

*Aus der Klinik für Kinder- und Jugendpsychiatrie und Psychotherapie
Universitätsklinikum des Saarlandes, Homburg/Saar*

Das Elternbelastungsscreening zur Kindeswohlgefährdung (EBSK)

*Validierung einer deutschsprachigen Version
des Child Abuse Potential Inventory (CAPI) von Joel S. Milner*

Dissertation zur Erlangung des Grades eines Doktors der Theoretischen Medizin
der Medizinischen Fakultät der Universität des Saarlandes

2008

vorgelegt von:

Nicolas Becker

geb.am 13.12.1981 in Zweibrücken

Inhaltsverzeichnis

Inhaltsverzeichnis.....	ii
Tabellenverzeichnis.....	v
Abbildungsverzeichnis.....	vi
Kurzzusammenfassung.....	vii
Abstract.....	viii
Publikationen.....	ix
A. Einleitung.....	1
1. Kindesmisshandlung im Wandel der Geschichte.....	1
2. Kinderschutzarbeit als interdisziplinäres Handlungsfeld.....	1
3. Psychologische Diagnostik in der Kinderschutzarbeit.....	2
4. Ziel und Aufbau der vorliegenden Arbeit.....	3
B. Kindesmisshandlung: Definition, Verbreitung, Ätiologie, empirische Befunde.....	5
1. Definition und Einteilung verschiedener Formen der Kindesmisshandlung.....	5
2. Häufigkeit von Kindesmisshandlung.....	7
2.1 Deutsche Dunkelfeldstudien zur Häufigkeit von Kindesmisshandlung.....	7
2.2 Kindesmisshandlung in amerikanischen Kriminalitätsstatistiken.....	9
3. Erklärungsmodelle zur Ätiologie von Kindesmisshandlung.....	12
3.1 Psychopathologische, sozial-situative und soziologische Ansätze.....	12
3.2 Integrative Ursachenmodelle.....	13
3.3 Metaanalytische Befunde zu Risiko- und Schutzfaktoren.....	16
4. Zusammenfassung der dargestellten Befunde und Überlegungen.....	24
C. Das Child Abuse Potential Inventory.....	26
1. Umfang und Skalenstruktur.....	27
2. Abuse Scale.....	28
2.1 Entwicklung.....	28
2.2 Reliabilität.....	31
2.3 Validität der Abuse Scale.....	32
2.3.1 Inhaltsvalidität.....	33
2.3.2 Faktorielle Validität.....	34
2.3.3 Kriteriumsvalidität.....	36
2.3.3.1 Konkurrente Validität.....	37
2.3.3.2 Prädiktive Validität.....	39
3. Validity Scales.....	41
3.1 Entwicklung der Lie Scale.....	41
3.2 Entwicklung der Random Response Scale.....	43

3.3 Entwicklung der Inconsistency Scale	46
4. Abschließende Bewertung	48
D. Validierung des Elternbelastungsscreenings zur Kindeswohlgefährdung	50
1. Übersetzung der englischsprachigen Originalversion	50
2. Akquise und Zusammenfassung der Validierungsstichprobe	52
2.1 Stichprobe von Eltern von Kindergarten- und Grundschulkindern	53
2.2 Stichprobe von Eltern von Patienten der Kinder- und Jugendpsychiatrie	53
2.3 Stichprobe von MitarbeiterInnen und Studierenden einer Polizeiakademie	54
2.4 Stichprobe von hochbelasteten und misshandelnden Eltern	54
2.5 Stichprobe von Begutachtungsfällen bei Verdacht auf Kindesmisshandlung	54
2.6 Stichproben von Studierenden der Universität des Saarlandes	55
2.7 Studierende der Otto-Friedrich-Universität Bamberg	56
2.8 Zusammenfassung der einzelnen Stichproben zu Belastungsgruppen	56
3. Validierung der Belastungsskala	57
3.1 Identifikation diskriminierender Items	57
3.2 Diskriminanzanalytische Validierung	58
3.3 Lineartransformation der Belastungsskala	60
3.4 Validität der Belastungsskala	60
3.4.1 Faktorielle Validität	60
3.4.2 Befunde zu konvergenter und diskriminanter Validität der Belastungsskala	64
3.4.3 Konkurrente Validität mit demographischen Variablen	65
3.4.3.1 Familienstand	66
3.4.3.2 Anzahl der Kinder	67
3.4.3.3 Bildungsstand der Probanden	68
4. Entwicklung der Validitätsskalen	69
4.1 Lügenskala	69
4.2 Zufalls-Antwort-Skala	72
4.3 Inkonsistenzskala	73
5. Wahl des Maximums zulässiger fehlender Antworten	75
6. Normierung	76
7. Vorbereitung von Test- und Auswertungsmaterial	77
8. Abschließende Bewertung	77
E. Diskussion und Ausblick	79
1. Unterschiede zwischen CAPI und EBSK	79
1.1 Unterschiede im Umfang der Verfahren	79
1.2 Unterschiede in klassifikatorischer Güte von Abuse Scale und Belastungsskala	80

1.3 Unterschiede in der Faktorenstruktur von Abuse Scale und Belastungsskala	82
2. Weitere Validierungsschritte	83
2.1 Generierung zusätzlicher Items	84
2.2 Differenziertere Erfassung von Misshandlungsformen und Schweregraden	85
2.3 Berücksichtigung längsschnittlicher Aspekte	86
3. Abschließende Bemerkung	88
F. Literatur	89
G. Anhang	98
1. Tabellen aus der Metaanalyse von Stith et al. (2004)	98
2. Items und Skalen der Übersetzung des CAPI	101
2.1 Wortlaut der Items nach Globisch und Spangler (o.J.)	101
2.2 Zugehörigkeit der Items zu den Skalen	106
3. Items und Skalen des EBSK	107
3.1 Wortlaut der Items des EBSK	107
3.2 Zugehörigkeit der Items zu den Skalen	109
H. Danksagung	110

Tabellenverzeichnis

Tabelle B.1: Ontogenetische Risikofaktoren für körperliche Misshandlung	18
Tabelle B.2: Ontogenetische Risikofaktoren für Kindesvernachlässigung	19
Tabelle B.3: Risikofaktoren für körperliche Misshandlung auf Mikrosystemebene	20
Tabelle B.4: Risikofaktoren für Kindesvernachlässigung auf Mikrosystemebene	21
Tabelle B.5: Risikofaktoren für körperliche Misshandlung auf Exosystemebene	22
Tabelle B.6: Risikofaktoren für Kindesvernachlässigung auf Exosystemebene.....	23
Tabelle D.1: Mittlere Ähnlichkeitseinschätzungen der beiden Übersetzungen	51
Tabelle D.2: Klassifikationsergebnisse der Diskriminanzanalyse	58
Tabelle D.3: Klassifikationsergebnisse in der Validierungsstichprobe	59
Tabelle D.4: Lineartransformierte Version der Belastungsskala	60
Tabelle D.5: Ladungsmatrix der Items der reduzierten Belastungsskala	63
Tabelle D.6: Korrelationen zwischen Belastungsskala und Skalen des FPI-R	64
Tabelle D.7: Mittelwerte in der Belastungsskala gruppiert nach Familienstand	66
Tabelle D.8: Mittelwerte in der Belastungsskala gruppiert nach Kinderzahl	67
Tabelle D.9: Mittelwerte in der Belastungsskala gruppiert nach Schulabschluss	68
Tabelle D.10: Ergebnisse der Validierung der Lügenskala.....	70
Tabelle D.11: Mittelwerte und Standardabweichungen der Lügenskala gruppiert nach Stichprobe..	71
Tabelle D.12: Ergebnisse der Validierung der Zufalls-Antwort-Skala.....	72
Tabelle D.13: Ergebnisse der Validierung der Inkonsistenzskala	74
Tabelle G.1: Risikofaktoren für körperliche Kindesmisshandlung.....	98
Tabelle G.2: Risikofaktoren für Kindesvernachlässigung	100
Tabelle G.3: Zuordnung der CAPI-Items zu den Skalen	106
Tabelle G.4: Zuordnung der EBSK-Items zu den Skalen	109

Abbildungsverzeichnis

Abbildung B.1: Integratives Ursachenmodell zur Entstehung von Kindesmisshandlung	15
Abbildung C.1: Skalenstruktur des Child Abuse Potential Inventory	27
Abbildung D.1: Eigenwertverlauf der MixFaktor-Analyse der Belastungsskala	61

Kurzzusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Validierung des Elternbelastungs-screenings zur Kindeswohlgefährdung (EBSK). Dieses stellt die deutsche Adaptation des Child Abuse Potential Inventory (CAPI), einem etablierten englischsprachigen Fragebogenverfahren zur Erfassung des Risikos zukünftiger Kindesmisshandlung, dar. Die deutsche Validierungsuntersuchung basiert hierbei auf einer Stichprobe von insgesamt 1581 Probanden, in der Daten von hinsichtlich Kindesmisshandlung unbelasteten, belasteten sowie von misshandelnden Elternteilen enthalten sind. Im Rahmen der Validierungsuntersuchung konnte zunächst festgestellt werden, dass die verwendete Übersetzung des Verfahrens gemäß unabhängiger Experten eine hohe inhaltliche Nähe zum Originalverfahren aufweist. Mittels diskriminanzanalytischer Methodik konnte eine Funktion geschätzt werden, auf deren Basis 75% der Probanden in der Validierungsstichprobe entsprechend des vorliegenden Belastungsgrades richtig eingeschätzt werden konnten. Die in der Funktion enthaltenen Items wurden zu der Belastungsskala zusammengefasst, die den primären Indikator für das Kindesmisshandlungsrisiko darstellt. Der Inhalt der Items ist konsistent zu theoretischen Überlegungen und empirischen Befunden zur Ätiologie von Kindesmisshandlung. Weiterhin zeigt die Skala eine hohe interne Konsistenz ($\alpha=0,907$) und kann gemäß faktoranalytischer Betrachtungen als eindimensional angenommen werden. Die Validität der Belastungsskala wurde zusätzlich mittels korrelativer und inferenzstatistischer Methodik überprüft und kann bezüglich anderer psychometrischer Verfahren sowie verschiedener demographischer Variablen angenommen werden. Durch Häufigkeitsauszählungen konnten neben der Belastungsskala ebenfalls Skalen zur Erfassung verfälschender Antworttendenzen konstruiert werden. Hierbei können durch die Betrachtung von Abweichungen von einer wahrheitsgemäßen Bearbeitung des Fragebogens sozial erwünschtes (Lügenskala), zufälliges (Zufalls-Antwort-Skala) sowie inkonsistentes Antwortverhalten (Inkonsistenzskala) erfasst werden. In einer abschließenden Diskussion werden Unterschiede zwischen dem EBSK und der amerikanischen Originalversion diskutiert und Ansätze zu einer Weiterentwicklung des Verfahrens besprochen.

Abstract

The Elternbelastungsscreening zur Kindeswohlgefährdung (EBSK). Validation of a German version of the Child Abuse Potential Inventory (CAPI) by Joel S. Milner

The present study deals with the validation of the Elternbelastungsscreening zur Kindeswohlgefährdung (EBSK). It represents a German adaptation of the Child Abuse Potential Inventory (CAPI), an established American questionnaire for the assessment of prospective child abuse risk. The German validation survey is based on a sample of 1581 subjects and contains data from non-abusing, hazarded and abusing parents. According to independent experts, the content of the American items was very similar to their German translations. Based on a function estimated by discriminant analysis, it was possible to correctly classify 75% of the subjects according to their actual abuse risk. The items included in the function were combined to an abuse scale, which is the primary indicator of abuse risk in the EBSK. The content of these items is consistent with theoretical considerations and empirical results concerning the etiology of child abuse. The abuse scale also shows a high internal consistency ($\alpha=0,907$) and is unidimensional according to factoranalytic examination. Using correlative and inferential statistic methodology additional evidence for the validity of the abuse scale could be found with respect to other psychometric questionnaires and demographic variables. By frequency count, three additional validity scales could be constructed in order to assess fake response tendencies. Assessing the deviation from veridical responses allowed the detection of socially desired (lie scale), random (random response scale) and inconsistent (inconsistency scale) response tendencies. In a final discussion differences between the EBSK and the American original are discussed. Furthermore, possibilities for a prospect validation of the questionnaire are suggested.

Publikationen

Im Zusammenhang mit der vorliegenden Arbeit entstand die folgende Publikation:

Deegener G, Spangler G, Körner W, Becker N (im Druck) Elternbelastungsscreening zur Kindeswohlgefährdung. Hogrefe, Göttingen

A. Einleitung

In diesem ersten Kapitel erfolgt eine allgemeine Einführung der Inhalte der vorliegenden Arbeit. In einem ersten Abschnitt wird beschrieben, inwiefern sich Kindesmisshandlung durch die Geschichte hinweg gezeigt hat. Ein zweiter Abschnitt stellt die Anfänge der Kinderschutzarbeit in westlichen Industrienationen dar, wobei ein Schwerpunkt auf die schon damals erkannte Notwendigkeit einer interdisziplinären Herangehensweise an das Thema gesetzt wird. Der Anteil, den die psychologische Disziplin der Diagnostik an der Kinderschutzarbeit leisten kann, wird in einem dritten Abschnitt aufgeführt. Ein abschließender Abschnitt befasst sich mit der diesbezüglichen Zielsetzung der Arbeit und stellt den Rahmen dar, in dem die Besprechung der wesentlichen Inhalte erfolgen soll.

1. Kindesmisshandlung im Wandel der Geschichte

Geschichtliche Belege für Gewalthandlungen gegen Kinder lassen sich in der Literatur in großem Umfang von der Antike bis in die Neuzeit feststellen. Um zwei diesbezügliche Eckpunkte aus Antike und Neuzeit aufzuzählen, sei zunächst erwähnt, dass Kindstötungen bis zum vierten Jahrhundert nach Christus in Rom und Griechenland weder gesetzlich verboten noch gesellschaftlich verpönt waren und in Deutschland in der Mitte des zwanzigsten Jahrhunderts 80% der Eltern das Schlagen mit der Hand und immerhin 35% der Eltern das Schlagen mit dem Rohrstock als akzeptable Form der Bestrafung billigten (Fürniss, 2005).

2. Kinderschutzarbeit als interdisziplinäres Handlungsfeld

Trotz seiner langen Geschichte gewann das Problem der Kindesmisshandlung erst mit den Arbeiten des amerikanischen Pädiaters C. Henry Kempe zu Beginn der 1960er weltweite Beachtung. Ein bedeutender Meilenstein kann hierbei in der Veröffentlichung des Artikels „The Battered Child Syndrome“ (Kempe et al., 1962) gesehen werden, der eine intensive internationale klinische und wissenschaftliche Beschäftigung mit den Folgen elterlicher Kindesmisshandlung und -vernachlässigung zur Folge hatte. Schon in dieser ersten Aufarbeitung der Thematik wurde die Wichtigkeit einer interdisziplinären Erkennung, Behandlung und Prävention von Kindes-

misshandlung und Vernachlässigung betont, wobei medizinischen, psychiatrischen und psychologischen Sichtweisen gleichermaßen Bedeutsamkeit zugewiesen wird. Diese multiprofessionelle Vorgehensweise zieht sich als roter Faden durch die Entstehung und Professionalisierung der internationalen Kindesmisshandlungsarbeit. So beschreibt Bange (2005), dass „die gemeinsame Arbeit von Kinderärzten, Psychiatern, Elterntherapeuten und Sozialarbeitern im Kinderschutz [...] die Interventionsgrundlage der frühen Arbeit bei körperlicher Misshandlung und Vernachlässigung“ (S.25) bildete und „eine intensivere sowohl monoprofessionelle als auch multiprofessionelle Weiterbildung [...] national wie international weiterhin zentrale Themenschwerpunkte zukünftiger Misshandlungsarbeit“ (S.31) darstelle.

3. Psychologische Diagnostik in der Kinderschutzarbeit

Psychologische Anteile in der Kinderschutzarbeit bestehen zunächst schon aufgrund der Tatsache, dass Misshandler sowie misshandelte Kinder üblicherweise bestimmten psychotherapeutischen Maßnahmen zugeführt werden. Ebenfalls bietet sich in der Kinderschutzarbeit auch die Möglichkeit eines sinnvollen Einsatzes einer weiteren psychologischen Disziplin, der psychologischen Diagnostik. Hierbei wird die psychologische Diagnostik als Methodenlehre im Dienste der angewandten Psychologie verstanden, deren Aufgabe unter anderem in der Feststellung interindividueller Differenzen liegt, auf deren Basis möglichst präzise Vorhersagen zukünftigen Verhaltens ermöglicht werden sollen (Amelang & Schmidt-Atzert, 2006). Die Möglichkeit von Verhaltensprognosen schließt in diesem Sinne ebenfalls die Vorhersage der Auftretenswahrscheinlichkeit von Kindesmisshandlung mit ein. Mögliche Handlungsfelder in der Kinderschutzarbeit stellen hierbei sowohl die Erkennung als auch die Behandlung und Prävention von Kindesmisshandlern dar. In Bezug auf die Erkennung von Kindesmisshandlungsfällen können diagnostische Instrumente zum Zwecke eines ökonomischen Screenings der Wahrscheinlichkeit des Vorliegens von Kindesmisshandlung eingesetzt werden. Im Bereich der Behandlung kann durch ihren Einsatz vor und nach einer Intervention abgeschätzt werden, inwiefern die Intervention zu einer Reduktion des Misshandlungsrisikos beigetragen hat. Hinsichtlich der Prävention von Kindesmisshandlung kann ein psychodiagnostisches Verfah-

ren eingesetzt werden, um risikobelastete Eltern einer entsprechenden Intervention zuzuführen, bevor ein tatsächlicher Misshandlungsfall auftritt.

Die Möglichkeit zu einem sinnvollen Einsatz psychologischer Diagnostika in der Kinderschutzarbeit wurde in den Vereinigten Staaten von Amerika bereits in den späten 1970ern in der Forschergruppe um Joel S. Milner erkannt. Das Ergebnis der infolge dieser Erkenntnis angestoßenen Entwicklungen wurde von Milner im Jahr 1986 unter der Bezeichnung „Child Abuse Potential Inventory“ publiziert und fasst die Befunde einer über zehn Jahre andauernden Konstruktionsphase zusammen.

4. Ziel und Aufbau der vorliegenden Arbeit

Obwohl ein Inventar zur Feststellung des elterlichen Misshandlungsrisikos im US-amerikanischen Bereich nun seit mehr als 20 Jahren besteht, lassen sich für den deutschen Sprachraum keine vergleichbaren Verfahren feststellen (Deegener & Körner, 2006). Über die Gründe, warum im deutschen Sprachraum bisher keine ähnlichen Konstruktionsversuche unternommen wurden, kann nur gemutmaßt werden. Festzuhalten ist jedoch, dass ein Einsatz einer deutschen Übersetzung des Fragebogens ohne vorhergehende Prüfung seiner Gültigkeit im Rahmen der oben genannten Einsatzbereiche unverantwortbar ist, da mit einer fälschlichen Feststellung bzw. Nicht-Feststellung des Misshandlungsrisikos ernsthafte Konsequenzen für Kind und Elternteil (bspw. eine grundlose Entziehung des Sorgerechts bzw. das Unterlassen einer Therapie bei einem misshandelnden Elternteil) erwartet werden können.

Die eben genannte Überprüfung der Validität, die für einen Einsatz einer deutschsprachigen Übersetzung des Verfahrens notwendig ist, wurde im Rahmen der vorliegenden Arbeit geleistet. Diesbezügliche Grundlagen und Vorgehensweisen sollen im Folgenden beschrieben werden. Inhaltlich gliedert sich die Arbeit in vier wesentliche Kapitel. In einem ersten Kapitel werden grundlegende theoretische Ansätze und Befunde zu der Ätiologie von Kindesmisshandlung beschrieben. Diese sind zunächst wichtig, um abschätzen zu können, wie gut sich das entwickelte Verfahren in wissenschaftliche Vorarbeiten und Erkenntnisse zu dem Thema einreihen lässt. Ein zweites Kapitel stellt die wesentlichen Befunde, die hinsichtlich der Konstruktion

der amerikanischen Originalversion berichtet werden, dar. Die Kenntnis dieser Befunde ist aus zweierlei Gründen wichtig. Zunächst geht hieraus hervor, welche spezifischen Strategien bei der Entwicklung des Originalverfahrens angewendet wurden. Weiterhin erlaubt dies die Möglichkeit eines Vergleiches zwischen der Originalversion und der deutsche Übersetzung. Die eigentlichen Befunde der Untersuchung zur Validierung der deutschen Version werden in einem dritten Kapitel dargestellt. Eine integrative Besprechung und Diskussion der in den vorhergehenden Kapiteln dargestellten Befunde wird in einem vierten Kapitel geleistet. Ein Schwerpunkt wird hierbei auf der Besprechung der Validität des Verfahrens vor dem Hintergrund des aktuellen wissenschaftlichen Kenntnisstandes zur Ätiologie von Kindesmisshandlung sowie dem Vergleich zwischen amerikanischer Originalversion und deutscher Übersetzung liegen.

B. Kindesmisshandlung: Definition, Verbreitung, Ätiologie, empirische Befunde

In den folgenden Abschnitten werden theoretische Grundlagen und empirische Forschungsbefunden zum Thema Kindesmisshandlung dargestellt. Ein erster Abschnitt befasst sich mit der Einführung einer allgemeinen Definition des Kindesmisshandlungsbegriffs und der Darstellung diesbezüglicher Subtypen. In einem zweiten Abschnitt werden Befunde zur Häufigkeit des Auftretens von Kindesmisshandlungsfällen besprochen. Modellhafte Vorstellungen zur Ätiologie von Kindesmisshandlung werden in einem dritten Abschnitt beschrieben. Ein letzter Abschnitt befasst sich mit der Darstellung empirischer Forschungsbefunde zur Validität der vorgestellten Modelle.

1. Definition und Einteilung verschiedener Formen der Kindesmisshandlung

Blum-Maurice, Knoller, Nitsch und Kröhnert (2000) definieren Kindesmisshandlung als eine „nicht zufällige, gewaltsame psychische und/oder physische Beeinträchtigung oder Vernachlässigung des Kindes durch Eltern/Erziehungsberechtigte oder Dritte, die das Kind schädigt, verletzt, in seiner Entwicklung hemmt oder zu Tode bringt“ (S.2).

Zusätzlich zu dieser allgemeinen Definition werden üblicherweise vier Untertypen der Kindesmisshandlungen unterschieden: Vernachlässigung, psychische und körperliche Misshandlung sowie sexueller Missbrauch (Deegener, 2005; Engfer, 2005). Unter dem Begriff der Vernachlässigung wird die Nichterfüllung elterlicher¹ Pflichten verstanden. Ein Kind wird demnach vernachlässigt, wenn es von den Eltern unzureichend ernährt, gepflegt, gefördert, gesundheitlich versorgt, beaufsichtigt oder vor Gefahren geschützt wird. Fälle von psychischer Kindesmisshandlung liegen dann vor, wenn das Kind durch nicht-physische Handlungen Erwachsener (z.B. Beschimpfen, Verspotten, Einsperren) geängstigt und überfordert oder ihm das Gefühl von Wertlosigkeit vermittelt wird. Körperliche Kindesmisshandlung umfasst alle physischen Formen der Gewaltanwendung, die bei einem Kind zu Verletzungen führen

¹ Hier und im weiteren Verlauf der vorliegenden Arbeit wird lediglich die Bezeichnung „Eltern“ verwendet. Damit sind gleichzeitig biologische Mütter bzw. Väter sowie erziehungsberechtigte Dritte gemeint.

können. Als Beispiele hierfür können Handlungen wie das Schlagen, Schütteln, Verbrennen oder Stechen eines Kindes angesehen werden. Sexueller Missbrauch bezeichnet jede sexuelle Handlung, die an einem Kind vorgenommen wird und der es sich aufgrund körperlicher Unterlegenheit nicht widersetzen kann bzw. zu der es aufgrund mangelnder emotionaler, geistiger oder sprachlicher Reife nicht einwilligen kann. Die sexuellen Handlungen des Misshandlers können hierbei passiver (z.B. Exhibitionismus, Beobachtung des Kindes beim Anziehen oder Baden) oder auch aktiver Art (z.B. Berühren der Genitalien des Kindes, Vollzug oraler, analer oder vaginaler Formen des Geschlechtsverkehrs) sein.

Engfer (2005) unterscheidet zusätzlich zwischen Misshandlung im engeren und der Misshandlung im weiteren Sinn, wobei die Unterteilung der beiden Formen anhand der Intensität (z.B. Häufigkeit, Dauer, Schweregrad der Misshandlungen) und der Abweichung von kulturellen Normen erfolgt (vgl. auch Barnett, Manley & Cicchetti, 1993). Als Misshandlungen im engeren Sinn werden Fälle verstanden, die zu körperlichen Verletzungen des Kindes (z.B. Blutergüssen, Knochenbrüchen) geführt haben oder in denen Geschlechtsverkehr mit dem Kind versucht oder erzwungen wurde. Misshandlungen im weiteren Sinn bezeichnen Handlungen, die nicht zu gravierenden physischen oder psychischen Beeinträchtigungen des Kindes führen und in einem geringeren Maße als Normabweichung gelten. Bezüglich der Kindesvernachlässigung sowie der körperlichen und psychischen Misshandlung könnten hierbei beispielsweise Fälle angeführt werden, in denen häufig mit dem Kind geschimpft wird, es mit der flachen Hand geschlagen oder durch Liebesentzug bestraft wird. Sexueller Missbrauch im weiteren Sinn bezeichnet Fälle, in denen kein Körperkontakt (z.B. Exhibitionismus) oder nur wenig schädigende Handlungen (z.B. sexualisierte Küsse, Berühren der Brüste eines Mädchens) vorzufinden sind.

Obwohl sexueller Missbrauch üblicherweise unter dem Oberbegriff der Kindesmisshandlung subsummiert wird, zeigt sich vor allem hinsichtlich der Ätiologie eine starke Heterogenität zu den anderen Misshandlungsformen (Engfer, 2005). Hieraus lässt sich ableiten, dass Fragebogeninventare, die zur Abschätzung der Wahrscheinlichkeit des Auftretens von körperlicher und psychischer Kindesmisshandlung bzw.

Kindesvernachlässigung konstruiert wurden, keine hinreichend genaue Vorhersage für die Auftretenswahrscheinlichkeit von sexuellem Kindesmissbrauch zulassen. Dies gilt gleichermaßen für das Verfahren, das im Rahmen der vorliegenden Arbeit entwickelt wurde. Eine weitere Auseinandersetzung mit der Häufigkeit und Ätiologie von sexuellem Missbrauch wird daher im weiteren Verlauf nicht erfolgen. Für weitere Informationen zu dieser Thematik sei an dieser Stelle lediglich auf einschlägige Übersichtsarbeiten (z.B. Engfer, 2005; Araji & Finkelhor, 1986; Bange & Deegener, 1996) verwiesen.

2. Häufigkeit von Kindesmisshandlung

In den folgenden Abschnitten werden Befunde zur Auftretenshäufigkeit von Kindesmisshandlungsfällen vorgestellt. Die Darstellung bezieht sich hierbei zunächst auf Zahlen aus der Bundesrepublik Deutschland (BRD), denen jedoch aufgrund untersuchungsrelevanter rechtlicher Rahmenbedingungen nur beschränkte Aussagekraft zugemessen werden kann (Kapitel B, Abschnitt 2.1). Aufgrund der Einschränkungen der deutschen Befunde sollen zusätzlich auch Häufigkeiten aus Kriminalitätsstatistiken einer anderen westlichen Industrienation besprochen werden. Hierbei wird auf Zahlen aus den Vereinigten Staaten von Amerika (USA) zurückgegriffen, denen aufgrund der in US-amerikanischen Kinderschutzverordnungen verankerten umfassenden Meldepflicht (Kapitel B, Abschnitt 2.2) eine höhere Validität zugesprochen werden kann.

2.1 Deutsche Dunkelfeldstudien zur Häufigkeit von Kindesmisshandlung

Verlässliche Kriminalstatistiken zur Prävalenz von Kindesmisshandlung liegen für den Bereich der BRD nicht vor. Begründet ist dies vor allem in dem bisherigen Fehlen einer Meldepflicht bei Verdacht auf oder Wissen über das Vorliegen eines Falls von Kindesmisshandlung, deren Einführung „vor allem mit dem Argument abgelehnt wurde, dass Eltern dann noch weniger bereit sein würden, misshandlungsbedingte kindliche Verletzungen von Ärzten behandeln zu lassen“ (Engfer, 2005; S.7). Die Verwendung polizeilicher Kriminalstatistiken wird vor allem aus drei Gründen heraus nicht als ausreichende Datenbasis für die Analyse der Häufigkeit von

Kindesmisshandlungen erachtet (vgl. Engfer, 1997; Wetzels & Pfeiffer, 1997, Buchner et al., 2001). Zunächst wird davon ausgegangen, dass die hohe soziale Nähe zwischen Täter und Opfer die Wahrscheinlichkeit einer Anzeige verringert (Raupp & Eggerts, 1993; Ruback, 1993). Weiterhin wird angenommen, dass Täter aus unteren sozialen Schichten in Kriminalstatistiken überproportional häufig erfasst werden (Wetzels & Pfeiffer, 1997). Da die Anzeige von Straftaten in der Regel durch die Opfer erfolgt, ist zuletzt eine altersspezifische Verzerrung der Daten zu erwarten, insofern, als Kinder nicht in der Lage sind, von sich selbst aus Anzeige zu erstatten (Buchner et al., 2001).

Aufgrund dieser Beschränkungen gehen die aussagekräftigsten Befunde zur Häufigkeit von Kindesmisshandlungsfällen in der BRD aus sozialwissenschaftlichen Dunkelfeldstudien hervor. Als Dunkelfeld wird dabei der Anteil tatsächlich vorliegender Misshandlungsfälle verstanden, der in amtlichen Kriminalstatistiken nicht aufgeführt wird. Hauptsächlich finden hierbei sogenannte Opferbefragungen statt, in denen eine repräsentative Stichprobe aus der Gesamtbevölkerung befragt wird, ob und welche Erfahrungen sie mit verschiedenen Formen der Kindesmisshandlung gemacht haben (Göppinger, 1997). Im Folgenden werden die wichtigsten Befunde derartiger Befragungen wiedergegeben. Für eine vertiefende Lektüre sei auf die aktuellen und umfassenden Überblicksartikel von Deegener (2005) und Engfer (2005) verwiesen.

