaben sowie schwach ausgeprägte und prägnante Kriterien gemeinsam, so waren in der Mehrzahl aller Schilderungen (61.4 %; n= 51) Äußerungen enthalten, die als Indikatoren für den Realitätsgehalt der Schilderungen gewertet wurden. Hingegen fanden sich nur in einem Drittel aller Schilderungen (33.7 %; n= 28) Äußerungen, die gegen den Realitätsgehalt sprachen.

Tabelle 6: Metakognitionen und Versuche der Wirklichkeitskontrolle

		alle Schilderungen (N= 83)		sugg. Schilderungen (N= 39)		erleb. Schilderungen (N= 44)	
		N	%	N	%	N	%
pro Realität, spontan	nicht vorhanden	81	97.6	38	97.4	43	97.7
	eher schwach	0	0	0	0	0	0
	eher stark	2	2.4	1	2.6	1	2.3
pro Realität, erfragt	nicht vorhanden	32	38.6	15	38.5	17	38.6
	eher schwach	50	60.2	24	61.5	26	59.1
	eher stark	1	1.2	0	0	1	2.3
contra Realität, spontan	nicht vorhanden	82	98.8	38	97.4	44	100.0
	eher schwach	1	1.2	1	2.6	0	0
	eher stark	0	0	0	0	0	0
contra Realität, erfragt	nicht vorhanden	55	66.3	24	61.5	31	70.5
	eher schwach	23	27.7	12	30.8	11	25.0
	eher stark	5	6.0	3	7.7	2	4.5

Schon bei Betrachtung der in Tabelle 6 dargestellten Ergebnisse zeigen sich kaum Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen. Dementsprechend ergaben weder Mann-Whitney-U-Tests für unabhängige Stichproben noch Wilcoxon-Tests für abhängige Stichproben für eine der vier Kategorien (spontane und erfragte Angaben, die für bzw. gegen den Realitätsgehalt sprechen) signifikante Unterschiede zwischen wahren und suggerierten Schilderungen hinsichtlich des Vorhandenseins bzw. der Ausprägung von Wirklichkeitskontrollkriterien und Metakognitionen.

5.2.5.3 Sonstige Besonderheiten im Antwortverhalten

Ein weiteres Phänomen im Antwortverhalten der Kinder wird nachfolgend erwähnt, weil es ebenfalls Informationen über die gedächtnismäßige Repräsentation und Reproduzierbarkeit der fiktiven Ereignisse verdeutlicht.

Wie bereits mehrfach erwähnt, wurde die fünfte Befragung von zwei Interviewern durchgeführt, die jeweils etwa die Hälfte der Kinder befragten. Dabei hatten die Interviewer insofern ein unterschiedliches Vorgehen, als Experte 1 den Kindern das jeweilige Ereignis konkret vorgab - und zwar in der Reihenfolge, in der die Verfasserin dies vorgeschlagen hatte -, während Experte 2 meistens (in 15 der insgesamt 21 durchgeführten Interviews) nur einen offenen Erzählstoß gab und den Kindern damit frei-

stellte, über welches der beiden relevanten Ereignisse sie zuerst berichten wollten. Auf entsprechende offene Frage schilderten fünf Kinder zunächst das reale Ereignis, fünf hingegen das fiktive. Drei Kinder nannten zunächst beide Ereignisse und fragten, welches davon sie schildern sollten. Weitere zwei Kinder schilderten zunächst das reale Ereignis, nannten dann aber auf erneute offene Frage, ob sie sich noch ein weiteres Mal wehgetan hätten, das fiktive Ereignis.

Diese Beobachtung, wenngleich nicht systematisch überprüft, legt die Vermutung nahe, daß die fiktiven und die realen Ereignisse gedächtnismäßig annähernd gleich stark repräsentiert sind bzw. für die Kinder eine ähnliche Relevanz haben.

5.3 Inhaltliche Qualität der kindlichen Schilderungen

5.3.1 Realkennzeichenanalyse

Wie in Abschnitt 4.5.2.1 dargestellt wurden die Schilderungen derjenigen Kinder, die von den Experten beim fünften Interviewtermin zu einem realen und einem fiktiven Ereignis befragt worden waren (N= 35), einer Realkennzeichenanalyse unterzogen. Die Realkennzeichen wurden jeweils in der Erstaussage und in der beim fünften Termin erhobenen Aussage der Kinder von zwei im Hinblick auf den Wahrheitsstatus uninformierten Beurteilerinnen kodiert. Geprüft wurde, ob sich suggerierte und erlebnisbegründete Schilderungen hinsichtlich des Vorhandenseins von Realkennzeichen voneinander unterscheiden. Darüber hinaus wurde untersucht, ob zwischen den Erstaussagen und den zum fünften Zeitpunkt erhobenen Aussagen qualitative Unterschiede bestehen. Die Erhebung und Auswertung dieses Untersuchungsabschnitts wurde von Böhm (1999) im Rahmen ihrer Diplomarbeit vorgenommen, so daß bezüglich der ausführlichen Ergebnisdarstellung auf diese Arbeit verwiesen wird. Im folgenden werden nur die zentralen Ergebnisse referiert.

Die Realkennzeichen „Schilderung von Komplikationen im Handlungsverlauf“ (7), „Korrekte Bezeichnung unverstandener Handlungselemente“ (10b), „Einwände gegen die Richtigkeit der eigenen Aussage“ (18) und „Ereignisspezifische Inhalte“ (19) wurden niemals signiert, so daß sie aus den weiteren Analysen ausgeschlossen wurden.¹⁰ Ferner ergaben sich bezüglich der Erstaussagen Bodeneffekte hinsichtlich vieler Realkennzeichen, d.h. sie wurden ausgesprochen selten signiert.

Die Interrater-Reliabilitäten, die mittels Cohen's Kappa und Pearson's Korrelationskoeffizient bestimmt wurden, wurden für die Aussagezeitpunkte und den Wahrheitsstatus der Schilderungen sowohl getrennt als auch gemeinsam sowie für jedes Realkennzeichen einzeln berechnet. Insgesamt ergaben sich mittlere bis hohe Übereinstimmungen, so daß nachfolgend nur die Ergebnisse, die unter Berücksichtigung der (gemittelten) Auswertungen beider Beurteilerinnen berechnet wurden, dargestellt werden (zu den für beide Beurteilerinnen getrennt ermittelten Ergebnissen vgl. Böhm, 1999).

Mittels t-Tests für abhängige Stichproben - verglichen wurden jeweils die erlebnisbasierte und die suggerierte Schilderung eines Kindes - wurde für beide Aussagezeitpunkte getrennt geprüft, ob sich erlebnisbegründete und suggerierte Schilderungen hin-

¹⁰ Die in Klammern stehenden Zahlenangaben hinter den einzelnen Glaubhaftigkeitsmerkmalen bezeichnen hier und im folgenden die Nummer des jeweiligen Merkmals in dem verwendeten Realkennzeichenkatalog, der sich im Anhang befindet.

sichtlich des Vorhandenseins und der Ausprägung der einzelnen Realkennzeichen voneinander unterscheiden.

Nahezu alle Realkennzeichen traten in den Erstaussagen in den Schilderungen über reale Ereignisse häufiger auf als in den Schilderungen über fiktive - ein umgekehrtes Verhältnis ergab sich lediglich für die Realkennzeichen „Schilderung von (vermuteten) Gefühlen Beteiligter“ (13b) und „Schilderung von Sanktionen“ (14b), die jedoch maximal einmal von jeder Auswerterin signiert wurden. Nachfolgend wird dargestellt, hinsichtlich welcher Realkennzeichen sich schwach bis hoch signifikante Unterschiede zwischen erlebnisbegründeten und suggerierten Aussagen ergaben.

Tabelle 7: Qualitative Unterschiede zwischen erlebnisbegründeten und suggerierten Erstaussagen

Realkennzeichen	p
Logische Konsistenz (1)	***
Sprunghaft-ungeordnete Darstellung (2)	**
Quantitativer Detailreichtum (3)	***
Schilderung von Ausgangssituationen (4a)	*
Zeitliche Verknüpfung I (objektive Zeitangaben) (4c)	*
Raum-zeitliche Verknüpfungen insgesamt (4)	***
Schilderung ausgefallener Einzelheiten (8)	*
Schilderung äußerlich beobachtbarer gefühlsbezogener motorischer oder physiologischer Abläufe (12b)	*
Schilderung eines Gefühlszustandes (12c)	*
Schilderung eigener psychischer Vorgänge insgesamt (12)	***

Bei der fünften suggestionsfreien Befragung, aber nach vorheriger wiederholter suggestiver Beeinflussung der Kinder, unterschieden sich realitätsbegründete und suggerierte Schilderungen nur noch hinsichtlich einiger sehr weniger Realkennzeichen bedeutsam voneinander.

Nur hinsichtlich der Glaubhaftigkeitsmerkmale „Quantitativer Detailreichtum“ (3) und „Entlastung anderer“ (15) ergaben sich zum fünften Befragungszeitpunkt bei einer Fehlerwahrscheinlichkeit von höchstens 5 % signifikante Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen - und zugleich in den Auswertungen beider Beurteilerinnen. Hinsichtlich der übrigen in Tabelle 8 aufgeführten Glaubhaftigkeitsmerkmale waren die Unterschiede bei Betrachtung der Auswertungen beider Beurteilerinnen nur schwach signifikant, und sie waren bei jeweils nur einer Auswerterin zu verzeichnen.

Wenngleich einige Realkennzeichen in suggerierten Schilderungen eine tendenziell höhere Ausprägung aufwiesen als in erlebnisbegründeten, war dieser Unterschied nur für das Realkennzeichen „Schilderung eigener Gedanken, die in bezug zur Handlung stehen“ (12d) und unter Berücksichtigung der Auswertungen nur einer Beurteilerin bedeutsam.

Tabelle 8: Qualitative Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen bei der fünften Befragung

Realkennzeichen	p
Logische Konsistenz (1)	*
Quantitativer Detailreichtum (3)	**
Zeitliche Verknüpfung I (objektive Zeitangaben) (4c)	*
Zeitliche Verknüpfung II (subjektiv bedeutsame Zeitangaben wie z.B. Geburtstage, Feiertage o.ä.) (4d)	*
Inhaltlich phänomengemäße Schilderung unverstandener Handlungselemente (10a)	*
Entlastung anderer (potentieller oder tatsächlicher) Schadensverursacher (15)	**

Um die Analyse der qualitativen Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen nicht nur auf die Betrachtung einzelner Realkennzeichen zu stützen, sondern um mögliche Qualitätsunterschiede zu erheben, die sich bei gleichzeitiger Einbeziehung aller Glaubhaftigkeitsmerkmale und auch unter Berücksichtigung der Interkorrelationen zwischen den Merkmalen ergeben, wurden für beide Aussagezeitpunkte getrennt simultane und schrittweise Diskriminanzanalysen durchgeführt.

Für die Erstaussagen ließen sich - unter Berücksichtigung der Auswertungen beider Beurteilerinnen und bei Zusammenfassung der im verwendeten Realkennzeichenkatalog ausdifferenzierten Merkmale - sowohl bei simultaner als auch bei schrittweiser Diskriminanzanalyse signifikante Diskriminanzfunktionen ermitteln: Die bei simultaner Diskriminanzanalyse theoretisch ermittelte Trefferquote lag bei 71.4 % (34.3 % falsch-negativ Klassifikationen; 22.9 % falsch-positiv Klassifikationen). Bei schrittweisem Vorgehen ging nur das Glaubhaftigkeitsmerkmal „Logische Konsistenz“ in die ermittelte Diskriminanzfunktion ein. Die theoretisch ermittelte Trefferquote lag hier mit 68.6 % etwas niedriger, wobei falsch-negativ Zuordnungen seltener vorkamen als bei simultanem Vorgehen (25.7 %), falsch-positiv Zuordnungen hingegen etwas häufiger (37.1 %). Demnach führt eine hohe logische Konsistenz in suggerierten Aussagen dazu, diese fälschlich als erlebnisbegründet zu klassifizieren.

Für die nach Beeinflussung im fünften Interview erhobenen Aussagen ergaben sich erheblich schlechtere Resultate als für die Erstaussagen. So ließen sich bei simultanem Vorgehen keine signifikanten Diskriminanzfunktionen ermitteln - weder bei gemeinsamer noch bei getrennter Berücksichtigung der Auswertung beider Beurteilerinnen sowie weder bei Zugrundelegung der ausdifferenzierten noch der zusammengefaßten Kriterien. Bei schrittweisem Vorgehen mit zusammengefaßten Kriterien ging nur das Merkmal „Entlastung anderer“ in die Ermittlung der signifikanten Diskriminanzfunktion ein, wobei die theoretisch ermittelte Trefferquote mit 55.7 % nur knapp über Zufallsniveau lag und es sich bei sämtlichen Fehlern um falsch-negativ Klassifikationen handelte; 88.6 % der realitätsbegründeten Schilderungen wurden fälschlich als suggeriert klassifiziert. Das Nichtvorhandensein des Merkmals „Entlastung anderer“, das ohnehin nur in realitätsbegründeten Schilderungen signiert wurde, führte bei Zugrundelegung der ermittelten Diskriminanzfunktion dazu, die Schilderung - unabhängig von ihrem tatsächlichen Realitätsgehalt - als unwahr zu klassifizieren. Legt man die ausdifferenzierten Kriterien zugrunde, so ging zusätzlich das Merkmal „Zeitliche Verknüpfungen II“ (subjektiv bedeutsame Zeitangaben wie Geburtstage, Feiertage o.ä.) in die Diskriminanzfunktion ein. Die Trefferquote erhöhte sich dann auf 62.9 %, wobei immer noch 51.4 % der erlebnis-

begründeten Schilderungen fälschlich als suggeriert klassifiziert wurden, hingegen nur 22.9 % der suggerierten fälschlich als erlebnisbegründet.¹¹

Um zu prüfen, ob und wenn ja, inwiefern sich die inhaltliche Qualität der erlebnisbegründeten und der suggerierten Schilderungen im Zeitverlauf verändert hat, wurden t-Tests für abhängige Stichproben mit dem Befragungszeitpunkt als unabhängige Variable und des Vorhandenseins der einzelnen Realkennzeichen als abhängige Variablen durchgeführt. Für die erlebnisbegründeten und suggerierten Schilderungen ist in Tabelle 9 und Tabelle 10 aufgeführt, hinsichtlich welcher Realkennzeichen sich Mittelwertsunterschiede ergeben haben.

Es wird deutlich, daß die suggerierten Schilderungen gegenüber den erlebnisbegründeten im Zeitverlauf insofern stärker an Qualität gewonnen haben, als sich bei einer größeren Anzahl von Realkennzeichen bedeutsame Unterschiede in Form einer Zunahme nachweisen ließen. Hinsichtlich einiger Merkmale war sowohl bei erlebnisbegründeten als auch bei suggerierten Schilderungen ein Anstieg zu verzeichnen, so daß sich der jeweilige Qualitätszuwachs möglicherweise auf unterschiedliche Interviewstrategien bei den beiden Befragungszeitpunkten zurückführen läßt. Andere Realkennzeichen zeigten weder bei erlebnisbegründeten noch bei suggerierten Schilderungen Veränderungen hinsichtlich ihrer Auftretenshäufigkeit. Dabei handelte es sich um Glaubhaftigkeitsmerkmale, die ohnehin nur sehr selten signiert wurden und die sich als wenig trennfähig erwiesen hatten („Interaktionsschilderung“, „Schilderung ausgefallener Einzelheiten“, „Schilderung nebensächlicher Einzelheiten“, „Indirekt handlungsbezogene Schilderung“, „Schilderung äußerlich nicht unbedingt beobachtbarer gefühlsbezogener motorischer oder physiologischer Abläufe“, „Selbstkritisches Einschätzen bzw. Zugeben von Fehlverhalten“, „Schilderung von Sanktionen“).

Die Durchführung univariater Varianzanalysen (mit Meßwiederholung) mit den Faktoren „Aussagezeitpunkt“ und „Realitätsgehalt der Aussage“ sollte Aufschluß über Interaktionseffekte zwischen diesen beiden Variablen, d.h. über möglicherweise unterschiedliche qualitative Veränderungen in Abhängigkeit vom Wahrheitsstatus der Aussagen geben. Signifikante Interaktionseffekte traten hinsichtlich der Realkennzeichen „Logische Konsistenz“ (nur auf dem 10 %-Niveau signifikant), „Inhaltlich phänomengemäße Darstellung unverstandener Handlungselemente“ sowie „Entlastung anderer (potentieller oder tatsächlicher) Schadensverursacher“ zutage. Unter Berücksichtigung der jeweiligen Mittelwerte sind die Ergebnisse dahin gehend zu interpretieren, daß das Merkmal „Logische Konsistenz“ in suggerierten Schilderungen im Zeitverlauf weit stärker anstieg als in erlebnisbegründeten, während sich bezüglich der anderen beiden Merkmale das umgekehrte Verhältnis ergab, wobei zu ergänzen ist, daß das Merkmal „Entlastung anderer“ in suggerierten Schilderungen zu keinem Zeitpunkt signiert wurde, das Merkmal „Inhaltlich phänomengemäße Darstellung unverstandener Handlungselemente“ in suggerierten Schilderungen nur zum ersten Zeitpunkt.

Aus den vorliegenden Ergebnissen läßt sich ableiten, daß eine qualitative Annäherung der suggerierten Aussagen an die erlebnisbegründeten stattgefunden hat und die suggerierten Schilderungen mit wiederholter Befragung an Qualität, insbesondere an logischer Konsistenz gewonnen haben.

¹¹ Schlechtere Trefferquoten beim fünften im Vergleich zum ersten Befragungszeitpunkt ergaben sich auch bei Analyse der Daten anhand des Summenmodells von Dahle (1997; Dahle & Wolf, 1997; zu den Ergebnissen im einzelnen vgl. Böhm, 1999).

Tabelle 9: Qualitative Unterschiede in erlebnisbegründeten Aussagen zwischen dem ersten und fünften Befragungszeitpunkt

Realkennzeichen	p
Logische Konsistenz (1)	**
Sprunghaft-ungeordnete Darstellung (2)	***
Quantitativer Detailreichtum (3)	*
Schilderung von Ausgangssituationen (4a)	***
Räumliche Verknüpfungen (4b)	*
Zeitliche Verknüpfungen II (subjektiv bedeutsame Zeitangaben wie z.B. Geburtstage oder Feiertage) (4d)	***
Raum-zeitliche Verknüpfungen insgesamt (4)	***
Wiedergabe von Gesprächen (6a)	*
Wiedergabe von Einzeläußerungen (6b)	*
Wiedergabe von Gesprächen insgesamt (6)	***
Inhaltlich phänomengemäße Schilderung unverstandener Handlungselemente (10a)	*
Schilderung eines Gefühlszustandes (12c)	*
Schilderung eigener psychischer Vorgänge insgesamt (12)	**
Schilderung psychischer Vorgänge von Beteiligten insgesamt (13)	*
Entlastung anderer (potentieller oder tatsächlicher) Schadensverursacher (15)	**
Spontane Verbesserungen der eigenen Aussage (16)	*
Eingeständnis von Wissenslücken (17b)	*

Tabelle 10: Qualitative Unterschiede in suggerierten Aussagen zwischen dem ersten und fünften Befragungszeitpunkt

Realkennzeichen	p
Logische Konsistenz (1)	***
Sprunghaft-ungeordnete Darstellung (2)	***
Quantitativer Detailreichtum (3)	***
Schilderung von Ausgangssituationen (4a)	***
Räumliche Verknüpfungen (4b)	**
Zeitliche Verknüpfungen I (objektive Zeitangaben) (4c)	**
Zeitliche Verknüpfungen II (subjektiv bedeutsame Zeitangaben wie z.B. Geburtstage oder Feiertage) (4d)	***
Raum-zeitliche Verknüpfungen insgesamt (4)	***
Wiedergabe von Gesprächen (6a)	*
Wiedergabe von Einzeläußerungen (6b)	***
Wiedergabe von Gesprächen insgesamt (6)	***
Schilderung äußerlich beobachtbarer gefühlsbezogener motorischer oder physiologischer Abläufe	**
Schilderung eines Gefühlszustandes (12c)	***
Schilderung eigener Gedanken (12d)	*
Schilderung eigener psychischer Vorgänge insgesamt (12)	***
Schilderung von (vermuteten) Gefühlen Beteiligter	*
Schilderung psychischer Vorgänge von Beteiligten insgesamt (13)	**
Eingeständnis von Erinnerungslücken im eigentlichen Sinne (17a)	**
Eingeständnis von Erinnerungslücken insgesamt (17)	***

5.3.2 Beurteilung globaler Qualitätsmerkmale der Schilderungen sowie des Aussageverhaltens der Kinder

Da in realen Begutachtungssituationen zwar grundsätzlich auch eine Qualitätseinschätzung der Aussage anhand von Realkennzeichen vorgenommen wird, jedoch die einzelnen Glaubhaftigkeitsmerkmale meist nicht signiert und ausgezählt werden, sondern häufig auch globale Qualitätseinschätzungen vorgenommen werden, sollte dieses Vorgehen auch in der eigenen Untersuchung simuliert werden. Daher wurden die Experten, die das fünfte Interview mit den Kindern geführt hatten, gebeten, im Anschluß an jedes Interview auf vierstufigen Ratingskalen einzuschätzen, wie stark verschiedene globale Merkmale (Detailliertheit, logische Konsistenz, Konstanz, Plausibilität und Qualität der Aussage sowie subjektive Sicherheit des Kindes) in den Aussagen bzw. im Verhalten der Kinder ausgeprägt waren und wie bedeutsam diese und weitere Merkmale (nonverbales Verhalten, Aussageverhalten, spezielle Äußerungen des Kindes, sonstige Merkmale) für ihre Einschätzung des Realitätsgehalts der Schilderungen waren. Ferner wurden dieselben Experten sowie ein dritter Experte aufgefordert, diese Einschätzungen noch einmal anhand der Videoaufzeichnungen und Transkripte der Schilderungen vorzunehmen. Die Erhebung dieser Daten diente insbesondere der Klärung der Frage, ob sich erlebnisbegründete und suggerierte Schilderungen in der Ausprägung der vorgenannten Merkmale unterscheiden und welche Merkmale Experten bevorzugt für ihre Beurteilung des Realitätsgehalts der Schilderungen heranziehen.

Nach Abwägung von meßtheoretischen und pragmatischen Überlegungen wird bei der Auswertung und Interpretation der erhobenen Daten Intervallskalenniveau angenommen, zumal im Rahmen einer Vorauswertung festgestellt wurde, daß Verfahren, die intervallskalierte Daten voraussetzen, und Verfahren, die für Ordinalskalen entwickelt wurden, zu weitgehend übereinstimmenden Aussagen führten (vgl. zum Ganzen Bortz, 1984, 1985).

Weitere Überlegungen sind zu der Frage anzustellen, ob bei den vorliegenden Daten die Anwendung von Verfahren für abhängige oder für unabhängige Stichproben indiziert ist. Sofern von jedem Kind eine erlebnisbegründete und eine suggerierte Schilderung erhoben und ausgewertet wurde, kann von einem Untersuchungsdesign mit Meßwiederholung bzw. von abhängigen bzw. verbundenen Stichproben ausgegangen werden. Wie in Abschnitt 4.3.2 dargestellt wurden von den Experten aber nicht nur Interviews mit den 38 Kindern geführt, die über je ein reales und ein fiktives Ereignis berichteten, sondern auch mit sechs Kindern, die über zwei reale Ereignisse berichteten, und mit einem Kind, das zwei fiktive Ereignisse schilderte. Aus der nachfolgenden Auswertung wurden die erlebnisbegründeten Schilderungen einiger Kinder ausgeschlossen, da hinsichtlich dieser Schilderungen erhebliche Zweifel am Realitätsgehalt bzw. der Zuverlässigkeit bestanden oder die Kinder angaben, sich an diese Ereignisse nicht mehr erinnern zu können und keine entsprechende Schilderung lieferten. Darüber hinaus verneinte ein Junge beim fünften Interviewtermin, das fiktive Ereignis erlebt zu haben, so daß auch seine Schilderung zum realen Ereignis ausgeschlossen wurde. Somit wurden die Schilderungen über insgesamt 83 Ereignisse in die nachfolgenden Auswertungen einbezogen, davon 39 Aussagen über fiktive Ereignisse, 44 über reale.

Bei Auswahl eines Verfahrens für abhängige Stichproben müßte eine deutliche Reduzierung der Stichprobengröße in Kauf genommen werden, da nur die Schilderungen derjenigen Kinder, die über je ein reales und ein fiktives Ereignis berichtet haben, be-

rücksichtigt werden könnten (Stichprobengröße von $N=83$ bei Annahme unabhängiger Stichproben versus $N=68$ bzw. 34 Paarvergleiche bei Annahme abhängiger Stichproben). Hinsichtlich der in Abschnitt 5.3.2.3 referierten Auswertungen, bei denen nur die von den Experten als zutreffend oder als falsch klassifizierten Schilderungen berücksichtigt werden, würde bei Annahme abhängiger Stichproben eine weitere erhebliche Stichprobenreduktion eintreten, da nur diejenigen Aussagen der Kinder einbezogen werden könnten, deren beide Aussagen (suggerierte und erlebnisbegründete) von den Experten entweder zutreffend oder falsch klassifiziert wurden.

Da in Verfahren mit Meßwiederholung individuelle, weitgehend situations- und zeit-unabhängige Besonderheiten der einzelnen Versuchspersonen berücksichtigt und eliminiert werden, werden derartige a priori-Unterschiede bei Verfahren für unabhängige Stichproben doppelt berücksichtigt, wenn von jeder Versuchsperson mehrfach Daten erhoben und miteinander verglichen werden. Dies hat zur Folge, daß bei abhängigen Stichproben Mittelwertsunterschiede von Verfahren, die das Meßwiederholungs-Design berücksichtigen, eher aufgedeckt werden als bei Anwendung von Verfahren ohne Meßwiederholung. Kommen im vorliegenden Fall - um eine Reduzierung der Stichprobengröße zu vermeiden - also Verfahren für unabhängige Stichproben ohne Berücksichtigung des Meßwiederholungs-Designs zur Anwendung, so sind die Ergebnisse dahin gehend zu interpretieren, daß aufgedeckte Unterschiede bei Anwendung von Verfahren mit Meßwiederholung noch deutlicher zutage treten würden.

Bezüglich der durchgeführten Diskriminanzanalysen ist darüber hinaus darauf hinzuweisen, daß diese nur für Designs ohne Meßwiederholung konzipiert sind, ein äquivalentes Verfahren für abhängige Stichproben hingegen nicht zur Verfügung steht. Aufgrund ihrer insgesamt hohen Aussagekraft erscheint aber auch die Durchführung von Diskriminanzanalysen im Sinne einer explorativen Strategie gerechtfertigt.

5.3.2.1 Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen hinsichtlich der Ausprägung der Merkmale

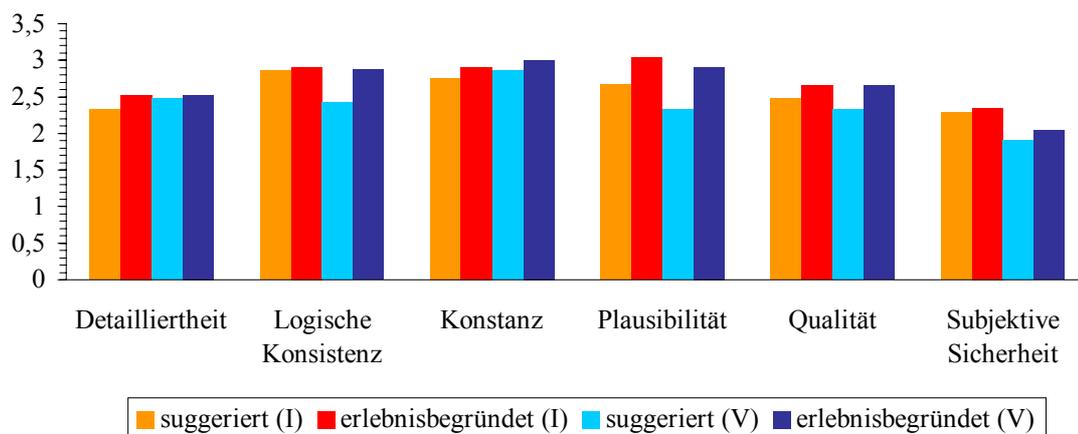
Um zu überprüfen, ob Unterschiede in den Einschätzungen der beiden Experten zur Ausprägung der Merkmale „Detailliertheit“, „logische Konsistenz“, „Konstanz“, „Plausibilität“, „Qualität“ sowie „subjektive Sicherheit des Kindes“ zwischen erlebnisbegründeten und suggerierten Schilderungen bestehen, werden in den folgenden Abbildungen die mittleren Ausprägungen (Minimum: 1, Maximum: 4) der einzelnen Merkmale für suggerierte und erlebnisbegründete Schilderungen sowie für alle drei Experten getrennt - für die Experten 1 und 2 zusätzlich differenziert nach Interview- und Video-bedingung (nachfolgend abgekürzt mit „I“ bzw. „V“) - dargestellt.¹² Um festzustellen, welche Merkmale sich als besonders geeignet zur Unterscheidung von erlebnisbegründeten und suggerierten Schilderungen erwiesen haben, wurden außerdem sowohl simultane (Einschluß aller Merkmale) als auch schrittweise Diskriminanzanalysen sowie einfaktorielle Varianzanalysen durchgeführt.

Bei Berücksichtigung der Auswertungen von Experte 1 in der Interviewsituation zeigen sich nur geringfügige Unterschiede in der Ausprägung der Merkmale zwischen erlebnisbegründeten und suggerierten Schilderungen. Dementsprechend führte weder die simultane noch die schrittweise Diskriminanzanalyse zu signifikanten Resultaten, d.h.

¹² Angaben zu Mittelwerten, Standardabweichung und Stichprobengröße finden sich in Tabelle 1 im Anhang.

die mittleren Werte der Diskriminanzfunktionen unterscheiden sich in beiden Gruppen nicht voneinander. Ebenso zeigen die Ergebnisse der einfaktoriellen Varianzanalyse bezüglich der von Experte 1 ausgewerteten Schilderungen keinerlei Unterschiede in der Ausprägung der Merkmale beim Vergleich von erlebnisbegründeten und suggerierten Schilderungen.