Große repräsentative Dunkelfeldstudien zur Prävalenz von körperlicher Misshandlung und Gewalt durch Eltern wurden von dem Kriminologischen Forschungsinstitut Niedersachsen (KFN) durchgeführt. In einer Befragung von 3214 Personen im Alter zwischen 16 und 59 Jahren konnte festgestellt werden, dass 74,9% der Probanden angaben, in ihrer Kindheit Formen körperliche Gewalt durch Eltern erlebt zu haben. In 10,8% der Fälle lagen dabei Formen tatsächlicher körperlicher Misshandlung (z.B. würgen, verprügeln, verbrennen) vor, die in 4,7% häufiger als selten auftraten (Wetzels & Pfeiffer, 1997; Pfeiffer & Wetzels, 1997). Von der selben Forschergruppe wurde ebenfalls eine weitere große Befragung an einer altershomogeneren Gruppe von 16190 Schülerinnen und Schüler der neunten und zehnten Jahrgangsstufen all-

gemeinbildender Schulen durchgeführt (Pfeiffer, Wetzels & Enzmann, 1999). Es zeigte sich hierbei, dass 43,3 % der Befragten angaben, nie Opfer körperlicher Züchtigung oder Misshandlung durch die Eltern geworden zu sein. 29,7% der Probanden berichteten von seltenen körperlichen Züchtigung (z.B. Ohrfeigen), 17,1% von häufigen derartigen Übergriffen. Tatsächliche Misshandlungen lagen in seltener Form bei 4,5%, gehäufte Misshandlungen bei 5,3% aller Probanden vor. Als zusammenfassendes Fazit der Sichtung dieser und ähnlicher Studien schließt Engfer (2000), dass „ca. die Hälfte bis zwei Drittel der deutschen Eltern ihre Kinder körperlich bestrafen, [wobei] schwerwiegende und relativ häufigere Bestrafungen [...] sich bei 10 bis 15% der deutschen Eltern“ (S.27) finden würden.

Hinsichtlich der Prävalenz von Kindesvernachlässigung und seelischer Misshandlung gibt es laut Deegener und Körner (2006) „in Deutschland keine wissenschaftlich seriös erhobenen Angaben zur Häufigkeit von [...] Vernachlässigung“ (S.86). Die einzige, wenn auch mittlerweile veraltete Ausnahme stellt die Studie von Frank (1993; Frank & Kopecky-Wenzel, 2002) dar. Dieser führte eine Untersuchung durch, in der 714 Ärzte und 685 Schwestern des Münchner Universitätsklinikums bei 3% bzw. 6% aller neu aufgenommen Kindern Vernachlässigung feststellten. Schätzungen der relativen Häufigkeit von Vernachlässigungsfällen werden ansonsten lediglich auf einem sehr undifferenzierten Niveau getroffen. Beispielsweise mutmaßt Engfer (2000), dass „Vernachlässigungen [...] offenbar wesentlich häufiger [...] als Misshandlungen“ (S.26) vorkommen.

2.2 Kindesmisshandlung in amerikanischen Kriminalitätsstatistiken

Verlässlichere Angaben zu der Häufigkeit verschiedener Misshandlungsformen finden sich in US-amerikanischen Kriminalstatistiken. Seit der Verabschiedung des Federal Child Abuse Prevention and Treatment Act (CAPTA) im Jahr 1974 sind das Wissen über bzw. der Verdacht auf das Vorliegen eines Falles von Kindesmisshandlung in den USA grundsätzlich meldepflichtig. Derartige Meldungen erfolgen bei den Child Protection Services der jeweiligen Bundesstaaten, durch die sie in einem zweistufigen Begutachtungsprozess weiter verfolgt werden. In diesem wird auf einer

ersten Stufe festgestellt, ob die Meldung bestimmten bundesstaatlichen Eingangskriterien entspricht. Liegen diese vor, so ordnet der zuständige Child Protection Service eine weitergehende Untersuchung des Falles an. Als Ergebnis der Untersuchung wird der Fall hinsichtlich des Vorliegens von Kindesmisshandlung einer von drei Kategorien zugeordnet. Als unbestätigt werden Fälle klassifiziert, bei denen kein hinreichender Verdacht besteht, dass tatsächliche Kindesmisshandlung oder ein erhöhtes Misshandlungsrisiko vorliegt. Fälle, in denen eine Misshandlung nicht nachgewiesen werden kann, aber dennoch gesicherter Verdacht oder ein erhöhtes Misshandlungsrisiko besteht, werden als indizierte Fälle erfasst. Als gesichert werden zuletzt Fälle klassifiziert, in denen Misshandlung eindeutig nachgewiesen bzw. mit an Sicherheit grenzender Wahrscheinlichkeit vermutet werden kann². Die landesweite Integration dieser Befunde findet in dem National Child Abuse and Neglect Data System statt. Eine differenzierte Auswertung dieses Datensatzes findet sich in dem Report „Child Maltreatment“, der jährlich von dem US Department of Health and Human Services (USDHHS) herausgegeben wird.

Die aktuellen Statistiken beziehen sich auf das Geschäftsjahr 2005 (USDHHS, 2007). In diesem gingen ca. 3,3 Millionen Meldungen bei den Child Protective Services ein, die sich auf eine Gesamtheit von ca. 6 Millionen Kindern beziehen. 37,9% der gemeldeten Fälle wurden nach der ersten Stufe des Begutachtungsprozesses nicht weiter verfolgt, da sie die bundesstaatlichen Kriterien einer weiteren Überprüfung nicht erfüllten. Bei den restlichen 62,1% der Meldungen, die einer weiteren Untersuchung unterzogen wurden, konnte in 71,4% der Fälle weder tatsächliche Kindesmisshandlung noch ein erhöhtes Misshandlungsrisiko festgestellt werden. In 3% der Fälle erschien der Verdacht der Kindesmisshandlung oder ein erhöhtes Misshandlungsrisiko als indiziert. In den übrigen 25,6% der Fälle wurde eine tatsächliche Misshandlung oder ein an Sicherheit grenzender Missbrauchsverdacht festgestellt. Ausgehend von diesen Befunden wird geschätzt, dass im Jahr 2005 in den USA ca.

² Die Kriterien, die für die Anordnung einer weitergehenden Untersuchung bzw. die Einordnung in die verschiedenen Kategorien angelegt werden unterscheiden sich zwischen den einzelnen Bundesstaaten. Eine detaillierte Aufstellung der in den einzelnen Staaten geltenden Kriterien findet sich in dem aktuellen Report „Child Maltreatment 2005“ des US Department of Health and Human Services (2007).

899000 Kinder Opfer verschiedener Formen der Kindesmisshandlung wurden. Umgerechnet auf die damalige Anzahl von Kindern in der Gesamtbevölkerung entspricht dies einer Häufigkeit von 12,1 Misshandlungsfällen pro 1000 Kindern.

Ebenfalls werden Angaben zu der Häufigkeit verschiedener Misshandlungsformen gemacht, wobei berücksichtigt wird, dass im Falle eines Kindes mehrere Misshandlungsformen auch gleichzeitig auftreten können, weswegen sich die im Folgenden dargestellten Prozentzahlen auch zu einem Wert größer als 100 aufsummieren. Mit einer relativen Häufigkeit von 62,8% stellen Formen der Vernachlässigung die häufigste Art der Kindesmisshandlung dar. Gefolgt wird diese von Formen der körperlichen Misshandlung, die in 16,6% aller Fälle festgestellt wurden und sexuellem Missbrauch mit einer relativen Häufigkeit von 9,3%. Psychische Misshandlung stellt mit einem Anteil von 7,1% die seltenste, der in der eingangs angeführten Definition enthaltenen Misshandlungsformen dar. Zusätzlich werden Häufigkeitsangaben zu gesundheitlicher Vernachlässigung und anderen Formen der Kindesmisshandlung (z.B. Kindesaussetzung, Drogenmissbrauch während der Schwangerschaft) gemacht, wobei erstere in 2% und letztere in 14,3% der erfassten Fälle auftraten.

Die Stabilität dieser Befunde lässt sich durch den Vergleich mit Zahlen aus den Vorjahren bewerten. Angaben zu der Auftretenshäufigkeit von Kindesmisshandlungen zwischen den Jahren 2001 und 2004 werden ebenfalls in dem Report des USDHHS (2007) dargestellt. Sie variieren zwischen 12,0 und 12,5 Misshandlungsfällen pro 1.000 Kindern und deuten damit auf eine hohe Stabilität der allgemeinen Prävalenz von Kindesmisshandlungen über die einzelnen Jahre hinweg hin. Befunde zur relativen Häufigkeit verschiedener Misshandlungsformen finden sich in einem Artikel von Emery und Laumann-Billings (1998). Diese berichten aus einer ähnlichen Statistik aus dem Jahr 1996 eine relative Häufigkeit von 54% für körperliche Vernachlässigung, 25% für körperliche Misshandlung, 11% für sexuellen Missbrauch, 3% für emotionale Vernachlässigung und 7% unklassifizierbarer Fälle. Es kann hierbei festgestellt werden, dass sich im Vergleich zu den Befunden aus 2005 gewisse Verschiebungen einzelner Prozentzahlen ergeben haben, die Rangreihe der relativen Häufig-

keiten jedoch identisch ist, was darauf hindeutet, dass auch die Verteilung der einzelnen Misshandlungsformen eine gewisse Stabilität aufweist.

3. Erklärungsmodelle zur Ätiologie von Kindesmisshandlung

Im Folgenden sollen prominente Ätiologiemodelle zur Entstehung von Kindesmisshandlung besprochen werden. Ein erster Abschnitt wird sich hierbei mit psychopathologischen, sozial-situativen und soziologischen Ansätzen beschäftigen, die die historische Grundlage derzeit bestehender Erklärungsansätze darstellen. In einem zweiten Abschnitt wird beschrieben, wie bestimmte Aspekte dieser Modelle aufgrund ihrer konkurrenten Gültigkeit in integrative Ursachenmodelle integriert wurden, die heutzutage als anerkannte theoretische Grundlage der Erklärung der Entstehung von Kindesmisshandlung gelten. Ein letzter Abschnitt befasst sich mit der Darstellung metaanalytischer Befunde, die die Gültigkeit der dargestellten Erklärungsmodelle in Bezug auf die Entstehung von Kindesmisshandlung bestätigen.

3.1 Psychopathologische, sozial-situative und soziologische Ansätze

Historisch gesehen gehen Erklärungsmodelle zur Ätiologie von Kindesmisshandlung vor allem aus drei Forschungstraditionen hervor. Engfer (2005, 1986; vgl. auch Belsky, 1993; Ziegler, 1994) unterscheidet hierbei drei prominente Erklärungsansätze, die als psychopathologisches, soziologisches und sozial-situatives Modell bezeichnet werden.

In psychopathologischen Ansätzen wird die Person des Misshandlers als Grundlage des Auftretens von Kindesmisshandlung betrachtet. Zentrale Bedeutung wird dabei Persönlichkeitseigenschaften und -problemen des Täters zugewiesen. Manche Autoren (z.B. Gelles & Straus, 1979; Schreiber, 1971) heben in diesem Zusammenhang vor allem das Vorhandensein pathologischer Störungsbilder (z.B. Angst, Depression, verminderte kognitive Leistungsfähigkeit; für einen Überblick vgl. Ziegler, 1994) hervor, während andere Autoren (z.B. Kempe, 1973; Brown & Daniels, 1968) im Zusammenhang der These der mehrgenerationalen Weitergabe der Gewalt vermuten, dass eigene kindliche Gewalterfahrungen der Täter ausschlaggebend für das Auftreten von Kindesmisshandlung sind.

Sozial-situationale Modelle zur Ätiologie von Kindesmisshandlung verstehen Gewalthandlungen gegen Kinder als Endpunkte eskalierender Konfliktsituationen. Wesentlich für die Entstehung derartiger Konflikte werden einerseits Merkmale des Kindes, wie z.B. gesundheitliche Probleme und Behinderungen (Belsky, 1993) sowie problematische Verhaltensweisen des Kindes (Lösel & Bliesener, 2003) verstanden. Andererseits wird ebenfalls die Bedeutung elterlicher Merkmale, wie Persönlichkeitsprobleme und unzureichendes Betreuungsverhalten hervorgehoben (Engfer, 1991). Hierbei wird angenommen, dass sich Konflikte in einer Wechselwirkung zwischen beiden Akteuren hochschaukeln und die Eltern aus Ärger und Ohnmacht ihre Kinder misshandeln, wenn andere pädagogische Maßnahmen fehlgeschlagen sind.

Soziologische Erklärungsmodelle konzeptionalisieren gesellschaftliche Faktoren als Grundlage von Gewalthandlungen gegenüber Kindern. Engfer (2005) beschreibt in diesem Zusammenhang drei Bedingungsfaktoren, die in verschiedenen derartigen Erklärungsmodellen als bedeutsam betrachtet werden. Ein erster Faktor befasst sich mit dem Ausmaß, in dem die umgebende Gesellschaft Gewalt gegenüber Kindern oder auch im Allgemeinen billigt. Die Überforderung von Familien durch ökonomische Lebensbelastungen (z.B. Armut, Arbeitslosigkeit) stellt einen zweiten Faktor dar. Zuletzt wird in soziologischen Modellvorstellungen ebenfalls angenommen, dass soziale Isolation sowie unzureichende Unterstützungssysteme zu einer erhöhten Belastung von Familien führen und damit auch für das Auftreten von Kindesmisshandlungsfällen verantwortlich sein können.

3.2 Integrative Ursachenmodelle

In den Ergebnissen empirischer Untersuchungen, die ausgehend von den eben genannten Erklärungsmodellen durchgeführt wurden, zeigt sich, dass sowohl psychopathologische als auch sozial-situative und soziologische Faktoren an der Entstehung von Kindesmisshandlungsfällen beteiligt sind (Black, Heyman & Slep, 2001; Schumacher, Slep & Heyman, 2001; Bender & Lösel, 2005a, 2005b). Da folglich keinem der drei eben genannten Modelle exklusive Gültigkeit unterstellt werden kann, werden die verschiedenen Erklärungsansätze als sinnvolle Konsequenz der Befund-

lage in integrativen Ursachenmodellen zusammengefasst (Ziegler, 1994). Diese Integration erfolgt vor dem Hintergrund ökosystemischer und transaktionaler Überlegungen (Belsky 1980, 1993; Cicchetti & Lynch, 1993; Cicchetti, Toth, & Maughan, 2000).

Von einer ökosystemischen Perspektive (Bronfenbrenner, 1981) ausgehend, wird argumentiert, dass sowohl die individuellen Einflüsse psychopathologischer Ansätze als auch sozial-situative und soziologische Faktoren der Umwelt an der Entstehung von Misshandlungsfällen gleichermaßen beteiligt sind. Diese Einflüsse werden entsprechend der Nähe, die sie zu dem tatsächlichen Misshandlungsgeschehen haben in proximale (von lat. proximus = der Nächste) und distale (von lat. distare = sich entfernen) Ebenen eingeteilt. Belsky (1980, 1993) unterscheidet vier diesbezügliche Ebenen: Auf der Makrosystemebene werden globalgesellschaftliche Einflüsse angesiedelt, die Kindesmisshandlung bedingen können (z.B. kulturelle Werte und Normen). Einflüsse des sozialen Umfeldes, also enger gefasster institutioneller und informeller sozialer Strukturen (z.B. Gemeinde, Nachbarschaft) werden unter dem Begriff der Exosystemebene subsummiert. Das Mikrosystem umfasst Einflüsse, die in dem unmittelbaren sozialen Umfeld (z.B. Familie, Freundeskreis) vorherrschen. Individuelle Faktoren, die an der Entstehung von Kindesmisshandlungsfällen beteiligt sein können (z.B. Eigenschaften von Kind und misshandelndem Elternteil), stellen die letzte Ebene dar.

Transaktionale Gesichtspunkte (Sameroff & Chandler, 1975; Cicchetti & Rizley, 1981) sind in integrativen Ursachenmodellen insofern enthalten, als hierbei die Entstehung von Kindesmisshandlung nicht lediglich durch ein Vorherrschen von pathogenetischen Einflüssen, die das Kindesmisshandlungsrisiko steigern können (Risikofaktoren), sondern ebenfalls auch durch die weitgehende Abwesenheit von Einflüssen, die das Misshandlungsrisiko absenken können (Ressourcen, Schutzfaktoren), erklärt werden (Bender & Lösel, 1998, 2005a, 2005b). Zusätzlich werden Risikofaktoren und Ressourcen nach ihrer zeitlichen Permanenz eingeteilt. Lediglich kurzfristig auftretende Risikofaktoren werden hierbei als vorübergehende Belastungen bezeichnet, kurzfristig auftretende Ressourcen als temporäre Puffer. Handelt es sich

um längerfristig wirkende Risikofaktoren, werden diese als dauerhafte Vulnerabilitäten bezeichnet. Handelt es sich um längerfristig vorherrschende Ressourcen, so wird von dauerhaft protektiven Faktoren gesprochen (Cicchetti & Lynch, 1993; Bender & Lösel, 2005b). Das Zusammenwirken der einzelnen Faktoren wird hier je nach Modellvorstellung als linear additiv, multiplikativ oder exponentiell konzeptualisiert (Masten, Best & Garmezy, 1990). Generell wird jedoch angenommen, dass eine Kumulation von Risiken die Wahrscheinlichkeit der Entstehung von Kindesmisshandlung erhöht und derartige Fälle vor allem dann auftreten, wenn es zu einem deutlichen Übergewicht von Risikofaktoren kommt, das durch die vorhandenen Ressourcen nicht mehr abgefangen werden kann.



Abbildung B.1: Integratives Ursachenmodell zur Entstehung von Kindesmisshandlung

Abbildung B.1 fasst die eben besprochenen Überlegungen integrativer Ursachenmodelle zur Ätiologie von Kindesmisshandlung überblicksmäßig zusammen. Zunächst kann festgestellt werden, dass die Bedingungsfaktoren gemäß transaktionaler Überlegungen in pathogenetische Risikofaktoren und salutogenetische Ressourcen aufgeteilt werden. Innerhalb dieser beiden Klassen wird, entsprechend ökosystemischer Überlegungen, zwischen der ontogenetischen Ebene sowie der Ebene des Mikrosystems, des Exosystems und des Makrosystems unterschieden. Das Misshandlungsrisiko, das in der Mitte der Abbildung dargestellt wird, ist gleichsam durch vorhandene Risikofaktoren und Ressourcen beeinflusst. Überwiegen vor allem die Risikofaktoren, so steigt die Auftretenswahrscheinlichkeit von Kindesmisshandlung, überwiegen vor allem die Ressourcen, so ist das Auftreten von Kindesmisshandlung eher unwahrscheinlich.

3.3 Metaanalytische Befunde zu Risiko- und Schutzfaktoren

In der aktuellen Forschungsliteratur findet sich eine Vielzahl von Untersuchungen, die sich mit verschiedenen misshandlungsrelevanten Risiko- und Schutzfaktoren auf der ontogenetischen sowie der Mikro- und Exosystemebene beschäftigen. Neben einigen narrativer Reviews (z.B. Belsky, 1993; Black, Heyman & Slep, 2001; Schumacher, Slep & Heyman, 2001) wurden ebenfalls zwei metaanalytische Aufarbeitungen dieser Befundlage durchgeführt (Stith et al., 2004; Stith, im Druck). Interessant sind sie an dieser Stelle vor allem, da sie einen umfassenden und integrativen Überblick über die Bedeutsamkeit der untersuchten Risiko- und Schutzfaktoren bieten. Die folgenden Ausführungen befassen sich mit der Metaanalyse von Stith et al. (2004). Für die vorliegende Arbeit sind die Ergebnisse dieser Arbeit relevant, da sie eine Abschätzung der Bedeutsamkeit der untersuchten Risiko- und Schutzfaktoren aufgliedert nach den eben besprochenen ökologischen Ebenen erlauben.³

In einer Literaturrecherche wurden insgesamt 211 Studien gefunden, die den von den Autoren im Vorfeld der Metaanalyse festgelegten Einschlusskriterien genügten (Stith et al., 2004). In diesen werden insgesamt 934 Effektstärken für den Zusammenhang zwischen der Ausprägung 80 verschiedener Risikofaktoren⁴ und dem Auftreten von Kindesmisshandlung berichtet. Gemäß inhaltlicher Gesichtspunkte wurden die Einflussfaktoren der ontogenetischen bzw. der Mikro- oder Exosystemebene zugeordnet.

Die Integration der Effektstärken der gefundenen Untersuchungen erfolgte entsprechend der metaanalytischen Strategie von Hedges und Olkin (1985). Diese besteht im Wesentlichen darin, die gefundenen Effektgrößen um den Einfluss der Stichprobengröße zu korrigieren und die korrigierte Effektgrößenverteilung mittels eines χ^2 -

³ Die Ergebnisse der Metaanalyse von Stith et. al. (im Druck) werden aus zwei Gründen nicht besprochen. Zunächst bezieht sie sich lediglich auf die Mikrosystemebene und macht daher keine Aussagen zu der Bedeutsamkeit von Faktoren auf der ontogenetischen und der Exosystemebene. Weiterhin werden auf der Mikrosystemebene nur wenige Studien berücksichtigt, die nicht bereits in die Metaanalyse von Stith et. al. (2004) einbezogen wurden.

⁴ Im Folgenden wird der Einfachheit der Darstellung halber ausschließlich von Risikofaktoren gesprochen. Faktoren, bei denen sich negative Effektstärken zeigen, können als Schutzfaktoren verstanden werden.

Tests auf Homogenität zu prüfen. Fällt dieser Test zugunsten der Homogenitätsannahme aus, so kann die mittlere korrigierte Effektstärke als „wahrer“ Effekt des Treatments, in diesem Falle des Risikofaktors angenommen werden. Kann die Homogenität der Effektgrößenverteilung nicht angenommen werden, so kann die Signifikanz des Effektes des Treatments überprüft werden, indem ein 95%-Konfidenzintervall um die mittlere korrigierte Effektstärke gelegt wird. Schließt dieses die Null nicht ein, so kann von der Signifikanz der „wahren“ Effektstärke ausgegangen werden. Schließt das Konfidenzintervall die Null ein, so kann kein signifikanter Effekt des Risikofaktors angenommen werden (Johnson, Mullen & Salas, 1995; Thomas & French, 1986).

Vor der eigentlichen Besprechung der Ergebnisse der Metaanalyse soll an dieser Stelle zunächst die von Cohen (1988, 1992) eingeführte Klassifikation von Effektstärken eingeführt werden, anhand derer die Bedeutsamkeit des Einflusses der im Folgenden besprochenen Risikofaktoren abgeschätzt werden kann. Gemäß dieser Klassifikation können Effekte mit einer Stärke zwischen 0,2 und 0,5 als schwache, Effekte mit einer Stärke zwischen 0,5 und 0,8 als mittlere und Effekte mit einer Stärke größer 0,8 als starke Effekte bezeichnet werden.⁵

Aus Gründen der Übersichtlichkeit wurden die bei Stith et al. (2004) dargestellten Tabellen modifiziert. Zunächst wurden die Angaben zu allen Risiko- und Schutzfaktoren, denen gemäß der metaanalytischen Überprüfung lediglich ein Effekt in der unteren Hälfte des Bereichs eines schwachen Effektes zugemessen werden kann (d.h. Effekte mit einer Größe kleiner 0,35), nicht berücksichtigt. Weiterhin wurden die Ergebnisse gemäß der jeweiligen ökologischen Systemebene verschiedenen Tabellen zugeordnet. Die originalen Tabellen aus der Metaanalyse von Stith et al. (2004) werden der Vollständigkeit halber im Anhang (Kapitel G, Abschnitt 1) aufgeführt.

⁵ In Bezug auf Mittelwertsunterschiede gibt die Effektstärke den Abstand der Werte zweier Gruppen gemessen in Standardabweichungen an. Für weitere Erläuterungen der Herleitung der Effektstärkenklassifikation soll auf die einschlägige Originalliteratur (Cohen, 1988, 1992) verwiesen werden.

Tabelle B.1: Ontogenetische Risikofaktoren für körperliche Misshandlung

Körperliche Misshandlung	d	CI	Q ^w	k	N
Parent Empathy	-0,94	-1,17/-0,71	5,96	6	355
Parenting Skills	-0,78	-1,17/-0,39	0,45	4	108
Parent Anger	0,78	0,35/1,20	12,82**	3	128
Parent Hyper-reactivity	0,65	0,36/0,93	1,84	5	214
Parent Anxiety	0,60	0,41/0,79	4,39	8	563
Parent Psychopathology	0,59	0,48/0,69	62,21***	13	8630
Child Aggression	0,59	0,45/0,73	43,12***	13	849
Parent Depression	0,55	0,45/0,67	46,18***	14	8258
Parent Self-esteem	-0,53	-0,67/-0,39	25,20**	10	2453
Child Social Competence	-0,53	-0,64/-0,42	27,46*	14	1527
Stress over Parenting	0,46	0,26/0,67	10,89*	5	379
Child Externalizing Behavior	0,42	0,33/0,52	68,51***	16	2022
Self-evaluation as Parent	-0,40	-0,64/-0,16	9,23	5	296
Child Non-Compliant/Difficult	0,41	0,24/0,59	34,69***	9	512

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Tabelle B.1 gibt einen Überblick über ontogenetische Risikofaktoren, die bei der Entstehung von körperlicher Kindesmisshandlung von Bedeutung sind. Bei der Betrachtung misshandlungsbezogener Risikofaktoren kann zunächst festgestellt werden, dass zehn der vierzehn aufgefassten Einflussgrößen elternbezogene Aspekte darstellen, die in diesem Zusammenhang als überwiegend bezeichnet werden können. Inhaltlich lassen sich verschiedene Bereiche elterlicher Risikofaktoren ausmachen, die sich auf klinische Störungsbilder (Parent Anger, Parent Hyper-reactivity, Parent Anxiety, Parent Psychopathology, Parent Depression), elterliche Persönlichkeitseigenschaften (Parent Empathy, Parent Self-esteem), Wahrnehmungen des eigenen Erziehungsverhaltens (Stress over Parenting, Self-evaluation as Parent) sowie spezifische Verhaltensweisen der Eltern (Parenting Skills) beziehen. Kindbezogene Einflussfaktoren auf das Misshandlungsrisiko beziehen sich auf spezifische Verhal-

tensweisen (Child Externalizing Behavior, Child Non-Compliant/Difficult), klinische Störungen (Child Aggression) sowie Persönlichkeitseigenschaften des Kindes (Child Social Competence). Von der Höhe der Effektgrößen lassen sich leicht stärkere Effekte bei den elterlichen Einflussgrößen feststellen, die von ihrem Betrag her einen mittleren Effekt von 0,63 zeigen. Entsprechend der Effektgrößenklassifikation von Cohen (1988, 1992) kann dieser als ein Effekt in der oberen Hälfte des Bereichs eines mittleren Effektes bezeichnet werden. Der betragsmäßige Effekt kindbezogener Risikofaktoren ist mit einem Mittelwert von 0,49 als schwächer zu beurteilen, fällt jedoch ebenfalls in den oberen Bereich eines mittleren Effektes.

Tabelle B.2: Ontogenetische Risikofaktoren für Kindesvernachlässigung

Kindesvernachlässigung	d	CI	Q ^w	k	N
Parenting Skills	-0,82	-1,23/-0,41	0,71	3	98
Parent Self-esteem	-0,69	-0,99/-0,38	1,98	4	184
Child Social Competence	-0,62	-0,81/-0,43	10,03	7	584
Child Aggression	0,59	0,38/0,81	20,08**	8	374
Child Non-Compliant/Difficult	0,59	0,26/0,93	7,24	4	149
Parent Psychopathology	0,52	0,39/0,66	14,45*	8	7652
Parent Depression	0,42	0,32/0,53	15,93*	8	8207
Child Externalizing Behavior	0,35	0,18/0,52	9,07	9	598

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Befunde zu ontogenetischen Risikofaktoren für Kindesvernachlässigung werden in Tabelle B.2 dargestellt. Hierbei lassen sich insgesamt acht bedeutsame Einflussgrößen feststellen. Vier dieser Faktoren beziehen sich auf elterliche, vier weitere auf kindbezogene Aspekte. Elternbezogene Risikofaktoren sind im Einzelnen Verhaltensweisen (Parenting Skills), Persönlichkeitseigenschaften (Parent Self-esteem) sowie klinische Störungen der Eltern (Parent Psychopathology, Parent Depression). Risikofaktoren der Kinder beziehen sich auf Verhaltensweisen (Child Externalizing Behavior), Verhaltensauffälligkeiten (Child Aggression, Child non-compliant/difficult)

sowie Persönlichkeitseigenschaften (Child Social Competence) des Kindes. Die Effekte der elterlichen Risikofaktoren können mit einer mittleren betragsmäßigen Effektstärke von 0,61 gemäß der Klassifikation von Cohen (1988, 1992) als Effekte in der oberen Hälfte des Bereichs eines mittleren Effektes bezeichnet werden. Der Effekt kindbezogener Einflussgrößen liegt mit einer mittleren betragsmäßigen Effektstärke von 0,54 nur wenig unter der der elternbezogenen und ist als mittlerer Effekt zu bezeichnen.

Tabelle B.3: Risikofaktoren für körperliche Misshandlung auf Mikrosystemebene

Körperliche Misshandlung	d	CI	Q ^w	k	N
Child Attachment to caregiver	-1,21	-1,55/-0,87	13,64**	4	117
Family Conflict	0,73	0,40/1,07	8,12*	4	154
Lack of Family Cohesion	-0,68	-0,98/-0,38	3,02	5	183
Parent perceives Problem Child	0,62	0,53/0,71	57,45***	24	3297
Parent Use of Corporal Punishment	0,55	0,38/0,72	4,65	7	703
Parent Unrealistic Expectations	0,49	0,35/0,62	42,12***	10	2325
Parent-child Relationship	-0,47	-0,58/-0,36	88,37***	30	1554
Spousal Violence	0,46	0,31/0,61	6,82	5	773
Authoritarian Parenting Style	0,42	0,27/0,57	33,37***	13	767

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Risikofaktoren für die Entstehung von körperlicher Kindesmisshandlung, die sich auf Mikrosystemebene ergeben können, werden in Tabelle B.3 dargestellt. Als bedeutsam für die Entstehung von körperlicher Kindesmisshandlung können neun Faktoren ausgemacht werden. Fünf dieser Einflussgrößen beziehen sich auf die Beziehung zwischen Kind und Elternteil. Inhaltlich wichtig sind dabei die Beziehung zwischen Kind und Elternteil (Child Attachment to Caregiver, Parent-child Relationship), Wahrnehmungen und Einstellungen zu dem Kind (Parent perceives Problem Child, Parent Unrealistic Expectations) sowie Aspekte des Erziehungsverhaltens (Authoritarian Parenting Style). Zwei weitere Faktoren befassen sich mit dem Klima des fa-

miliären Umfeldes, wobei Konflikte innerhalb (Family Conflict) und Zusammenhalt der Familie (Lack of Family Cohesion) hervorzuheben sind. Zuletzt ist auch das generelle Auftreten von Gewaltakten innerhalb der Familie von Bedeutung für die Entstehung von Kindesmisshandlungsfällen. Hierzu lassen sich Einflüsse der Akzeptanz körperlicher Züchtigung (Parent Use of Corporal Punishment) sowie das Vorhandensein von Gewalt zwischen den Elternteilen (Spousal Violence) zählen. Die höchste Bedeutung kann dem Klima des familiären Umfeldes zugemessen werden, dessen mittlere betragsmäßige Effektgröße sich in einer Höhe von 0,71 bewegt und daher, gemäß der Klassifikation von Cohen (1988, 1992), beinahe als starker Effekt bezeichnet werden kann. Mit einer mittleren betragsmäßigen Effektstärke von 0,64 sind die Einflüsse der Beziehung zwischen Kind und Elternteil leicht weniger bedeutsam. Dennoch kann dieser Faktorengruppe ein Effekt im oberen Bereich der mittleren Größe zugesprochen werden. Zuletzt zeigen die Einflüsse des Vorhandenseins von Gewalt im familiären Umfeld eine mittlere betragsmäßige Effektgröße von 0,51, die als ein Effekt mittlerer Größe betrachtet werden kann.