Abbildung 4: Einschätzungen von Experte 1 zur Ausprägung der Merkmale in suggerierten und erlebnisbegründeten Schilderungen

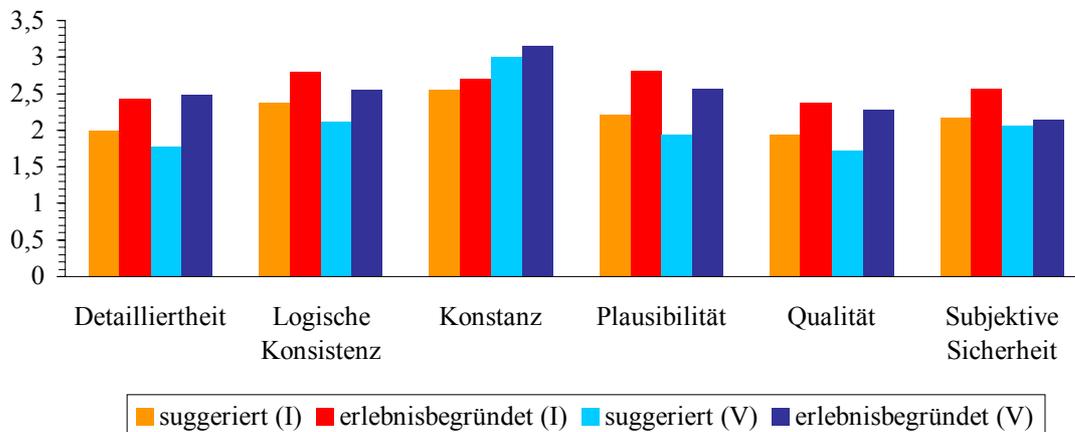


Betrachtet man diejenigen Auswertungen, die von Experte 1 in der Videobedingung vorgenommen wurden, so zeigen sich deutliche Unterschiede sowohl hinsichtlich des Merkmals „Plausibilität“ als auch hinsichtlich des Merkmals „Logische Konsistenz“ (zu den Ergebnissen der univariaten Varianzanalyse vgl. Tabelle 12). Entsprechend führte die schrittweise Diskriminanzanalyse zu einem signifikanten Resultat, wobei nur das Merkmal „Plausibilität“ in die ermittelte Diskriminanzfunktion einbezogen wurde (vgl. Tabelle 11 und Tabelle 12).

Hinsichtlich der von Experte 2 in der Interviewsituation vorgenommenen Auswertungen zeigten sich mit Ausnahme des Merkmals „Konstanz“ bei allen anderen Merkmalen deutlich höhere Ausprägungen bei den erlebnisbegründeten im Vergleich zu den suggerierten Schilderungen. In die schrittweise Diskriminanzanalyse ist jedoch wiederum nur das Merkmal „Plausibilität“ eingegangen (die simultane Diskriminanzanalyse lieferte kein signifikantes Resultat).

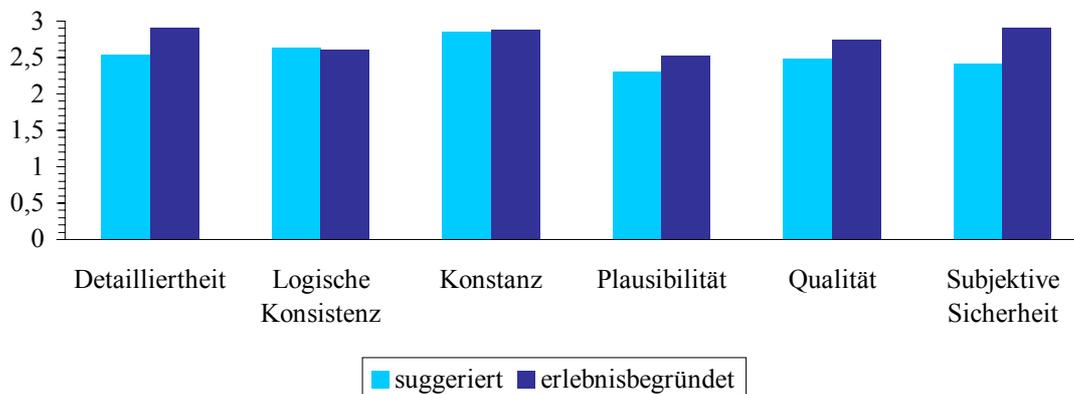
Bezüglich der von Experte 2 in der Videobedingung vorgenommenen Auswertungen zeigten sich in bezug auf die Merkmale „Konstanz“ und „subjektive Sicherheit“ kaum, ansonsten deutliche Unterschiede. Sowohl die simultane als auch die schrittweise Diskriminanzanalyse lieferten (schwach) signifikante Resultate. In die schrittweise Diskriminanzanalyse ist nur das Merkmal „Qualität“ eingegangen.

Abbildung 5: Einschätzungen von Experte 2 zur Ausprägung der Merkmale in suggerierten und erlebnisbegründeten Schilderungen



Bezüglich Experte 3 lieferten sowohl die schrittweise als auch die simultane Diskriminanzanalyse signifikante Resultate. In die schrittweise Diskriminanzanalyse wurde lediglich das Merkmal „subjektive Sicherheit“ aufgenommen.

Abbildung 6: Einschätzungen von Experte 3 zur Ausprägung der Merkmale in suggerierten und erlebnisbegründeten Schilderungen



Mit Ausnahme der simultanen Diskriminanzanalyse auf Grundlage der von Experte 2 in der Videobedingung vorgenommenen Auswertungen (allerdings ergab sich hier nur ein schwach signifikantes Resultat hinsichtlich der Diskriminationsfähigkeit der ermittelten Funktion) fallen die Eigenwerte der ermittelten Diskriminanzfunktionen (Verhältnis der Quadratsumme zwischen den Gruppen zu der Quadratsumme innerhalb der Gruppen) ebenso wie die Korrelationen zwischen den Werten der Diskriminanzfunktion und der Gruppenzugehörigkeit recht unbefriedigend aus.

Tabelle 11: Kennwerte der ermittelten Diskriminanzfunktionen

		Eigenwert	Kanonische Korrelation	Wilks' Lambda	χ^2	df	p
E1-V	schritt.	.199	.408	.834	7.549	1	.006
E2-I	schritt.	.189	.399	.841	5.809	1	.016
E2-V	simultan	.434	.550	.697	11.182	6	.083
	schritt.	.170	.381	.855	5.292	1	.022
E3	simultan	.177	.388	.850	12.069	6	.060
	schritt.	.100	.301	.909	7.288	1	.007

Die folgende Tabelle gibt Aufschluß darüber, wie hoch die verschiedenen - im Falle schrittweiser Diskriminanzanalysen auch die nicht einbezogenen Merkmale - mit den Werten der Diskriminanzfunktionen korrelieren (Strukturkoeffizient, in nachfolgender Tabelle abgekürzt mit „S-K“). Darüber hinaus sind in der Tabelle die mittels einer ein-faktoriellen Varianzanalyse ermittelten F-Werte angegeben, die unabhängig von der Diskriminanzanalyse Auskunft über die Diskriminationsfähigkeit einzelner Merkmale geben.

Tabelle 12: Strukturkoeffizienten und F-Werte

		Detail- liertheit	Logische Konsist.	Konstanz	Plausibi- lität	Qualität	Subj. Sicherh.
E1-V	F-Wert	.050	7.630***	.383	8.378***	1.962	.358
	S-K (schritt.)	.198	.593	.188	1.000	.427	.129
E2-I	F-Wert	2.846*	3.651*	.482	7.566***	3.434*	3.330*
	S-K (schritt.)	.367	.391	.341	1.000	.495	.370
E2-V	F-Wert	.670**	3.169*	.189	5.578**	5.723**	.087
	S-K (simultan)	.611	.534	.076	.597	.626	.088
	S-K (schritt.)	.342	.144	.192	.569	1.000	.511
E3	F-Wert	4.482**	.008	.010	1.299	2.066	7.567***
	S-K (simultan)	.520	-.050	.063	.264	.312	.751
	S-K (schritt.)	.137	.148	-.026	.170	.059	1.000

Zusammenfassend ist festzuhalten, daß die Ausprägung aller Merkmale von den Experten in den erlebnisbegründeten Schilderungen zumindest tendenziell höher eingeschätzt wurde als in den suggerierten (Ausnahme: Experte 3 hinsichtlich des Merkmals „Logische Konsistenz“). Den Einschätzungen von Experte 1 zufolge kommt insbesondere den Merkmalen „Plausibilität“ und „logische Konsistenz“ Bedeutung für die Unterscheidung zwischen erlebnisbegründeten und suggerierten Schilderungen zu. Betrachtet man die Auswertungen von Experte 2, so weisen ebenfalls die Merkmale „Plausibilität“ und „logische Konsistenz“, aber auch die Merkmale „Qualität“ und „Detailiertheit“ und in begrenztem Maße auch das Merkmal „subjektive Sicherheit“ Trennfähigkeit auf. Legt man die Auswertungen von Experte 3 zugrunde, so kommt neben dem Merkmal „subjektive Sicherheit des Kindes“ dem Merkmal „Detailiertheit“ besondere Aussagekraft zu. Bei keinem Auswerter erwies sich die „Konstanz“ als geeignet, um zwischen erlebnisbegründeten und suggerierten Schilderungen zu differenzieren.

Nachfolgend wird aufgeführt, inwieweit die Schilderungen aufgrund der ermittelten Diskriminanzfunktionen zutreffend bzw. falsch klassifiziert werden.¹³ Auf Grundlage aller Schilderungen lagen die Trefferquoten, die aufgrund der verschiedenen Diskriminanzanalysen ermittelt wurden, bei etwa zwei Dritteln. Betrachtet man die Ergebnisse für suggerierte und erlebnisbegründete Schilderungen getrennt, so zeigen die simultanen Diskriminanzanalysen kaum Unterschiede, während bei Einbeziehung von jeweils nur einem Merkmal im Rahmen schrittweiser Diskriminanzanalysen deutliche Unterschiede in den Trefferquoten bei erlebnisbegründeten und suggerierten Schilderungen sichtbar wurden (Ausnahme: Auswertungen von Experte 2 in der Interviewsituation).

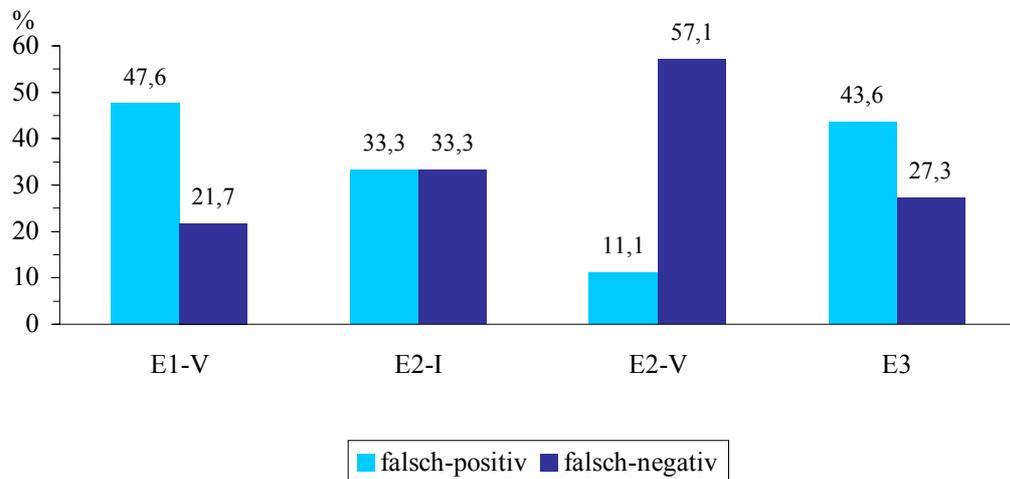
Betrachtet man beispielsweise die Ergebnisse, die sich auf Grundlage der Auswertungen von Experte 1 in der Plausibilität fälschlich als erlebnisbegründet zu klassifizieren (hoher Anteil an falsch-positiv-Klassifikationen). Die Auswertungen von Experte 2 in der Videobedingung zeigen, daß die ausschließliche Berücksichtigung der Aussagequalität dazu führt, daß erlebnisbegründete Schilderungen, wenn sie eine geringe Qualität aufweisen, nicht als solche erkannt werden (hoher Anteil an falsch-negativ-Klassifikationen). Legt man die Auswertungen Videobedingung ergeben, so zeigt sich, daß bei alleiniger Berücksichtigung der Plausibilität einer Schilderung die Gefahr besteht, suggerierte Schilderungen aufgrund ihrer von Experte 3 zugrunde, so führt eine hohe subjektive Sicherheit der Kinder offenbar zur Klassifikation ihrer Schilderungen als erlebnisbegründet, unabhängig von deren tatsächlichem Realitätsgehalt (hoher Anteil an falsch-positiv-Klassifikationen).

Tabelle 13: Theoretische Treffer- und Fehlerquoten aufgrund simultaner und schrittweiser Diskriminanzanalysen

		alle Schilderungen		suggerierte Schilderungen		erlebnisbegründete Schilderungen	
		falsch	richtig	falsch	richtig	falsch	richtig
E1-V	schrittweise (Plaus.)	34.1 % (15)	65.9 % (29)	47.6 % (10)	52.4 % (11)	21.7 % (5)	78.3 % (18)
E2-I	schrittweise (Plaus.)	33.3 % (13)	66.7 % (26)	33.3 % (6)	66.7 % (12)	33.3 % (7)	66.7 % (14)
E2-V	simultan	30.7 % (12)	69.2 % (27)	27.8 % (5)	72.2 % (13)	33.3 % (7)	66.7 % (14)
	schrittweise (Qual.)	35.9 % (14)	64.1 % (25)	11.1 % (2)	88.9 % (16)	57.1 % (12)	42.9 % (9)
E3	simultan	34.9 % (29)	65.1 % (54)	33.3 % (13)	66.7 % (26)	36.4 % (16)	63.6 % (28)
	schrittweise (subj. S.)	34.9 % (29)	65.1 % (54)	43.6 % (17)	56.4 % (22)	27.3 % (12)	72.7 % (32)

¹³ Da die Diskriminanzfunktion so ermittelt wird, daß die Trefferquote in der verwendeten Stichprobe maximal wird, ergeben sich bei Anwendung auf eine andere Stichprobe niedrigere Trefferquoten. Eine bereinigte Trefferquote läßt sich durch eine Kreuzvalidierung gewinnen, d.h. wenn jeder Fall durch die Funktion klassifiziert wird, die von allen anderen Fällen außer dem zu klassifizierenden abgeleitet wird. Im vorliegenden Fall sowie in allen nachfolgenden Diskriminanzanalysen führte die Kreuzvalidierung bei schrittweisem Vorgehen zu identischen Resultaten, bei simultanem Vorgehen zu schlechteren Trefferquoten.

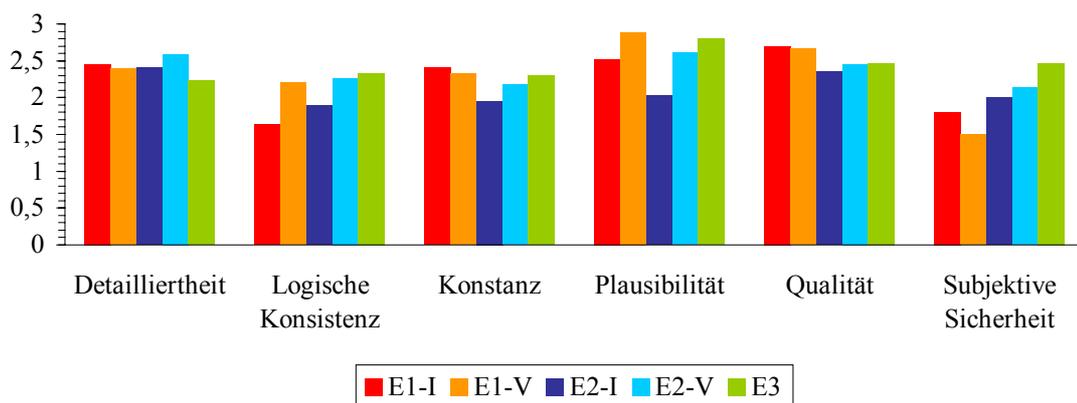
Abbildung 7: Falsch-positiv- und falsch-negativ-Klassifikationen auf Grundlage schrittweiser Diskriminanzanalysen



5.3.2.2 Einschätzungen zur Bedeutsamkeit der Merkmale

Die Experten wurden nicht nur gebeten, die Ausprägung der einzelnen Merkmale in den Schilderungen einzuschätzen, sondern auf ebenfalls vierstufigen Ratingskalen anzugeben, für wie bedeutsam sie die einzelnen Merkmale für ihre Einschätzung des Realitätsgehalts gehalten haben. Nachfolgend sind für die drei Experten getrennt die Mittelwerte bezüglich der Bedeutsamkeit der einzelnen Merkmale angegeben (vgl. auch Tabelle 14).

Abbildung 8: Bedeutsamkeit der Merkmale



Die Ergebnisse zeigen, daß sich die Einschätzungen der Experten bezüglich der Merkmale „Detailliertheit“ und „Qualität“ am ähnlichsten waren, während sich in bezug auf das Merkmal „subjektive Sicherheit“ die größten Diskrepanzen ergaben. So maß Experte 1 - besonders in der Videobedingung - dem Merkmal „subjektive Sicherheit“ geringe Bedeutung bei, während Experte 3 dieses Merkmal für eher bedeutsam hielt. Dies

deckt sich insofern mit den Einschätzungen zur Ausprägung der Merkmale, als sich das Merkmal „subjektive Sicherheit“ bei Experte 3 als das beste Merkmal zur Unterscheidung zwischen erlebnisbegründeten und suggerierten Schilderungen erwies, bei Experte 1 als eher schlechtes. Hingegen war die „Plausibilität“ einer Aussage in den Videoauswertungen von Experte 1 ein geeignetes Differenzierungsmerkmal und wurde von Experte 1 auch subjektiv für (vergleichsweise) recht bedeutsam gehalten. Auch bei Experte 2 stimmen die Einschätzungen zur Bedeutsamkeit der einzelnen Merkmale mit ihrer tatsächlichen Trennfähigkeit überein (verhältnismäßig große Aussagekraft der Merkmale „Detailliertheit“, „Plausibilität“ (Videobedingung) und „Qualität“).

Tabelle 14: Bedeutsamkeit der Merkmale

		Detail- liertheit	Logische Konsistenz	Konstanz	Plausibili- tät	Qualität	Subjektive Sicherheit
E1-I	M	2.45	1.64	2.41	2.52	2.70	1.80
	SD	.82	.61	.79	.70	.76	.67
	N	44	44	44	44	44	44
E1-V	M	2.39	2.20	2.32	2.89	2.66	1.50
	SD	.87	.76	.77	.72	.81	.70
	N	44	44	44	44	44	44
E2-I	M	2.41	1.89	1.95	2.03	2.35	2.00
	SD	.98	.97	.87	.80	.82	.86
	N	37	37	38	37	37	39
E2-V	M	2.59	2.26	2.18	2.61	2.45	2.13
	SD	.98	.97	.87	.80	.82	.86
	N	39	39	39	39	39	39
E3	M	2.23	2.33	2.30	2.80	2.47	2.47
	SD	.90	.98	1.00	.97	.82	1.16
	N	83	83	83	83	83	83

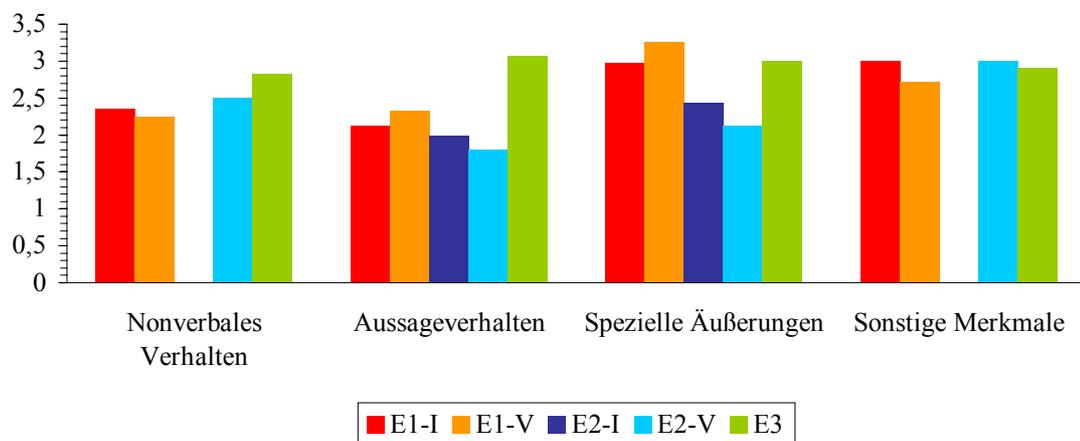
Weitere Merkmale, deren Bedeutsamkeit für die Beurteilung des Realitätsgehalts die Experten angeben sollten, sind „nonverbales Verhalten“, „Aussageverhalten“, „spezielle Äußerungen“ und „sonstige Merkmale“. Bezüglich dieser Merkmale sollten die Experten die entsprechenden Verhaltensweisen bzw. Äußerungen der Kinder - wenn vorhanden - nennen und deren Bedeutsamkeit ebenfalls auf vierstufigen Ratingskalen einschätzen.

Die Experten 2 und 3 (vor allem E2-I) lieferten kaum Angaben zu den vier vorgenannten Merkmalen, während Experte 1 allein zum Merkmal „spezielle Äußerungen“ unter beiden Auswertungsbedingungen in über der Hälfte der Fälle präzisierende Angaben lieferte und dieses Merkmal für recht bedeutsam hielt. Ebenso maß er den „sonstigen Merkmalen“ große Bedeutung bei. Experte 3 schätzte die Bedeutsamkeit aller vier Merkmale eher hoch ein.

Tabelle 15: Bedeutsamkeit weiterer Aussage- und Verhaltensmerkmale

		Nonverbales Verhalten	Aussageverhalten	Spezielle Äußerungen	Sonstige Merkmale
E1-I	M	2.35	2.13	2.97	3.00
	SD	.49	.35	.61	.67
	N	17 (38.6%)	15 (34.1%)	30 (68.2%)	10 (22.7%)
E1-V	M	2.25	2.33	3.26	2.71
	SD	.46	.58	.45	.49
	N	8 (18.2 %)	3 (6.8 %)	23 (52.3 %)	7 (15.9 %)
E2-I	M	---	2.00	2.44	---
	SD	---	---	.73	---
	N	0	1 (2.6%)	9 (23.1%)	0
E2-V	M	2.50	1.80	2.13	3.00
	SD	.71	.84	.99	---
	N	2 (5.1 %)	5 (12.8 %)	8 (20.5 %)	1 (2.6 %)
E3	M	2.83	3.08	3.00	2.90
	SD	.94	.95	1.00	1.20
	N	12 (14.5 %)	13 (15.7 %)	7 (8.4 %)	10 (12.0 %)

Abbildung 9: Bedeutsamkeit weiterer Aussage- und Verhaltensmerkmale



Besonders bei Experte 1, aber auch bei Experte 2 zeigte sich, daß die Schilderungen meist zutreffend klassifiziert wurden, wenn zu mindestens einem der vier Merkmale - insbesondere aber zu den Merkmalen „spezielle Äußerungen“ und „sonstige Merkmale“ - Angaben vorlagen. Offenbar haben somit einzelne, besonders prägnante Äußerungen der Kinder zu einer zutreffenden Klassifikation beigetragen. Bei Experte 3 war dieser Effekt weniger stark ausgeprägt.

5.3.2.3 Interaktionseffekte zwischen dem tatsächlichen Realitätsgehalt der Schilderungen und der Klassifikation (zutreffend-falsch) durch die Experten

Nachfolgend soll die Annahme überprüft werden, daß eine hohe Ausprägung der Merkmale bei erlebnisbegründeten Schilderungen zu zutreffenden Klassifizierungen durch die Experten führt, bei suggerierten zu falschen; bezüglich niedriger Ausprägungen der Merkmale wird das umgekehrte Verhältnis erwartet.

Betrachtet man nur die Schilderungen, die von den Experten zutreffend oder falsch klassifiziert wurden, und differenziert jeweils zwischen erlebnisbegründeten und suggerierten Schilderungen, so zeigt sich, daß bei Betrachtung nur der von den Experten zutreffend klassifizierten Schilderungen die Ausprägungen aller zu beurteilenden Merkmale bei den erlebnisbegründeten Schilderungen tendenziell höher eingeschätzt wurden als bei den suggerierten Aussagen. Hinsichtlich der falsch klassifizierten Schilderungen ergab sich - insbesondere unter Berücksichtigung der verschiedenen Experten - ein uneinheitliches Bild (vgl. Tabelle 2 im Anhang).

Betrachtet man nur die suggerierten Schilderungen so ergeben sich überwiegend höhere Ausprägungen der Einzelmerkmale bei den falschen verglichen mit den zutreffenden Klassifizierungen. Bezüglich der erlebnisbegründeten Schilderungen zeigt sich tendenziell das umgekehrte Verhältnis.

Zur Überprüfung der statistischen Bedeutsamkeit dieser Wechselwirkungseffekte wurden zweifaktorielle Varianzanalysen mit den Faktoren „Realitätsgehalt“ und „Klassifikation durch die Experten“ durchgeführt. Es wurden sowohl univariate Varianzanalysen durchgeführt, um die Effekte in bezug auf jedes einzelne Merkmal festzustellen, als auch multivariate Varianzanalysen mit gleichzeitiger Berücksichtigung aller Merkmale.

Tabelle 16: Interaktionseffekte (F-Werte)

	E1-I	E2-I	E2-V	E3
Detailliertheit	9.554***	4.127**	3.077*	n.s.
Logische Konsistenz	11.578***	n.s.	10.834***	n.s.
Konstanz	20.029***	n.s.	n.s.	3.992**
Plausibilität	18.120***	6.851**	7.940***	n.s.
Qualität	20.361***	5.363**	4.109**	n.s.
Subjektive Sicherheit	6.378**	9.484***	n.s.	n.s.
MANOVA	4.879***	n.s.	6.000**	n.s.

Bezogen auf die Auswertungen von Experte 1 in der Interviewsituation ergaben sowohl die univariaten Varianzanalysen für die Einzelmerkmale als auch die multivariate Varianzanalyse signifikante Interaktionseffekte, hingegen keine signifikanten Resultate hinsichtlich der Haupteffekte. Somit kann die Annahme, daß eine hohe Ausprägung der Merkmale bei erlebnisbegründeten Schilderungen zu zutreffenden Klassifizierungen führt, bei suggerierten zu falschen - bezüglich niedriger Ausprägungen der Merkmale wird das umgekehrte Verhältnis erwartet -, als bestätigt angesehen werden. Berücksichtigt man, daß in den Einschätzungen von Experte 1 in der Interviewsituation aber kaum Unterschiede in der Ausprägung der Merkmale zwischen erlebnisbegründeten und suggerierten Schilderungen zutage traten (vgl. Abschnitt 5.3.2.1, Abbildung 4), so ist zu

erwarten, daß Experte 1 den Realitätsgehalt der suggerierten und erlebnisbegründeten Schilderungen relativ häufig falsch einschätzte (zu den entsprechenden Ergebnissen vgl. Abschnitt 5.4.1).

Hingegen zeigte sich hinsichtlich der Auswertungen von Experte 1 in der Videobedingung, daß bei den zutreffend klassifizierten Schilderungen alle Merkmale in realen Aussagen höher ausgeprägt waren als in suggerierten, während sich bei den falsch klassifizierten Schilderungen nur hinsichtlich der Merkmale „Detailliertheit“ und „Qualität“ erwartungskonforme Unterschiede zwischen erlebnisbegründeten und suggerierten Aussagen zeigten. Da Experte 1 darüber hinaus nur jeweils zwei reale und fiktive Schilderungen falsch eingeschätzt hatte, lieferten weder univariate Varianzanalysen noch eine multivariate Varianzanalyse signifikante Haupt- oder Interaktionseffekte.

Die multivariate Varianzanalyse bezüglich der Auswertungen von Experte 2 in der Interviewsituation lieferte weder signifikante Haupteffekte noch einen bedeutsamen Interaktionseffekt. Die univariaten Varianzanalysen zeigten nur in bezug auf die Merkmale „Detailliertheit“, „Plausibilität“, „Qualität“ und „subjektive Sicherheit“ signifikante Interaktionseffekte. Bezüglich der Einschätzungen von Experte 2 in der Videobedingung ergaben die univariaten Varianzanalysen für die Merkmale „Detailliertheit“, „logische Konsistenz“, „Plausibilität“ und „Qualität“ signifikante Interaktionseffekte; hingegen zeigten sich keine signifikanten Resultate hinsichtlich der Haupteffekte. Ebenso zeigte auch die multivariate Varianzanalyse lediglich einen signifikanten Interaktionseffekt, jedoch keine signifikanten Resultate hinsichtlich der Haupteffekte.

Berücksichtigt man die Auswertungen von Experte 2 zur Ausprägung der Einzelmerkmale in suggerierten und erlebnisbegründeten Schilderungen (vgl. Abschnitt 5.3.2.1, Abbildung 5 sowie Tabelle 12), so wird deutlich, daß Unterschiede zwischen den beiden Gruppen von Schilderungen mit dem Auftreten von Interaktionseffekten korrespondieren. Somit dürfte insbesondere die Berücksichtigung der Merkmale „Detailliertheit“, „Plausibilität“ und „Qualität“ bei Experte 2 in beiden Auswertungsbedingungen zu guten Trefferquoten geführt haben (vgl. Abschnitt 5.4.1).

Hinsichtlich der Einschätzungen von Experte 3 zeigte die multivariate Varianzanalyse einen (schwach) signifikanten Haupteffekt für den Faktor „Realitätsgehalt“ ($F=6.000$; $p=.092$). Ferner ergaben univariate Varianzanalysen bezüglich der Merkmale „Detailliertheit“ ($F=3.098$; $p=.082$) und „subjektive Sicherheit“ ($F=6.144$; $p=.015$) signifikante Haupteffekte für den Faktor „Realitätsgehalt“, d.h. ihre Ausprägung in erlebnisbegründeten Schilderungen wurde höher beurteilt als die in suggerierten. Da jedoch keine Interaktionseffekte auftraten, hatte die Ausprägung dieser Merkmale in suggerierten und erlebnisbegründeten Aussagen offenbar keine Bedeutung für die Zuordnung des Realitätsgehalts der Schilderungen, d.h. Experte 3 hat die potentielle Diskriminierungsfähigkeit der Merkmale nicht ausgeschöpft. Lediglich für das Merkmal „Konstanz“ ergab sich ein signifikanter Interaktionseffekt. Da sich hier jedoch keine signifikanten Resultate hinsichtlich der Haupteffekte zeigten, d.h. suggerierte und erlebnisbegründete Schilderungen unterschieden sich nicht hinsichtlich ihrer Konstanz, kommt der Berücksichtigung der Konstanz keine Aussagekraft für die Unterscheidung zwischen erlebnisbegründeten und suggerierten Schilderungen zu.

5.3.2.4 Übereinstimmung zwischen Einschätzungen, die von verschiedenen Experten bzw. zu verschiedenen Zeitpunkten vorgenommen wurden

Die vorgenannten Ergebnisse zeigen, daß die verschiedenen Experten die einzelnen Aussagemerkmale offenbar recht unterschiedlich ausgelegt, beurteilt und verwertet haben. Um weiter zu überprüfen, inwieweit zwischen den Einschätzungen von Experte 3 und denen der Experten 1 bzw. 2 Übereinstimmungen bestehen, wurden Korrelationskoeffizienten nach Spearman sowie Kappa-Koeffizienten berechnet.¹⁴ Die Ergebnisse zeigen kaum Übereinstimmungen zwischen den Einschätzungen der Experten zur Ausprägung der einzelnen Merkmale - unabhängig davon, ob suggerierte und erlebnisbegründete Geschichten getrennt oder gemeinsam betrachtet wurden. Mittlere (auf dem 5 %-Niveau signifikante) Korrelationskoeffizienten nach Spearman ergaben sich beim Vergleich der Einschätzungen von Experte 1 und Experte 3 lediglich für das Merkmal „Detailliertheit“ ($r_s = .55$), für den Vergleich der Beurteilungen von Experte 2 und Experte 3 nur für das Merkmal „Konstanz“ ($r_s = .60$).