Tabelle B.4: Risikofaktoren für Kindesvernachlässigung auf Mikrosystemebene

Kindesvernachlässigung	d	CI	Q ^w	k	N
Parent Perceives Problem Child	1,06	0,54/1,58	3,61	3	67
Parent-child Relationship	-0,91	-1,13/-0,69	47,20***	10	380
Family Size	0,54	0,45/0,64	75,19***	12	8546
Parent Unrealistic Expectations	0,39	0,25/0,53	14,95*	6	878

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Tabelle B.4 stellt vier Risikofaktoren der Mikrosystemebene dar, die als bedeutsam für die Entstehung von Kindesvernachlässigung erachtet werden. Drei dieser Einflussgrößen beziehen sich auf das Verhältnis zwischen Kind und Elternteil. Inhaltlich sind hierunter kindbezogene Wahrnehmungen (Parent Perceives Problem Child) und Erwartungen (Parent Unrealistic Expectations) des vernachlässigenden Elternteils sowie dessen Verhältnis zu dem vernachlässigten Kind (Parent-child Relation-

ship) zu verstehen. Ein vierter Risikofaktor bezieht sich auf die Familiengröße (Family Size), wobei in größeren Familien ein erhöhtes Vernachlässigungsrisiko zu erwarten ist. Mit einer mittleren betragsmäßigen Effektgröße von 0,79 kann dem Einfluss der Beziehung zwischen Kind und Elternteil auf die Entstehung von Kindesvernachlässigung gemäß der Klassifikation von Cohen (1988, 1992) beinahe ein starker Effekt zugemessen werden. Der Effekt des Einflusses der Familiengröße liegt mit einer Effektstärke von 0,54 unter dem der Beziehungsaspekte, kann aber dennoch als mittlerer Effekt klassifiziert werden.

Tabelle B.5: Risikofaktoren für körperliche Misshandlung auf Exosystemebene

Körperliche Misshandlung	d	CI	Q ^w	k	N
Parent received Harsh Discipline	0,59	0,41/0,77	40,92***	5	552
Parent Poor Relationship	0,44	0,34/0,54	20,38*	11	2997
Parent Stress	0,42	0,32/0,53	47,78***	20	1674
Parent Financial Stress	0,41	0,20/0,62	4,72	4	397
Parent Criminal Record	0,40	0,20/0,59	0,34	3	563
Parent Experienced Abuse	0,38	0,26/0,49	33,94***	10	3170
Parent Social Support	-0,36	-0,45/-0,28	65,32***	20	10315
Socio-economic Status	-0,37	-0,54/-0,21	22,04***	5	7423

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

In Tabelle B.5 werden Faktoren der Exosystemebene aufgelistet, die für die Entstehung von körperlicher Kindesmisshandlung von Bedeutsamkeit sind. Hierbei können acht Faktoren als bedeutsam angenommen werden. Zwei dieser Faktoren beziehen sich auf Misshandlungen und Gewalterfahrungen, die die Eltern bei ihren eigenen Eltern miterlebt haben (Parent received Harsh Discipline, Parent Experienced Abuse). Drei weitere Faktoren spiegeln die Ungünstigkeit der familiären Lebensumstände wieder, wobei hierunter sowohl eine allgemeine Stressbelastung der Eltern (Parent Stress), als auch finanzielle Probleme der Familie (Parent Financial Stress, Socio-economic Status) zu verstehen sind. Die Unterstützung der Familie durch

Dritte stellt eine weitere Gruppe von Einflussfaktoren dar. Inhaltlich lassen sich hierunter Einflussfaktoren fassen, die einen Mangel an außerfamiliären Kontakten (Parent Poor Relationship) bzw. mangelnde gesellschaftliche Unterstützung (Parent Social Support) widerspiegeln. Ein letzter Risikofaktor befasst sich mit einer kriminellen Vergangenheit der Eltern (Parent Criminal Record). Den stärksten Einfluss auf die Entstehung von Kindesmisshandlung haben Gewalterfahrungen, die das misshandelnde Elternteil erlebt hat. Mit einer mittleren betragsmäßigen Effektgröße von 0,49 kann dieser Einfluss als Effekt mittleren Ausmaßes betrachtet werden. Leicht schwächer stellen sich Effekte der Ungünstigkeit der Lebensumstände, des Mangels an Unterstützung sowie der kriminellen Vergangenheit der Eltern dar. Mit einer mittleren betragsmäßigen Effektgröße von 0,4 kann diesen drei Faktorengruppen jedoch ebenfalls ein Einfluss auf die Entstehung von Kindesmisshandlung in der unteren Hälfte des Bereichs eines mittleren Effektes zugemessen werden.

Tabelle B.6: Risikofaktoren für Kindesvernachlässigung auf Exosystemebene

Kindesvernachlässigung	d	CI	Q ^w	k	N
Parent Stress	0,81	0,60/1,02	24,94***	3	386
Unemployment	0,51	0,35/0,67	8,69*	4	719
Parent Education Level	-0,41	-0,54/-0,28	44,83***	8	7743
Income	-0,38	-0,52/-0,24	31,50***	7	861
Parent Poor Relationship	0,39	0,25/0,54	15,40*	7	855

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Fünf bedeutsame Risikofaktoren für die Entstehung von Kindesvernachlässigung auf Exosystemebene werden in Tabelle B.6 angeführt. Inhaltlich beziehen sich diese zunächst auf die Ungünstigkeit der familiären Lebensumstände, wobei sowohl die allgemeine Stressbelastung der Eltern (Parent Stress), als auch finanzielle Probleme der Familie von Bedeutsamkeit sind (Unemployment, Income). Ebenfalls bedeutsam für die Entstehung von Kindesvernachlässigung ist in diesem Zusammenhang der Bildungsstand der Eltern (Parent Educational Level), wobei ein niedriger Bildungs-

stand mit einer erhöhten Vernachlässigungswahrscheinlichkeit einhergeht. Ein letzter bedeutsamer Risikofaktor ist die Reichhaltigkeit außerfamiliärer Kontakte der Eltern (Parent Poor Relationship). Bei der Betrachtung der Effektstärken lässt sich feststellen, dass die allgemeine Stressbelastung der Eltern mit einer mittleren betragsmäßigen Effektgröße von 0,57 den bedeutsamsten Risikofaktor für die Entstehung von Kindesvernachlässigung darstellt. Dieser Effekt ist entsprechend der Effektgrößenklassifikation von Cohen (1988, 1992) in der oberen Hälfte des Bereiches eines Effektes mittleren Ausmaßes einzuordnen. Einen geringeren Einfluss hat der Bildungsstand der Eltern. Mit einer Effektgröße von 0,41 kann der Einfluss dennoch als mittlerer Effekt klassifiziert werden. Zuletzt lässt sich hinsichtlich der außerfamiliären Beziehungen der Eltern ein Einfluss auf die Entstehung von Kindesvernachlässigung in Höhe von 0,39 feststellen, was ebenfalls einem mittleren Effekt entspricht.

4. Zusammenfassung der dargestellten Befunde und Überlegungen

In den vorherigen Abschnitten wurde auf Definitionen, Häufigkeitsabschätzungen und Ätiologiemodelle von körperlicher und psychischer Kindesmisshandlung sowie Kindesvernachlässigung eingegangen. Hinsichtlich der dargestellten Definitionsversuche lässt sich feststellen, dass der Bereich stimmig definiert und untergliedert werden kann (Kapitel B, Abschnitt 1). In der Darstellung der Häufigkeitsabschätzungen wurde zunächst aufgezeigt, dass sich für den Bereich der BRD keine zuverlässigen Kriminalstatistiken zur Prävalenz verschiedener Misshandlungsformen feststellen lassen. Befunde aus Dunkelfeldstudien lassen darauf schließen, dass Erfahrungen mit körperlicher Gewalt durch Eltern bei einem Großteil (ca. 70%) der Bevölkerung vorliegen, während tatsächliche Misshandlungen bei einem kleineren, aber dennoch bedeutsamen Anteil von ca. 10% der Bevölkerung zu vermuten sind (Kapitel B, Abschnitt 2.1). Verlässlichere Zahlen aus Kriminalitätsstatistiken der USA schätzen das Auftreten tatsächlicher Kindesmisshandlungen mit einer relativen Häufigkeit von ca. 1% weitaus kleiner ein, wobei jedoch zu vermuten ist, dass ein hierbei bestehendes Dunkelfeld sowie strenge Klassifikationskriterien eine Unterschätzung der tatsächlichen Auftretenswahrscheinlichkeit zur Folge hat (Kapitel B, Abschnitt 2.2). Aus der Darstellung der genannten Erklärungsansätze zur Ätiologie

von Kindesmisshandlung geht hervor, dass Modelle aus verschiedenen Forschungstraditionen inhaltlich stimmig zu einem integrativen Gesamtmodell zusammengefasst werden können, das verschiedene Risiko- und Schutzfaktoren auf unterschiedlichen ökologischen Ebenen als Grundlage der Entstehung bzw. des Ausbleibens von Kindesmisshandlung konzipiert (Kapitel B, Abschnitt 3.2). Die Gültigkeit eines derartigen Modells kann hierbei durch die dargestellte metaanalytische Aufarbeitung bestehender Forschungsbefunde angenommen werden (Kapitel B, Abschnitt 3.3).

Zusammenfassend lässt sich festhalten, dass Kindesmisshandlung aufgrund definitorischer Vorarbeiten als inhaltlich kohärentes Konstrukt aufgefasst werden kann. Prävalenzeinschätzungen zeigen, dass es sich bei Kindesmisshandlung um ein gesellschaftliches Randphänomen handelt, wobei der Anteil belasteter Personen in der Gesellschaft nicht unterschätzt werden sollte. Theoretische Überlegungen und konkrete Forschungsbefunde zu Bedingungskonstellationen, die das Misshandlungsrisiko erhöhen, bieten letztlich die Möglichkeit, die Ursachen von Kindesmisshandlung zu verstehen und weiter zu untersuchen.

C. Das Child Abuse Potential Inventory

Das Child Abuse Potential Inventory (CAPI; Milner, 1986) stellt die US-amerikanische Originalversion des Verfahrens dar, das im Rahmen der vorliegenden Arbeit validiert wurde. Erste Schritte zur Konstruktion des Verfahrens wurden bereits Mitte der 1970er Jahre unternommen. Ziel hierbei war es, ein kurzes, leicht administrierbares Fragebogeninstrument zu entwickeln, mithilfe dessen misshandelnde Probanden mit hinreichender Genauigkeit von nicht-misshandelnden unterschieden werden können. Die Ergebnisse der zehnjährigen Entwicklungsphase veröffentlichte Milner im Jahr 1986 in Form eines zusammenfassenden Manuals, das auch heute noch die zentrale Referenz des CAPI darstellt.

Obwohl zu dem Zeitpunkt der Veröffentlichung des Manuals bereits eine Vielzahl verschiedener Checklisten und Instrumente zur Feststellung des Kindesmisshandlungsrisikos, wie zum Beispiel das Michigan Screening Profile of Parenting (MSPP; Helfer, Hoffmeister & Schneider, 1978) oder der Parenting Stress Index (PSI; Abidin, 1983) bestand, zeigte sich in den USA schon bald eine zunehmende Verbreitung des CAPI im Bereich von Sozialfürsorge, Familienhilfe und Forschung (Milner, 1989). Heutzutage wird das CAPI in der Forschungsliteratur als der „weltweit am besten empirisch überprüfte Fragebogen [...] zur Erfassung des Risikos zukünftiger körperlicher Misshandlung“ (Deegener & Körner, 2006; S. 54; vgl. auch Milner, 1998; Ondersma, Chaffin, Mullins & LeBreton, 2005) besprochen.

Im Folgenden werden die bedeutsamsten Befunde aus den Entwicklungs- und Validierungsuntersuchungen, die im Rahmen der Konstruktion des CAPI berichtet wurden, dargestellt. Die Präsentation der Ergebnisse orientiert sich hierbei an der Skalenstruktur des CAPI, die in einem ersten Abschnitt eingehend besprochen werden soll. Die darauf folgenden Abschnitte befassen sich mit den Oberbereichen der Skalenstruktur, der Abuse Scale und den Validity Scales. In dem letzten Abschnitt wird eine abschließende Bewertung des CAPI erfolgen, die sich im Wesentlichen mit der Betrachtung der klassischen Testgütekriterien sowie dem Bezug zwischen inhaltli-

chen Gesichtspunkten des CAPI und den im ersten Abschnitt besprochenen wissenschaftlichen Befunden zur Entstehung von Kindesmisshandlung befasst.

1. Umfang und Skalenstruktur

Der Fragebogen des CAPI besteht insgesamt aus 160 einfach formulierten Selbstaussagen, die von den Probanden in einem dichotomen Antwortformat („agree“ vs. „disagree“) hinsichtlich des Zutreffens auf die eigene Person beantwortet werden. 40 Items wurden zum Zweck einer Weiterentwicklung des Verfahrens als „Experimentalitems“ aufgenommen. Die übrigen 120 Items werden bei der Auswertung der Fragebögen zu insgesamt 10 Skalenkennwerten aggregiert, die das Endergebnis einer Risikoerfassung mittels des CAPI darstellen.

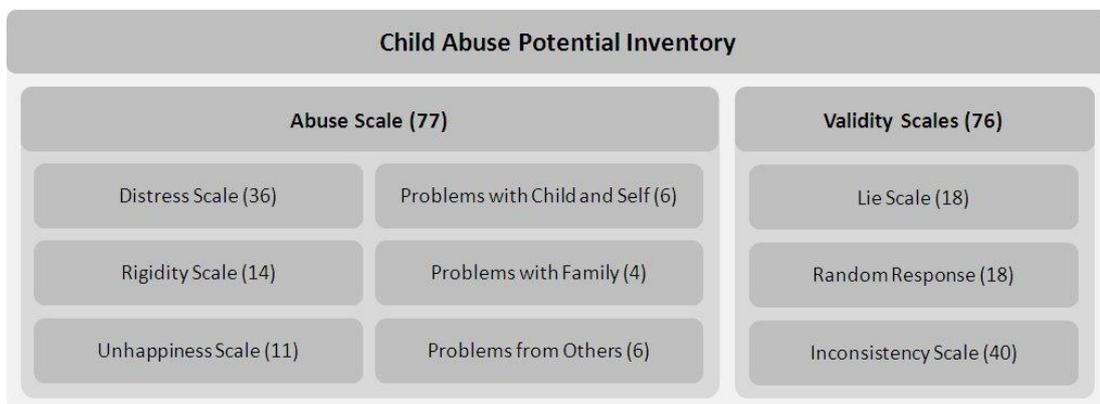


Abbildung C.1: Skalenstruktur des Child Abuse Potential Inventory. In den Klammern hinter den Skalenbezeichnungen wird die Anzahl der Items angegeben, die in die Berechnung der Skalen eingehen. 33 Items der Abuse Scale gehen ebenfalls in die Berechnung der Inconsistency Scale ein, weswegen sich die Itemzahlen im unteren Bereich der Tabelle zu einem Wert größer 120 aufsummieren.

Die Struktur dieser Skalen wird in Abbildung C.1 überblicksmäßig dargestellt. Hier können zunächst zwei Oberbereiche ausgemacht werden. Die eigentliche Abschätzung des Kindesmisshandlungsrisikos erfolgt durch den ersten Oberbereich, der Abuse Scale. Die zugehörigen Subskalen (Distress, Rigidity, Unhappiness, Problems with Child and Self, Problems with Family, Problems with Others) ermöglichen zusätzlich eine differenzierte Betrachtung spezifischer Gesichtspunkte des Misshandlungsrisikos. Die Detektion verfälschender Antworttendenzen der Probanden erfolgt durch die Validity Scales, die den zweiten Oberbereich der Skalenstruktur des CAPI darstellen. Inhaltlich erfassen diese sozial erwünschtes (Lie Scale), unreflektiertes

und zufälliges (Random Response Scale) sowie widersprüchliches Antwortverhalten (Inconsistency Scale).

2. Abuse Scale

Die folgenden Abschnitte befassen sich mit der Abuse Scale, die die primäre klinische Skala des CAPI darstellt und als Indikator des zu messenden Misshandlungsrisikos dient. In einem ersten Abschnitt soll zunächst die Strategie, die zur Entwicklung dieser Skala angewandt wurde, dargestellt werden. Ein zweiter Abschnitt befasst sich mit der Reliabilität der Skala, wobei einerseits auf Befunde zur internen Konsistenz, andererseits auf Befunde zur zeitlichen Stabilität von Werten in der Skala eingegangen werden soll. In einem abschließenden Abschnitt werden Befunde zur Validität der Abuse Scale dargestellt. Hierbei werden Aspekte der Inhaltsvalidität, der faktoriellen Validität sowie der Kriteriumsvalidität besprochen.

2.1 Entwicklung

Da zum Zeitpunkt der Konstruktion des CAPI kein allgemein anerkanntes Ätiologiemodell zur Entstehung von Kindesmisshandlung vorlag, bestanden erste Schritte zur Generierung eines Itempools in einer extensiven Literaturrecherche. In dieser konnten Milner und Williams (1978) über 700 Artikel und Bücher finden, in denen theoretische Überlegungen bzw. empirische Befunde zur Entstehung verschiedener Misshandlungsformen dargestellt wurden. Milner (1986) fasst die folgenden Bereiche als zentrales Ergebnis auf: „unrealistic child-rearing attitudes and expectations, anxiety over a child’s behavior, problems in interpersonal relationships, feelings of inadequacy, feelings of isolation and loneliness, depression, vulnerability, insecurity, inability to handle stress, rigid attitudes, impulsivity, dependency, immaturity, negative childhood experiences including abuse and neglect, [...] problems in parental relationships, [...] poverty, lack of education, drug abuse and psychosis“ (S.20).

Zu den einzelnen Risikofaktoren, die in der Literaturrecherche gefunden wurden, wurde eine Anzahl von 15 bis 20 Items generiert. Dieser Itempool wurde von Sozialarbeitern und Mitarbeitern von Child Protective Services im Hinblick auf die Fähigkeit zur Erfassung von Kindesmisshandlungsrisiko begutachtet. Der bestehende

Itempool wurde hinsichtlich der Anmerkungen der Experten überarbeitet. Hierbei wurden einige Items ausgeschlossen, weitere neu aufgenommen und andere in ihrer Formulierung verändert. Nach Abschluss der Generierungsphase bestand ein Itempool von insgesamt 334 Items.

Die konstruierten Items wurden anschließend auf ihre Fähigkeit zur Trennung zwischen Gruppen misshandelnder Eltern und Kontrollprobanden überprüft. Milner (1986) beschreibt hierzu drei Validitätsstudien. Eine erste Studie befasste sich mit der Identifikation diskriminierender Items. Hierbei wurden die 334 konstruierten Items einer Stichprobe von 19 misshandelnden Eltern und 19 gemäß verschiedener demographischer Variablen parallelisierten Kontrollprobanden vorgegeben. Die Identifikation diskriminierender Items erfolgte durch die Durchführung univariater t-Tests zwischen den beiden Gruppen. Es zeigten sich bei 44 Items signifikante Mittelwertsunterschiede mit Irrtumswahrscheinlichkeiten kleiner 5% zwischen den beiden Gruppen. Bei 28 weiteren Items zeigten sich Mittelwertsunterschiede, die mit Irrtumswahrscheinlichkeiten kleiner 10% noch als marginal signifikant bezeichnet werden können. Bei 39 Items fanden sich zuletzt Mittelwertsunterschiede, die mit einer Irrtumswahrscheinlichkeit kleiner 20% als signifikant angenommen werden können. 28 Items zeigten Decken- bzw. Bodeneffekte, d.h. die Antworten aller Probanden konzentrierten sich auf eine der beiden Antwortalternativen.

In der Folge der ersten Validitätsstudie wurde ein revidierter Fragebogen erstellt. Die 44 Items, die signifikant zwischen den beiden Gruppen trennten, sowie die 28 Items, die Decken- oder Bodeneffekte zeigten, wurden in ihrem originalen Wortlaut übernommen. Die 67 Items, deren Mittelwertsunterschiede zwischen den beiden Gruppen lediglich marginale Signifikanz mit Irrtumswahrscheinlichkeiten im Bereich von 10% bis 20% zeigten, wurden in einer eindeutigeren Weise umformuliert. Weiterhin wurden 21 zusätzliche Items in den neuen CAPI-Fragebogen aufgenommen, weshalb der revidierte CAPI-Fragebogen einen Umfang von insgesamt 160 Items hatte.

Der revidierte Fragebogen wurde einer weiteren Probandenstichprobe vorgegeben. Diese bestand aus insgesamt 130 Probanden, wobei es sich in 65 Fällen um misshandelnde Eltern, in 65 Fällen um parallelisierte Kontrollprobanden handelte. Die Fähigkeit der Items, zwischen den beiden Stichproben zu unterscheiden, wurde wiederum mittels t-Tests auf Itemebene überprüft. Hierbei zeigten sich bei 77 Items signifikante Mittelwertsunterschiede zwischen den Gruppen der misshandelnden Eltern und der Kontrollprobanden mit einer Irrtumswahrscheinlichkeit kleiner 5%. In einer Diskriminanzanalyse wurde getestet, inwiefern diese 77 Items zusammen in der Lage sind, zwischen den beiden Gruppen zu trennen. In den Ergebnissen dieser Analyse zeigte sich, dass 125 der 130 Probanden korrekt als Fälle aus der Gruppe der Misshandler bzw. der Kontrollgruppe klassifiziert werden konnten. Dies entspricht einem Anteil von 96% korrekt klassifizierter Fälle.

Zur Absicherung gegen mögliche Stichprobeneffekte wurde weiterhin eine dritte Validierungsstudie durchgeführt. Hierbei sollte überprüft werden, inwiefern sich die guten Klassifikationsergebnisse, die sich in der zweiten Validierungsstudie gezeigt haben, auch in einer hiervon unabhängigen weiteren Stichprobe zeigen würden. Hierzu wurden die 77 Items, die in der zweiten Validierungsstudie signifikante Mittelwertsunterschiede zwischen misshandelnden Eltern und Kontrollprobanden zeigten, verwendet. Diese wurden einer Stichprobe von 220 Probanden vorgegeben, die zur einen Hälfte aus misshandelnden Eltern, zur anderen Hälfte aus gemäß bestimmter demographischer Variablen parallelisierten Kontrollprobanden bestand. Bei der Betrachtung der Mittelwertsunterschiede der Items zeigten sich ähnliche Ergebnisse wie bereits in der zweiten Validierungsstudie. 64 der 77 Items zeigten in t-Tests signifikante Unterschiede zwischen den misshandelnden Eltern und den Kontrollprobanden. Marginale Signifikanz mit Irrtumswahrscheinlichkeiten kleiner 10% zeigte sich bei fünf der übrigen dreizehn Items. Bei den restlichen sieben Items zeigte sich, dass die Mittelwertsdifferenzen zwischen den beiden Gruppen in die gleiche Richtung deuteten, wie bereits in der zweiten Validierungsstudie. Bei der Durchführung einer Diskriminanzanalyse zeigte sich zunächst eine schlechtere Rate korrekt klassifizierter Probanden von 85,4%. Bei der weiteren Untersuchung des Datensatz

wurde jedoch festgestellt, dass 45 misshandelnde Eltern und 29 Kontrollprobanden deutlich erhöhte Werte auf der zu diesem Zeitpunkt bereits entwickelten Lie Scale (Kapitel C, Abschnitt 3.1) zeigten. In Folge dieser Feststellung wurden die Probanden mit erhöhten Lügenskalenwerten aus der Stichprobe ausgeschlossen und eine weitere Diskriminanzanalyse für die übrigen Probanden durchgeführt. Es zeigte sich, dass in dieser Stichprobe 93,2% der Probanden richtig klassifiziert werden konnten. Dieses Ergebnis stellt eine ähnlich gute Klassifikation, wie bereits in der zweiten Validierungsstudie, dar. Die gefundenen 77 Items wurden daher als endgültige Form der Abuse Scale gewählt.

2.2 Reliabilität

Die aussagekräftigsten Befunde zur Reliabilität der Abuse Scale lassen sich der Darstellung in dem technischen Manual des CAPI (Milner, 1986) entnehmen. Die dargestellten Werte beziehen sich hierbei auf Maße der internen Konsistenz und Maße der zeitlichen Stabilität der Werte in der Abuse Scale.

Zur internen Konsistenz der Abuse Scale werden bei Milner (1986) Ergebnisse in Form von Kennwerten der Kuder-Richardson-Formel 20 (KR20; Kuder & Richardson, 1937) sowie in Form von mittels der Spearman-Brown-Formel korrigierter Split-Half-Reliabilitäten (SH; Lienert, 1989) dargestellt. Die Angaben beruhen auf einer Stichprobe von 2610 Probanden, wobei es sich in 548 Fällen um tatsächliche Misshandler, in den übrigen 2062 Fällen um Kontrollprobanden handelt. Die angegebenen Reliabilitätskoeffizienten werden nach verschiedenen Eigenschaften der Probanden (Misshandler vs. Kontrollgruppe, Geschlecht, Alter, Bildungsstand, ethnische Abstammung) aufgeschlüsselt. Diesbezügliche SH-Werte bewegen sich in einem Bereich von .93 und .99, während die KR20-Werte in einem Bereich von .85 bis .96 rangieren. Insgesamt deuten diese Befunde auf eine sehr hohe interne Konsistenz der Abuse Scale hin, weshalb die diesbezügliche Reliabilität als hinreichend gesichert angesehen werden kann.

Milners (1986) Angaben zur zeitlichen Stabilität von Werten auf der Abuse Scale beziehen sich auf eine Substichprobe von 549 Kontrollprobanden aus der Stichprobe,

die bereits zur Überprüfung der internen Konsistenz verwandt wurde⁶. Hierbei wurde der CAPI-Fragebogen 125 Probanden nach einem Tag, 162 Probanden nach einer Woche, 112 Probanden nach einem Monat und 150 Probanden nach drei Monaten unter sonst gleichbleibenden Bedingungen erneut vorgegeben. Die Ermittlung der Retest-Reliabilität erfolgte durch die Berechnung von Produkt-Moment-Korrelationen zwischen den Werten auf der Abuse Scale zum ersten und zweiten Messzeitpunkt. Hierbei konnte für den eintägigen Zeitraum eine Korrelation von .91, für den einwöchigen eine Korrelation von .90, für den einmonatigen eine Korrelation von .83 und für den dreimonatigen eine Korrelation von .75 festgestellt werden. Diese Werte werden zusätzlich, auch aufgeschlüsselt nach den oben genannten Eigenschaften der Probanden, dargestellt. Hierbei können Korrelationen im Bereich zwischen .65 und .94 festgestellt werden. Zusätzlich wurde ebenfalls überprüft, ob sich signifikante Mittelwertsunterschiede zwischen den beiden Messzeitpunkten ergaben, was jedoch in keiner der betrachteten Konstellation der Fall war. Vor diesem Hintergrund kann der Abuse Scale ebenfalls im Hinblick auf zeitliche Stabilität eine gute bis sehr gute Reliabilität zugemessen werden.

Zusätzlich zu den hier dargestellten Befunden findet sich in der Literatur eine Vielzahl weiterer Untersuchungen, die die Reliabilität der Abuse Scale gleichermaßen stützen (z.B. Burrell, Thompson & Sexton, 1992; Caliso & Milner, 1992; Milner & Robertson, 1990). Eine Zusammenfassung dieser Befundlage findet sich in einem Überblicksartikel von Milner (1994).

2.3 Validität der Abuse Scale

Die folgenden Abschnitte befassen sich mit spezifischen Aspekten der Validität, die der Abuse Scale des CAPI zugemessen werden können. Schwerpunkte werden hierbei auf die Inhaltsvalidität, die faktorielle Validität sowie die Kriteriumsvalidität der Abuse Scale gelegt.

⁶ Der ausschließliche Rückgriff auf Kontrollprobanden ist hierbei dadurch zu erklären, dass die misshandelnden Eltern in dieser Stichprobe bereits in misshandlungsbezogenen Therapieprogrammen aufgenommen wurden, weshalb zu vermuten ist, dass bei ihnen Merkmalsveränderungen aufgetreten sind, die eine Unterschätzung der zeitlichen Stabilität der Abuse Scale zur Folge haben.

2.3.1 Inhaltsvalidität

Die Inhaltsvalidität der Abuse Scale kann durch einen Vergleich der durch die Items abgefragten Inhalte und den aus der Forschungsliteratur bekannten Theorien und Befunden zur Entstehung von Kindesmisshandlung besprochen werden. Wie bereits in dem Abschnitt zur Entwicklung der Abuse Scale beschrieben wurde, wurden die Items des CAPI auf der Basis eines extensiven Reviews der damals aktuellen Forschungsliteratur entwickelt (Milner & Williams, 1978). Die Inhaltsvalidität bezüglich des damaligen Forschungsstands kann damit als hinreichend gesichert betrachtet werden.

Bezüge zu der aktuellen Forschungsliteratur können durch den Rückgriff auf integrative Ursachenmodelle zur Ätiologie von Kindesmisshandlung (Kapitel B, Abschnitt 3.2) sowie auf diesbezügliche metaanalytische Befunde zu Risiko- und Schutzfaktoren (Stith et al., 2004; vgl. Kapitel B, Abschnitt 3.3), die in dem zweiten Kapitel der vorliegenden Arbeit besprochen werden, festgestellt werden. Hinsichtlich ontogenetischer Risiko- und Schutzfaktoren werden zunächst klinische Störungsbilder und Persönlichkeitseigenschaften der Eltern wie z.B. Depression (CAPI 49: „Ich bin manchmal sehr traurig“), Erregbarkeit (CAPI 17: „Ich bin oft innerlich ärgerlich“) und Ängstlichkeit (CAPI 93: „Ich habe Ängste, von denen niemand etwas weiß“) abgefragt. Ebenfalls beziehen sich die Items des CAPI auf problematische Eigenschaften des Kindes wie z.B. Verhaltensauffälligkeit (CAPI 45: „Ich habe ein Kind, das böse ist“), Devianz (CAPI 69: „Ich habe ein Kind, das oft in Schwierigkeiten gerät“), Langsamkeit (CAPI 128: „Ich habe ein Kind, das langsam ist“) und sonstige Probleme (CAPI 113: „Mein Kind hat besondere Probleme“) des Kindes. Bezüglich Risiko- und Schutzfaktoren auf Mikrosystemebene werden zunächst elterliche Wahrnehmungen und Einstellungen zu dem Kind, wie z.B. übersteigerte Erwartungen an Sauberkeit (CAPI 122: „Ein gutes Kind hält seine Spielsachen und seine Kleidung sauber und ordentlich“) und Verträglichkeit (CAPI 54: „Ein Kind sollte nie widersprechen“) erfasst. Weiterhin stellt das CAPI ebenfalls Fragen zu autoritärem Erziehungsverhalten (CAPI 129: „Eltern müssen Strafen benutzen, wenn sie das Verhalten eines Kindes kontrollieren wollen“), dem Klima des familiären Umfeldes (CAPI 148: „Meine Fami-

lie hat viele Probleme“) sowie zu dem Auftreten von Streits innerhalb der Familie (CAPI 83: „Meine Familie streitet viel“). Hinsichtlich der Risiko- und Schutzfaktoren auf Exosystemebene werden die Eltern zu negativen Erfahrungen in der eigenen Kindheit (CAPI 29: „Manchmal wünsche ich mir, dass mein Vater mich mehr geliebt hätte“), zur Ungünstigkeit der familiären Lebensumstände und finanziellen Problemen der Familie (CAPI 94: „Meine Familie hat Probleme zurechtzukommen“), zur Unterstützung der Familie durch Dritte (CAPI 74: „Heutzutage weiß man nicht so recht, auf wen man zählen kann“) sowie zur Reichhaltigkeit außerfamiliärer Kontakte (CAPI 81: „Ich habe einige enge Freunde in meiner Nachbarschaft“) befragt.