Bedeutsame Übereinstimmungen zwischen den Experten hinsichtlich der Einschätzung der Bedeutsamkeit der einzelnen Merkmale für die Beurteilung des Realitätsgehalts waren nicht nachweisbar.

Vergleicht man die Einschätzungen der Experten 1 und 2, die sie im Rahmen der Interviewsituation abgegeben haben, mit denen, die sie im Rahmen der Auswertung der Videoaufzeichnungen und Transkripte vorgenommen haben, so ergeben sich hinsichtlich der Ausprägung aller Merkmale insgesamt niedrige bis mittlere Korrelationen nach Spearman (r_s im Bereich von .34 bis .73) sowie eher niedrige Kappa-Koeffizienten. Die ermittelten Übereinstimmungsmaße für das Merkmal „logische Konsistenz“ lagen bei beiden Experten deutlich unter denen betreffend die anderen Merkmale.

Bezüglich der Einschätzungen zur Bedeutsamkeit der Merkmale für die Beurteilung des Realitätsgehalts der Schilderungen ergaben sich nur bezüglich der Auswertungen von Experte 2 und nur hinsichtlich der Merkmale „Detailliertheit“, „Konstanz“ und „subjektive Sicherheit“ (auf dem 5 %-Niveau) signifikante, wenn auch eher niedrige Zusammenhänge (r_s im Bereich von .42 bis .50) zwischen den Einschätzungen in der Interviewsituation und bei Auswertung der Videoaufzeichnungen und Transkripte.

Zusammenfassend ist festzuhalten, daß die verschiedenen Experten die einzelnen Merkmale offenbar unterschiedlich ausgelegt und eingeschätzt haben, was sich zum einen darin zeigt, daß sich bei den verschiedenen Experten unterschiedliche Merkmale als geeignet erwiesen haben, zwischen erlebnisbegründeten und suggerierten Schilderungen zu trennen, zum anderen darin, daß die verschiedenen Experten die Bedeutsamkeit der Merkmale für die Unterscheidung zwischen erlebnisbegründeten und suggerierten Schilderungen dementsprechend auch unterschiedlich bewertet haben. Interindividuelle Unterschiede in der Beurteilung der Einzelmerkmale spiegeln auch die geringen Übereinstimmungen in den Einschätzungen von Experte 3 und denen der Experten 1 und 2 wider. Auch die Beobachtung, daß zwischen den beiden Auswertungszeitpunkten jeweils nur mittlere Übereinstimmungen in den Einschätzungen der Experten 1 und 2 bestehen, deutet auf eine breite Auslegbarkeit der Einzelmerkmale hin.

¹⁴ Da die Experten 1 und 2 jeweils unterschiedliche Schilderungen bewertet haben, ließ sich ein entsprechender Vergleich zwischen den Einschätzungen dieser beiden Experten nicht durchführen.

5.4 Beurteilung des Realitätsgehalts der Schilderungen

Nachfolgend wird der Frage nachgegangen, inwieweit verschiedene Beurteiler unter unterschiedlichen Auswertungsbedingungen dazu in der Lage sind, den Realitätsgehalt der erlebnisbegründeten und der suggerierten Schilderungen, die zum fünften Erhebungszeitpunkt erhoben worden waren, zutreffend einzuschätzen. Dazu wurden zum einen die beiden Auswerterinnen, die anhand der Transkripte der Schilderungen Realkennzeichen kodiert hatten, sowohl nach dem ersten Durchlesen der Transkripte zu jeder der beiden Schilderungen eines Kindes als auch nach Abschluß der Realkennzeichenanalyse über beide Schilderungen zu ihren Einschätzungen des Realitätsgehalts befragt. Ihnen wurde eine fünfstufige Ratingskala mit den Polen *sehr unwahrscheinlich* versus *sehr wahrscheinlich* vorgegeben. Wie in Abschnitt 4.5.2.1 dargestellt wurden von beiden Auswerterinnen Schilderungen von 35 Kindern zu je einem realen und einem fiktiven Ereignis ausgewertet.¹⁵

Des weiteren wurden die Experten, die beim fünften Befragungszeitpunkt die Interviews geführt hatten, im Rahmen der Interviewsituation gebeten, den Wahrheitsstatus der Aussagen zu beurteilen. Außerdem wurden dieselben Experten sowie ein dritter Experte gebeten, zu einem späteren Zeitpunkt eine erneute Einschätzung anhand der Videoaufzeichnungen und Transkripte der Interviews vorzunehmen. Die Einschätzungen wurden wiederum sowohl unmittelbar nach Erheben bzw. Betrachten der Videoaufzeichnung jeder einzelnen Schilderung vorgenommen als auch nach ausführlicher Beschäftigung mit beiden Schilderungen eines Kindes (vgl. Abschnitt 4.5.3).

Während die Experten 1 und 2 jeweils unterschiedliche Schilderungen beurteilten, traf Experte 3 Einschätzungen sowohl bezüglich der von Experte 1 erhobenen und ausgewerteten Schilderungen als auch bezüglich der von Experte 2 beurteilten Schilderungen. Bezüglich der Beschreibung der Stichprobe von Schilderungen, die von diesen drei Experten beurteilt und in die nachfolgenden Analysen einbezogen wurden, wird auf die Ausführungen in den Abschnitten 4.3.2 und 5.3.2 verwiesen.¹⁶ Die Einschätzungen aller Auswerter zur Wahrscheinlichkeit des Realitätsgehalts der Schilderungen, die zu den verschiedenen Erhebungszeitpunkten und unter den unterschiedlichen Bedingungen (Transkripte, Interviewsituation, Videoaufzeichnungen und Transkripte) erhoben worden waren, wurden dahin gehend analysiert, inwieweit sie mit dem tatsächlichen Wahrheitsstatus der Schilderungen übereinstimmen. Dabei gilt als eindeutig zutreffende Zuordnung, wenn eine Extremeinschätzung in der Form vorgenommen wurde, daß der Realitätsgehalt einer erlebnisbegründeten Schilderung als *sehr wahrscheinlich*, der einer suggerierten als *sehr unwahrscheinlich* beurteilt wurde. Als tendenziell zutreffende Einschätzung gilt die Einschätzung des Realitätsgehalts einer erlebnisbegründeten Schilderung als *eher wahrscheinlich*, die des Realitätsgehalts einer suggerierten Schilderung als *eher unwahrscheinlich*. Bei eindeutig und tendenziell falschen Zuordnungen gilt jeweils

¹⁵ Die Schilderung über das reale Ereignis eines dieser Kinder wird nicht mit in die nachfolgenden Berechnungen einbezogen, da aufgrund der Angaben der Eltern und der starken Inkonsistenzen und mangelnden Kohärenz der Aussage erhebliche Zweifel am Realitätsgehalt bzw. der Zuverlässigkeit bestehen. Somit beziehen sich die nachfolgenden Auswertungen auf die Einschätzungen von 35 fiktiven und 34 erlebnisbegründeten Schilderungen seitens der beiden Raterinnen.

¹⁶ Aus den Analysen betreffend die Videoauswertungen von Experte 2 mußten zwei Schilderungen ausgeschlossen werden, da Experte 2 im Zeitverlauf zwischen Interviewsituation und Auswertung der Videos Kenntnis über den tatsächlichen Wahrheitsstatus der Schilderungen dieser beiden Kinder erhalten hatte.

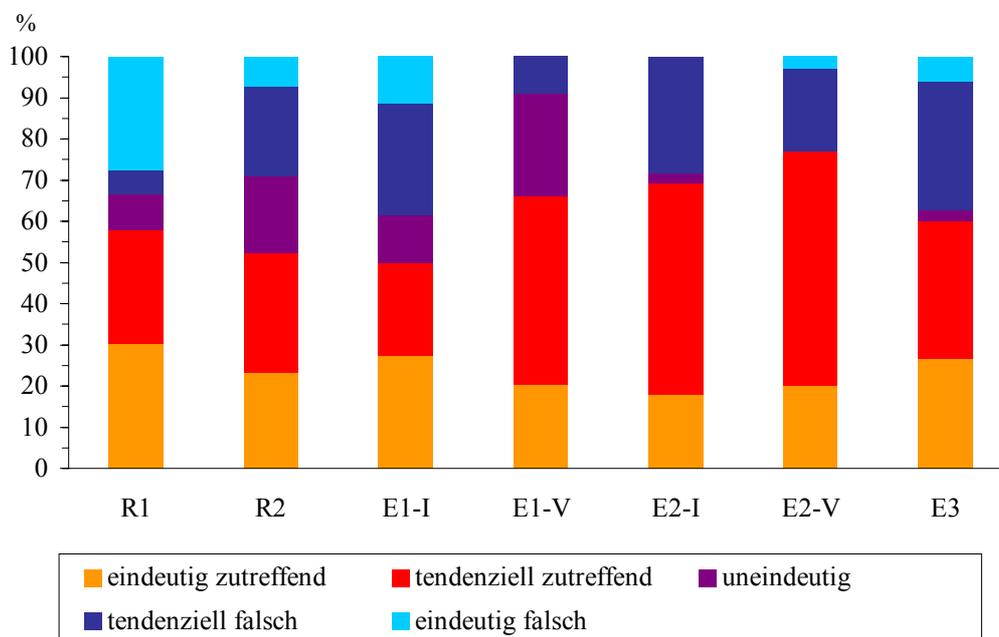
das umgekehrte Verhältnis. Uneindeutige Klassifikationen bestehen in der Wahl der Mittelkategorie.

Ferner wurde untersucht, inwieweit bei jedem einzelnen Auswerter Übereinstimmungen in den Beurteilungen bestehen, die zu den verschiedenen Einschätzungszeitpunkten (bei den beiden Raterinnen, die die Realkennzeichen in den Transkripten der Aussagen signiert hatten - in nachfolgenden Tabellen und Abbildungen als R1 und R2 bezeichnet - und dem an der Befragung der Kinder unbeteiligten Experten je zwei Zeitpunkte, bei den beiden Interviewer-Experten, die die Kinder auch befragt hatten, je vier) vorgenommen wurden, und inwieweit Übereinstimmungen zwischen den einzelnen Auswertern bestehen.

5.4.1 Treffer- und Fehlerquoten auf Grundlage aller Schilderungen

Nachfolgend wird dargestellt, inwieweit die verschiedenen Auswerter zutreffende und falsche Klassifizierungen der Schilderungen vorgenommen haben. Dargestellt werden zunächst die Einschätzungen auf Grundlage aller Schilderungen (erlebnisbegründeter und suggerierter), die nach ausführlicher Beschäftigung mit jeweils beiden Aussagen eines Kindes vorgenommen wurden.

Abbildung 10: Treffer- und Fehlerquoten aller Auswerter auf Grundlage aller von den einzelnen Auswertern beurteilten Schilderungen



Die Ergebnisse zeigen, daß alle Beurteiler unter allen Auswertungsbedingungen höhere Treffer- als Fehlerquoten erzielten (jeweils eindeutig und tendenziell zutreffende bzw. falsche Zuordnungen zusammen betrachtet). So lagen die Trefferquoten zwischen 50 % und 77 %, die Fehlerquoten zwischen 9 % und 39 %, die Anteile uneindeutiger Zuordnungen zwischen 0 % und 25 % (vgl. auch Tabelle 17).

Betrachtet man nur die eindeutigen Klassifikationen so liegen die Trefferquoten im Bereich zwischen 18 % und 30 %, die Fehlerquoten im Bereich zwischen 0 % und 28 %. Sofern Extremeinschätzungen vorgenommen wurden, waren diese überwiegend zutreffend (R1: 53 %, R2: 76 %, E1-I: 71 %, E1-V: 100 %, E2-I: 100 %, E2-V: 88 %, E3: 82 % zutreffende Extremeinschätzungen).

Die Auswerter, die ihre Einschätzungen auf die Videoaufzeichnungen und Transkripte der Schilderungen stützen konnten, sowie Experte 2 in der Interviewsituation erzielten zumindest tendenziell bessere Ergebnisse (hinsichtlich Experte 3 nur geringfügig) als die beiden Auswerterinnen, die nur die Transkripte der Schilderungen zur Verfügung hatten.

Dementsprechend lagen die von den beiden Raterinnen und die von Experte 1 in der Interviewsituation vorgenommenen Trefferquoten (tendenziell und eindeutig richtige Zuordnungen) nur im Zufallsbereich, während die drei Experten bei der nachträglichen Auswertung der Videoaufzeichnungen sowie Experte 2 in der Interviewsituation Trefferquoten über Zufallsniveau erzielten (E1-V: $\chi^2(1)= 4.455$; $p= .035$; E2-I: $\chi^2(1)= 5.769$, $p= .016$; E2-V: $\chi^2(1)= 10.314$, $p= .001$; E3: $\chi^2(1)= 3.482$, $p= .062$).

Während Experte 1 sich in seinen Einschätzungen bei der nachträglichen Auswertung der Videoaufzeichnungen und Transkripte deutlich gegenüber seinen Einschätzungen in der Interviewsituation verbesserte, war dieser Effekt bei Experte 2 nur schwach ausgeprägt.

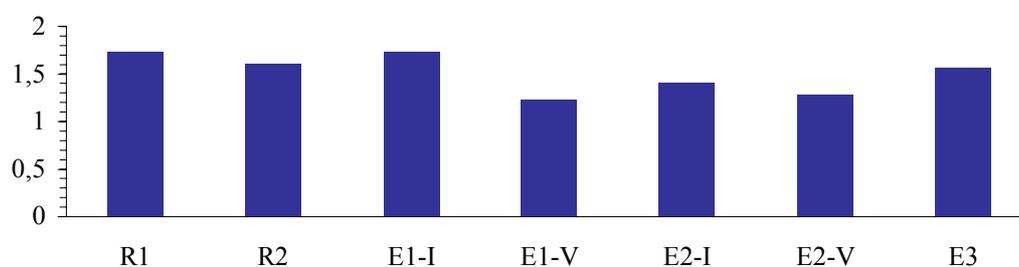
Tabelle 17: Treffer- und Fehlerquoten aller Auswerter auf Grundlage aller von den einzelnen Auswertern beurteilten Schilderungen

	R1 (N= 69)	R2 (N= 69)	E1-I (N= 44)	E1-V (N= 44)	E2-I (N= 39)	E2-V (N= 35)	E3 (N= 83)
eindeutig zutreffend	30.4 % (21)	23.2 % (16)	27.3 % (12)	20.5 % (9)	17.9 % (7)	20 % (7)	26.5 % (22)
tendenziell zutreffend	27.5 % (19)	29 % (20)	22.7 % (10)	45.5 % (20)	51.3 % (20)	57.1 % (20)	33.7 % (28)
uneindeutig	8.7 % (6)	18.8 % (13)	11.4 % (5)	25.0 % (11)	2.6 % (1)	0 % (0)	2.4 % (2)
tendenziell falsch	5.8 % (4)	21.7 % (15)	27.3 % (12)	9.1 % (4)	28.2 % (11)	20 % (7)	31.3 % (26)
eindeutig falsch	27.5 % (19)	7.2 % (5)	11.4 % (5)	0 % (0)	0 % (0)	2.9 % (1)	6.0 % (5)

Da die Festlegung, wann die Klassifikation einer Schilderung (noch) als zutreffend, wann als uneindeutig und wann als falsch gilt, letztlich per Definition erfolgt ist, besteht eine andere Herangehensweise, um zu überprüfen, inwieweit die einzelnen Auswerter den Realitätsgehalt der Schilderungen zutreffend eingeschätzt haben und inwieweit ihre Einschätzungen vom tatsächlichen Realitätsgehalt der Schilderungen abweichen, in der Berechnung eines sogenannten Abweichungsscores. Setzt man zwischen den einzelnen Stufen der Ratingskala mit den Polen *sehr wahrscheinlich* und *sehr unwahrscheinlich* Äquidistanz voraus, so kann zunächst für jede einzelne Schilderung bestimmt werden,

inwieweit die Einschätzungen der verschiedenen Auswerter vom tatsächlichen Realitätsgehalt (*sehr wahrscheinlich* bei erlebnisbegründeten Schilderungen, *sehr unwahrscheinlich* bei suggerierten) abweichen und dann ein mittlerer Abweichungsscore über alle Schilderungen berechnet werden. Die Abweichung kann null bis vier Punkte betragen. Wird der Realitätsgehalt einer erlebnisbegründeten Schilderung beispielsweise für wahrscheinlich gehalten, so ergibt sich eine Abweichung um einen Punkt vom tatsächlichen Realitätsgehalt, bei einer Einschätzung als sehr unwahrscheinlich eine Abweichung von vier Punkten, bei einer Einschätzung als sehr wahrscheinlich eine Abweichung von null. Somit läßt sich für jeden Auswerter ein durchschnittlicher Abweichungsscore berechnen, wobei ein niedriger Score auf ein gutes Ergebnis, ein hoher Score auf ein schlechtes Resultat verweist.