Die Betrachtung der Items des CAPI in Bezug auf theoretische Modelle zur Ätiologie von Kindesmisshandlung und diesbezügliche metaanalytische Befunde zeigt eine gute Passung zwischen der Forschungsliteratur und den in dem Inventar abgefragten Inhalten. Die inhaltliche Validität des CAPI kann daher auch in Bezug auf den aktuellen Forschungsstand als gegeben betrachtet werden.

2.3.2 Faktorielle Validität

Die faktorielle Struktur, die bei der Auswertung der Abuse-Skala angenommen wird, basiert auf der Untersuchung von Milner, Gold und Wimberley (1986). Hierbei wurden die 77 Items der Abuse-Skala einer Stichprobe von insgesamt 220 Probanden vorgegeben, die zu einer Hälfte aus tatsächlichen Kindesmisshandlern und zur anderen aus gemäß Alter und Geschlecht parallelisierten Kontrollprobanden bestand.

Zur explorativen Überprüfung der faktoriellen Struktur der Items wurde eine Hauptachsenanalyse durchgeführt. Die Bestimmung der Anzahl bedeutsamer Faktoren aufgrund der Ergebnisse dieser Analyse gestaltete sich hierbei schwierig. Einerseits zeigte sich, dass der nach dem Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser & Dickman, 1959) kritische Eigenwert von eins erst vor dem neunten Faktor unterschritten wurde, was eine achtfaktorielle Lösung nahelegen würde. Andererseits kann der „Knick“ des Eigenwerteverlaufs, der das Extraktionskriterium des Scree-Tests (Cattell, 1966) darstellt, entweder nach dem dritten oder nach dem siebten Faktor interpretiert werden. Die dreifaktorielle Lösung wäre hierbei zunächst argu-

mentierbar, da allein mit dieser schon 70% der Varianz des Gesamtdatensatzes aufgeklärt werden können. Ebenfalls wäre jedoch auch die Extraktion von sieben Faktoren plausibel, da sich mitunter bedeutsame Itemladungen auf den letzten vier Faktoren zeigten (Milner, 1986).

Aufgrund dieser wenig eindeutigen Ergebnisse wurde in der Folge der ersten Hauptachsenanalyse eine weitere Hauptachsenanalyse des Datensatzes mit anschließender Promax-Rotation durchgeführt. Ziel dieser Analyse war es, eine Faktorenlösung zu finden, die inhaltlichen Gesichtspunkten gemäß gut interpretierbar sein sollte. Nach einer Gegenüberstellung potentiell denkbarer Faktorstrukturen beschreibt Milner (1986), dass eine sechsfaktorielle Lösung als angemessen zu bewerten ist. Entsprechend dem Inhalt der zugehörigen Items werden die Faktoren (bzw. Skalen) als Distress, Rigidity, Unhappiness, Problems with Child and Self, Problems from Others und Problems with Family bezeichnet. Im Folgenden wird die Bedeutung der einzelnen Faktoren unter Einbezug einiger Beispielitems näher veranschaulicht.

Die Items der Distress-Skala erfassen Aspekte einer allgemeinen Belastung der Probanden. Hierbei werden unter anderem Gefühle wie Frustration (z.B. CAPI 25 „Ich fühle mich häufig sehr frustriert“), Traurigkeit (CAPI 49 „Ich bin manchmal sehr traurig“), Einsamkeit (z.B. CAPI 18 „Manchmal fühle ich mich ganz allein in dieser Welt“), Angst (z.B. CAPI 154 „Ich fühle mich oft ängstlich“), Verwirrtheit (z.B. CAPI 5 „Ich bin ein verwirrter Mensch“), Wertlosigkeit (z.B. CAPI 47 „Ich fühle mich manchmal wertlos“), Zurückweisung (z.B. CAPI 22 „Ich fühle mich häufig abgelehnt“) und Ärger (z.B. CAPI 17 „Ich bin oft innerlich ärgerlich“) erfasst.

In der Rigidity-Skala wird erfragt, inwiefern bei einem Probanden rigide Überzeugungen bezüglich verschiedener Verhaltensweisen, die Kinder zeigen sollten, vorliegen. Inhaltlich werden vor allem Überzeugungen hinsichtlich der Gepflegtheit von Kindern (z.B. CAPI 68 „Kinder sollten sauber sein“), der Gehorsamkeit (z.B. CAPI 26 „Kinder sollten immer gehorchen“), Aufmerksamkeit (z.B. CAPI 80 „Kinder sollten still sein und zuhören“) sowie Ruhe (z.B. CAPI 115 „Kinder sollte man sehen und nicht hören“) abgefragt.

Mit der Unhappiness-Skala wird ermittelt wie glücklich bzw. unglücklich ein Proband mit seinen allgemeinen Lebensumständen ist. Erfasst wird dies durch allgemeine Fragen zur Lebenszufriedenheit (z.B. CAPI 75 „Mein Leben ist glücklich“), Fragen zur Partnersituation (z.B. CAPI 147 „Im Moment bin ich total verliebt“) und zum Sexualleben (z.B. CAPI 141 „Ich habe ein gutes Sexualleben“) sowie Fragen zu freundschaftlichen Beziehungen (z.B. CAPI 81 „Ich habe einige enge Freunde in meiner Nachbarschaft“).

Die Skala Problems with Child and Self erfasst negative Einstellungen gegenüber dem eigenen Kind. Hierbei wird nach der Wahrnehmung spezieller Probleme mit dem Kind wie Bösartigkeit (z.B. CAPI 45 „Ich habe ein Kind, das böse ist“), Problemverhalten (z.B. CAPI 69 „Ich habe ein Kind, das oft in Schwierigkeiten gerät“) oder Langsamkeit des Kindes (z.B. CAPI 128 „Ich habe ein Kind das langsam ist“) gefragt.

Mittels der Skala Problems with Family werden be- und entlastende Aspekte des familiären Umfeld des Probanden abgefragt. Die Items der Skala beziehen sich dabei unter anderem auf den sozioökonomischen Status der Familie (z.B. CAPI 94 „Meine Familie hat Probleme über die Runden zu kommen“), auf die Häufigkeit von Streitigkeiten (z.B. CAPI 83 „Meine Familie streitet viel“) und allgemeine Probleme der Familie (z.B. CAPI 148 „Meine Familie hat viele Probleme“).

Zuletzt erfasst die Skala Problems from others Probleme, die Probanden mit ihrem sozialen Umfeld haben. Hierbei wird nach Aspekten sozialer Unterstützung (z.B. CAPI 13 „Man kann sich nicht auf andere verlassen“) und zurückliegenden negativen Erfahrungen mit anderen Menschen (z.B. CAPI 67 „Man hat mir viel Schmerz zugefügt“) gefragt.

2.3.3 Kriteriumsvalidität

Die Kriteriumsvalidität eines Verfahrens zeigt sich in dem Ausmaß, in dem ein Test ein bestimmtes Kriterium vorhersagen kann. Wichtig ist diese Form der Validität vor allem, wenn die Gruppenzugehörigkeit eines Probanden anhand der Testwerte bestimmt werden soll. Auch wenn sich in einem Test signifikante Unterschiede zwischen einer Kriteriums-Gruppe (z. B. Kindesmisshandler) und einer Kontrollgruppe

(z.B. keine Kindesmisshandler) zeigen, so genügt dieses Ergebnis nicht als Beweis für die Kriteriumsvalidität eines Tests. Im Folgenden werden vor allem Befunde zu zwei Bereichen der Kriteriumsvalidität dargestellt: Die konkurrente Validität bezieht sich auf Güte der Vorhersage eines zeitlich koexistierenden Merkmals. Die prädiktive Validität bezieht sich auf die Güte der Vorhersage eines Merkmals, das zeitlich versetzt zu der Testung auftreten kann.

2.3.3.1 Konkurrente Validität

In Untersuchungsdesigns, in denen eine Gruppe tatsächlicher Misshandler von einer parallelisierten Kontrollgruppe nicht misshandelnder Eltern mittels einer Diskriminanzanalyse über die Items der Misshandlungs-Skala getrennt werden sollen, kann üblicherweise ein Großteil der Probanden korrekt klassifiziert werden. Für das CAPI werden in der Literatur Anteile korrekter Klassifikationen im Bereich von ca. 85 bis 90% richtig zugeordneter Probanden berichtet (Milner & Wimberley, 1980; Milner, Gold & Wimberley, 1986; Milner & Robertson, 1989).

Einschränkend muss hierbei bemerkt werden, dass die eben geschilderten Befunde zur klassifikatorischen Güte die Obergrenze der Diskriminationsfähigkeit des CAPI darstellen. In diesen Studien herrschen zwei Bedingungen vor, die die Diskriminationsfähigkeit der Abuse Scale erhöhen. Einerseits wird in diesen Studien eine optimal trennende Diskriminanzfunktion an die Untersuchungsstichprobe angepasst. Andererseits werden in diesen Studien maximal heterogene Gruppen (d.h. tatsächliche Misshandler vs. Kontrollgruppe) verwendet. Im Folgenden werden daher Studien dargestellt, in denen überprüft wurde, ob die Diskriminationsfähigkeit der Abuse Scale auch unter ungünstigeren Bedingungen noch besteht.

Interessant ist zunächst die Betrachtung von Studien, in denen die von Milner (1986) angegebene Diskriminanzfunktion zur Klassifikation der Probanden verwendet wird und daher nicht auf die in der Untersuchung vorliegende Stichprobe angepasst wird. Der Anteil richtiger Klassifikationen in derartigen Untersuchungsdesigns ist zwar etwas geringer als in den weiter oben beschriebenen Designs, deutet aber doch auf eine gute Diskriminationsfähigkeit der Misshandlungs-Skala hin. Milner

(1986) berichtet beispielsweise einen Anteil von 90,2% richtig klassifizierter Probanden. Caliso und Milner (1992) berichten von 87,7% korrekt klassifizierten Misshandlern, 73,3% korrekt klassifizierten nicht-misshandelnden Eltern mit einer Misshandlungsvorgeschichte und 100% korrekt klassifizierten nicht-misshandelnden Eltern ohne Misshandlungsvorgeschichte. Zusätzlich berichtet Milner (1994) von zahlreichen Studien, in denen die Spezifität der Misshandlungs-Skala unter Verwendung der von Milner (1986) angegebenen Diskriminanzfunktion bei der Erkennung nicht-misshandelnder Eltern überprüft wurde. Beispielsweise konnten 100% korrekter Klassifikationen in Studien mit Müttern (Lamphear, Stets, Whitaker & Ross, 1985), Pflegemüttern (Milner, 1986, 1989) und Pflegeeltern (Couron, 1982), die von Experten als nicht-misshandelnd eingestuft wurden, festgestellt werden.

Ein Design, in dem eine eher belastete und daher weniger heterogen zu der Gruppe der Misshandler erscheinende Gruppe untersucht wurde, findet sich bei Milner (1991). Dieser überprüfte die Beeinflussungen der Spezifität der Misshandlungs-Skala bei der Erkennung nicht-misshandelnder Eltern durch gesundheitsbedingte Stressfaktoren (gebärende Mütter mit und ohne Komplikationen bei der Geburt). Hierbei wurden 94,8% der Mütter korrekterweise als nicht-misshandelnd klassifiziert. Im Hinblick auf die klassifikatorische Güte gab es in dieser Studie keine signifikanten Unterschiede zwischen den Müttern ohne und mit Geburtskomplikationen. Einschränkungen der Spezifität der Misshandlungs-Skala bei der Erkennung nicht-misshandelnder Eltern lassen sich in dieser Studie lediglich feststellen, sofern das Kind bestimmte gesundheitliche Probleme hat. Beispielsweise werden lediglich 66,7% der Mütter korrekterweise als nicht-misshandelnd klassifiziert, sofern das Kind unter Mukoviszidose leidet und lediglich 79%, sofern das Kind unter Meningitis leidet.

Couron (1982) berichtet von einer Studie, in der die Heterogenität zwischen den zu trennenden Stichproben noch geringer war als in der eben beschriebenen. Hier sollten physisch misshandelnde Eltern von anderen nicht-physisch misshandelnden Eltern (z.B. vernachlässigenden Eltern) getrennt werden. Aufgrund dieser mangelnden Trennschärfe zwischen den beiden Gruppen sinkt die Klassifikationsgüte der

Misshandlungs-Skala auf einen Anteil von 72,6% korrekt klassifizierten Probanden. Wurden jedoch zusätzlich zur Misshandlungs-Skala auch demographische Variablen und verschiedene Stressindikatoren betrachtet, so konnten mittels Diskriminanzanalyse 90,3% der Probanden richtig klassifiziert werden.

Von einer Testung der diskriminanten Validität der Misshandlungs-Skala unter denkbar ungünstigen Bedingungen berichtet Matthews (1984). Dieser entfernte „stark“ und „moderat“ misshandelnde Eltern aus seiner Stichprobe und prüfte die Misshandlungs-Skala auf ihre Fähigkeit, „geringfügig“ misshandelnde Eltern von einer Kontrollstichprobe nicht-misshandelnder Eltern zu trennen, die jedoch ebenfalls Kinder mit emotionalen oder verhaltensbezogenen Problemen hatten. Weiterhin befanden sich beide Stichproben bereits in Therapie. Trotz dieser wenig heterogenen Zusammensetzung der Gesamtstichprobe berichtet Matthews (1984) von 72,2% korrekt klassifizierter Probanden. Aufgrund der ungünstigen Bedingungen, die in dieser Studie galten, schätzt Milner (1994) die Untergrenze der klassifikatorischen Güte der Misshandlungs-Skala ungefähr bei diesem Anteil korrekt klassifizierter Probanden.

2.3.3.2 Prädiktive Validität

Es liegen nur wenige Untersuchungen zur prädiktiven Validität der Abuse Scale, d.h. ihrer Fähigkeit zur Vorhersage zukünftiger Kindesmisshandlung, vor. Dies liegt vor allem an den ethisch-moralischen Grenzen, die durch ein Design, das zur Untersuchung dieser Form der Validität nötig wäre, überschritten werden müssten. So ist es nicht verantwortbar zu beobachten, ob ein Elternteil, bei dem ein bestimmtes Ausmaß an Misshandlungsrisiko besteht, sein Kind auch wirklich misshandelt, ohne bestimmte Interventionen einzuleiten, die die Misshandlung letztendlich abwenden sollen. Als problematisch für die Untersuchung der prädiktiven Validität erweist sich der Gesichtspunkt, dass durch die Durchführung derartiger Interventionen das Misshandlungsrisiko verändert wird und damit auch Maße prädiktiver Validität sinken.

Eine empirische Überprüfung der prädiktiven Validität der Misshandlungs-Skala wird von Milner, Gold, Ayoub und Jacewitz (1984) dargestellt. Sie berichten von ei-

ner Längsschnittstudie mit 200 Eltern, die im Rahmen eines „at-risk“ Programms den CAPI-Fragebogen ausgefüllt hatten. Ziel dieser Untersuchung war es festzustellen, ob höhere Werte in der Abuse-Skala systematisch mit zukünftiger physischer Kindesmisshandlung, Kindesvernachlässigung bzw. mit Gedeihstörungen des Kindes ohne somatische Ursachen zusammenhängen. Nach der Bearbeitung des Fragebogens wurden die Probanden durch die Mitarbeiter des Projektes sechs Monate lang therapeutisch begleitet. Weiterhin fassten die Mitarbeiter Berichte ab, in denen aufgeführt wurde, ob der jeweilige Proband weitere Fälle von Kindesmisshandlung begeht. Zum Ende des Beobachtungszeitraumes wurde bei 42 Probanden die erneute Misshandlung als bewiesen angesehen. In 11 dieser Fälle lag physische Kindesmisshandlung vor, in 15 dieser Fälle Formen von Vernachlässigung und in 16 Fällen Gedeihstörungen ohne somatische Ursache. Für diese 42 Fälle wurden nun Zusammenhangsmaße zwischen Werten in der Misshandlungs-Skala und Form der Misshandlung berechnet. Für Formen der physischen Kindesmisshandlung ergab sich Cramers V zu 0,34 mit einer Irrtumswahrscheinlichkeit kleiner 1% und für Kindesvernachlässigung ein Cramers V von 0,19 mit einer Irrtumswahrscheinlichkeit kleiner 5%. Der Zusammenhang zwischen Werten in der Misshandlungs-Skala und Gedeihstörungen ohne somatische Ursache konnte nicht signifikant größer als Null geschätzt werden.

Zusammenfassend unterstützt diese Studie die Annahme der prädiktiven Validität der Misshandlungs-Skala bei der Vorhersage zukünftiger Kindesmisshandlung. Am wichtigsten erscheint das Ergebnis, dass anscheinend eine positive Beziehung zwischen den Werten in der Abuse-Skala und späterer physischer Kindesmisshandlung besteht. Da das CAPI eben zur Vorhersage dieser Form von Kindesmisshandlung entwickelt wurde, entspricht es ebenfalls den Erwartungen, dass physische Kindesmisshandlung besser vorhergesagt werden kann als Kindesvernachlässigung sowie Gedeihstörungen.

3. Validity Scales

Bei der Konstruktion der Abuse Scale zeigte sich, dass Fehlklassifikationen hauptsächlich im Bereich der sogenannten „False Negatives“ auftraten. Hierbei liegt eine falsch negative Klassifikation vor, wenn es sich bei einem Probanden eigentlich um einen misshandelnden Elternteil handelt, der dennoch aber als nicht-misshandelnder Elternteil klassifiziert wird. Von den Autoren wurde daher angenommen, dass misshandelnde Probanden in der Lage sind, die Items der Abuse Scale so zu beantworten, dass sie trotz vorhandenem Misshandlungsrisiko nicht der Gruppe der Misshandler zugerechnet werden. Daher war es wünschenswert, zusätzlich zu der Abuse Scale ebenfalls Skalen zu konstruieren, mithilfe derer das Vorhandensein verfälschender Antworttendenzen identifiziert werden kann.

Im Rahmen der Konstruktion des CAPI wurden daher Skalen zur Erfassung von drei spezifischen Verfälschungstendenzen konstruiert. Mithilfe der Lie Scale soll positiv verfälschendes Antwortverhalten im Sinne sozialer Erwünschtheit erfasst werden. Die Random Response Scale befasst sich mit der Feststellung von zufälligem Antwortverhalten. Die Inconsistency Scale wurde konstruiert, um das Vorhandensein widersprüchlichen Antwortverhaltens von Probanden anzuzeigen. Die folgenden Abschnitte befassen sich mit der Konstruktion dieser drei Validity Scales, wobei neben der Beschreibung der testtheoretischen Güte der Skalen ebenfalls strategische Überlegungen, die der Entwicklung der jeweiligen Skalen zugrunde liegen, besprochen werden sollen.

3.1 Entwicklung der Lie Scale

Die Strategie zur Konstruktion der Lie Scale orientierte sich an der üblichen Vorgehensweise zur Erstellung von Skalen, mithilfe derer positiv verfälschendes Antwortverhalten im Sinne sozialer Erwünschtheit festgestellt werden soll (Seiwald, 2003). Zunächst sollten Items gefunden werden, die von ihrem Inhalt her in besonderer Weise sozial erwünscht, aber nur bei den wenigsten Probanden vorzufinden sind (z.B. CAPI 160 : „Ich sage immer die Wahrheit“) bzw. Items, die von ihrer Konnotation her negativ sind, aber bei den meisten Probanden vorliegen (z.B. CAPI 57

„Manchmal habe ich schlechte Gedanken“). Zentrale Annahme hierbei war, dass Probanden, die sich in einer besonders erwünschten Art und Weise darstellen wollen, bei derartigen Items häufiger die positiv konnotierte Antwortalternative wählen, während ehrlich antwortende Probanden eher die Antwortalternative mit der negativen Konnotation wählen. Bei der Überprüfung der Eignung der Items wurde ein sog. „85/15-Split“ vorausgesetzt, was bedeutet, dass sich mindestens 85% der Probanden für die sozial unerwünschte Antwortalternative und maximal 15% der Probanden für die sozial erwünschte Antwortalternative entscheiden sollten.

Die Vorgehensweise zur Konstruktion der Skala wird bei Milner (1986) beschrieben. Zunächst wurde ein Pool von 47 Items generiert, die in der Lage sein sollten, soziale Erwünschtheit in der eben besprochenen Form zu erfassen. Diese Items wurden einer studentischen Stichprobe von 111 Probanden ausgeteilt, die die Items ohne weitere Hinweise beantworten sollten. Bei der Auszählung der Häufigkeiten der Verwendung der Antwortalternativen zeigte sich, dass 30 Items den geforderten „85/15-Split“ erreichten. In einer zweiten Studie wurden die 30 Items aus der ersten Studie zusammen mit 70 zufällig ausgewählten Items der Abuse Scale einer Stichprobe von 83 misshandelnden Elternteilen bearbeitet. Hierbei zeigte sich, dass 18 Items weiterhin die geforderte Trennung von 15% zu 85% erreichten. Zur Bestimmung eines sinnvollen Cut-offs, ab dem einem Probanden verfälschendes Antwortverhalten unterstellt werden kann, wurden einer Stichprobe von 132 misshandelnden Elternteilen ein Fragebogen, der aus den 18 Items, die in der zweiten Studie als geeignet befunden wurden, sowie 142 zufällig aus dem Konstruktionspool der Abuse Scale ausgewählten Items vorgegeben. Der Mittelwert von Fragen, die im Sinne sozialer Erwünschtheit beantwortet wurden, lag in dieser Stichprobe bei 2,8. Bei der Betrachtung der kumulativen Häufigkeiten der Verteilung zeigte sich, dass 95% der Probanden maximal sechs Antworten in sozial erwünschter Art und Weise ausgefüllt haben. Daher wurde der Cut-off der Lügenskala bei einem Wert von sieben Punkten festgelegt.

Befunde zur Reliabilität der Lie Scale finden sich im technischen Manual des CAPI (Milner, 1986), wobei diese Angaben ebenfalls aus der Studie stammen, die bereits

im Zusammenhang mit der Reliabilität der Abuse Scale besprochen wurde (Kapitel C, Abschnitt 2.2). Angaben zur internen Konsistenz der Lie Scale werden in Form von KR20-Werten sowie in Form von SH-Werten gemacht und aufgeschlüsselt nach verschiedenen Eigenschaften der Probanden (Misshandler vs. Kontrollgruppe, Geschlecht, Alter, Bildungsstand, ethnische Abstammung) dargestellt. Hinsichtlich der berichteten KR20-Werte kann festgestellt werden, dass sie in einem Bereich von .60 bis .88 liegen, während sich die entsprechenden SH-Werte zwischen .60 und .92 bewegen. Insgesamt können diese Werte als passabel bis sehr gut beurteilt werden. Die zeitliche Stabilität der Lie Scale wird in Form von Produkt-Moment-Korrelationen hinsichtlich verschiedener Test-Retest-Abstände angegeben. Für einen eintägigen Abstand betragen die Zusammenhänge .81, für einen einwöchigen .84, für einen einmonatigen .82 und für einen dreimonatigen .65. Ebenfalls wurde der Unterschied der Mittelwerte der Probandengruppen zwischen den beiden Messzeitpunkten auf Signifikanz getestet, wobei der Unterschied in keinem der Vergleiche mit einer hinreichend kleinen Irrtumswahrscheinlichkeit angenommen werden konnte. Aufgrund der guten Test-Retest-Reliabilitäten sowie der ausbleibenden Mittelwertsunterschiede zwischen den Messzeitpunkten kann daher auch die zeitliche Stabilität der Lie Scale als gesichert betrachtet werden.

3.2 Entwicklung der Random Response Scale

Die Random Response Scale wurde in das CAPI zu Erfassung wahllosen Antwortverhaltens, wie es beispielsweise bei akuten Psychosen, Drogenintoxikation oder Unlust der Probanden entstehen kann, aufgenommen. Die Strategie, die zur Konstruktion dieser Skala gewählt wurde, ist der der Lie Scale sehr ähnlich ist. Es sollten ebenfalls Items gefunden werden, bei denen wahrheitsgemäß und reflektiert antwortende Probanden stark zu einer Antwortalternative tendieren. Im Gegensatz zu den Items der Lie Scale sollten diese Items jedoch keine Tendenzen sozialer Erwünschtheit hervorrufen, sondern von ihrem Inhalt her eher wertneutral sein (z.B. CAPI 61: „Kinder sollten nicht Schwimmen lernen“; CAPI 116: „Die meisten Kinder sind gleich“). Hierbei wurde davon ausgegangen, dass Probanden, die in einer reflektierten Art und Weise antworten, eher die Antwortalternativen wählen, die die

meisten anderen Probanden auch wählen, während zufällig antwortende Probanden die unwahrscheinlichen Antwortalternativen in einem weitaus höheren Maße wählen. Bei der Konstruktion sollte weiterhin sichergestellt werden, dass die Items weitestgehend unabhängig von den Werten in der Abuse Scale sind, um eine mögliche Konfundierung der Werte in dieser Skala mit dem tatsächlichen Misshandlungsrisiko zu vermeiden.

Zur Identifikation von Items, die für die Aufnahme in eine derartige Skala geeignet sind, wurde zunächst ein Fragebogen konstruiert, der aus den 77 Items der Abuse Scale, aus den 18 Items der Lie Scale sowie aus 65 Konstruktionsitems bestand. Dieser Bogen wurde einer Stichprobe von 234 nicht-misshandelnden Elternteilen ausgegeben. In den Ergebnissen dieser Erhebung zeigte sich bei neun Items eine deutliche Häufung von Items auf der zustimmenden Antwortalternative, bei weiteren neun Items eine deutliche Häufung der Antwortalternativen auf der ablehnenden Antwortalternative. Für die 18 gefundenen Items zeigten sich lediglich geringe Korrelationen mit der Abuse Scale, wobei der maximale diesbezügliche Zusammenhang sich auf einen Wert von .158 belief. Im Mittel ließen sich bei den Probanden der Stichprobe 2,42 Antworten in der unwahrscheinlichen Antwortalternative finden. Bei einer rein zufälligen Beantwortung der Skala wäre ein Mittelwert von 9 mit einer Standardabweichung von 2 zu erwarten gewesen. Basierend auf dem Vergleich dieser beiden Verteilungen wurde ein Cut-off von 5 ausgewählt, ab dem auf ein zufälliges Antwortverhalten geschlossen werden kann.

Vor dem Hintergrund einer spezifischen Besonderheit der Random Response Scale muss vor der Besprechung der diesbezüglichen Reliabilitätskennwerte einleitend erwähnt werden, dass die Befunde zur internen Konsistenz der Skala allenfalls vorsichtig interpretiert werden können. Da bei einer aufmerksamen Beantwortung der Skala sowohl zutreffende als auch nicht zutreffende Antwortalternativen gewählt werden sollten, kann das klassische Konzept der internen Konsistenz, das eine gleichförmige Beantwortung der Items einer Skala vorsieht, an dieser Stelle nicht in der üblichen Art und Weise interpretiert werden. Für die im Folgenden dargestellten Befunde zur internen Konsistenz der Random Response Scale muss vielmehr berücksich-

tigt werden, dass niedrige bzw. negative diesbezügliche Kennwerte auf eine aufmerksame und wahrheitsgemäße Beantwortung hindeuten, während hohe positive Kennwerte auf ein eher zufälliges Antwortverhalten schließen lassen. Befunde zur zeitlichen Stabilität der Random Response Scale sind ohne die eben dargestellten Vorbehalte interpretierbar, da in diesen Fällen lediglich Skalensummenwerte zu den beiden Messzeitpunkten miteinander korreliert werden.

Angaben zur Reliabilität der Random Response Scale stammen aus der bereits im Zusammenhang mit der Reliabilität der Abuse Scale beschriebenen Studie (Kapitel C, Abschnitt 2.2) und werden im Rahmen des technischen Manuals des CAPI (Milner, 1986) besprochen. Ergebnisse zur internen Konsistenz werden in Form von KR20-Werten sowie in Form von SH-Werten angegeben und nach verschiedenen Eigenschaften der Probanden (Misshandler vs. Kontrollgruppe, Geschlecht, Alter, Bildungsstand, ethnische Abstammung) aufgeschlüsselt. Die KR20-Werte für die Random Response Scale bewegen sich dabei in einem Bereich zwischen $-.03$ und $.53$ während diesbezügliche SH-Werte zwischen $-.83$ und $.77$ rangieren. Es kann festgestellt werden, dass gemäß der eben geschilderten Annahmen lediglich geringe bis moderate interne Konsistenzen zeigen, was jedoch auf ein wahrheitsgemäßes und reflektiertes Antwortverhalten der meisten Probanden schließen lässt. Von Probandengruppen, in denen sich die höheren internen Konsistenzen zeigen, kann angenommen werden, dass Probanden hier häufiger zufälliges Antwortverhalten zeigen. Beispielweise zeigt sich in der Gruppe stationär aufgenommener misshandlunggefährdeter Probanden (At-Risk Hospital) ein SH-Wert von $.77$ und ein KR20-Wert von $.47$. Die besser interpretierbaren Maße der zeitlichen Stabilität werden in Form von Produkt-Moment-Korrelationen zwischen Test- und Retest-Zeitpunkten angegeben. Für den Zeitraum von einem Tag betragen sie $.84$ für den einer Woche $.74$, für den eines Monats $.54$ und für den von drei Monaten $.55$. Mittelwertsunterschiede zwischen den beiden Messzeitpunkten wurden ebenfalls überprüft, konnten jedoch in keinem Fall mit einer hinreichenden Signifikanz angenommen werden. Die zeitliche Stabilität der Random Response Scale kann aufgrund der eben dargestellten Kennwerte als hinreichend gesichert betrachtet werden.