Abbildung 11: Abweichungsscores der einzelnen Beurteiler



Die ermittelten Abweichungsscores bestätigen die zuvor dargestellten Ergebnisse insofern, als die Experten 1 und 2 in der Videobedingung die besten Resultat erzielten, die Raterinnen sowie Experte 1 in der Interviewsituation die schlechtesten.

Tabelle 18: Mittlere Abweichungsscores und Standardabweichungen

	R1	R2	E1-I	E1-V	E2-I	E2-V	E3
M	1.74	1.61	1.73	1.23	1.41	1.29	1.57
SD	1.63	1.26	1.42	0.89	1.09	1.10	1.34
n	69	69	44	44	39	35	83

Nachfolgend wird geprüft, inwieweit zwischen den Einschätzungen der Experten 1 und 2, die sie im Rahmen der Interviewsituation vorgenommen haben, und denen, die sie aufgrund der Auswertung der Videoaufzeichnungen und Transkripte getroffen haben, Übereinstimmungen bestehen. Dabei wurde insofern eine dreistufige Skala zugrunde gelegt, als die Stufen *sehr wahrscheinlich* und *wahrscheinlich* sowie die Stufen *sehr unwahrscheinlich* und *unwahrscheinlich* zu jeweils einer Stufe zusammengefaßt wurden, während die Mittelkategorie als solche erhalten blieb.

Experte 1 beurteilte den Realitätsgehalt von insgesamt 22 Schilderungen (50 %) in beiden Auswertungsbedingungen gleich. Dabei war der Anteil übereinstimmender Zuord-

nungen bei den erlebnisbegründeten Schilderungen mit 65.2 % deutlich höher als bei den suggerierten Schilderungen (33.3 %).

Da bei bloßer Betrachtung des Anteils an Übereinstimmungen zwischen zwei Beurteilern oder zwei Zeitpunkten im Verhältnis zur Gesamtzahl der Urteile die Tatsache unberücksichtigt bleibt, daß auch bei völlig zufälliger Klassifizierung einige Beobachtungen zufällig übereinstimmen, führt die Ermittlung von Prozentanteilen zu einer Überschätzung der Übereinstimmung (Bortz, 1984). Daher wurden als Übereinstimmungsmaße zusätzlich Kappa-Koeffizienten und Korrelationskoeffizienten nach Spearman bestimmt. Diese wurden sowohl auf Grundlage der dreistufigen als auch auf Grundlage der fünfstufigen Skala sowie sowohl für suggerierte und erlebnisbegründete Schilderungen gemeinsam als auch getrennt berechnet.

Die ermittelten Korrelationskoeffizienten nach Spearman und die Kappa-Koeffizienten zeigen eher niedrige Übereinstimmungen zwischen den Einschätzungen, die Experte 1 im Rahmen der Interviewsituation und anhand der Videoaufzeichnungen und Transkripte vorgenommen hat, bezüglich der suggerierten Schilderungen überhaupt keine signifikanten Zusammenhänge.

Experte 2 beurteilte den Realitätsgehalt von insgesamt 29 Schilderungen (82.9 %) in beiden Auswertungsbedingungen gleich. Dementsprechend zeigen die ermittelten Korrelationskoeffizienten nach Spearman mittlere Übereinstimmungen um .60. Bedeutsame Unterschiede zwischen erlebnisbegründeten und suggerierten Schilderungen hinsichtlich der Anteile an übereinstimmenden Zuordnungen traten nicht zutage.

Während die Raterinnen die Aussagen von 35 Kindern, die über je ein reales und ein fiktives Ereignis berichtet hatten (wobei die reale Aussage eines Kindes aus der Auswertung ausgeschlossen wurde), beurteilt haben, wurden von den Experten darüber hinaus jeweils eine reale und eine fiktive Schilderung von zwei weiteren Kindern sowie die Schilderungen von Kindern, die zwei reale (bzw. fiktive) Ereignisse geschildert hatten, ausgewertet. Vergleicht man die Ergebnisse der Experten, die sie hinsichtlich der auch von den Raterinnen ausgewerteten Schilderungen erzielten (reduzierte Stichprobe), mit denen, die sie bezüglich aller auswertbaren Schilderungen erreichten, so finden sich kaum Abweichungen hinsichtlich der Treffer- und Fehlerquoten sowie der Anteile uneindeutiger Zuordnungen.

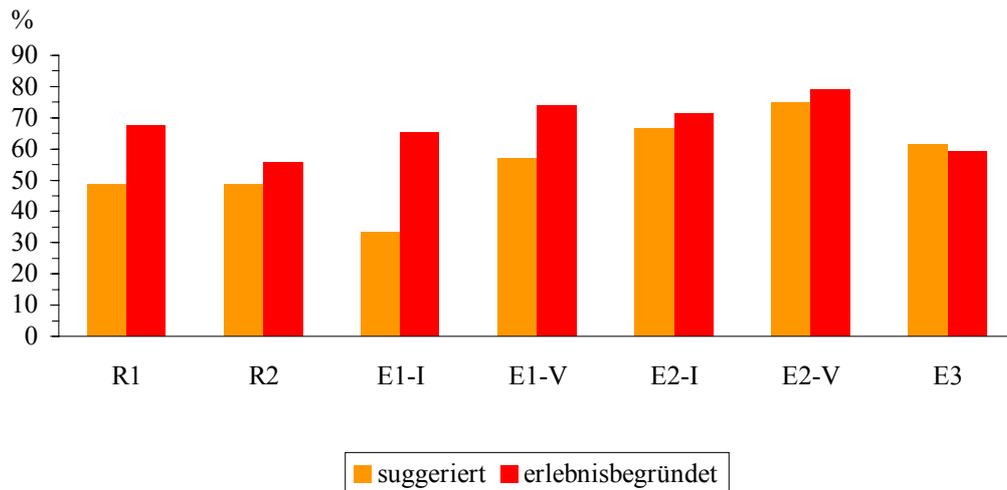
Da die Schilderungen teils von Experte 1, teils von Experte 2 erhoben worden waren (die Experten 1 und 2 haben nur die jeweils von ihnen selbst erhobenen Schilderungen ausgewertet), wurde auch geprüft, ob in den Einschätzungen der beiden Raterinnen und von Experte 3 Unterschiede hinsichtlich der Treffer- und Fehlerquoten zwischen diesen beiden Gruppen von Schilderungen bestehen. Wengleich sich bei den beiden Raterinnen, nicht jedoch bei Experte 3, etwas bessere Ergebnisse bezüglich der von Experte 2 erhobenen Schilderungen ergaben, waren diese Unterschiede nicht signifikant.

5.4.2 Treffer- und Fehlerquoten differenziert nach suggerierten und erlebnisbegründeten Schilderungen

In einem weiteren Auswertungsschritt wurde geprüft, ob sich die Anteile zutreffender und falscher Zuordnungen bei suggerierten und erlebnisbegründeten Schilderungen voneinander unterscheiden.

Mit Ausnahme von Experte 3 erzielten alle übrigen Auswerter höhere Trefferquoten (tendenziell und eindeutig zutreffende Zuordnungen) bei den erlebnisbegründeten Schilderungen als bei den suggerierten, wenngleich diese Unterschiede nicht signifikant waren.

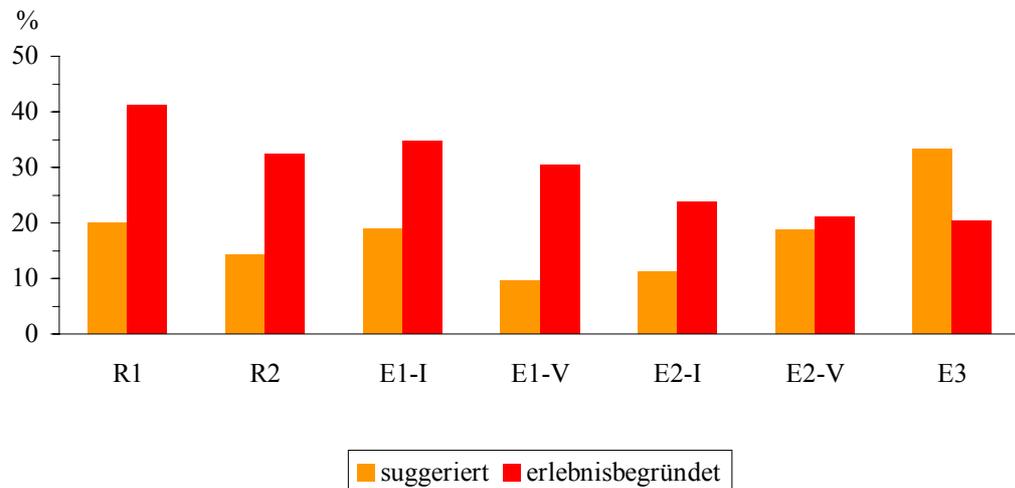
Abbildung 12: Eindeutig und tendenziell zutreffende Klassifikationen von suggerierten und erlebnisbegründeten Schilderungen



Dementsprechend ergaben sich Trefferquoten über Zufallsniveau nur bezüglich der erlebnisbegründeten Schilderungen von Raterin 1 ($\chi^2(1)= 4.235$, $p= .040$), von Experte 1 in der Videobedingung ($\chi^2(1)= 5.261$, $p= .050$) und von Experte 2 unter beiden Auswertungsbedingungen (I: $\chi^2(1)= 3.857$, $p= .050$; V: $\chi^2(1)= 6.368$, $p= .012$). Darüber hinaus erzielte Experte 2 in der Videobedingung auch bezüglich der suggerierten Schilderungen Trefferquoten über Zufallsniveau ($\chi^2(1)= 4.000$, $p= .046$).

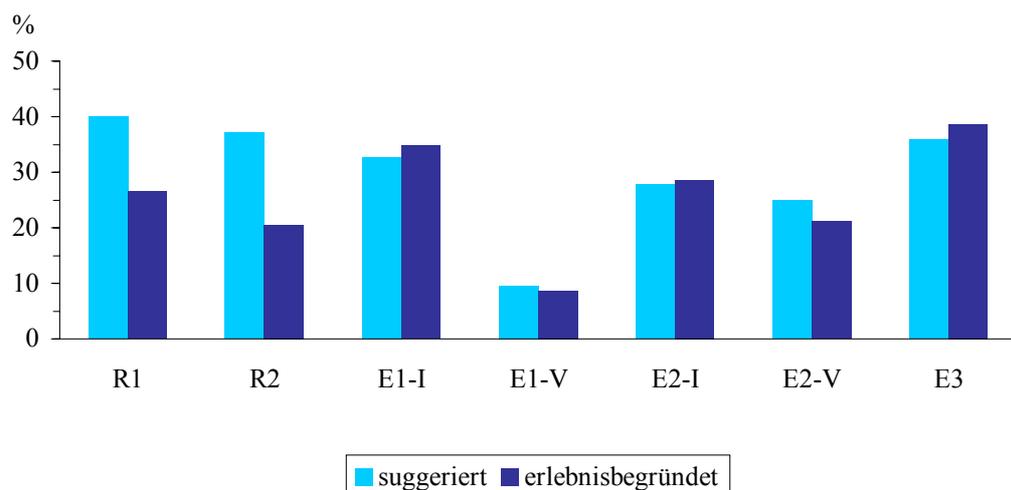
Die gleichen Effekte - mit Ausnahme von Experte 3 bei allen anderen Auswertern deutlich höhere Trefferquoten bei den erlebnisbegründeten im Vergleich zu den suggerierten Schilderungen - treten noch deutlicher zutage, wenn nur die eindeutig zutreffenden Zuordnungen betrachtet werden.

Abbildung 13: Eindeutig zutreffende Klassifikationen von suggerierten und erlebnisbegründeten Schilderungen



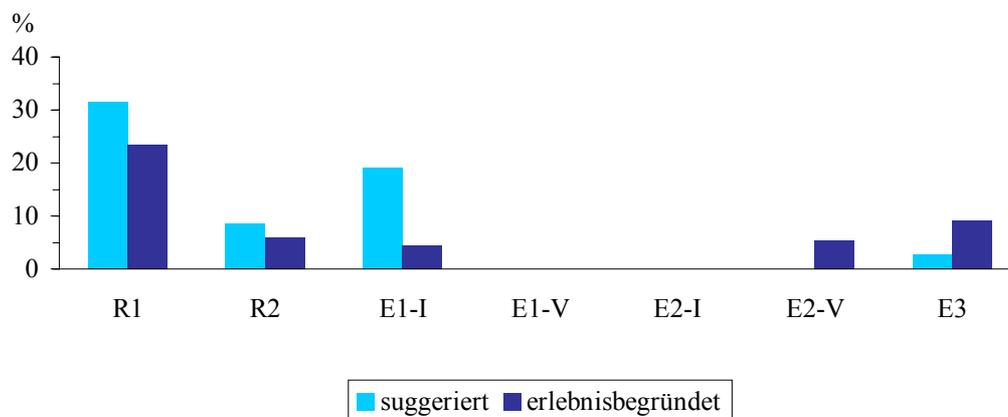
Hinsichtlich der eindeutig und tendenziell falschen Zuordnungen zeigt sich, daß die beiden Raterinnen deutlich mehr Fehleinschätzungen bei den suggerierten Schilderungen vornahmen (falsch-positiv-Klassifikationen) als bei den erlebnisbegründeten (falsch-negativ-Klassifikationen), während sich bei den übrigen Auswertern nur sehr geringfügige Unterschiede, teils in umgekehrter Richtung, ergaben.

Abbildung 14: Eindeutig und tendenziell falsche Klassifikationen von suggerierten und erlebnisbegründeten Schilderungen



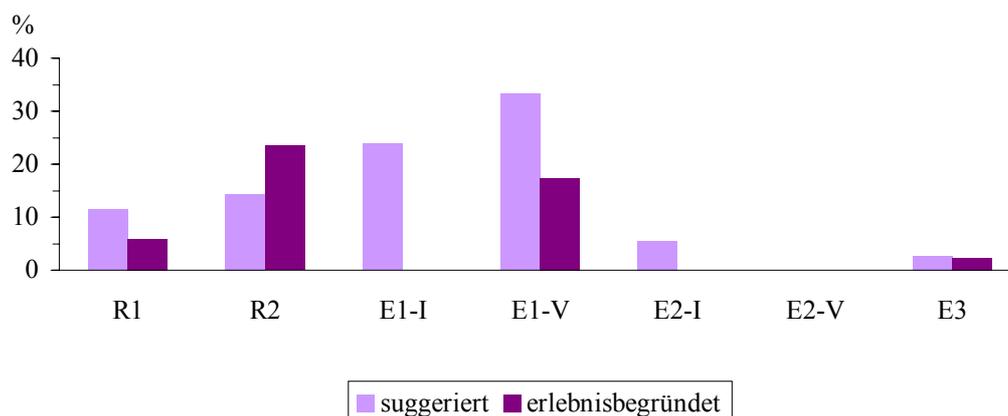
Auch wenn nur die eindeutigen Fehleinschätzungen betrachtet werden, ergibt sich ein ähnliches Bild, wobei hier auch Experte 1 in der Interviewsituation - ebenso wie die Raterinnen - weit mehr suggerierte als erlebnisbegründete Schilderungen eindeutig falsch klassifizierte.