3.3 Entwicklung der Inconsistency Scale

Die Inconsistency Scale stellt die dritte Validitätsskala des CAPI dar und wurde zur Erfassung inkonsistenten und widersprüchlichen Antwortverhaltens aufgenommen. Die Strategie zur Konstruktion dieser Skala sah es vor, Itempaare zu generieren, die konsistente Inhalte (z.B. CAPI 78: „Andere Leute verstehen nicht, wie ich mich fühle“ und CAPI 98: „Die Leute verstehen mich nicht“) oder inkonsistente Inhalte (z.B. CAPI 75: „Mein Leben ist glücklich“ und CAPI 118: „Ich bin häufig niedergeschlagen“) erfassen. Grundlegende Idee hierbei war es, dass Probanden, die wahrheitsgemäß und reflektiert antworten, im Falle konsistenter Itempaare gleiche und im Falle inkonsistenter Itempaare die entgegengesetzten Antwortalternativen wählen. Als Kriterium für die Eignung der Items wurde ein Anteil von 80% der Probanden festgelegt, die das Itempaar in einer konsistenten Art und Weise beantworten.

Im Vorfeld der Konstruktion der Inconsistency Scale wurden Itempaare gesucht, die inhaltlich ähnliche bzw. widersprüchliche Sachverhalte abfragen. Im Gegensatz zu den bisher beschriebenen Skalen wurden hierbei keine neuen Items generiert, sondern auf die bereits bestehenden 160 Fragen des CAPI zurückgegriffen. Weiterhin wurde keine Rücksicht auf die Skalenzugehörigkeit der Items genommen, sondern vielmehr Items aus allen bereits bestehenden Skalen (Abuse, Lie und Random Response Scale) sowie den Forschungsskalen benutzt.

Nach der Identifikation geeigneter Itempaare wurde der CAPI-Fragebogen einer Stichprobe von 100 nicht-misshandelnden Elternteilen ausgegeben. In dem so erhaltenen Datensatz wurde die Anzahl konsistent beantworteter Itempaare ausgezählt. Auf Basis dieser Auszählung konnten 20 Itempaare bestimmt werden, die mindestens von einem Anteil von 80% der Probanden in konsistenter Art beantwortet wurden. Zehn dieser Itempaare fragten ähnliche, zehn weitere widersprüchliche Sachverhalte ab. Die mittlere Anzahl inkonsistent beantworteter Itempaare lag bei 2,2 mit einer Standardabweichung von 0,5. Hätten die Probanden rein zufällig geantwortet, so wäre eine Mittelwert von 10 mit einer Standardabweichung von 2 zu er-

warten gewesen. Basierend auf diesen beiden Verteilungen wurde der Cut-off der Inconsistency Scale bei 6 gewählt.

Die Ergebnisse dieser Studie wurden in einer zweiten Studie auf Stichprobenunabhängigkeit hin überprüft. Hierbei wurde der CAPI-Fragebogen von einer Stichprobe von 133 nicht-misshandelnden Elternteilen ausgefüllt. Bei der Auszählung der inkonsistent bearbeiteten Itempaare zeigte sich ein durchschnittlicher Wert von 2,7 mit einer Standardabweichung von 1,9. Aufgrund der ähnlichen Verteilung von Fehlern auf der Inconsistency Scale kann daher von einer weitestgehenden Stichprobenunabhängigkeit der Validität der Inconsistency Scale ausgegangen werden.

Befunde zur Reliabilität der Inconsistency Scale stammen aus der Studie, die bereits im Zusammenhang mit der Reliabilität der Abuse Scale besprochen wurde (Kapitel C, Abschnitt 2.2) und werden im technischen Manual des CAPI (Milner, 1986) besprochen. Bei der Interpretation der im Folgenden dargestellten Kennwerte müssen die bereits im Zusammenhang mit der Reliabilität der Random Response Scale vorgebrachten Vorbehalte (Kapitel C, Abschnitt 3.2) gleichermaßen berücksichtigt werden. Es werden sowohl Angaben zur internen Konsistenz als auch zur zeitlichen Stabilität der Skala dargestellt. Interne Konsistenzen werden in Form von KR20-Werten sowie in Form von SH-Werten angegeben und nach verschiedenen Eigenschaften der Probanden (Misshandler vs. Kontrollgruppe, Geschlecht, Alter, Bildungsstand, ethnische Abstammung) aufgeschlüsselt dargestellt. Die KR20-Werte rangieren hierbei zwischen .00 und .57 während die SH-Werte im Bereich zwischen .07 und .75 liegen. Die geringen bis moderaten internen Konsistenzen deuten hierbei darauf hin, dass die Fragen der Inconsistency Scale in einer richtigen, d.h. bei konsistenten Itempaaren in gleicher und bei inkonsistenten Paaren in entgegengesetzter Weise beantwortet wurden. Die besser interpretierbaren Werte zur zeitlichen Stabilität der Inconsistency Scale werden in Form von Produkt-Moment-Korrelationen angegeben. Für einen Zeitraum zwischen Test und Retest von einem Tag betragen sie .84, für einen Zeitraum von einer Woche .74, für einen Zeitraum von einem Monat .54 und für einen Zeitraum von drei Monaten .55. Mittelwertsunterschiede der Werte in der Inconsistency Scale zwischen den beiden Messzeitpunkten wurden auf

Signifikanz geprüft, die jedoch in keinem Fall festgestellt werden konnte. Aufgrund der Zusammenhänge zwischen den beiden Messzeitpunkten, sowie den ausbleibenden Mittelwertsunterschieden kann die zeitliche Stabilität der Inconsistency Scale daher als gesichert betrachtet werden.

4. Abschließende Bewertung

In den vorherigen Abschnitten wurde die Vorgehensweise bei der Konstruktion des CAPI sowie diesbezügliche Befunde beschrieben. An dieser Stelle soll abschließend eine integrative Zusammenfassung und Bewertung dieser Befundlage erfolgen.

Zunächst kann der Abuse Scale sowohl im Hinblick auf interne Konsistenz, als auch im Hinblick auf zeitliche Stabilität eine gute bis sehr gute Reliabilität unterstellt werden (Kapitel C, Abschnitt 2.2). Befunde aus den Konstruktionsstudien der Abuse Scale zeigen, dass Gruppen von misshandelnden Eltern von Gruppen von nicht-misshandelnden Eltern mithilfe dieser Skala gut unterschieden werden können (Kapitel C, Abschnitt 2.1). Diese Diskriminationsfähigkeit besteht auch dann, wenn weniger heterogene Gruppen, z.B. stärker misshandelnde Eltern und Eltern, bei denen sich nur leichte Formen der Misshandlung finden lassen, unterschieden werden sollen (Kapitel C, Abschnitt 2.3.3.1). Weiterhin erlaubt die Kenntnis der Werte in der Abuse Scale den Schluss auf die zukünftige Auftretenswahrscheinlichkeit von Kindesmisshandlungen (Kapitel C, Abschnitt 2.3.3.2). Inhaltlich wurden die Items des Verfahrens konsistent zu bestehenden Theorien und Befunden zur Ätiologie von Kindesmisshandlung ausgewählt, wobei sich diese inhaltliche Passung auch für den aktuellen Stand der Forschung zu dieser Thematik feststellen lässt (Kapitel C, Abschnitt 2.3.1).

Während der Entwicklung der Abuse Scale zeigte sich, dass das Verfahren nicht resistent gegen mögliche Verfälschungstendenzen ist. Dieser Problematik kann beim Einsatz des Verfahrens jedoch durch die Betrachtung der Validity Scales begegnet werden. Im Rahmen der Validierungsuntersuchungen wurden Skalen zur Identifikation von drei spezifischen Verfälschungstendenzen entwickelt. Im Einzelnen können Tendenzen sozialer Erwünschtheit mithilfe der Lie Scale (Kapitel C, Abschnitt 3.1),

zufälliges und unreflektiertes Antwortverhalten mithilfe der Random Response Scale (Kapitel C, Abschnitt 3.2) sowie inkonsistentes Antwortverhalten mithilfe der Inconsistency Scale (Kapitel C, Abschnitt 3.3) erfasst werden. Befunde aus den Konstruktionsstudien deuten hierbei auf eine passable interne Konsistenz und eine hohe zeitliche Stabilität von Werten in den Skalen hin, weshalb die diesbezügliche Reliabilität als gegeben betrachtet werden kann. Die Validität der Skalen kann aufgrund der gewählten Konstruktionsstrategien sowie aufgrund der Befunde, dass bei Berücksichtigung der Werte in den Validity Scales eine bessere Zuordnung von Probanden möglich ist, ebenfalls als gesichert betrachtet werden.

Abschließend lässt sich daher festhalten, dass das CAPI ein ökonomisches Fragebogenverfahren darstellt, das zu einer hinreichend genauen Bestimmung des Kindesmisshandlungsrisikos eingesetzt werden kann, wobei möglicherweise zu erwartende Verfälschungstendenzen ebenfalls berücksichtigt werden.

D. Validierung des Elternbelastungsscreenings zur Kindeswohlgefährdung

In diesem Kapitel wird die Validierung der deutschsprachigen Version des CAPI, dem Elternbelastungsscreening zur Kindeswohlgefährdung (EBSK; Deegener, Spangler, Körner & Becker, im Druck) beschrieben. In einem ersten Abschnitt wird dargestellt, wie die Items der amerikanischen Originalversion ins Deutsche übersetzt wurden und inwiefern die inhaltliche Passung der beiden Versionen als gegeben betrachtet werden kann. Ein zweiter Abschnitt stellt sowohl die Akquise, als auch spezifische Charakteristika der Stichproben dar, die im Rahmen der deutschen Validierungsuntersuchung betrachtet wurden. Die beiden darauf folgenden Abschnitte befassen sich mit den eigentlichen Ergebnissen der Validierungsstudie. Diese werden dabei aufgetrennt nach der Belastungsskala, die die primäre klinische Skala des EBSK darstellt und den Validitätsskalen, mithilfe derer Verfälschungstendenzen im Antwortverhalten der Probanden erfasst werden sollen, besprochen. In einem letzten Abschnitt erfolgt eine zusammenfassende Darstellung der Ergebnisse der Validierung des EBSK.

1. Übersetzung der englischsprachigen Originalversion

Als Vorbereitung zur Validierung des EBSK wurden die Items der englischsprachigen Originalversion zunächst ins Deutsche übersetzt. Um die inhaltliche Nähe zu der Originalversion sicherzustellen, sollte die Bedeutung der deutschen Items der der englischen so gut wie möglich entsprechen. Die Strategie, die zur Überprüfung dieser inhaltlichen Nähe gewählt wurde, beruht auf folgender Grundidee: Sofern Items besonders gut, d.h. mit einer hohen inhaltlichen Nähe übersetzt werden können, sollten sie von unabhängigen Übersetzern weitestgehend identisch übersetzt werden. Können die Items bspw. aufgrund von semantischer Inkompatibilität zwischen der deutschen und der englischen Sprache nur schlecht übersetzt werden, so sollten sich die unabhängigen Übersetzungen weniger stark ähneln.

Für die 160 Items des CAPI lagen zwei unabhängige Übersetzungen vor, eine von Deegener (o. J.) sowie eine von Globisch und Spangler (o. J.). Beide Übersetzungen wurden vier in der psychologischen Diagnostik tätigen unabhängigen Experten vor-

gelegt. Diese wurden gebeten, die Ähnlichkeit der beiden deutschsprachigen Versionen der einzelnen Items einzuschätzen. Diese Einschätzung erfolgte hierbei auf einer vierstufigen Antwortskala, wobei die 1 bei Items gewählt werden sollte, die in beiden Übersetzungen den identischen Wortlaut haben, eine 2 bei Items, bei denen sich minimale Unterschiede zwischen den beiden Übersetzungen zeigen, eine 3 bei Items, die in unterschiedlicher Weise übersetzt wurden, aber dennoch den gleichen Sinn haben und eine 4, falls die beiden Übersetzungen eines Items einen unterschiedlichen Sinn haben.

Vor der eigentlichen Auswertung der Experteneinschätzung wurde zunächst die Konsistenz der Bewertungen zwischen den Experten (im Sinne von Interrater-Reliabilität) überprüft. Hierzu wurde der Übereinstimmungskoeffizient (Ü; Fricke, 1974) über Items und Experten hinweg bestimmt, der im Falle einer identischen Einschätzung aller Experten einen Wert von 1 und im Falle einer vollkommenen Unabhängigkeit ihrer Einschätzungen einen Wert von 0 annehmen würde. Bezogen auf die von den Experten abgegebenen Ähnlichkeitseinschätzungen der beiden Übersetzungen ergab sich ein Ü von 0,96, was auf eine sehr hohe Übereinstimmung der Experten bei der Einschätzung der Ähnlichkeit der Items hindeutet. Die Interrater-Reliabilität der Experten kann damit als sehr gut bezeichnet werden.

Tabelle D.1: Mittlere Ähnlichkeitseinschätzungen der beiden Übersetzungen

Bereich	160 Items	63 Items
1 - 1,49 (identisch)	36,87% (59)	38,1% (24)
1,5 - 2,49 (ähnlich)	53,13% (85)	52,38% (33)
2,5 - 3,49 (sinngleich)	8,75% (14)	7,94% (5)
3,5 - 4 (sinnentstellt)	1,25% (2)	1,58% (1)

Anmerkung: In der Spalte „160 Items“ werden die Ergebnisse für alle Items der Originalversion, in der Spalte „63 Items“ die Ergebnisse für die Items der deutschen Version dargestellt. Die Zahlen in den Zellen stellen die relativen Häufigkeiten über die Spalten hinweg dar. In den Klammern dahinter werden die absoluten Häufigkeiten angegeben.

Für die folgende Besprechung wurden die Einschätzungen der Experten über die einzelnen Items hinweg gemittelt und entsprechend der Skala, auf der die Einschätzungen abgegeben wurden, klassifiziert. Items mit einem Mittelwert von 1 bis 1,49

können daher als identische Übersetzungen, Items mit einem Mittelwert von 1,5 bis 2,49 als weitgehend identische Übersetzungen, Items mit einem Mittelwert von 1,5 bis 2,49 als vom Sinn her identische Übersetzungen und Items mit einem Mittelwert von 3,5 bis 4 als sinnentstellte Übersetzungen verstanden werden. Die so klassifizierten Ergebnisse der Einschätzungen der Experten werden in Tabelle D.1 sowohl für alle 160 Items des CAPI, als auch für die 63 Items, die in die deutschsprachige Version aufgenommen wurden, dargestellt. Es kann festgestellt werden, dass sowohl in der Spalte, in der die Ergebnisse für alle 160 Items, als auch in der Spalte in der die Ergebnisse für die 63 Items der deutschsprachigen Version dargestellt werden, jeweils ca. 90% der beiden deutschen Übersetzungen der Items als identisch oder lediglich minimal verschieden eingeschätzt wurden. Als vom Sinn her ähnlich wurden 8,75% der Items des originalen Itemsatzes und 7,94% der Items des Itemsatzes der deutschsprachigen Version eingeschätzt. Als sinnentstellt eingeschätzt wurden im Mittel lediglich 1,25% der Items des CAPI und 1,58% der Items der deutschsprachigen Version. Aufgrund dieser deutlich im Bereich der identisch bis annähernd identisch liegenden Einschätzungen, kann von einer hohen inhaltlichen Nähe zum Originalverfahren ausgegangen werden.

Für den weiteren Verlauf der Validierungsuntersuchung wurde die im Anhang (Kapitel G, Abschnitt 2) abgedruckte Übersetzung von Globisch und Spangler (o. J.) eingesetzt. Die Verwendung dieser Version geschah dabei weniger aus inhaltlichen Überlegungen, als aus dem praktischen Grunde heraus, dass bereits eine Vielzahl von Probanden mithilfe dieser Version erfasst wurde.

2. Akquise und Zusammenfassung der Validierungsstichprobe

Für die Validierung der deutschsprachigen Version des CAPI konnten in acht einzelnen Stichproben insgesamt 1581 Probanden akquiriert werden. In den folgenden Abschnitten soll zunächst beschrieben werden, in welcher Art und Weise die Akquise der Probanden geleistet wurde und weiterhin spezifische demographische Charakteristika (Geschlechterverhältnis, Alter) der jeweiligen Stichproben dargestellt werden. In einem letzten Abschnitt soll abschließend die Integration der einzelnen

Datensätze in Gruppen unterschiedlicher Belastungsgrade besprochen werden, die für die eigentliche Validierung des Fragebogens notwendig war.

2.1 Stichprobe von Eltern von Kindergarten- und Grundschulkindern

Die größte Stichprobe zur Validierung der deutschsprachigen Version des CAPI wurde im Rahmen eines Forschungsprojekts an der Justus-Liebig-Universität Gießen und der Friedrich-Alexander-Universität Erlangen-Nürnberg erhoben. Im Zeitraum zwischen den Jahren 2000 und 2002 konnten hierbei Fragebögen von insgesamt 1004 Müttern und Vätern von Kindergarten- und Grundschulkindern erhoben werden. Das Alter der Probanden zeigt einen Mittelwert von 36,91 Jahre mit einer Standardabweichung von 5,79. Hinsichtlich des Geschlechterverhältnisses lässt sich ein Anteil von 74,7% weiblicher und 25,3% männlicher Probanden feststellen. Die Akquise der Eltern erfolgte über die Klassenleiter in den Grundschulen bzw. die Erziehungskräfte in den Kindergärten. Diese teilten neben zwei Exemplaren des Fragebogens ebenfalls auch einen zugehörigen Begleitbrief an die Eltern aus. In diesem wurde den Eltern mitgeteilt, dass es sich um einen Fragebogen zur Familien- und Lebenssituation handele, in dem familiäre Umstände der kindlichen Entwicklung, die Einstellungen von Eltern zu Kindern und Erziehung sowie insbesondere elterliche Belastungen abgefragt werden. Es wurde weiterhin darum gebeten, dass beide Elternteile den Fragebogen unabhängig voneinander ausfüllen. Die Fragebogen konnten in einem verschlossenen Umschlag wieder in den Kindergärten bzw. Schulen abgegeben werden.

2.2 Stichprobe von Eltern von Patienten der Kinder- und Jugendpsychiatrie

Diese Stichprobe wurde in der Klinik für Kinder- und Jugendpsychiatrie und Psychotherapie des Universitätsklinikums des Saarlandes in Homburg/Saar erhoben. Es handelt sich um 211 Eltern minderjähriger Patienten, die in dem Zeitraum zwischen Juli 2004 und Februar 2006 in die ambulante oder stationäre Therapie aufgenommen wurden. Die Probanden dieser Stichprobe wiesen ein mittleres Alter von 39,6 Jahren mit einer Standardabweichung von 5,86 auf. 62,6 % der Probanden waren weiblichen und 37,4% männlichen Geschlechts. Die Fragebögen wurden den Eltern

zusammen mit einem Begleitbrief im Rahmen der Anmeldung bzw. Anamnese ausgeteilt. In dem Begleitbrief wurde neben einer allgemeinen Darstellung des Fragebogens ebenfalls hervorgehoben, dass die Beantwortung des Fragebogens freiwillig und die Anonymität der Verwendung der Daten garantiert sei.

2.3 Stichprobe von MitarbeiterInnen und Studierenden einer Polizeiakademie

Im Rahmen einer Befragung von MitarbeiterInnen und Studierenden der Polizeiführungsakademie Münster im April 2006 konnten 31 weitere Fragebögen erhoben werden. Im Mittel waren die Probanden 40,65 Jahre alt, wobei die Altersverteilung eine Standardabweichung von 6,38 aufwies. Die Geschlechterverteilung in der Stichprobe zeigte einen Anteil von 25,8% weiblicher und 74,2% männlicher Probanden. Die Probanden wurden in einem Anschreiben informiert, dass es sich um eine Untersuchung zur Überprüfung der Eignung eines Screening-Fragebogens zur Erfassung des Risikos körperlicher Kindesmisshandlung im deutschen Sprachraum handele. Ebenfalls wurde betont, dass die Beantwortung des Fragebogens freiwillig sei und die Daten ausschließlich in anonymisierter Form verwendet werden.

2.4 Stichprobe von hochbelasteten und misshandelnden Eltern

Im Zeitraum von Juli 2004 bis April 2007 konnten Fragebögen von insgesamt 43 MisshandlerInnen bzw. hochbelasteten Elternteilen in Jugendhilfeeinrichtungen (Kinderschutz-Zentren, Familienhilfe-Zentren, therapeutischen Wohngruppen, ärztlichen Beratungsstellen gegen Vernachlässigung und Misshandlung, Kinderschutzdiensten, Mutter-Kind-Heimen) im Saarland sowie in Rheinland-Pfalz und Nordrhein-Westfalen erhoben werden. Die Probanden in der Stichprobe hatten ein mittleres Alter von 39,84 Jahren mit einer Standardabweichung von 8,94. Die Stichprobe bestand weiterhin zu 62,8% aus Frauen und zu 37,2% aus Männern.

2.5 Stichprobe von Begutachtungsfällen bei Verdacht auf Kindesmisshandlung

Im Rahmen einer anonymisierten Auswertung von Akten einer rechtspsychologischen Praxis in Süddeutschland konnten Angaben von insgesamt 47 Fragebögen aus den Jahren 2003 bis 2006 in den Validierungsdatensatz aufgenommen werden. Der Fragebogen wird in der Praxis seit mehreren Jahren im Rahmen der Begutachtung

des Vorhandenseins von Kindesmisshandlungsfällen eingesetzt. Das Alter der Probanden zeigt einen Mittelwert von 34,66 Jahre mit einer Standardabweichung von 9,6. Hinsichtlich des Geschlechterverhältnisses lässt sich ein Anteil von 36,2% weiblicher und 63,8% männlicher Probanden feststellen. Aus den Akten ließen sich weiterhin Angaben zu Fallmerkmalen der Begutachtungsfälle entnehmen, hinsichtlich derer sich die Fragebögen bezüglich der vorhandenen Wahrscheinlichkeit des Vorliegens eines Misshandlungsfalles klassifizieren ließen. Hierbei konnte festgestellt werden, dass in 14 Fällen ein eindeutiger Misshandlungsfall vorlag, in weiteren 14 Fällen als wahrscheinlich erschien, in 8 Fällen nicht ausgeschlossen werden konnte und in 11 Fällen wahrscheinlich nicht vorlag.

2.6 Stichproben von Studierenden der Universität des Saarlandes

Zur Validierung der Lügenskala wurden zwei Stichproben von Studierenden der Universität des Saarlandes erhoben, bei denen gezielt Antworttendenzen in eine sozial erwünschte Richtung induziert wurden. Eine erste Gruppe, die im Folgenden als „Fake-Good 1“ bezeichnet werden soll, bestand aus 96 Studierenden der Psychologie, wobei 59,4% der Probanden weiblichen und 40,6% männlichen Geschlechts waren. Es kann hierbei ein mittleres Alter von 26,13 Jahren mit einer Standardabweichung von 4,42 festgestellt werden. Ihnen wurde die Übersetzung der 160 Items der amerikanischen Originalversion mit der Anweisung ausgeteilt, sie sollen sich vorstellen, sie seien ein Elternteil, das wegen Problemen mit seinem Kind das Jugendamt aufsuche. Eine zweite Gruppe, die im Folgenden als „Fake-Good 2“ bezeichnet werden soll, bestand aus 83 Studierenden der Humanmedizin. Im Mittel wiesen diese ein Alter von 22,57 Jahren mit einer Standardabweichung von 3,09 auf. 63,9% der Probanden waren männlichen und 36,1% weiblichen Geschlechts. Ihnen wurde der gleiche Fragebogen mit der Anweisung ausgeteilt, sie sollen sich vorstellen, sie seien ein Elternteil, das sein Kind misshandelt hat, und ihnen würde beim Jugendamt der Fragebogen vorgelegt werden. Abschließend wurden beide Gruppen aufgefordert: „Füllen Sie den Fragebogen bitte so aus, wie sie meinen, einen besonders guten Eindruck zu hinterlassen“.

2.7 Studierende der Otto-Friedrich-Universität Bamberg

Zur Bestimmung von konvergenter und diskriminanter Validität der Werte in der Belastungsskala mit Werten in einem Persönlichkeitsinventar wurden die Übersetzung der 160 Items der amerikanischen Originalversion sowie das Freiburger Persönlichkeitsinventar in seiner revidierten Form (FPI-R; Fahrenberg, Hampel & Selg, 1994) an eine Stichprobe von 66 Studierenden der Otto-Friedrich-Universität Bamberg (vorwiegend Lehramtskandidaten) ausgeteilt. Hinsichtlich Alters- und Geschlechterverteilung liegen keine Daten für diese Stichprobe vor.

2.8 Zusammenfassung der einzelnen Stichproben zu Belastungsgruppen

Für den weiteren Verlauf der Validierungsuntersuchung wurden die einzelnen Stichproben in einen Gesamtdatensatz integriert. Hierbei wurden die einzelnen Fälle einer von drei Belastungsgruppen zugeordnet. In die Gruppe der Unbelasteten wurden alle Probanden aufgenommen, deren Misshandlungsrisiko als gering einzuschätzen ist. Im Einzelnen wurden hierbei die Stichprobe von Eltern von Kindergarten- und Grundschulkindern (N=1004), die Stichprobe von MitarbeiterInnen und Studierenden der Polizei-Führungsakademie Münster (N=31) sowie die Probanden der Stichprobe von Begutachtungsfällen, bei denen wahrscheinlich kein Misshandlungsfall vorlag (N=11) zusammengefasst. Insgesamt ergibt sich daher ein Umfang der Gruppe der Unbelasteten von 1046 Probanden.

Zu der Risikogruppe wurden die Daten aller Probanden zusammengefasst, denen ein erhöhtes Misshandlungsrisiko zu unterstellen ist. Hierbei wurde die Stichprobe von Eltern von Patienten der Kinder- und Jugendpsychiatrie (N=211) sowie die Probanden der Stichprobe von Begutachtungsfällen, bei denen das Vorhandensein eines Falls von Kindesmisshandlung wahrscheinlich erschien bzw. nicht ausgeschlossen werden konnte (N=22), zusammengefasst. Nach der Integration der Datensätze ergab sich ein Fallzahl von 233 Probanden in der Risikogruppe.

In die Gruppe der Misshandler und Hochbelasteten wurden zuletzt alle Probanden aufgenommen, bei denen ein gesicherter Misshandlungsfall vorlag. Im Einzelnen wurden hierbei die Stichprobe von hochbelasteten und misshandelnden Eltern

(N=43) sowie die Stichprobe der Begutachtungsfälle, in denen ein gesicherter Misshandlungsfall vorlag (N=14), zusammengefasst. Für die Gruppe der Misshandler und Hochbelasteten ergibt sich daher eine Fallzahl von 57 Probanden.

Der Datensatz, in dem die nach Belastungsgruppen aufgetrennten Personen enthalten sind, hat damit einen Umfang von insgesamt 1336 Probanden. Nicht in die Belastungsgruppen aufgenommen wurden die Stichproben von Studierenden der Universität des Saarlandes sowie die Stichprobe von Studierenden der Otto-Friedrich-Universität Bamberg. Vorwiegender Grund hierfür war, dass in diesen Stichproben nicht feststellbar war, ob es sich bei den Probanden um Eltern handelt.

3. Validierung der Belastungsskala

In den folgenden Abschnitten wird die Vorgehensweise, die zur Validierung der Belastungsskala gewählt wurde, dargestellt. Die einzelnen Unterabschnitte beziehen sich hierbei auf die Identifikation diskriminierender Items, die diskriminanzanalytische Validierung sowie die Lineartransformation der Belastungsskala. Weiterhin werden zusätzliche Befunde zur Validität der Belastungsskala dargestellt, die im Hinblick auf die faktorielle Struktur, die Beziehung zu demographischen Variablen sowie die konvergente und diskriminative Validität zu anderen psychometrischen Verfahren angenommen werden kann.

3.1 Identifikation diskriminierender Items

Ein erster Schritt der Validierungsuntersuchung stellt der Versuch der Identifikation von Items, die signifikant zwischen den drei Belastungsgruppen trennen können, dar. Dazu wurde für jedes der 77 Items der Abuse Scale von Milner (1986) eine Varianzanalyse über die drei Belastungsgruppen durchgeführt. Hierbei konnte festgestellt werden, dass sich lediglich bei 47 Items signifikante Mittelwertsunterschiede zwischen den drei Gruppen zeigten. Alle folgenden Betrachtungen zur Validierung der Belastungsskala beschäftigten sich lediglich mit diesen 47 Items.

3.2 Diskriminanzanalytische Validierung

Die Konstruktion einer Funktion zur Gruppenzuordnung erfolgte in gleicher Form wie bei Milner (1986). Hierbei wurde eine Diskriminanzanalyse mithilfe des Programmes SPSS durchgeführt (Backhaus, Erichson, Plinke & Weiber, 2006). Ziel war es, eine Funktion zu finden, mithilfe derer eine hinreichend genaue Prognose der Gruppenzugehörigkeit ausfüllender Probanden ermöglicht wird.

Aufgrund der geringen Stichprobengröße in der Gruppe der tatsächlichen Misshandler bezog sich diese Betrachtung zunächst lediglich auf den Unterschied zwischen der Gruppe der Unbelasteten und der Gruppe der Belasteten. Um eine Verzerrung der Schätzung zugunsten der größeren Gruppe zu verhindern, wurden in der Analyse gleich große Belastungsgruppen verwendet. Hierzu wurden zufallsmäßig 233 Probanden aus der Gruppe der Unbelasteten bestimmt, die in die eigentliche Analyse eingingen.

Tabelle D.2: Klassifikationsergebnisse der Diskriminanzanalyse

		Vorhergesagt	
		Unbelastet	Belastet
Original	Unbelastet	191 (82,0%)	42 (18,0%)
	Belastet	66 (28,3%)	167 (71,7%)

Anmerkung: In den Zellen abgetragen sind Fallzahlen, in den Klammern dahinter die prozentualen Anteile in der Zeile

Tabelle D.2 stellt die Klassifikationsergebnisse der in der Diskriminanzanalyse bestimmten Funktion dar. Auf der Basis der gefundenen Funktion konnten insgesamt 76,8% der Fälle korrekt klassifiziert werden. Die testtheoretischen Gütekriterien dieser Funktion ergaben sich zu einem Wilk's Lambda von 0,625 und einem kanonischen Korrelationskoeffizienten von 0,612.

Im Rahmen der Diskriminanzanalyse wurde auch eine Kreuzvalidierung durchgeführt, bei der jeder einzelne Fall anhand einer Diskriminanzfunktion klassifiziert wurde, die aus den Daten der übrigen Probanden berechnet wurde. In dieser Kreuzvalidierten Diskriminanzanalyse konnten insgesamt 70,2% der ursprünglichen Fälle

korrekt klassifiziert werden, was auf eine Stichprobenunabhängigkeit der gefundenen Funktion hindeutet.