Abbildung 15: Eindeutig falsche Klassifikationen von suggerierten und erlebnisbegründeten Schilderungen



Die größeren Schwierigkeiten der Auswerter bei der Klassifikation der suggerierten Schilderungen im Vergleich zu den erlebnisbegründeten spiegeln sich auch in den Anteilen uneindeutiger Zuordnungen (Wahl der Mittelkategorie auf der fünfstufigen Ratingskala) wider. So trafen die meisten Auswerter mehr uneindeutige Zuordnungen bei den suggerierten Schilderungen als bei den erlebnisbegründeten. Ein umgekehrtes Verhältnis zeigte sich lediglich bei Raterin 1.

Abbildung 16: Uneindeutige Klassifikationen von suggerierten und erlebnisbegründeten Schilderungen



Schließlich zeigen auch die ermittelten Abweichungsscores (vgl. Abschnitt 5.4.1), daß alle Auswerter mit Ausnahme von Experte 3 bezüglich der Zuordnung der suggerierten Schilderungen zumindest tendenziell schlechtere Resultate (höhere Abweichungsscores) erzielten als bei der Klassifikation der erlebnisbegründeten.

Abbildung 17: Abweichungsscores der einzelnen Auswerter bei suggerierten und erlebnisbegründeten Schilderungen

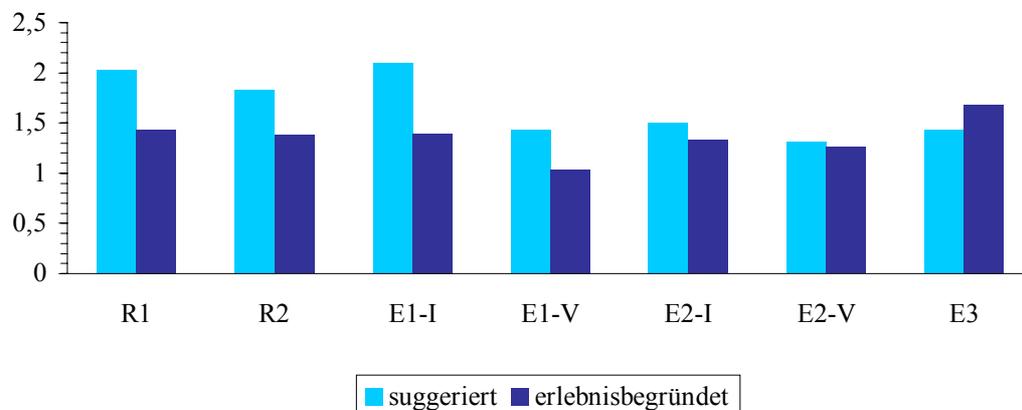


Tabelle 19: Mittlere Abweichungsscores und Standardabweichungen

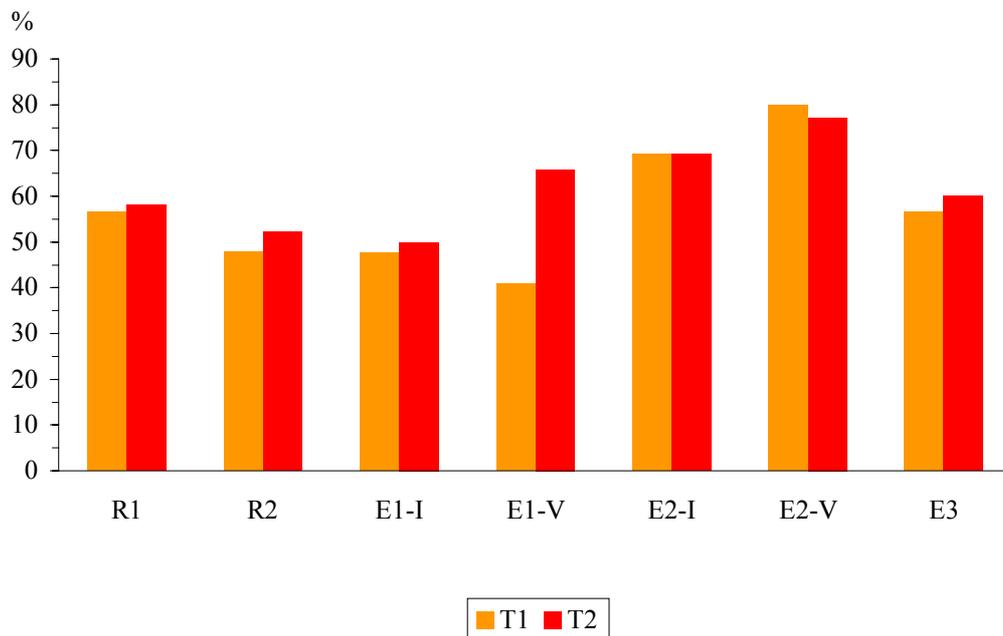
		R1	R2	E1-I	E1-V	E2-I	E2-V	E3
fiktiv	M	2.03	1.83	2.10	1.43	1.50	1.31	1.44
	SD	1.58	1.25	1.41	0.81	1.04	1.08	1.33
	n	35	35	21	21	18	16	39
real	M	1.44	1.38	1.39	1.04	1.33	1.26	1.68
	SD	1.65	1.26	1.37	0.93	1.15	1.15	1.34
	n	34	34	23	23	21	19	44

5.4.3 Treffer- und Fehlerquoten bei der ersten (spontanen) im Vergleich zur zweiten (fundierten) Einschätzung

Wie bereits erwähnt nahmen alle Beurteiler unter den verschiedenen Auswertungsbedingungen Einschätzungen der Wahrscheinlichkeit des Realitätsgehalts der einzelnen Schilderungen zu jeweils zwei Zeitpunkten vor: zunächst unmittelbar nach dem Lesen des Transkripts, dem Erheben bzw. nach Betrachten der Videoaufzeichnung jeder einzelnen Schilderung eines Kindes, ein zweites Mal nach anschließender ausführlicher Beschäftigung mit beiden Aussagen eines Kindes (bei den Raterinnen nach Realkennzeichenanalyse, bei den übrigen Auswertern nach Beurteilung der globalen Aussage- und Verhaltensmerkmale). Nachfolgend wird dargestellt, inwieweit sich die zu den beiden Einschätzungszeitpunkten vorgenommenen Auswertungen der verschiedenen Beurteiler voneinander unterscheiden.

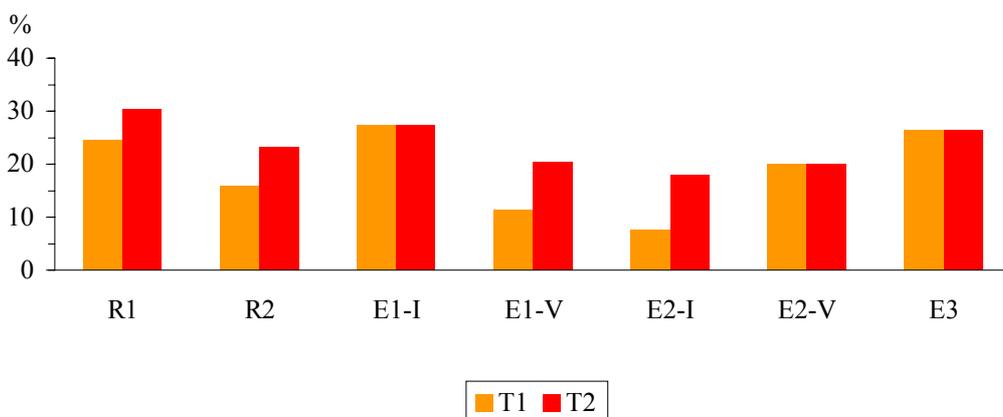
Hinsichtlich der Trefferquoten (eindeutig und tendenziell zutreffende Einschätzungen) ergaben sich folgende Resultate: Während Experte 1 sich in der Interviewsituation bei den zweiten gegenüber den ersten spontanen Einschätzungen deutlich verbesserte, zeigten sich bei den anderen Auswertern nur geringfügige Unterschiede, bei Experte 2 in der Videobedingung in umgekehrter Richtung.

Abbildung 18: Eindeutig und tendenziell zutreffende Klassifikationen beim ersten und zweiten Einschätzungszeitpunkt



Betrachtet man nur die eindeutig richtigen Zuordnungen, so trafen auch die beiden Raterinnen sowie Experte 2 in der Interviewsituation bei der zweiten Einschätzung mehr zutreffende Einschätzungen, während sich unter den übrigen Auswertungsbedingungen keinerlei Unterschiede zeigten.

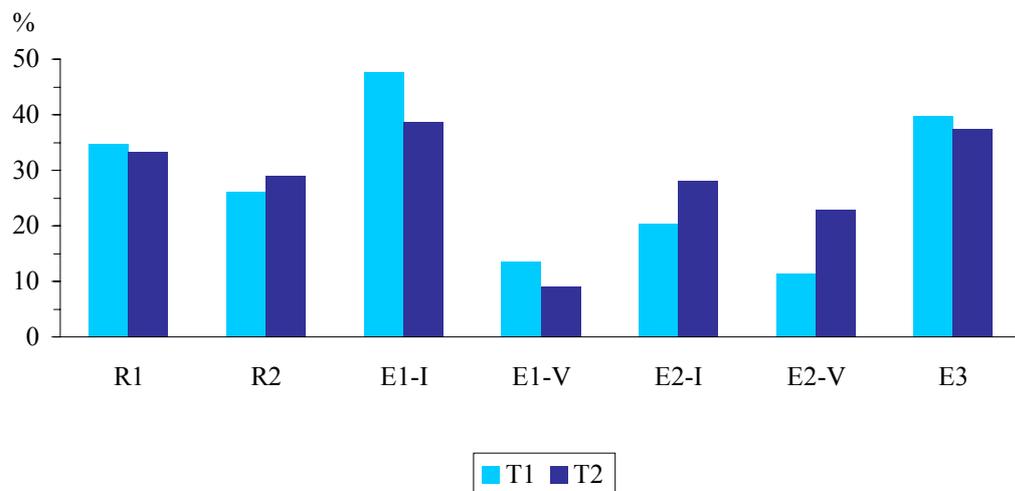
Abbildung 19: Eindeutig zutreffende Klassifikationen beim ersten und zweiten Einschätzungszeitpunkt



Die nachfolgende Abbildung zeigt einen Vergleich der Fehlerquoten (eindeutig und tendenziell falsche Zuordnungen) zum ersten und zweiten Einschätzungszeitpunkt.

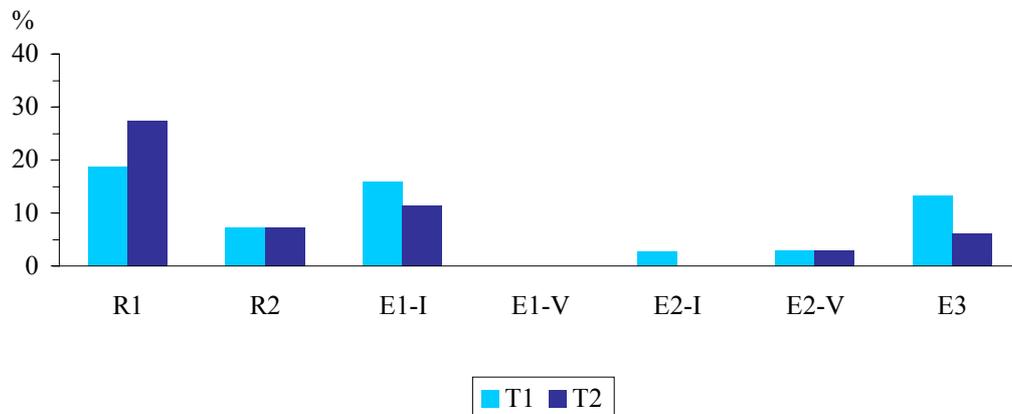
Während Experte 2 unter beiden Auswertungsbedingungen - ebenso Raterin 2, wenngleich der Unterschied hier geringer war - bei der zweiten Einschätzung mehr eindeutig und tendenziell falsche Zuordnungen vornahm als bei der ersten spontanen Beurteilung der einzelnen Aussagen, zeigte sich bei den übrigen Auswertern eher das umgekehrte Verhältnis, wobei dieser Unterschied nur bei Experte 1 deutlich ausgeprägt war.

Abbildung 20: Eindeutig und tendenziell falsche Klassifikationen beim ersten und zweiten Einschätzungszeitpunkt



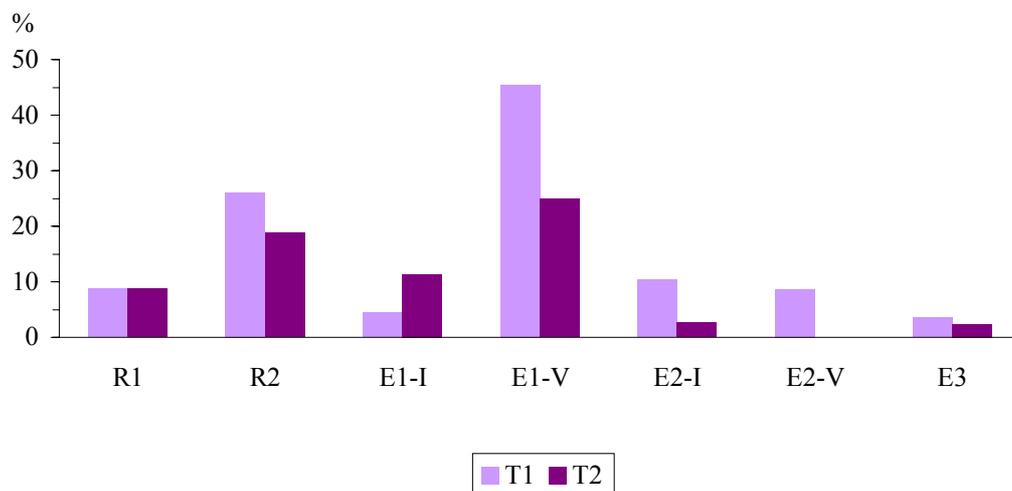
Betrachtet man nur die eindeutig falschen Zuordnungen, so zeigt sich bei Raterin 1 eine deutliche Verschlechterung von der ersten zur zweiten Einschätzung, bei den Experten 1 und 2 in der Interviewsituation sowie bei Experte 3 eine Verbesserung, unter den übrigen Bedingungen kein Unterschied zwischen den beiden Einschätzungsbedingungen.

Abbildung 21: Eindeutig falsche Klassifikationen beim ersten und zweiten Einschätzungszeitpunkt



Schließlich werden die uneindeutigen Zuordnungen zu den beiden Einschätzungszeitpunkten miteinander verglichen.

Abbildung 22: Uneindeutige Klassifikationen beim ersten und zweiten Einschätzungszeitpunkt



Lediglich Experte 1 in der Interviewsituation nahm bei der zweiten Einschätzung mehr uneindeutige Klassifikationen vor als bei der ersten, während sich unter den anderen Bedingungen bzw. Auswertern das umgekehrte Verhältnis, bei Raterin 1 kein Unterschied zeigte.