Tabelle D.3: Klassifikationsergebnisse in der Validierungsstichprobe

		Vorhergesagt	
		Unbelastet	Belastet
Original	Unbelastet	793 (75,8%)	253 (24.2%)
	Belastet	66 (28,3%)	167 (71,7%)
	Misshandler	15 (26,3%)	42 (73,7%)

Anmerkung: In den Zellen abgetragen sind Fallzahlen, in den Klammern dahinter die prozentualen Anteile in der Zeile

Zur Überprüfung der Validität der gefundenen Diskriminanzfunktion in der Gesamtstichprobe wurden alle Fälle des Validierungsdatensatzes anhand der gefundenen Diskriminanzfunktion klassifiziert. Das Ergebnis dieser Betrachtung wird in Tabelle D.3 dargestellt. Wie gesehen werden kann, kann auch ein Großteil der Fälle, die nicht in die eigentliche Diskriminanzanalyse eingegangen sind, auf Basis der Diskriminanzfunktion korrekt klassifiziert werden. Der prozentuale Anteil der Probanden in der Gesamtstichprobe, die richtig klassifiziert wurden (wobei die Misshandler als korrekt klassifiziert verstanden werden, sofern sie der Gruppe der Belasteten zugeordnet wurden), ergibt sich zu 75%.

Die Wahl des Cut-Off-Wertes, ab dem ein bestimmter Proband als belastet einzustufen ist, ergibt sich aus dem Cut-Off der Diskriminanzfunktion. Zusätzlich wurden die Centroide der Gruppe der Unbelasteten und der Belasteten als weitere Cut-Offs innerhalb der Gruppen gewählt. Probanden mit Diskriminanzwerten unter dem Gruppencentroid der Unbelasteten werden als nicht belastet klassifiziert. Probanden mit Werten oberhalb des Gruppencentroides der Unbelasteten, aber unterhalb des Cut-Offs, werden als eher belastet eingestuft. Probanden oberhalb des Cut-Offs, aber unterhalb des Gruppencentroides der Belasteten, werden als stark belastet eingestuft. Schließlich werden Probanden, deren Testwerte oberhalb des Gruppencentroides der Belasteten liegen, als sehr stark belastet klassifiziert. Da die Belastungsskala einer Lineartransformation unterzogen wurde, um so eine einfachere

Auswertung und Interpretation des EBSK zu ermöglichen, sollen die genauen Cut-Off-Werte an dieser Stelle nicht besprochen und lediglich auf die folgende Darstellung (Kapitel D, Abschnitt 3.3) verwiesen werden.

3.3 Lineartransformation der Belastungsskala

Da eine Version der Belastungsskala mit einem absoluten Nullpunkt und möglichst ganzzahligen Cut-Off-Werten zwischen den Belastungsgruppen wünschenswert erschien, wurde die in der Diskriminanzanalyse ermittelte Funktion einer Lineartransformation unterzogen. Zunächst wurden die Gewichte aller Items sowie die Konstante mit einem festen Faktor multipliziert, um die gewünschte Ganzzahligkeit der Werte zu erreichen. Danach wurde die Konstante so adjustiert, dass der minimale Wert, der in der Skala erreicht werden kann, genau Null beträgt.

Tabelle D.4: Lineartransformierte Version der Belastungsskala

	keine	geringe	starke	sehr starke	
Min		C₁	C₂	C₃	Max
0		161	185	207	422

Anmerkung: Min = Absolutes Skalenminimum; C₁ = Gruppencentroid der Unbelasteten; C₂ = Cut-Off zwischen Unbelasteten und Belasteten; C₃ = Gruppencentroid der Belasteten; Max = Absolutes Skalenmaximum

Tabelle D.4 stellt die lineartransformierte Version der Belastungsskala dar, die auch bei der Auswertung und Interpretation der Fragebögen eingesetzt wird. Wie bereits beschrieben wurde, liegt das absolute Minimum, das sich bei der Auswertung der Skala ergeben kann, genau bei Null. Das absolute Maximum der Skala liegt bei 422. Der Cut-Off zwischen der Gruppe der Unbelasteten und der Belasteten liegt bei dieser Version bei 185, der Gruppencentroid der Unbelasteten bei 161, der der Belasteten bei 207.

3.4 Validität der Belastungsskala

3.4.1 Faktorielle Validität

Im Gegensatz zu der bei Milner (1986) angewendeten Hauptachsenanalysen wurde im Rahmen der Validierung des EBSK eine Faktorenanalyse für dichotome Daten

durchgeführt. Diese faktoranalytische Strategie erschien entsprechend der Überlegungen von Kubinger (2003) geeigneter, um die strukturelle Validität der reduzierten Belastungsskala zu überprüfen, da bei der Anwendung klassischer Faktoranalysen, die eigentlich für die Analyse intervallskalierte Daten entwickelt wurden, die Anzahl der bedeutsamen Faktoren in dichotomen Datensätzen zumeist überschätzt wird (Kubinger, 2003). Die Analyse wurde daher mittels des Programmes MixFactor (Häusler, 2003; Häusler & Kubinger, 2004) durchgeführt.

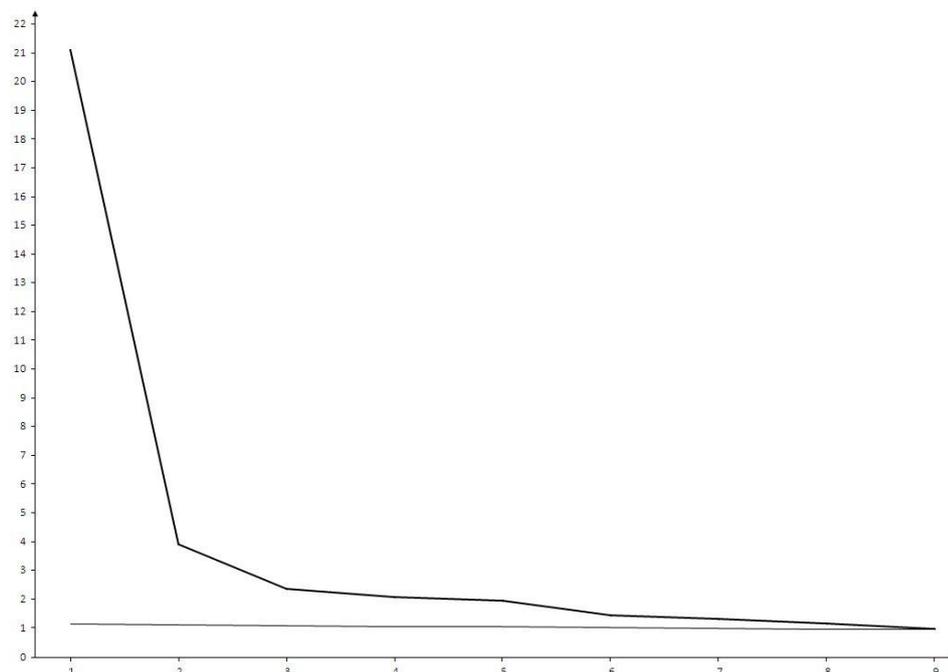


Abbildung D.1: Eigenwertverlauf der MixFactor-Analyse der Belastungsskala. Eigenwerte der Faktoren der reduzierten Belastungsskala werden durch die dicke Linie dargestellt. Die dünne stellt den Verlauf der mittleren Eigenwerte der Faktoren aus 100 MixFactor-Analysen in gleichgroßen Datensätzen dar. Auf der Ordinate sind die Eigenwerte abgetragen, auf der Abzisse die Faktoren.

Die Feststellung der Anzahl bedeutsamer Faktoren soll anhand der Darstellung in Abbildung D.1 erfolgen. Hier werden sowohl die Eigenwerte der ersten neun Faktoren, die bei der Analyse des gesamten Validierungsdatensatzes extrahiert wurden, als auch die Mittelwerte der Eigenwerte der ersten neun Faktoren, die in 100 Faktoranalysen gleichgroßer Zufallsdatensätze extrahiert werden konnten, abgetragen. Zunächst zeigt sich in der Abbildung ein deutlicher Eigenwertesprung zwischen dem ersten Faktor (mit einem Eigenwert von 21,12) und dem zweiten Faktor (mit einem Eigenwert von 3,93). Der „Knick“ des Eigenwerteverlaufes, der das zentrale Extrakti-

onskriterium des Scree-Tests (Cattell, 1966) darstellt, kann damit eindeutig nach dem ersten Faktor interpretiert werden. Bei einem Vergleich der Eigenwerte aus der Analyse des Validierungsdatensatzes und der Eigenwerte der Zufallsdatensätze zeigt sich, dass die des Validierungsdatensatzes erst nach dem achten Faktor überschritten werden. Eine Betrachtung des Eigenwerteverlaufs gemäß der Parallelanalyse von Horn (1965) deutet daher bei einer strengen Interpretation auf eine achtfaktorielle Struktur des Datensatzes hin. Dennoch kann festgestellt werden, dass sich der größte Sprung der Abstände zwischen dem Verlauf der Zufallseigenwerte und der Eigenwerte des Validierungsdatensatzes zwischen dem ersten und dem zweiten Faktor zeigen, was wiederum eine einfaktorielle Struktur des Datensatzes nahelegt. Zuletzt zeigt sich, dass die Eigenwerte des Validierungsdatensatzes den Eigenwert von 1, der gemäß dem Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser & Dickman, 1959) zur Beurteilung der Anzahl bedeutsamer Faktoren herangezogen wird erst bei dem neunten Faktor unterschritten wird. Eine strenge Interpretation hinsichtlich dieses Kriteriums würde daher auf eine achtfaktorielle Struktur hindeuten. Dennoch kann im Hinblick auf das Ausmaß, in dem die Eigenwerte über dem kritischen Wert von 1 liegen, festgestellt werden, dass der größte Sprung dieses Abstandes sich zwischen dem ersten und dem zweiten Faktor zeigt, was auf einen einzigen bedeutsamen Faktor hindeutet.

Da somit sowohl gemäß des Scree-Tests, als auch gemäß des Kaiser-Guttman-Kriteriums sowie der Parallelanalyse eine unidimensionale Struktur der reduzierten Belastungsskala angenommen werden kann, wurde diese als die Grundlage der weiteren Betrachtungen gewählt. Dieser einzige bedeutsame Faktor ist in der Lage, 44,9% der Varianz in den Items aufzuklären.

Tabelle D.5: Ladungsmatrix der Items der reduzierten Belastungsskala

EBSK-Item	Ladung	EBSK-Item	Ladung	EBSK-Item	Ladung	EBSK-Item	Ladung
1	0,59	15	0,71	30	0,25	46	0,56
2	0,79	16	-0,81	31	0,6	47	0,9
3	0,46	17	0,86	33	0,81	48	0,87
5	0,34	18	0,79	34	0,25	49	0,9
7	0,63	19	0,64	35	0,6	54	0,59
8	0,5	20	0,83	38	0,88	55	-0,25
9	-0,22	23	0,62	39	-0,8	56	0,75
10	0,75	24	0,82	41	0,24	57	0,37
11	-0,62	25	0,69	42	0,7	58	0,95
12	0,72	26	0,34	43	0,6	61	0,95
13	0,87	27	0,34	44	0,67	62	0,6
14	0,73	28	0,85	45	0,26		

Anmerkung: Die Nummerierung der Items entspricht der in der endgültigen Version des Fragebogens. EBSK-Item = Itemnummer im EBSK-Fragebogen; Ladung = Ladung auf dem ersten Faktor

Tabelle D.5 stellt die Ladungsmatrix der durchgeführten Analyse dar. Es kann festgestellt werden, dass alle Items der reduzierten Belastungsskala substantielle Ladungen auf dem ersten Faktor haben. Von ihrem Betrag her liegt die minimale Ladung bei -0,22 (Item 9). Das Maximum der Verteilung von Faktorladungen beträgt 0,95 (Item 61). Der Mittelwert der Verteilung von Faktorladungen liegt bei 0,64 mit einer Standardabweichung von 0,22. Damit kann festgestellt werden, dass der unidimensionalen Lösung auch von der betrachteten Ladungsmatrix her eine gute Passung zugemessen werden kann.

Zusätzlich zu der faktoranalytischen Betrachtung wurde eine Reliabilitätsanalyse zur Bestimmung der internen Konsistenz der Gesamtskala durchgeführt. Hierzu wurden zunächst die Items umgepolt, die negative Ladungen auf dem ersten Faktor hatten. Das Cronbach- α , das nach dieser Manipulation des Datensatzes ermittelt wurde, ergab sich zu einem sehr guten Wert von 0,907.

3.4.2 Befunde zu konvergenter und diskriminanter Validität der Belastungsskala

Im Rahmen der Validierung des EBSK fand zusätzlich eine Untersuchung der konvergenten und der diskriminanten Validität der Belastungsskala in Bezug auf ein etabliertes Persönlichkeitsinventar statt. Hierbei wurden die Daten der Studierenden der Otto-Friedrich-Universität Bamberg (Kapitel D, Abschnitt 2.7), die zusätzlich zu der Übersetzung der 160 Items der amerikanischen Originalversion ebenfalls die revidierte Form des Freiburger Persönlichkeitsinventars ausgefüllt hatten (FPI-R; Fahrenberg, Hampel & Selg, 1994) analysiert. Die im Folgenden dargestellten Betrachtungen beziehen sich lediglich auf die Zusammenhänge zwischen den Skalen des FPI-R und der reduzierten Form der Belastungsskala.

Tabelle D.6: Korrelationen zwischen Belastungsskala und Skalen des FPI-R

Skala	r	p
Lebenszufriedenheit	-0,323	0,008
Soziale Orientierung	0,023	0,853
Leistungsorientierung	-0,220	0,076
Gehemmtheit	0,418	0,000
Erregbarkeit	0,167	0,184
Aggressivität	-0,005	0,970
Beanspruchung	0,174	0,163
Körperliche Beschwerden	0,375	0,002
Gesundheitssorgen	0,059	0,636
Offenheit	0,201	0,106
Extraversion	-0,353	0,004
Emotionalität	0,309	0,012

Anmerkung: r = Produkt-Moment-Korrelation; p = zweiseitige Signifikanz

In Tabelle D.6 werden die Korrelationen zwischen den Werten in der Belastungsskala und den Werten in den Subskalen des FPI-R aufgeführt. Es kann zunächst festgestellt werden, dass fünf der zwölf FPI-Skalen signifikant mit der Belastungsskala korrelieren. Der Korrelation zwischen der Belastungsskala und der FPI-Skala Leistungsorientierung kann mit einer Irrtumswahrscheinlichkeit von 7,6% immerhin noch

marginale Signifikanz zugesprochen werden. Nichtsignifikante Korrelation lassen sich zwischen den übrigen sechs Skalen des FPI-R und der Belastungsskala finden.

Inhaltlich können alle signifikanten Korrelationen zwischen der Belastungsskala und den jeweiligen Subskalen des FPI-R stimmig interpretiert werden. So fügen sich negative Zusammenhänge zwischen Belastung und Lebenszufriedenheit, Leistungsorientierung sowie Extraversion und positive Zusammenhänge zwischen Belastung und Gehemmtheit, körperlichen Beschwerden und Emotionalität stimmig in die dargestellten Befunde und theoretischen Überlegungen zur Ätiologie von Kindesmisshandlung ein. Nichtsignifikante Korrelationen, deren Richtung auf Grund der Größe der Abweichung von Null noch sinnvoll interpretiert werden kann, erscheinen für die FPI-Skalen Erregbarkeit und Beanspruchung ebenfalls als stimmig. Auf den ersten Blick schwer interpretierbar ist der positive Zusammenhang zwischen Belastung und Offenheit. Er kann jedoch als ein verstärktes Mitteilungsbedürfnis im Sinne eines „Hilferufes“ belasteter Eltern verstanden werden. Die Unabhängigkeit der Werte in der Belastungsskala von den Werten der FPI-Skalen soziale Orientierung und Gesundheitsorgen sind nicht interpretierbar. Interessant ist zuletzt die Nullkorrelation zwischen Belastung und Aggressivität. Diese zeigt, dass hohe Belastung zwar zu Aggressivität und damit auch zu Tendenzen zur tatsächlichen Kindesmisshandlung führen kann, es aber nicht unweigerlich muss.

3.4.3 Konkurrente Validität mit demographischen Variablen

Aus dem Validierungsdatensatz heraus ergab sich eine zusätzliche Möglichkeit zur Überprüfung der Validität der konstruierten Belastungsskala. Zusätzlich zu den Antworten auf die Items der deutschen Übersetzung des Originalverfahrens liegen, neben den bereits vorher beschriebenen demographischen Variablen (Kapitel D, Abschnitt 2), für bestimmte Probanden ebenfalls Daten zu Familien- und Bildungsstand sowie zur Anzahl der Kinder vor. In diesem Validierungsschritt wurde überprüft, ob sich signifikante Unterschiede der Werte in der Belastungsskala zwischen Gruppen, denen aufgrund ihrer Ausprägung in diesen demographischen Variablen mehr oder weniger große Belastung unterstellt werden kann, feststellen lassen. Die Annah-

men, warum sich unterschiedliche Belastungsgrade in Abhängigkeit von diesen Variablen zeigen sollten, sowie die inferenzstatistische Überprüfung dieser Hypothesen werden in den folgenden Unterabschnitten beschrieben.

3.4.3.1 Familienstand

Hinsichtlich des Familienstandes der Probanden lässt sich annehmen, dass Probanden, die in einer festen Beziehung leben, weniger belastet sind als Probanden, die dies nicht tun. Dies kann unter anderem dadurch begründet werden, dass ein Partner, sofern dieser vorhanden ist, bestimmte Aufgaben der Kindeserziehung mit übernehmen kann und den Probanden dadurch entlastet. Steht dem Probanden kein Partner zur Seite, so konzentriert sich die Belastung, die sich durch die Kindeserziehung ergibt, auf den Probanden alleine. Zur Überprüfung dieser Annahme wurde eine einfaktorielle Varianzanalyse der Werte der Belastungsskala über die nach Familienstand gruppierten Probanden hinweg durchgeführt. Als Ergebnis lässt sich festhalten, dass sich die Mittelwerte in der Belastungsskala zwischen den einzelnen Gruppen hochsignifikant voneinander unterscheiden ($F(4,1144)=9,79$; $p<0,001$).

Tabelle D.7: Mittelwerte in der Belastungsskala gruppiert nach Familienstand

Familienstand	N	MW	SD
ledig	91	174,44	31,70
verheiratet	924	174,28	35,14
geschieden	93	195,77	37,55
getrennt	36	192,30	44,23
verwitwet	5	189,78	53,51

Anmerkung: N = Anzahl der Probanden; MW = Mittelwerte in der Belastungsskala; SD = Standardabweichung der Werte in der Belastungsskala

Tabelle D.7 stellt die Mittelwerte und Standardabweichungen der Belastungsskala in den einzelnen Gruppen dar. Es kann festgestellt werden, dass die verheirateten Probanden im Mittel deutlich geringere Werte in der Belastungsskala aufweisen als die geschiedenen, getrennten und verwitweten Probanden. Dieser Befund bestätigt die oben angestellte Annahme. Inkonform erscheint zunächst die Ähnlichkeit der Mittelwerte zwischen den verheirateten und den ledigen Probanden. Eine Erklärung

hierfür könnte sein, dass die Gruppe der ledigen Probanden nicht mehr in alleinstehende Probanden und Probanden in festen Beziehungen aufgeteilt wurde. Haben sich in dieser Gruppe viele Probanden, die in fester Beziehung leben, befunden, so könnte dies die Belastungswerte dieser Gruppe reduziert haben.

3.4.3.2 Anzahl der Kinder

Eine weitere Annahme bezieht sich auf die Anzahl der Kinder der Probanden. Es wird vermutet, dass die elterliche Belastung mit einer zunehmenden Anzahl von Kindern ansteigt. Die Überlegung hierbei ist, dass die Anforderungen, die an einen Probanden durch Erziehungsaufwand gestellt werden, mit einer zunehmenden Kinderzahl bei gleichbleibenden Ressourcen steigen. Diese Vermutung wurde mit einer einfaktoriellen Varianzanalyse der Werte in der Belastungsskala über die nach Anzahl der Kinder gruppierten Probanden überprüft. Ergebnis dieser Analyse war ein signifikanter Unterschied zwischen den einzelnen Gruppe ($F(5,1119)=2,23$; $p=0,05$).

Tabelle D.8: Mittelwerte in der Belastungsskala gruppiert nach Kinderzahl

Kinderzahl	N	MW	SD
1	341	173,11	33,79
2	559	178,64	36,47
3	163	174,97	37,6
4	45	175,56	42,18
5	13	197,87	39,32
6	4	197,12	31,68

Anmerkung: N = Anzahl der Probanden; MW = Mittelwerte in der Belastungsskala; SD = Standardabweichung der Werte in der Belastungsskala

Tabelle D.8 stellt die Mittelwerte und Standardabweichungen der Belastungsskala in den nach Anzahl der Kinder gruppierten Validierungsstichprobe dar. Konform zu den Annahmen scheint zunächst der Befund zu sein, dass die Probandengruppen mit einer Anzahl von bis zu vier Kindern geringere Werte in der Belastungsskala aufweisen als Probanden mit einer Anzahl von fünf oder sechs Kindern. Als problematisch erweisen sich hierbei zwei Punkte: Zunächst unterscheiden sich die Mittelwerte zwischen den ersten vier Gruppen lediglich minimal und steigen zwischen der

vierten und der fünften Gruppe sprunghaft an, was dem implizit linearen Zusammenhang zwischen Kinderzahl und Belastung widerspricht. Weiterhin sind in den Gruppen mit den beiden höchsten Kinderzahlen nur äußerst wenige Probanden enthalten, weshalb nicht von einer Repräsentativität dieser Stichproben ausgegangen werden kann. Die konkurrente Validität mit der Anzahl der Kinder erweist sich daher als sehr fraglich.

3.4.3.3 Bildungsstand der Probanden

Der Bildungsstand der Probanden stellt eine weitere Variable dar, die Zusammenhänge mit der elterlichen Belastung zeigen kann. Diese Annahme beruht auf Überlegungen, die sich mit Wissen und allgemeiner kognitiver Fähigkeit der Probanden beschäftigen. Probanden, die einen qualifizierteren Bildungsabschluss haben, verfügen in der Regel über ein größeres Wissen und eine höhere allgemeine kognitive Fähigkeit als Probanden mit einem weniger qualifizierten Bildungsabschluss. Durch Wissen und kognitive Fähigkeit können Probanden, im Sinne eines breiteren Fundus an Problemlösestrategien, erziehungsspezifische Belastungspotentiale abbauen. Aus diesem Grund kann angenommen werden, dass Probanden mit einem höheren Bildungsabschluss niedrigere Werte in der Belastungsskala zeigen. Die Überprüfung dieser Annahme wurde wieder mittels einer einfaktoriellen Varianzanalyse der Werte in der Belastungsskala über die nach Bildungsabschluss gruppierten Probanden überprüft. Ergebnis dieser Analyse ist ein hochsignifikanter Unterschied zwischen den Bildungsgruppen ($F(3,1095)=44,806; p<0,001$).

Tabelle D.9: Mittelwerte in der Belastungsskala gruppiert nach Schulabschluss

Schulabschluss	N	MW	SD
Gymnasium	454	164,16	29,85
Realschule	356	177,46	32,39
Hauptschule	252	193,67	40,18
keiner	37	194,66	49,50

Anmerkung: N = Anzahl der Probanden; MW = Mittelwerte in der Belastungsskala; SD = Standardabweichung der Werte in der Belastungsskala

Tabelle D.9 gibt einen Überblick über Mittelwerte und Standardabweichungen der Belastungsskala in den einzelnen Bildungsgruppen. Es kann festgestellt werden, dass die Mittelwerte der Belastungsskala in der Gruppe der Abiturienten am geringsten ausgeprägt ist. Der nächsthöhere Mittelwert findet sich in der Gruppe von Probanden mit Realschulabschluss, gefolgt von dem der Probanden mit Hauptschulabschluss. Der höchste Mittelwert lässt sich in der Gruppe der Probanden ohne Schulabschluss finden. Diese Befunde sind in höchster Weise konsistent zu den vorher getroffenen Annahmen. Eine konkurrente Validität der Belastungsskala mit der Qualifikation des Schulabschlusses kann somit als gesichert gelten.

4. Entwicklung der Validitätsskalen

Da sich in den Validierungsuntersuchungen des CAPI zeigte, dass bei einem Ausschluss von Probanden, die verfälschende Antworttendenzen zeigen, eine höhere Validität der Belastungsskala angenommen werden kann, wurde bei der Validierung des EBSK eine Lügenskala, eine Zufalls-Antwort-Skala sowie eine Inkonsistenzskala konstruiert. Diesbezügliche Befunde werden in den folgenden Abschnitten dargestellt.

4.1 Lügenskala

In der deutschsprachigen Version des CAPI sollte ebenfalls eine Lügenskala zur Erfassung von Tendenzen sozial erwünschten Antwortverhaltens aufgenommen werden. Hierbei wurde auf die bereits von Globisch und Spangler (o. J.) übersetzten Originalitems zurückgegriffen. Im Rahmen der Entwicklung der Lügenskala wurde auf die Strategie zurückgegriffen, die bereits bei Milner (1986; vgl. Kapitel C, Abschnitt 3.1) verwendet wurde. Es sollten daher Items ausgewählt werden, deren Inhalt zwar in einem hohen Maße erwünscht, aber bei den Probanden nur in wenigen Fällen anzufinden ist (z.B. EBSK 50 „Ich tue immer das was richtig ist“), bzw. Items, deren Inhalt weniger erwünscht, aber in den meisten Fällen vorzufinden ist (z.B. CAPI 57: „Manchmal habe ich schlechte Gedanken“). Zur Feststellung der Eignung der Items zur Aufnahme in die Skala wurde hierbei auf das von Milner (1986) vorgeschlagene Kriterium des „85/15-Splits“ zurückgegriffen. Es sollten also lediglich die

Items in die Lügenskala aufgenommen werden, bei denen das Verhältnis zwischen zustimmenden und ablehnenden Antworten in allen Belastungsgruppen zumindest 85% zu 15% betragen.

Tabelle D.10: Ergebnisse der Validierung der Lügenskala

EBSK-Item	Unbelastet		Belastet		Misshandler	
	S	SN	S	SN	S	SN
4	6,0%	94,0	6,1	93,9	10,5	89,5
40	13,6	86,4	14,6	85,4	29,8	70,2
50	12,7	87,3	13,4	86,6	27,3	72,7
52	13,1	86,9	11,6	88,4	21,1	78,9
59	5,9	94,1	9,1	90,9	14,0	86,0

Anmerkung: Kursiv gedruckt sind Zellen, in denen das Verhältnis zwischen Zustimmung und Ablehnung zumindest 85% zu 15% beträgt. S = Stimmt, SN = Stimmt nicht

Tabelle D.10 stellt die Ergebnisse dieser Betrachtungen dar. In der Gruppe der Unbelasteten und der Belasteten erreichten fünf Items das vorgegebene Verhältnis. In der Gruppe der Misshandler erreichen nur 2 Items (Item 4, Item 59) das Verhältnis von 85% zu 15%. Dennoch zeigte sich auch bei den anderen Items eine starke Verschiebung zu der ablehnenden Antwortalternative. Zur Überprüfung der Signifikanz dieser Abweichungen von der Gleichverteilung wurden χ^2 -Tests durchgeführt. Für alle drei Items zeigten sich eine hochsignifikante Verschiebung in Richtung der ablehnenden Antwortalternative (Item 40: $\chi^2 = 9,28$, $p = 0,002$; Item 50: $\chi^2 = 11,36$, $p = 0,001$; Item 52: $\chi^2 = 19,11$, $p < 0,001$). Weiterhin lässt sich feststellen, dass alle Items in die gleiche Richtung gepolt sind, d.h. die geforderte Verteilung ergab sich lediglich für Items, in denen erwünschte Inhalte abgefragt werden.

Der nächste Validierungsschritt befasste sich mit der Festlegung eines sinnvollen Wertes, ab dem die Ergebnisse des Fragebogens auf Grund einer möglichen Verzerrung durch sozial erwünschtes Antwortverhalten vorsichtig interpretiert werden müssen. Hierzu wurden die fünf Items für jeden Probanden aufsummiert. Die so gewonnene Verteilung hat einen Mittelwert von 0,534 mit einer Standardabweichung von 0,911. Als Cut-Off der Lügenskala wurde der Wert 3 gewählt. Er entspricht einer

Erhöhung des Summenscores um mehr als 2 Standardabweichungen und schneidet 95,2% der Validierungsstichprobe ab.

Zur Überprüfung der Fähigkeit der Lügenskala zur Identifikation von Probanden, die in einer sozial erwünschten Art und Weise antworten, wurde zusätzlich ein letzter Validierungsschritt unternommen. Hierbei wurden die Stichproben „Fake-Good 1“ und „Fake-Good 2“ (Kapitel D, Abschnitt 2.6) herangezogen, bei denen Tendenzen der Antwortverzerrung in eine sozial erwünschte Richtung gezielt induziert wurden. Die Überlegung hinter dieser Strategie war die Folgende: Sollte die Lügenskala geeignet sein Probanden zu identifizieren, die in einer sozial erwünschten Art und Weise antworten, so sollte sich die in Mittelwertsunterschieden der Lügenskala zwischen der Validierungsstichprobe und den Stichproben „Fake-Good 1“ sowie „Fake-Good 2“ zeigen. Die Überprüfung dieser Annahme erfolgte über eine Varianzanalyse der Werte in der Lügenskala zwischen der Validierungsstichprobe und den Gruppen „Fake-Good 1“ sowie „Fake-Good 2“. Als Ergebnis dieser Analyse stellte sich ein hochsignifikanter Effekt zwischen den drei Stichproben heraus ($F(2,1489)=181$; $p<0,001$).

Tabelle D.11: Mittelwerte und Standardabweichungen der Lügenskala gruppiert nach Stichprobe

Stichprobe	N	MW	SD
Validierungsstichprobe	1313	0,54	0,96
Fake-Good 1	96	1,92	1,69
Fake-Good 2	83	2,4	1,67

Anmerkung: N = Anzahl der Probanden in der Stichprobe; MW = Mittelwert der Werte in der Lügenskala; SD = Standardabweichung der Werte in der Lügenskala; Die Darstellung für die Ergebnisse der Validierungsstichprobe bezieht sich lediglich auf die Probanden, die alle Items, die in die Lügenskala eingehen, beantwortet haben

Tabelle D.11 stellt die Mittelwerte und Standardabweichung der Werte der Lügenskala in den drei betrachteten Gruppen dar. Es kann festgestellt werden, dass der Mittelwerte in der Validierungsstichprobe am geringsten ausgeprägt ist, gefolgt von der Stichprobe „Fake-Good 1“ und danach erst von der Stichprobe „Fake-Good 2“. Dieser Befund ist als konsistent zu den getroffenen Annahmen zu werten. Die Validi-

tät der Lügenskala als ein Indikator von Probanden, die sozial erwünschtes Antwortverhalten zeigen, kann somit als gegeben betrachtet werden.

4.2 Zufalls-Antwort-Skala

Die Zufalls-Antwort-Skala sollte in die deutschsprachigen Version ebenfalls als Indikator eines Antwortverhaltens, das zufällig und damit ohne Bezug zu dem Inhalt der Items geschieht, aufgenommen werden. Zur Konstruktion der Skala wurde auf die bei Milner (1986) beschriebene Strategie (Kapitel C, Abschnitt 3.2) zurückgegriffen. Es sollten daher Items gefunden werden, die von aufmerksam antwortenden Probanden in gleicher Weise bearbeitet werden sollten (z.B. EBSK 22 „Die meisten Kinder sind gleich“). Von ihrem Inhalt her unterscheiden sich die Items jedoch von denen der Lügenskala insofern, als sie sozial neutrale Verhaltensweisen bzw. Einstellungen repräsentieren.

Die Validierung dieser Skala erfolgte in einer ähnlichen Art und Weise wie die der Lügenskala. Items wurden in die Zufalls-Antwort-Skala aufgenommen, wenn zustimmende und ablehnende Antworten sich mindestens in einem Verhältnis von 85% zu 15% verteilten.

Tabelle D.12: Ergebnisse der Validierung der Zufalls-Antwort-Skala

Item	Unbelastet		Belastet		Misshandler	
	S	SN	S	SN	S	SN
6	3,0	97,0	2,6	97,4	16,1	83,9
21	0,9	99,1	0,9	99,1	3,5	96,5
22	10	90	13	87	19,3	80,7
37 (R)	88,9	10,1	89,7	10,3	91,2	8,8
51	5,2	94,8	3,9	96,1	5,5	94,5
53	2,1	97,9	2,6	97,4	7	93
60 (R)	88,2	11,8	90,4	9,6	87,5	12,5

Anmerkung: Kursiv gedruckt sind Zellen in denen das Verhältnis zwischen Zustimmung und Ablehnung zumindest 85% zu 15% beträgt. Mit R markierte Items müssen vor der Bildung des Summenscores umcodiert werden. S = Stimmt, SN = Stimmt nicht

Tabelle D.12 gibt einen Überblick über die Items, die die vorausgesetzte Verteilung erfüllten. In den Gruppen der Unbelasteten und der Belasteten erreichen alle Items das vorgegebene Verhältnis. In der Gruppe der Misshandler erreichen lediglich fünf der sieben Items das Verhältnis von 85% zu 15% (Item 21, Item 37, Item 51, Item 53, Item 60). Bei den beiden übrigen Items der Zufalls-Antwort-Skala (Item 6, Item 22) lässt sich jedoch ebenfalls eine deutliche Verschiebung des Verhältnisses in erwarteter Richtung feststellen. Bei der Überprüfung der Abweichung von der Gleichverteilung mittels χ^2 -Tests zeigten sich für beide Items hochsignifikante Ergebnisse (Item 6: $\chi^2 = 25,77$, $p < 0,001$; Item 22: $\chi^2 = 21,49$, $p < 0,001$). Da sich bei zwei Items (Item 37, Item 60) eine Anhäufung bei der ablehnenden Antwortalternative zeigte, müssen diese vor der Berechnung des Summenscores umcodiert werden.

Zur Bestimmung eines sinnvollen Cut-Off-Wertes, ab dem eine Beantwortung des Fragebogens nach dem Zufallsprinzip wahrscheinlich ist, wurden zunächst alle Items der Skala in die gleiche Richtung gepolt und danach aufsummiert. Der Mittelwert der Verteilung von Werten in der Zufalls-Antwort-Skala liegt bei 0,44 mit einer Standardabweichung von 0,71. Der Cut-Off dieser Skala wurde ebenfalls etwas mehr als zwei Standardabweichungen über dem Mittelwert bei zwei Skalenpunkten gewählt und schneidet damit 95,2% der Gesamtstichprobe ab.

4.3 Inkonsistenzskala

Die Inkonsistenzskala, als Indikator für widersprüchliches Antwortverhalten, sollte ebenfalls in die deutschsprachige Version aufgenommen werden. Die Strategie, die zur Konstruktion der Skala gewählt wurde, orientiert sich an der von Milner (1986; vgl. Kapitel C, Abschnitt, 3.3). Hierbei sollten Itempaare gefunden werden, deren Inhalt entweder ähnlich (z.B. EBSK 17 „Die Dinge in meinem Leben gingen normalerweise gegen mich“ und EBSK 47 „Ich habe immer Pech“) oder gegensätzlich (z.B. EBSK 11 „Ich lache fast jeden Tag“ und EBSK 62 „Ich lache nicht sehr viel“) formuliert ist. Bei einer konsistenten Beantwortung des Fragebogens müssten die ähnlichen Itempaare gleich, die gegensätzlichen Itempaare unterschiedlich beantwortet werden.

Die Vorgehensweise, die zur Konstruktion dieser Skala verwendet wurde, ist ähnlich der bei den beiden anderen Validitätsskalen. Itempaare wurden dann in die Skala aufgenommen, wenn sie von mindestens 85% der Probanden in den einzelnen Belastungsgruppen in einer konsistenten Weise beantwortet wurden.

Tabelle D.13: Ergebnisse der Validierung der Inkonsistenzskala

Itempaar	Unbelastet		Belastet		Misshandler	
	K	I	K	I	K	I
11-62	85,8	14,2	85,8	14,2	82,5	17,5
12-24	<i>91,5</i>	<i>8,5</i>	<i>89,2</i>	<i>10,8</i>	<i>86</i>	<i>14</i>
17-47	<i>95,8</i>	<i>4,2</i>	<i>89,7</i>	<i>10,3</i>	<i>78,9</i>	<i>21,1</i>
29-36	<i>91,5</i>	<i>8,5</i>	<i>90</i>	<i>10</i>	<i>84,2</i>	<i>15,8</i>
32-49	<i>93,8</i>	<i>6,2</i>	<i>91,8</i>	<i>8,2</i>	<i>87,7</i>	<i>12,3</i>
39-61	<i>91,8</i>	<i>8,2</i>	<i>82,7</i>	<i>17,3</i>	<i>78,6</i>	<i>21,4</i>

Anmerkung: Kursiv gedruckt sind Zellen, in denen mindestens 85% der Probanden in einer konsistenten Weise geantwortet haben. K = Konsistent, I = Inkonsistent

Tabelle D.13 stellt die Ergebnisse der überprüfenden Betrachtung der Inkonsistenzskala dar. In der Gruppe der Unbelasteten erreichten lediglich das Itempaar 11-62 nicht das vorgegebene Kriterium. Dennoch kann eine deutliche Verschiebung hin zu der Seite der konsistenten Antworten festgestellt werden. Die Verteilung weicht hierbei hochsignifikant von der Gleichverteilung ab ($\chi^2 = 533,17$, $p < 0,001$). In der Gruppe der Belasteten erreichten fünf Itempaare das erwartete Verhältnis von mindestens 85% zu 15%. Auch bei dem Itempaar 39-61 lässt sich eine hochsignifikante Verschiebung hin zu dem konsistenten Antwortverhalten feststellen. Der χ^2 -Test auf Abweichung von der Gleichverteilung ist hierbei mit einem χ^2 von 98,71 mit einer Irrtumswahrscheinlichkeit kleiner 0,01% anzunehmen. In der Gruppe der tatsächlichen Misshandler erreichten lediglich zwei Itempaare (Itempaar 12-24, Itempaar 32-49) den vorgegebenen Anteil konsistenter Antworten. Bei den übrigen Itempaaren zeigte sich jedoch ebenfalls eine Verschiebung der Verteilung zu konsistentem Antwortverhalten hin, die in allen Fällen als hochsignifikant zu werten ist (Itempaar

17-47: $\chi^2 = 19,11$, $p < 0,001$; Itempaar 11-62: $\chi^2 = 24,02$, $p < 0,001$; Itempaar 39-61: $\chi^2 = 18,29$, $p < 0,001$; Itempaar 29-36: $\chi^2 = 26,68$, $p < 0,001$).

Zur Bestimmung eines sinnvollen Cut-Offs, ab dem inkonsistentes Antwortverhalten angenommen werden kann, wurde für jeden Probanden in der Validierungsstichprobe die Anzahl der inkonsistenten Antworten bestimmt. Der Mittelwert dieser Verteilung ergab sich zu 0,55 mit einer Standardabweichung von 0,81. Als Cut-Off wurde die Zahl von drei inkonsistent beantworteten Paaren festgesetzt. Diese entspricht einer Erhöhung der Summenscores von mehr als 2 Standardabweichungen über dem Mittelwert und schneidet insgesamt 96,6% der Verteilung ab.

5. Wahl des Maximums zulässiger fehlender Antworten

Im Rahmen der deutschen Validierungsuntersuchung wurde kein Cut-Off bestimmt, ab dem eine Interpretation des Tests aufgrund einer zu großen Anzahl fehlender Werte abgebrochen werden sollte. Für die Anzahl fehlender Werte in der Belastungsskala wird auf den Cut-Off des amerikanischen Originalverfahrens zurückgegriffen. Dieser schreibt vor, dass die Interpretation der Belastungsskala ab einem Anteil 10% fälschlich beantworteter bzw. ausgelassener Items in der jeweiligen Skala nicht mehr vollzogen werden sollte. Milner (1986, S. 10) schreibt hierzu: „the 10% rule appears functional in eliminating those with excessive blanks, yet allows retention of most of the examinees who have one or several blanks“. Aus diesem Grunde wird ein Abbruch der Interpretation der Belastungsskala ab einer Zahl von sechs fehlenden Items empfohlen. Dies entspricht einem Anteil von 12,77% fehlender Werte in den 47 Items der Belastungsskala.

Für die Validitätsskalen eignet sich der von Milner (1986) vorgeschlagene Wert von 10% nicht. Da diese Skalen in jedem Fall weniger als zehn Items umfassen, dürfte eine Interpretation dieser Werte schon bei einem fehlenden Item nicht mehr erfolgen. Für die Auswertung des EBSK werden daher für diese Skalen weniger konservative Cut-Offs empfohlen. Die Interpretation der Lügen- und Inkonsistenzskala sollte ab einer Anzahl drei fehlender Items, bzw. unvollständiger Itempaare nicht mehr erfolgen. Für die Lügenskala entspricht dies einem Anteil von 60% der insgesamt fünf

Items, für die Inkonsistenzskala einem Anteil von 50% der insgesamt sechs Paare. Für die Zufalls-Antwort-Skala wurde ein Cut-Off von vier Items gewählt, ab dem eine Interpretation der Skala nicht mehr erfolgen sollte. Dies entspricht einem Anteil von 50% der insgesamt acht Items der Skala.

6. Normierung

Letzter Schritt in der Untersuchung zur Validierung des EBSK war die Normierung der Belastungsskala. Zunächst stellte sich die Frage, welche Variablen sich neben der der Belastungsgruppe als sinnvolle Grundlage zur Stratifizierung der Validierungsstichprobe anbieten. Infolge diesbezüglicher Überlegungen wurden zwei zusätzliche Variablen hinsichtlich ihres Einflusses auf die Werte der Belastungsskala untersucht: Geschlecht und Alter.

Zur gemeinsamen Überprüfung des Einflusses dieser drei Variablen (Belastungsgruppe, Geschlecht und Alter) auf die Werte in der Belastungsskala wurde eine dreifaktorielle Varianzanalyse durchgeführt⁷. Als Ergebnis dieser Analyse lässt sich festhalten, dass lediglich der Haupteffekt der Belastungsgruppe einen signifikanten Einfluss auf die Werte in der Belastungsskala hat ($F(2,1127)=45,74$; $p<0,001$). Für die übrigen beiden Faktoren sowie alle Interaktionen konnte kein signifikanter Einfluss auf die Werte in der Belastungsskala festgestellt werden.

Aufgrund der Ergebnisse dieser Analyse wurde die Validierungsstichprobe lediglich nach den Belastungsgruppen stratifiziert. Aufgrund der relativ geringen Fallzahl und der damit verbundenen großen Streuung der Belastungswerte in der Substichprobe der Misshandler werden keine gesonderten Normwerte für diese Belastungsgruppe angegeben. Mittelwerte und Standardabweichungen der beiden übrigen Substichproben (Unbelastete, Belastete) wurden als Grundlage der Erstellung von T-Normen genutzt.

⁷ Der Faktor Belastungsgruppe wurde gestuft in Unbelastet, Belastet und Misshandler, der Faktor Geschlecht in männlich und weiblich, der Faktor Alter in weniger als 20 Jahre, 20 bis 30 Jahre, 30 bis 40 Jahre, 40 bis 50 Jahre und mehr als 50 Jahre.

7. Vorbereitung von Test- und Auswertungsmaterial

Zur ökonomischen und validen Durchführung und Auswertung des EBSK wurde eine Reihe von Materialien hergestellt. Zunächst wurden Fragebögen, Auswertungsschablonen und Auswertungsblätter angefertigt, die für einen klassischen Einsatz des Verfahrens z.B. in psychiatrischen Kliniken, psychologischen Praxen, Jugendhilfeeinrichtungen gedacht sind. Weiterhin wurde eine SPSS-Syntax erstellt, mithilfe derer große Fragebogenmengen, wie sie beispielsweise bei einem Einsatz in Forschungsprojekten zu bearbeiten sind, ökonomisch ausgewertet werden können. Um eine korrekte Benutzung des Verfahrens zu gewährleisten, werden in dem Manual ausführliche Informationen für den Anwender dargestellt. Eine weitere Besprechung des Test- und Auswertungsmaterials soll an dieser Stelle nicht erfolgen. Ausführliche weitere Informationen hierzu finden sich im Manual (Deegener, Spangler, Körner & Becker, im Druck).

8. Abschließende Bewertung

Zusammenfassend kann festgestellt werden, dass das EBSK als eine ökonomische, reliable und valide deutschsprachige Version des CAPI entwickelt werden konnte. Zunächst zeigte sich, dass die englischsprachigen Items mit einer hohen inhaltlichen Nähe ins Deutsche übersetzt werden konnten (Kapitel D, Abschnitt 1). Weiterhin kann einem gewissen Anteil der Items diskriminative Fähigkeit in Bezug auf die Trennung von Gruppen unbelasteter und belasteter Probanden zugemessen werden (Kapitel D, Abschnitt 3.1). Mithilfe diskriminanzanalytischer Betrachtungen konnte auf Basis der Items eine Funktion geschätzt werden, mithilfe derer ein großer Anteil der Probanden der Validierungsstichprobe entsprechend des vorliegenden Belastungsgrades richtig klassifiziert werden konnte (Kapitel D, Abschnitt 3.2). Hinsichtlich der faktoriellen Struktur der so konstruierten Belastungsskala zeigt sich ein einziger bedeutsamer Faktor, dessen Validität ebenfalls aufgrund der Betrachtung der internen Konsistenz der Items als gesichert anzusehen ist (Kapitel D, Abschnitt 3.4.1). Weiterhin konnte die konvergente und diskriminante Validität des Verfahrens sowohl in Bezug auf ein anderes psychometrisches Verfahren (Kapitel D,

Abschnitt 3.4.2) als auch im Hinblick auf bestimmte demographische Variablen (Kapitel D, Abschnitt 3.4.3) stimmig interpretiert werden. Zusätzlich zu der Belastungsskala konnten ebenfalls drei Validitätsskalen zur Identifikation spezifischer verfälschender Antworttendenzen konstruiert werden, die den von Milner (1986) festgelegten Gütekriterien in einer guten Weise entsprechen (Kapitel D, Abschnitt 4). Zuletzt konnte durch die Wahl des Maximums fehlender Antworten, die Festlegung der Normwerte und die Vorbereitung des Test- und Auswertungsmaterials die Grundlage für einen ökonomischen und sinnvollen Einsatz des Verfahrens in der Praxis geschaffen werden (Kapitel D, Abschnitt 5 bis 7).

E. Diskussion und Ausblick

In den folgenden Abschnitten soll eine abschließende Diskussion der Befunde der vorliegenden Arbeit erfolgen sowie ein Ausblick auf Ansatzpunkte gegeben werden, an denen eine weitergehende Validierung des Verfahrens anknüpfen kann. Ein erster Abschnitt wird sich in diesem Zusammenhang mit der Darstellung eines Vergleichs zwischen der englischsprachigen Originalversion und der deutschen Version beschäftigen. Hierbei wird vor allem auf Unterschiede eingegangen, die sich im Rahmen der beiden Validierungsuntersuchungen zeigten. In einem zweiten Abschnitt werden diese Unterschiede aufgegriffen und im Hinblick auf weitere Schritte zur Validierung des EBSK diskutiert.

1. Unterschiede zwischen CAPI und EBSK

Im Rahmen der Konstruktion von CAPI und EBSK konnten zwei einander relativ ähnliche Verfahren entwickelt werden. Unterschiede zwischen den beiden Versionen zeigen sich dennoch in der Anzahl der in den Fragebögen verwendeten Items, der klassifikatorischen Güte von Abuse Scale und Belastungsskala sowie in der Interpretation der faktoriellen Struktur, die der Auswertung dieser beiden Skalen zugrunde gelegt wird. In den folgenden Abschnitten sollen diese Unterschiede eingehender besprochen und hinsichtlich möglicher Ursachen diskutiert werden.

1.1 Unterschiede im Umfang der Verfahren

Bezüglich des Umfangs der beiden Verfahren lässt sich feststellen, dass in die Auswertung des CAPI insgesamt 120, in die Auswertung des EBSK lediglich 63 Items eingehen. Als Grundlage dieser Reduktion des Itemsatzes können kulturelle Unterschiede zwischen der US-amerikanischen und der deutschen Gesellschaft diskutiert werden. Da das CAPI an Stichproben amerikanischer Probanden validiert wurde, sind die Inhalte des Fragebogens besonders gut an dort vorherrschende kulturelle Gegebenheiten (z.B. Wertvorstellungen, Lebensweisen) angepasst. Da die gleichen Fragen einer Stichprobe deutscher Probanden vorgegeben wurden, ist anzunehmen, dass lediglich Items, deren Inhalt in den beiden Kulturen eine ähnliche Bedeutung hat auch ähnlich gute Indikatoren der zu erfassenden Konstrukte (Misshand-

lungsrisiko, Verfälschungstendenzen) darstellen. Items, die lediglich in der US-amerikanischen Gesellschaft eine Bedeutung für die zu messenden Konstrukte haben, erfüllen die Kriterien für eine Aufnahme in den Fragebogen nicht und werden daher auch nicht in die deutsche Version übernommen.

1.2 Unterschiede in klassifikatorischer Güte von Abuse Scale und Belastungsskala

Hinsichtlich der Güte der Klassifikation kann festgestellt werden, dass in der englischsprachigen Version die relativen Anteile korrekter Klassifikationen (d.h. Misshandler werden als Misshandler klassifiziert, Kontrollprobanden als Kontrollprobanden) zwischen 96% in der Validierungsuntersuchung (Milner, 1986), ca. 85 bis 90% in vergleichbaren Studien in der Literatur (z.B. Milner & Wimberley, 1980; Milner, Gold & Wimberley, 1986; Milner & Robertson, 1989) und 72,7% sofern die zu trennenden Gruppen sich nur wenig hinsichtlich des Misshandlungsrisikos unterscheiden (Matthews, 1984), variieren. Der Vergleich mit Befunden aus der deutschen Validierungsuntersuchung zeigt, dass das EBSK bezüglich dieses Validitätsaspekts etwas schlechter zu bewerten ist. Mit einem Anteil von 76,2% korrekt klassifizierter Probanden in der Stichprobe, die zur Konstruktion der Diskriminanzfunktion verwendet wurde, einem Anteil von 75% korrekt klassifizierter Probanden in der gesamten Validierungsstichprobe und einem Anteil von 70,2% korrekt klassifizierter Probanden in der Kreuzvalidierungsstichprobe liegt die klassifikatorische Güte des EBSK unter der des CAPI.

Als Grundlage des Abfalls der Genauigkeit der Zuordnung können zunächst die bereits vorher (Kapitel E, Abschnitt 1.1) besprochenen Unterschiede zwischen der Bedeutsamkeit bestimmter Items in der deutschen und der US-amerikanischen Gesellschaft diskutiert werden. Sollten misshandlungsrelevante kulturelle Inkompatibilitäten zwischen den USA und der BRD bestehen, so sind die auf die US-amerikanische Stichprobe angepassten Items in der BRD keine hinreichend genauen Indikatoren des Misshandlungsrisikos. Die klassifikatorische Güte des Verfahrens sinkt somit im Bezug auf deutsche Stichproben, da teilweise Aspekte abgefragt werden, die für diesen Kulturkreis irrelevant sind, während andererseits bedeutsame Aspekte nicht

abgefragt werden. Ein Indiz hierfür zeigt sich in den Befunden aus der Validierungsuntersuchung der Belastungsskala, in der für 30 Items der Abuse Scale keine Mittelwertsunterschiede zwischen den Belastungsgruppen festgestellt werden konnten, obwohl sie in der US-amerikanischen Validierungsuntersuchung signifikant zwischen Misshandlern und Kontrollprobanden trennen konnten.

Eine weitere Ursache für Unterschiede in der klassifikatorischen Güte zwischen der Originalversion und der deutschsprachigen Übersetzung kann in der Zusammensetzung der Stichproben, die zur Validierung der Verfahren verwendet wurden, vermutet werden. Die Validierungsstichproben des CAPI bestanden ausschließlich aus Probanden, bei denen eindeutig festgestellt werden konnte, ob sie Misshandler oder nicht-misshandelnde Kontrollprobanden waren. Eine derart hohe Trennschärfe ist zwischen den verschiedenen Belastungsgruppen in der Validierungsstichprobe des EBSK nicht auszumachen. So besteht die Stichprobe der Unbelasteten im Wesentlichen aus Probanden aus einer Feldstudie an Eltern von Kindergarten- und Grundschulkindern (Kapitel D, Abschnitt 2.1). Zwar ist davon auszugehen, dass es sich bei einem überwiegenden Anteil dieser Stichprobe um nicht-misshandelnde Elternteile handelt. Eine definitive Aussage, alle Probanden der Stichprobe seien tatsächlich unbelastet, ist jedoch nicht zulässig. Hinsichtlich der Stichprobe der Belasteten und der Misshandler (Kapitel D, Abschnitt 2.2; Kapitel D, Abschnitt 2.4 bis 2.5) zeigt sich eine ähnliche Schwierigkeit. Zwar hat ein Großteil dieser Probanden große Probleme mit ihren Kindern, die Aussage, die Stichproben beständen ausschließlich aus tatsächlichen Misshandlern, ist aber dennoch nicht möglich. Es kann damit festgestellt werden, dass die Heterogenität der verglichenen Gruppen in der deutschen Validierungsstichprobe in einem geringeren Maße festzustellen ist, als die Heterogenität in der US-amerikanischen Stichprobe. Aufgrund dieser unzureichenden Trennschärfe zwischen den beiden Gruppen ist anzunehmen, dass sich die Verteilungen der Belastungswerte in den beiden Gruppen zu einem gewissen Grade überlappen. Es ist daher wahrscheinlich, dass es besonders bei Probanden, die am Rande der beiden Verteilungen liegen (d.h. besonders belastete Probanden aus der Grup-

pe der Unbelasteten sowie besonders unbelastete aus der Gruppe der Belasteten) zu Fehlklassifikationen kommen kann.

1.3 Unterschiede in der Faktorenstruktur von Abuse Scale und Belastungsskala

Ein letzter Unterschied zwischen der englischsprachigen Originalversion und der deutschen Übersetzung ist in der Faktorenstruktur zu sehen, die der Auswertung der Abuse Scale bzw. der Belastungsskala zugrunde gelegt wird. Während die im Manual des CAPI (Milner, 1986) beschriebenen Befunde aus den Validierungsuntersuchungen eine drei- bis sechsfaktorielle Faktorenstruktur nahelegen (Kapitel C, Abschnitt 2.3.2), deuten die Ergebnisse aus dem deutschen Datensatz auf eine ein-faktorielle Struktur der Belastungsskala hin (Kapitel D, Abschnitt 3.4.1).

Erklären lässt sich dieser Unterschied vor dem Hintergrund des Einsatzes unterschiedlicher faktoranalytischer Strategien. Während in den Validierungsstudien des CAPI Hauptachsenanalysen mit anschließender Promax-Rotation durchgeführt wurden, erfolgte die Ermittlung der Faktorenstruktur des EBSK mithilfe einer MixFactor-Analyse. Diese beiden Faktoranalysen unterscheiden sich zunächst im Hinblick auf die Skalendignität, die in dem zu analysierenden Datensatz vermutet wird. Während die Durchführung einer Hauptachsenanalysen nur bei intervallskalierten Daten zulässig ist (Überla, 1971), wurde die MixFactor-Analyse explizit als Strategie zur Faktorisierung dichotom-ordinaler Datensätze entwickelt (Kubinger & Häusler, 2004; Kubinger, 2003). Da sowohl die Items des CAPI, als auch die Items des EBSK in einem dichotomen Antwortformat beantwortet wurden, kann die MixFactor-Analyse als die geeignetere Strategie zur Bestimmung der Skalenstruktur von Abuse Scale bzw. Belastungsskala angesehen werden. Die einfaktorielle Struktur, die gemäß der Ergebnisse der deutschen Validierungsuntersuchung angenommen wird, kann damit als die plausiblere angenommen werden. Weiterhin ist entsprechend der Überlegungen von Kubinger (2003) zu vermuten, dass es sich bei den zusätzlichen Faktoren, die in der US-amerikanischen Validierungsstudie gefunden wurden, um Methodenartefakte handelt.

Zusätzlich zu den eben dargestellten methodischen Überlegungen deuten auch Forschungsbefunde aus der Validierungsuntersuchung des CAPI auf eine bestehende Überschätzung der Anzahl bedeutsamer Faktoren hin. Zunächst lässt sich feststellen, dass die Faktoren der sechsfaktoriellen Struktur hoch interkorreliert sind. Ayoub, Pan, Rocha und Emerson (2000) berichten hierbei von durchgängig hochsignifikanten Korrelationen zwischen den einzelnen Subskalen und der Abuse Scale. Größenmäßig liegen diese in einem Bereich von .34 bis .98 und weisen einen Mittelwert von .70 auf. Ebenfalls zeigen die Befunde zur Reliabilität der Abuse Scale, dass die interne Konsistenz der Skala mit SH-Werten im Bereich von .93 und .99 und KR20-Werten in einem Bereich von .85 bis .96 ebenfalls auf einem sehr hohen Niveau angenommen werden kann, was wiederum auf eine sehr hohe Interkorrelation der Items der Skala schließen lässt. Aufgrund dieser bedeutsamen Korrelationen können die eigenständigen Varianzanteile der Subskalen bzw. Items, die ausschlaggebend für die Sinnhaftigkeit der Annahme zusätzlicher Faktoren wären, als vernachlässigbar gering betrachtet werden.

Zusammenfassend kann festgestellt werden, dass sowohl hinsichtlich methodischer Überlegungen als auch hinsichtlich empirischer Forschungsbefunde die Vorgehensweise, die im Rahmen der Entwicklung des EBSK verwendet wurde, eine plausiblere Interpretation der Struktur der Abuse Scale bzw. der Belastungsskala erlaubt. Insgesamt liegt somit die Vermutung nahe, dass die einfaktorielle Lösung, die als Grundlage des EBSK gewählt wurde, eher anzunehmen ist als die sechsfaktorielle Lösung des CAPI.

2. Weitere Validierungsschritte

Im Folgenden werden Schritte zur weiteren Validierung des EBSK besprochen, die in diesem Zusammenhang als Anknüpfungspunkte für weitergehende Forschung aufgegriffen werden können. Ein erster Abschnitt wird sich hierbei mit der Generierung zusätzlicher Items zur Verbesserung der Diskriminationsfähigkeit zwischen Gruppen misshandelnder und nicht-misshandelnder Probanden befassen. In einem zweiten Abschnitt wird beschrieben, inwiefern die Validität des Verfahrens durch die Analy-

se von Daten spezieller Stichproben zusätzlich verbessert werden kann. Ein letzter Abschnitt befasst sich mit der Erweiterung der Betrachtungen durch eine längsschnittliche Perspektive, mithilfe derer zusätzliche Aussagen zu Reliabilität und Validität des Verfahrens gemacht werden können.

2.1 Generierung zusätzlicher Items

Wie bereits oben dargestellt zeigte sich, dass eine Reihe von Items der amerikanischen Originalversion nicht für eine Aufnahme in die Skalen der deutschsprachigen Version geeignet war. In diesem Zusammenhang wurde diskutiert, dass sich die mangelnde Passung der Items wahrscheinlich auf kulturelle Inkompatibilitäten zwischen der deutschen und der US-amerikanischen Gesellschaft zurückführen lässt. Weiterhin wird vermutet, dass der Unterschied der klassifikatorischen Güte zwischen dem CAPI und dem EBSK unter anderem in diesem Sachverhalt begründet liegt. Da eine Erhöhung des Anteils richtig zugeordneter Probanden als wünschenswert zu erachten ist, können erste Schritte zur Weiterentwicklung des Verfahrens in der Generierung zusätzlicher Items gesehen werden, die sich für eine Erfassung misshandlungsrelevanter bzw. validitätsmindernder Inhalte eignen, die spezifische Gültigkeit in der deutschen Gesellschaft besitzen. Hierfür bieten sich zunächst vor allem Inhalte an, die in den von Milner beschriebenen Literaturrecherchen (Milner & Williams, 1978) gefunden wurden, aufgrund mangelnder Gültigkeit in der US-amerikanischen Gesellschaft jedoch im CAPI nicht enthalten sind.

Ebenfalls ist festzustellen, dass die Generierung der Items des CAPI auf der Grundlage von Literaturrecherchen erfolgte, die mittlerweile mehr als 20 Jahre alt sind. Neuere Erkenntnisse zu Risiko- und Schutzfaktoren für die Entstehung von Kindesmisshandlung konnten daher nicht berücksichtigt werden. Wie bereits dargestellt wurde, kann den Items des CAPI dennoch eine hohe inhaltliche Nähe zu aktuellen Befunden zur Ätiologie von Kindesmisshandlung zugemessen werden (Kapitel C, Abschnitt 2.3.1). Trotzdem könnten in der neueren Forschungsliteratur Anregungen für die Generierung zusätzlicher Items gefunden werden, deren Einsatz zu einer Verbesserung der klassifikatorischen Güte des Verfahrens dienen könnte. Eine gute

Grundlagen hierfür ist in den angeführten aktuellen Übersichtsarbeiten (Black, Heyman & Slep, 2001; Black, Slep & Heyman, 2001) und Metaanalysen (Stith et al., 2004; Stith et al., im Druck) zu sehen.

2.2 Differenziertere Erfassung von Misshandlungsformen und Schweregraden

Als einen weiteren Grund für den Unterschied zwischen der klassifikatorischen Güte zwischen CAPI und EBSK wurden Unterschiede in der Zusammensetzung der bei den Validierungsuntersuchungen verwendeten Stichproben besprochen (Kapitel E, Abschnitt 1.2). Ein sinnvoller Schritt zur Verbesserung dieses Validitätsaspekts im EBSK ist daher zunächst in der Akquise von Probanden zu sehen, bei denen das Vorliegen von Kindesmisshandlungsfällen eindeutig festgestellt werden kann, d.h. Probanden, bei denen ausgeschlossen werden kann, dass sie in irgendeiner Form misshandelt haben und Probanden, bei denen Misshandlungen eindeutig vorlagen. Aufgrund der Heterogenität dieser Stichproben könnte zunächst ein Vergleich der Diskriminationsfähigkeit des CAPI und des EBSK unter gleichartig günstigen Voraussetzungen erfolgen. Weiterhin könnte aufgrund dieser Stichproben eine neue Diskriminanzfunktion zur Gruppenzuordnung bestimmt werden, die eine schärfere Abgrenzung von misshandelnden und nicht-misshandelnden Eltern erlaubt und daher weniger Fehlklassifikationen produziert.