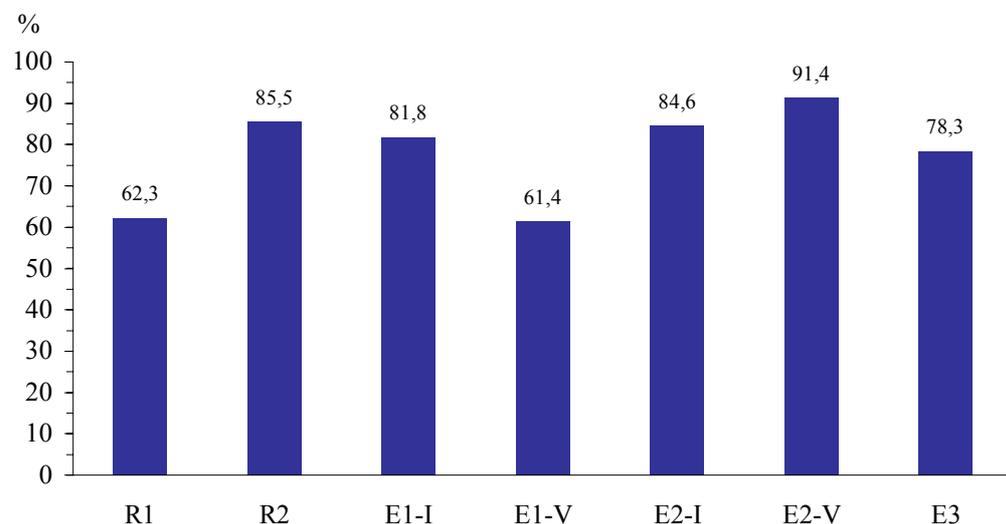
Nachfolgend wird dargestellt, wie groß die Anteile an Schilderungen sind, die von den einzelnen Auswertern zu beiden Einschätzungszeitpunkten übereinstimmend beurteilt wurden. Dabei wurde wiederum eine dreistufige Skala zugrunde gelegt, und die Stufen *sehr wahrscheinlich* und *wahrscheinlich* sowie die Stufen *sehr unwahrscheinlich* und

unwahrscheinlich wurden zu jeweils einer Stufe zusammengefaßt, während die Mittelkategorie als solche erhalten blieb (vgl. Abschnitt 5.4.1).

Die Anteile übereinstimmender Zuordnungen liegen zwischen 61 % und 91 %. Mit Ausnahme von Experte 3, bei dem sich das umgekehrte Verhältnis zeigte, fanden sich bezüglich der erlebnisbegründeten Schilderungen größere Übereinstimmungen als hinsichtlich der suggerierten (z.B. E1-I: 67 % bei den suggerierten gegenüber 96 % bei den erlebnisbegründeten Schilderungen).

Entsprechend ergaben sich auf Grundlage aller Schilderungen überwiegend mittlere bis sehr hohe Korrelationen nach Spearman sowie mittlere bis sehr hohe Kappa-Koeffizienten im Bereich von etwa .35 bis 1.00.¹⁷

Abbildung 23: Anteile an Schilderungen, die zu den Zeitpunkten T1 und T2 gleich beurteilt wurden



Zusammenfassend ist somit festzuhalten, daß zwischen den beiden Einschätzungszeitpunkten große Übereinstimmung besteht, d.h. die ausführliche Beschäftigung mit beiden Schilderungen eines Kindes - insbesondere in Form der inhaltlichen Qualitätsanalyse - führte bei den meisten Auswertern (Ausnahme: Experte 1 in der Videobedingung) zu keinen nennenswerten Verbesserungen hinsichtlich der Einschätzungen des tatsächlichen Realitätsgehalts der Schilderungen.

¹⁷ Die vorgenannten Koeffizienten wurden zum einen auf Grundlage der fünfstufigen, zum anderen auf Grundlage der dreistufigen Skala bestimmt. Des weiteren wurden die Koeffizienten auf Grundlage aller Schilderungen sowie getrennt für erlebnisbegründete und suggerierte Schilderungen berechnet.

5.4.4 Übereinstimmungen und Abweichungen in den Einschätzungen der verschiedenen Auswerter

Ein weiterer Analyseschritt diente der Klärung der Frage, inwieweit die verschiedenen Auswerter in ihren Einschätzungen des Realitätsgehalts der Schilderungen übereinstimmen. Dabei interessieren insbesondere die Einschätzungen, die unter jeweils gleichen Auswertungsbedingungen nach ausführlicher Beschäftigung (vor allem inhaltliche Qualitätsanalyse) mit beiden Aussagen eines Kindes vorgenommen wurden (T2).¹⁸

In Tabelle 20 wird dargestellt, inwieweit zwischen den Einschätzungen der beiden Raterinnen Übereinstimmungen und Abweichungen bestehen. Bezogen auf alle Schilderungen gaben die beiden Raterinnen in etwa zwei Drittel der Fälle gleiche Einschätzungen ab. In etwa 20 % der Fälle nahmen beide falsche Einschätzungen vor, in 5.8 % der Fälle kamen sie zu einem eindeutig unterschiedlichen Urteil.

Tabelle 20: Konkordanzen und Diskordanzen in den Einschätzungen der Raterinnen

	alle Schilderungen (N= 69)		sugerierte Schilderungen (N= 35)		erlebnisbegr. Schilderungen (N= 34)	
	n	%	n	%	n	%
beide Raterinnen eindeutig oder tendenziell richtige Zuordnungen	32	46.4	14	40.0	18	52.9
beide Raterinnen eindeutig oder tendenziell falsche Zuordnungen	14	20.3	11	31.4	3	8.8
beide Raterinnen unentschieden	0	0	0	0	0	0
eindeutig abweichende Zuordnungen (1 Raterin richtig, 1 falsch)	4	5.8	1	2.9	3	8.8
tendenziell abweichende Zuordnungen (1 Raterin richtig oder falsch, 1 Raterin uneindeutig)	19	27.5	9	25.7	10	29.4

Beim Vergleich zwischen den Einschätzungen der suggerierten und der erlebnisbegründeten Schilderungen ergab sich eine höhere übereinstimmende Trefferquote bei den erlebnisbegründeten Schilderungen. Hingegen traten übereinstimmende falsche Zuordnungen häufiger bei den suggerierten Schilderungen auf.

Als weitere Maße der Übereinstimmung in den Einschätzungen beider Raterinnen wurden - auf Grundlage der dreistufigen und der fünfstufigen Skala sowie für suggerierten und erlebnisbegründete Schilderungen sowohl getrennt als auch gemeinsam - Kappa-Koeffizienten sowie Rangkorrelationskoeffizienten nach Spearman bestimmt. Letztere lagen etwa bei .70, die Kappa-Koeffizienten bei .40.

Des weiteren wurde geprüft, inwieweit die Einschätzungen von Experte 3 mit denen der Experten 1 und 2, die diese im Rahmen der Videoauswertung getroffen haben, übereinstimmen.

¹⁸ An dieser Stelle sei noch einmal erwähnt, daß die Experten 1 und 2 jeweils unterschiedliche Schilderungen ausgewertet haben, so daß ein Vergleich zwischen den Einschätzungen dieser beiden Experten nicht möglich ist.

Tabelle 21: Konkordanzen und Diskordanzen in den Einschätzungen der Experten 1 und 3

	alle Schilderungen (N= 44)		sugerierte Schilderungen (N= 21)		erlebnisbegr. Schilderungen (N= 23)	
	n	%	n	%	n	%
beide Experten eindeutig oder tendenziell richtige Zuordnungen	17	38.6	7	33.3	10	43.5
beide Experten eindeutig oder tendenziell falsche Zuordnungen	2	4.5	1	4.8	1	4.4
eindeutig abweichende Zuordnungen (1 Experte richtig, 1 falsch)	14	31.8	6	28.6	8	34.8
tendenziell abweichende Zuordnungen (1 Experte richtig oder falsch, 1 Experte uneindeutig)	11	25.0	7	33.3	4	17.4

Bezogen auf alle Schilderungen, die von beiden Experten ausgewertet worden waren, gaben Experte 1 und Experte 3 in 43 % der Fälle gleiche Einschätzungen ab. Dabei trafen die Experten in zwei Fällen übereinstimmend falsche Zuordnungen. In fast einem Drittel der Fälle kamen sie zu einem eindeutig unterschiedlichen Urteil. Beim Vergleich zwischen den Einschätzungen der suggerierten und der erlebnisbegründeten Schilderungen zeigte sich eine höhere Übereinstimmungsquote zwischen den Experten bei den erlebnisbegründeten Schilderungen als bei den suggerierten.

Die geringen Übereinstimmungen in den Angaben der beiden Experten spiegeln auch die ermittelten Maße der Übereinstimmung (Korrelationskoeffizienten nach Spearman sowie Kappa-Koeffizienten) wider, die - sofern sich überhaupt signifikante Zusammenhänge zeigten - als niedrig zu bezeichnen sind.

Die Diskrepanzen in den Einschätzungen zwischen den beiden Experten sind offenbar darauf zurückzuführen, daß Experte 3 im Vergleich zu Experte 1 in bezug auf die suggerierten Schilderungen höhere Trefferquoten erzielte, in bezug auf die erlebnisbegründeten hingegen niedrigere. Des weiteren waren die Fehlerquoten bei Experte 3 höher als bei Experte 1, während letzterer weit häufiger uneindeutige Zuordnungen traf.

Zwischen den Einschätzungen der Experten 2 und 3 ergaben sich nachfolgende Übereinstimmungen und Abweichungen: Bezogen auf alle Schilderungen gaben die beiden Experten in etwa 54 % der Fälle gleiche Einschätzungen ab. Dabei trafen sie in drei Fällen übereinstimmend falsche Zuordnungen. In 43 % der Fälle kamen beide zu einem eindeutig unterschiedlichen Urteil. Beim Vergleich zwischen den Einschätzungen der suggerierten und der erlebnisbegründeten Schilderungen zeigten sich kaum Unterschiede hinsichtlich der Übereinstimmungsraten.

Tabelle 22: Konkordanzen und Diskordanzen in den Einschätzungen der Experten 2 und 3

	alle Schilderungen (N= 35)		suggestierte Schilderungen (N= 16)		erlebnisbegr. Schilderungen (N= 19)	
	n	%	n	%	n	%
beide Experten eindeutig oder tendenziell richtige Zuordnungen	16	45.7	7	43.8	9	47.7
beide Experten eindeutig oder tendenziell falsche Zuordnungen	3	8.6	1	6.3	2	10.5
eindeutig abweichende Zuordnungen (1 Experte richtig, 1 falsch)	15	42.8	7	43.8	8	42.1
tendenziell abweichende Zuordnungen (1 Experte richtig oder falsch, 1 Experte uneindeutig)	1	2.9	1	6.3	0	0

Die niedrigen Übereinstimmungen in den Angaben der beiden Experten, die auf die deutlich niedrigeren Treffer- und höheren Fehlerquoten von Experte 3 im Vergleich zu Experte 2 zurückzuführen sind, spiegeln auch die ermittelten Übereinstimmungsmaße (Korrelationskoeffizienten nach Spearman sowie Kappa-Koeffizienten) wider, die - sofern überhaupt signifikante Zusammenhänge erkennbar waren - niedrig waren.

Zur weiteren Klärung der Frage, inwieweit die Einschätzungen der Experten mit denen der beiden Raterinnen übereinstimmen, wurden sowohl auf Grundlage der dreistufigen als auch auf Grundlage der fünfstufigen Skala Korrelationskoeffizienten nach Spearman sowie Kappa-Koeffizienten ermittelt. Es wurden die Einschätzungen jedes einzelnen Experten mit denen jeder einzelnen Raterin verglichen. Die Koeffizienten wurden jeweils auf Grundlage aller Schilderungen sowie getrennt für suggerierte und erlebnisfundierte Schilderungen ermittelt. Sofern überhaupt signifikante Zusammenhänge zwischen den Einschätzungen der Experten und denen der Raterinnen zutage traten, waren diese als sehr niedrig einzustufen.

5.4.5 Zusammenfassende Darstellung der Einschätzungen aller Auswerter

Faßt man die Einschätzungen aller Auswerter - die der beiden Raterinnen, der beiden Interviewer-Experten und des dritten Experten -, die unter den verschiedenen Bedingungen (Transkripte, Interviewsituation, Transkripte und Videoaufzeichnungen) nach jeweils ausführlicher Beschäftigung (vor allem inhaltliche Qualitätsanalyse) mit beiden Schilderungen eines Kindes vorgenommen wurden, zusammen, so ergeben sich bezogen auf alle Schilderungen Trefferquoten (eindeutig und tendenziell zutreffende Zuordnungen) im Bereich von 50 % bis 77 %, Fehlerquoten (eindeutig und tendenziell falsche Zuordnungen) im Bereich von 9 % bis 39 %, uneindeutige Zuordnungen im Bereich von 0 % bis 25 %. Differenziert man zusätzlich zwischen suggerierten und erlebnisbegründeten Schilderungen, so dehnen sich die Schwankungsbreiten hinsichtlich der Trefferquoten (zwischen 33 % und 79 %) und der uneindeutigen Zuordnungen (zwischen 0 % und 33 %) weiter aus.

Extreme bzw. eindeutige Einschätzungen wurden in durchschnittlich knapp einem Drittel aller Zuordnungen (31.5 %) vorgenommen (bei den verschiedenen Auswertern in 18 % bis 58 % der Fälle), und sie waren in der überwiegenden Mehrheit zutreffend (in 53 % bis 100 % der Fälle, durchschnittlich in 81 %). Damit wurde durchschnittlich etwa ein Viertel der Schilderungen eindeutig zutreffend klassifiziert (bei den verschiedenen Auswertern zwischen 18 % und 30 %).

Bei den ersten spontanen Einschätzungen lagen die Anteile (eindeutig und tendenziell) zutreffender Zuordnungen zwischen 41 % und 80 %, die Fehleranteile zwischen 11 % und 48 % und die Anteile uneindeutiger Zuordnungen zwischen 4 % und 46 %. Die ersten spontanen Einschätzungen der einzelnen Schilderungen waren damit bei den meisten Auswertern nur unwesentlich schlechter als die zweiten Einschätzungen, die nach ausführlicher Beschäftigung mit beiden Aussagen eines Kindes vorgenommen wurden. Während Experte 1 sich in seinen Beurteilungen bei der nachträglichen Auswertung der Videoaufzeichnungen und Transkripte gegenüber der Auswertung im Rahmen der Interviewsituation deutlich verbesserte, war dieser Effekt bei Experte 2 unbedeutend.

Die beiden Raterinnen und Experte 1 zeigten bessere Ergebnisse bei der Beurteilung der erlebnisbegründeten Schilderungen als bei der Beurteilung der suggerierten. Bei Experte 3 ergab sich hingegen das umgekehrte Verhältnis, bei Experte 2 traten kaum Unterschiede zutage.

Beim Vergleich der Einschätzungen jeweils eines Auswertern mit denen aller anderen Beurteiler zeigten sich nur sehr geringe Übereinstimmungen. Lediglich die beiden Raterinnen stimmten in vergleichsweise hohem Maße in ihren Einschätzungen des Realitätsgehalts der Schilderungen überein.

Betrachtet man die Zuordnungen von allen fünf Auswertern im Hinblick auf die absolute Anzahl an übereinstimmenden und abweichenden Einschätzungen, so ist festzustellen, daß in bezug auf die 34 erlebnisbegründeten Aussagen, die von allen Auswertern beurteilt wurden, zwölf Schilderungen (35.3 %) übereinstimmend und zutreffend als wahr beurteilt wurden. Eine erlebnisbegründete Aussage wurde von allen Auswertern fälschlich als suggeriert beurteilt.

Von den 35 suggerierten Schilderungen beurteilten alle Auswerter zu beiden Zeitpunkten sechs Aussagen (17.1%) zutreffend als suggeriert. Drei weitere Schilderungen (8.6 %) beurteilten die Auswerter teils zutreffend als suggeriert, teils trafen sie eine uneindeutige Zuordnung. Eine Aussage wurde von allen Auswertern fälschlich für erlebnisbegründet gehalten.