Zusätzlich ist eine sinnvolle Erweiterung der Stichprobe in der Akquise von Probanden zu sehen, bei denen sich eindeutig unterschiedliche Formen der Kindesmisshandlung (physische und psychische Misshandlung, Vernachlässigung, sexueller Missbrauch) sowie unterschiedliche diesbezügliche Schweregrade (Misshandlung im engeren und weiteren Sinn) feststellen lassen. Zunächst erlaubt die Erfassung des Schweregrades vorliegender Misshandlungen eine Abschätzung der Fähigkeit des EBSK zur Diskrimination der Belastungsgrade der Eltern entlang des Belastungskontinuums. Hierbei ließe sich feststellen, mit welcher Genauigkeit Probanden unterschieden werden können, bei denen sich lediglich geringe Abweichungen in dem Schweregrad der Misshandlung vorfinden lassen. Mithilfe von hinsichtlich unterschiedlicher Formen der Misshandlung gruppierter Stichproben kann zunächst die

Fähigkeit des EBSK zur Abgrenzung derselben überprüft werden. Weiterhin kann auf der Basis einer derartigen Datenlage eine multifaktorielle Diskriminanzfunktion geschätzt werden, die eine Einschätzung der Wahrscheinlichkeit der Zugehörigkeit der Probanden zu Gruppen verschiedener Misshandlungsformen erlaubt.

Ein erster Schritt zur Akquise der dargestellten Stichprobentypen ist in dem „Aufruf zur zukünftigen Forschung mit dem EBSK“ zu sehen, der das letzte Kapitel des Testmanuals (Deegener, Spangler, Körner & Becker, im Druck; S.93 ff) darstellt. Hierbei werden Anwender des EBSK gezielt um eine Zusendung von anonymisierten EBSK-Fragebögen nebst Einstufungsbögen zum Schweregrad von Formen der Kindesmisshandlung sowie Einstufungsbögen zum Schweregrad von Risikofaktoren, die ebenfalls in dem Testmanual enthalten sind, gebeten. In dem Einstufungsbogen zur Erfassung des Schweregrades von Formen der Kindesmisshandlung wird erfasst, welche Misshandlungsform wie stark vorliegt. In dem Einstufungsbogen zum Schweregrad von Risikofaktoren wird erfasst, welche Risikofaktoren im Umfeld des misshandelnden Elternteils mit welchem Schweregrad vorliegen. Eine Abschätzung des voraussichtlichen Rücklaufs ist zu diesem Zeitpunkt noch nicht möglich, wäre jedoch aufgrund der Möglichkeit zur Verbesserung des Verfahrens hinsichtlich der angesprochen Punkte als sehr wünschenswert zu betrachten.

2.3 Berücksichtigung längsschnittlicher Aspekte

Die dargestellten Befunde aus der Untersuchung zur Validierung des EBSK beziehen sich lediglich auf Daten, die zu einem Messzeitpunkt erhoben wurden. Somit sind zum aktuellen Stand sowohl im Hinblick auf die Reliabilität als auch im Hinblick auf die Validität des Verfahrens ausschließlich querschnittliche Aussagen zulässig. Für die Weiterentwicklung des Verfahrens ist daher die Erweiterung der vorgenommenen Betrachtungen um eine längsschnittliche Perspektive von hoher Bedeutsamkeit. Hinsichtlich der Reliabilität der Validitätsskalen sind zunächst Befunde zur Retest-Reliabilität interessant. Wichtig sind diese vor allem, da aufgrund der beschriebenen Probleme hinsichtlich der Bestimmung der internen Konsistenz dieser Skalen (Kapitel C, Abschnitt 3.2 und 3.3) eine sinnvolle Abschätzung der Reliabilität erst im

Hinblick auf die zeitliche Stabilität geleistet werden kann. Weiterhin können auf der Basis längsschnittlicher Daten ebenfalls auch Aussagen zur zeitlichen Stabilität der Belastungsskala getroffen werden. Die Bedeutung dieser Befunde ist jedoch nur dann als hoch einzuschätzen, wenn in dem Zeitraum zwischen Test und Retest keine Merkmalsveränderungen, d.h. Veränderungen in dem Belastungsgrad der Probanden, beispielsweise durch therapeutische Interventionen, entstanden sind. Die Unterlassung einer Intervention bei der Feststellung eines erhöhten Belastungsgrades ist jedoch ethisch nicht verantwortbar, weshalb lediglich durch Artefakte der Intervention verunreinigte Zusammenhänge zwischen den Belastungswerten zu den beiden Messzeitpunkten festgestellt werden können. Da bei der Ermittlung der Retest-Reliabilität der Validitätsskalen der Einsatz des gesamten EBSK-Fragebogens zweckmäßig ist, können Befunde zur zeitlichen Reliabilität der Belastungsskala mit einem geringen zusätzlichen Erhebungsaufwand erfasst werden und sollten daher, wenn auch nur vorsichtig, auf jeden Fall betrachtet und interpretiert werden.

Die Akquise einer längsschnittlichen Stichprobe erlaubt ebenfalls Aussagen zur prädiktiven Validität des EBSK, die in der bisherigen Validierungsuntersuchung nicht geleistet werden konnten. Hierbei sind jedoch wiederum ethische Beschränkungen zu berücksichtigen. Die Betrachtung der Fragestellung, ob ein Elternteil nachdem ein erhöhter Belastungsgrad bei ihm festgestellt wird, sein Kind misshandelt, sofern keine Intervention bei ihm durchgeführt wird, ist hierbei als nicht verantwortbar auszuschließen. Einen Ausweg aus dieser Problematik ist in der Betrachtung von belasteten Probanden zu sehen, die im Rahmen von Therapiemaßnahmen behandelt werden. Hierbei könnte festgestellt werden, ob Eltern, die zu Beginn der Therapie besonders hohe Belastungswerte aufweisen, in der Folge häufiger misshandeln als Probanden, bei denen sich ein geringerer Belastungsgrad zeigt. Die Einschätzungen der prädiktiven Validität der Ergebnisse des EBSK sind damit zwar als verunreinigt durch den Einfluss der therapeutischen Maßnahme zu betrachten, in diesem Zusammenhang jedoch die einzige Möglichkeit zur Abschätzung dieses Aspektes der Validität.

3. Abschließende Bemerkung

Abschließend lässt sich festhalten, dass mit dem EBSK in Deutschland erstmals ein empirisch gut validiertes Fragebogenverfahren zum Screening bei Kindesmisshandlungsfällen vorliegt, das die in der Fachliteratur geforderte Unterstützung der Kindesmisshandlungsarbeit durch geeignete psychodiagnostische Instrumente realisiert (Deegener & Körner, 2006). Obwohl der Einsatz des Verfahrens in keinsten Weise als ein Ersatz für die derzeitig stattfindende Misshandlungsarbeit darstellt, kann er doch als eine sinnvolle Ergänzung derselben angesehen werden. Diese bezieht sich auf eine Vielfalt möglicher praktische Einsatzbereiche in präventiven, anamnestischen, therapeutischen und diagnostisch-evaluativen Kontexten. Das Verfahren richtet sich damit an eine breite Masse potentieller Anwender, unter anderem an klinische und forensische Psychologen, Allgemeinmediziner, Internisten, Pädiater und Psychiater sowie an Sozialarbeiter und Mitarbeiter von Jugendberatungsstellen. Ebenfalls ist eine Verwendung des EBSK in der Forschung denkbar, beispielsweise zur weiteren Untersuchung von Bedingungsfaktoren für die Entstehung von Kindesmisshandlungsfällen. Aufgrund der dargestellten Vielfalt sinnvoller Einsatzbereiche bleibt zu hoffen, dass diese Möglichkeiten in der Zukunft auch genutzt werden. Obwohl derzeit keine diesbezügliche Einschätzung getroffen werden kann, bleibt festzuhalten, dass die Validierungsuntersuchung sowie die Erstellung und Publikation des Manuals einen ersten Schritt in diese Richtung darstellen.

F. Literatur

1. Abidin RR (1983) Parenting Stress Index. Pediatric Psychology Press, Charlottesville
2. Amelang M, Schmidt-Atzert L (2006) Psychologische Diagnostik und Intervention. Springer, Heidelberg
3. Araji S, Finkelhor D (1986) Abusers: A review of research. In: Finkelhor D (ed) A sourcebook on child sexual abuse. Sage, Beverly Hills, pp 89-118
4. Ayoub C, Pan BA, Rocha V, Emerson K (2000) Initial characteristics and service needs of rural families with young children: An assessment of risk and resilience. Poster auf der Fifth National Head Start Research Conference, Washington
5. Backhaus K, Erichson B, Plinke W, Weiber R (2006) Multivariate Analysemethoden. Springer, Heidelberg
6. Bange D (2005) Gewalt gegen Kinder in der Geschichte. In: Deegener G, Körner W (eds) Kindesmisshandlung und Vernachlässigung. Hogrefe, Göttingen, pp 13-18
7. Bange D, Deegener G (1996) Sexueller Missbrauch an Kindern. Beltz, Weinheim
8. Barnett D, Manley JT, Cicchetti D (1993) Defining child maltreatment. In: Cicchetti D Toth SL (eds): Child abuse, child development and social policy. Ablex, Norwood, pp 7-74
9. Belsky J (1980) Child maltreatment: An ecological integration. Am Psychol 75: 320-335
10. Belsky J (1993) Etiology of child maltreatment: A developmental-ecological analysis. Psychol Bull 114: 413-434

11. Bender D, Lösel F (1998) Protektive Faktoren der psychisch gesunden Entwicklung junger Menschen. In: Margraf J, Siegrist J, Neumer S (eds) Gesundheits- oder Krankheitstheorie. Springer, Heidelberg, pp 117-145
12. Bender D, Lösel F (2005a) Misshandlung von Kindern: Risikofaktoren und Schutzfaktoren. In: Deegener G, Körner W (eds) Kindesmisshandlung und Vernachlässigung. Hogrefe, Göttingen, pp 317-346
13. Bender D, Lösel F (2005b) Risikofaktoren, Schutzfaktoren und Resilienz bei Misshandlung und Vernachlässigung. In: Egle UT, Hoffmann SO, Joraschky P (eds) Sexueller Missbrauch, Misshandlung, Vernachlässigung. Schattauer, Stuttgart, pp 85-104
14. Black DA, Heyman RE, Slep AMS (2001) Risk factors for physical child abuse. *Aggression and Violent Behavior* 6: 121-188
15. Black DM, Slep AMS, Heyman RE (2001) Risk factors for child psychological abuse. *Aggression and Violent Behavior* 6: 189-201
16. Blum-Maurice R, Knoller EC, Nitsch M, Kröhnert A (2000) Qualitätsstandards für die Arbeit eines Kinderschutz-Zentrums. Eigenverlag der Bundesarbeitsgemeinschaft der Kinderschutz-Zentren e.V., Köln
17. Bronfenbrenner U (1981) Die Ökologie der menschlichen Entwicklung. Klett, Stuttgart
18. Brown J, Daniels R (1968) Some observations on abusive parents. *Child Welfare* 47: 89-94
19. Buchner G, Cizek B, Gössweiner V, Kapella O, Pfliegerl J, Steck M (2001) Gewalt gegen Kinder. In: Bundesministerium für Soziale Sicherheit und Generationen (ed) Gewalt in der Familie. Bundesministerium für Soziale Sicherheit und Generationen, Wien, pp 75-259

20. Burrell B, Thompson B, Sexton D (1992) The measurement integrity of data collected using the Child Abuse Potential Inventory. *Educational and Psychological Measurement* 52: 993-1001
21. Caliso JA, Milner JS (1992) Childhood history of abuse and child abuse screening. *Child Abuse and Negl* 16: 647-659
22. CAPTA, Child Abuse Prevention and Treatment Act, 42 U.S.C.A. §5106g
23. Cattell RB (1966) The scree test for the number of factors. *Multivariate behavioral Research* 1: 245-276
24. Cicchetti D, Lynch M (1993) Toward an ecological/transactional model of community violence and child maltreatment: Consequences for children's development. *Psychiatry* 53: 96-118
25. Cicchetti D, Rizley R (1981) Developmental perspectives on the etiology, intergenerational transmission and sequelae of child maltreatment. *New Dir Child Dev* 11: 31-55
26. Cicchetti D, Toth SL, Maughan A (2000). An ecological-transactional model of child maltreatment. In: Sameroff AJ, Lewis M, Miller SM (eds) *Handbook of Developmental Psychopathology*. Kluwer, New York, pp 689-722
27. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. Erlbaum, Hillsdale
28. Cohen J (1992) A power primer. *Psych Bull* 112: 115-119
29. Couron BL (1982) Assessing parental potentials for child abuse in contrast to nurturing. *Dissertation Abstracts International* 43: 3412B
30. Deegener G (o. J.) Deutsche Übersetzung des CAP-Inventory VI. Unveröffentlichter Fragebogen, Universität des Saarlandes
31. Deegener G (2005) Formen und Häufigkeiten der Kindesmisshandlung. In: Deegener G, Körner W (eds) *Kindesmisshandlung und Vernachlässigung*. Hogrefe, Göttingen, pp 37-58

32. Deegener G, Körner W (2006) Risikoerfassung bei Kindesmisshandlung und Vernachlässigung. Pabst, Lengerich
33. Deegener G, Spangler G, Körner W, Becker N (im Druck) Elternbelastungsscreening zur Kindeswohlgefährdung (EBSK). Hogrefe, Göttingen
34. Emery RE, Laumann-Billings L (1998) An Overview of the Nature, Causes and Consequences of Abusive Family Relationships. *Am Psychol* 53: 121-135
35. Engfer A (1986) Kindesmisshandlung. Enke, Stuttgart
36. Engfer A (1991) Prospective identification of violent mother-child relationships. In: Kaiser G, Kury H, Albrecht HJ (eds) *Victims of criminal justice*. Max-Planck-Institut für Ausländisches und Internationales Strafrecht, Freiburg, pp 418-458
37. Engfer A (1997) Gewalt gegen Kinder in der Familie. In: Egle UT, Hoffmann SO, Joraschky P (eds) *Sexueller Missbrauch, Misshandlung, Vernachlässigung*. Schattauer, Stuttgart, pp 21-34
38. Engfer A (2000) Gewalt gegen Kinder in der Familie. In: Egle UT, Hoffmann SO, Joraschky P (eds) *Sexueller Missbrauch, Misshandlung, Vernachlässigung*. Schattauer, Stuttgart, pp 23-39
39. Engfer A (2005) Formen der Misshandlung von Kindern - Definitionen, Häufigkeiten und Erklärungsansätze. In: Egle UT, Hoffmann SO, Joraschky P (eds) *Sexueller Missbrauch, Misshandlung, Vernachlässigung*. Schattauer, Stuttgart, pp 3-19
40. Fahrenberg J, Hampel R, Selg H (1994) *Das Freiburger Persönlichkeitsinventar*. Hogrefe, Göttingen
41. Frank R (1993) *Kinderärztlich-kinderpsychiatrische Untersuchungen an misshandelten und vernachlässigten Kindern und deren Familien*. Habilitationsschrift, Universität München

42. Frank R, Kopecky-Wenzel M (2002) Vernachlässigung von Kindern. Monatsschrift Kinderheilkunde 150: 1339-1343
43. Fricke R (1974) Kriteriumsorientierte Leistungsmessung. Kohlhammer, Stuttgart
44. Fürniss T (2005) Geschichtlicher Abriss zur Kindesmisshandlung und Kinderschutzarbeit von C.H. Kempe bis heute. In: Deegener G, Körner W (eds) Kindesmisshandlung und Vernachlässigung. Hogrefe, Göttingen, pp 19-33
45. Gelles RJ, Straus MA (1979) Determinants of violence in the family. In: Burr WR, Hill R, Nye FI, Reiss IL (eds) Contemporary theories about the family. Free Press, New York, pp 35-47
46. Globisch J, Spangler G (o. J.) Deutsche Version des CAPI Form VI. Unveröffentlichter Fragebogen, Universität Erlangen
47. Göppinger H (1997) Kriminologie. Beck, München
48. Guttman L (1954) Some necessary conditions for common factor analysis. *Psychometrika* 19: 149-161
49. Häusler J (2003) MixFactor 1.3.75. Institut für Psychologie an der Universität Wien, Abteilung für Methodik, Wien
50. Häusler J, Kubinger KD (2004) Zur artifiziellen Anwendung der Faktorenanalyse auf dichotome Daten: Empirische Beispiele. Referat Österreichische Gesellschaft für Psychologie
51. Hedges LV, Olkin I (1985) Statistical methods for meta-analysis. Academic Press, San Diego
52. Helfer RE, Hoffmeister JK, Schneider CJ (1978) MSSP: A manual for the use of the Michigan Screening Profile of Parenting. Express Press, Boulder
53. Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika* 30: 179-185

54. Johnson BT, Mullen B, Salas E (1995) Comparison of Three Meta-Analytic Approaches. *J Appl Psychol* 80: 94-106
55. Kaiser HF, Dickman K (1959) Analytic determination of common factors. *American psychological Reports* 14: 425-435
56. Kempe CH (1973) A practical approach to the protection of the abused child and rehabilitation of the abusing parent. *Pediatrics* 51: 804-809
57. Kempe CH, Silverman FN, Steele BF, Droegenmuller W, Silver HK (1962) The Battered Child Syndrome. *JAMA* 181: 17-24
58. Kubinger KD (2003) On artificial results due to using factor analysis on dichotomous variables. *Psychology Science* 45: 106-110
59. Kuder GF, Richardson W (1937) The theory of the estimation of test reliability. *Psychometrika* 2: 151-160
60. Lamphear VS, Stets JP, Whitaker P, Ross AO (1985) Maladjustment in at-risk for physical child abuse and behavior problem children: Differences in family environment and marital discord. Paper auf dem Meeting of the American Psychological Association, Los Angeles.
61. Lienert GA (1989) Testaufbau und Testanalyse. Psychologie Verlags Union, München
62. Lösel F, Bliesener T (2003) Aggression und Delinquenz unter Jugendlichen: Untersuchungen von kognitiven und sozialen Bedingungen. Luchterhand, München
63. Masten AS, Best KM, Garmezy N (1990) Resilience and development: Contributions from the study of children who overcome adversity. *Dev Psychopathol* 2: 425-444
64. Matthews RD (1984) Screening and identification of child abusing parents through self-report inventories. Dissertationsschrift am Florida Institute of Technology, Melbourne

65. Milner JS (1986) The Child Abuse Potential Inventory. Psytech, DeKalb
66. Milner JS (1989) Applications of the Child Abuse Potential Inventory. *J Clin Psychol* 3: 450-454
67. Milner JS (1991) Medical conditions and the Child Abuse Potential Inventory specificity. *J Consul Clin Psychol* 3: 208-212
68. Milner JS (1994) Assessing Physical Child Abuse Risk: The Child Abuse Potential Inventory. *Clin Psychol Rev* 14: 547-583
69. Milner JS (1998) Assessment Issues in Child Abuse Evaluations. In: Lutzker JR (ed) *Handbook of Child Abuse Research and Treatment*. Plenum Press, New York, pp 75-117
70. Milner JS, Gold RG, Ayoub C, Jacewitz MM (1984) Predictive validity of the Child Abuse Potential Inventory. *J Consul Clin Psychol* 52: 879-884
71. Milner JS, Gold RG, Wimberley RC (1986) Prediction and explanation of child abuse: Cross-validation of the Child Abuse Potential Inventory. *J Consul Clin Psychol* 54: 865-866
72. Milner JS, Robertson KR (1989) Inconstant response patterns and the prediction of child maltreatment. *Child Abuse Negl* 13: 59-64
73. Milner JS, Robertson KR (1990) Comparison of physical child abusers, intrafamilial sexual child abusers and child neglecters. *Journal of Interpersonal Violence* 5: 37-48
74. Milner JS, Williams PP (1978) Child abuse and neglect: A bibliography. *Journal Supplemental Abstract Services* 8: Ms. No. 1690
75. Milner JS, Wimberley RC (1980) Prediction and explanation of child abuse. *J Clin Psychol* 36: 875-884
76. Ondersma SJ, Chaffin MJ, Mullins SM, LeBreton JM (2005) A Brief Form of the Child Abuse Potential Inventory. *Journal of Child and Adolescent Psychology* 34: 301-311

77. Pfeiffer C, Wetzels P (1997) Kinder als Täter und Opfer. Eine Analyse auf Basis der PKS und einer repräsentativen Opferbefragung (Forschungsbericht Nr. 68). Kriminologisches Forschungsinstitut Niedersachsen, Hannover
78. Pfeiffer C, Wetzels P, Enzmann D (1999) Innerfamiliäre Gewalt gegen Kinder und Jugendliche und ihre Auswirkungen (Forschungsbericht Nr. 80). Kriminologisches Forschungsinstitut Niedersachsen, Hannover
79. Raupp U, Eggerts C (1993) Sexueller Missbrauch von Kindern. Eine regionale Studie über Prävalenz und Charakteristik. Monatsschrift für Kinderheilkunde 141: 316-322
80. Ruback BR (1993) The victim-offender relationship does affect victims decision to report sexual assaults. Criminal Justice and Behavior 20: 271-279
81. Sameroff AJ, Chandler MJ (1975) Reproductive risk and the continuum of caretaking causality. In: Horowitz FD, Hetherington M, Scarr-Salapatek S, Siegel G (eds) Review of child development research. University Press, Chicago, pp 187-243
82. Schreiber LH (1971) Misshandlung von Kindern und alten Menschen. Kriminalistik-Verlag, Hamburg
83. Schumacher JA, Slep AMS, Heyman RE (2001) Risk factors for child neglect. Aggression and Violent Behavior 6: 231-254
84. Seiwald BB (2003) Lügenskala. In: Kubinger KD, Jäger RS (eds) Schlüsselbegriffe der psychologischen Diagnostik. Beltz, Weinheim, pp 271-273
85. Stith SM, Davies LC, Boykin EL, Alder MC, McPherson M, Dees JE, Harris JM, Som A (2004) Risk Factors in Child Maltreatment: A Meta-analytic Review of the Literature. Unveröffentlichtes Manuskript, Virginia Polytechnic Institute and State University, Blacksburg

86. Stith SM, Liu T, Davies C, Boykin EL, Alder MC, Harris JM, Som A, McPherson M, Dees JE (im Druck) Risk Factors in Child Maltreatment: A Meta-analytic Review of the Literature. *Journal of Aggression, Maltreatment and Trauma*
87. Straus MA (1994) *Beating the Devil Out of Them: Corporal Punishment in American Families*. Lexington, New York
88. Thomas JR, French KE (1986) The Use of Meta-Analysis in Exercise and Sport: A Tutorial. *Research Quarterly for Exercise and Sport* 57: 196-204
89. Überla K (1971) *Faktorenanalyse: Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler*. Springer, Heidelberg
90. USDHHS, US Department of Health and Human Services, Administration on Children, Youth and Families (2007) *Child Maltreatment 2005*. US Government Printing Office, Washington
91. Wetzels P, Pfeiffer C (1997) *Kindheit und Gewalt: Täter- und Opferperspektiven aus Sicht der Kriminologie*. *Prax Kinderpsychol Kinderpsychiatr* 46: 143-152
92. Ziegler F (1994) *Kinder als Opfer von Gewalt: Ursachen und Interventionsmöglichkeiten*. Huber, Bern

G. Anhang**1. Tabellen aus der Metaanalyse von Stith et al. (2004)****Tabelle G.1:** Risikofaktoren für körperliche Kindesmisshandlung

Risikofaktor	d	CI	Q ^w	k	N
(M) Child Attachment to caregiver	-1,21	-1,55/-0,87	13,64**	4	117
(O) Parent Empathy	-0,94	-1,17/-0,71	5,96	6	355
(O) Parenting Skills	-0,78	-1,17/-0,39	0,45	4	108
(O) Parent Anger	0,78	0,35/1,20	12,82**	3	128
(M) Family Conflict	0,73	0,40/1,07	8,12*	4	154
(M) Lack of Family Cohesion	-0,68	-0,98/-0,38	3,02	5	183
(O) Parent Hyper-reactivity	0,65	0,36/0,93	1,84	5	214
(M) Parent perceives Problem Child	0,62	0,53/0,71	57,45***	24	3297
(O) Parent Anxiety	0,60	0,41/0,79	4,39	8	563
(E) Parent received Harsh Discipline	0,59	0,41/0,77	40,92***	5	552
(O) Parent Psychopathology	0,59	0,48/0,69	62,21***	13	8630
(O) Child Aggression	0,59	0,45/0,73	43,12***	13	849
(O) Parent Depression	0,55	0,45/0,67	46,18***	14	8258
(O) Parent Self-esteem	-0,53	-0,67/-0,39	25,20**	10	2453
(O) Child Social Competence	-0,53	-0,64/-0,42	27,46*	14	1527
(M) Parent Use of Corporal Punishment	0,55	0,38/0,72	4,65	7	703
(M) Parent Unrealistic Expectations	0,49	0,35/0,62	42,12***	10	2325
(O) Stress over Parenting	0,46	0,26/0,67	10,89*	5	379
(M) Parent-child Relationship	-0,47	-0,58/-0,36	88,37***	30	1554
(M) Spousal Violence	0,46	0,31/0,61	6,82	5	773
(E) Parent Poor Relationship	0,44	0,34/0,54	20,38*	11	2997
(O) Child Externalizing Behavior	0,42	0,33/0,52	68,51***	16	2022
(M) Authoritarian Parenting Style	0,42	0,27/0,57	33,37***	13	767
(E) Parent Stress	0,42	0,32/0,53	47,78***	20	1674
(O) Self-evaluation as Parent	-0,40	-0,64/-0,16	9,23	5	296
(E) Parent Financial Stress	0,41	0,20/0,62	4,72	4	397

Tabelle G.1 (Fortsetzung)

Risikofaktor	d	CI	Q ^w	k	N
(O) Child Non-Compliant/Difficult	0,41	0,24/0,59	34,69***	9	512
(E) Parent Criminal Record	0,40	0,20/0,59	0,34	3	563
(E) Parent Experienced Abuse	0,38	0,26/0,49	33,94***	10	3170
(E) Parent Social Support	-0,36	-0,45/-0,28	65,32***	20	10315
(E) Socio-economic Status	-0,37	-0,54/-0,21	22,04***	5	7423
(M) Parent Alcohol Abuse	0,34	0,19/0,50	8,06*	3	654
(M) Marital Satisfaction	-0,32	-0,47/-0,18	14,45*	8	840
(M) Family Size	0,31	0,24/0,38	65,53***	23	11224
(O) Child Internalizing Behavior	0,31	0,22/0,40	50,62***	23	2282
(E) Unemployment	0,30	0,19/0,42	29,57***	8	1263
(O) Parent Coping Skills	-0,28	-0,54/-0,03	7,45*	3	285
(E) Income	-0,26	-0,36/-0,17	22,55*	11	2920
(M) Single Parent	0,24	0,19/0,30	108,23***	22	14223
(M) Parent Hostile Attributions	0,22	-0,04/0,48	2,74	3	241
(O) Parent Health Problems	0,22	-0,01/0,45	3,17	3	286
(O) Parent Age	-0,21	-0,27/-0,14	215,42***	30	12118
(O) Parent Drug Abuse	0,16	0,01/0,32	2,18	3	654
(O) Parent Education Level	-0,06	-0,14/0,02	28,28**	12	11621
(M) Approval of Corporal Punishment	0,09	-0,11/0,30	5,65	5	1674
(O) Child Pre- or Neo-natal Problems	0,08	-0,03/0,19	15,34	10	1432
(O) Child Gender	0,08	-0,04/0,19	6,10	13	1702
(M) Non-biological Parent	-0,05	-0,28/0,17	3,25	3	302
(O) Child Age	-0,05	-0,14/0,04	12,63	14	3332
(O) Child Disability	0,02	-0,20/0,24	0,80	4	325

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

Tabelle G.2: Risikofaktoren für Kindesvernachlässigung

Risikofaktor	d	CI	Q ^w	k	N
(M) Parent Perceives Problem Child	1,06	0,54/1,58	3,61	3	67
(M) Parent-child Relationship	-0,91	-1,13/-0,69	47,20***	10	380
(O) Parenting Skills	-0,82	-1,23/-0,41	0,71	3	98
(E) Parent Stress	0,81	0,60/1,02	24,94***	3	386
(O) Parent Self-esteem	-0,69	-0,99/-0,38	1,98	4	184
(O) Child Social Competence	-0,62	-0,81/-0,43	10,03	7	584
(O) Child Aggression	0,59	0,38/0,81	20,08**	8	374
(O) Child Non-Compliant/Difficult	0,59	0,26/0,93	7,24	4	149
(M) Family Size	0,54	0,45/0,64	75,19***	12	8546
(E) Unemployment	0,51	0,35/0,67	8,69*	4	719
(O) Parent Psychopathology	0,52	0,39/0,66	14,45*	8	7652
(O) Parent Depression	0,42	0,32/0,53	15,93*	8	8207
(E) Parent Education Level	-0,41	-0,54/-0,28	44,83***	8	7743
(E) Income	-0,38	-0,52/-0,24	31,50***	7	861
(E) Parent Poor Relationship	0,39	0,25/0,54	15,40*	7	855
(M) Parent Unrealistic Expectations	0,39	0,25/0,53	14,95*	6	878
(O) Child Externalizing Behavior	0,35	0,18/0,52	9,07	9	598
(E) Parent Social Support	-0,33	-0,43/-0,24	55,11***	13	8582
(E) Parent Experienced Abuse	0,31	0,17/0,45	25,21***	6	1417
(O) Parent Age	-0,25	-0,36/-0,13	19,94*	9	8120
(O) Child Internalizing Behavior	0,22	0,08/0,36	48,22***	11	922
(M) Single Parent	0,16	0,03/0,28	13,41	9	7751
(O) Child Age	-0,02	-0,23/0,19	13,49	8	369
(O) Child Gender	0,03	-0,17/0,22	0,25	5	961

Anmerkung: Die Angaben in der Tabelle sind der Metaanalyse von Stith et al. (2004) entnommen. d=mittlere korrigierte Effektstärke; CI=95%-Konfidenzintervall des wahren Effekts; Q^w=Homogenitätsmaß der Effektstärkenverteilung nach Hedges und Olkin (1985); k=Anzahl der berücksichtigten Studien; N=Anzahl der Probanden in den berücksichtigten Studien; * p<0,05; ** p<0,01; *** p<0,001

2. Items und Skalen der Übersetzung des CAPI

2.1 Wortlaut der Items nach Globisch und Spangler (o.J.)

CAPI1. Andere tun mir nie leid.

CAPI2. Ich habe gerne Haustiere.

CAPI3. Ich bin immer stark und gesund gewesen.

CAPI4. Ich mag die meisten Menschen.

CAPI5. Ich bin ein verwirrter Mensch.

CAPI6. Den meisten Menschen traue ich nicht.

CAPI7. Man erwartet zu viel von mir.

CAPI8. Kinder sollten nie böse sein.

CAPI9. Ich bin häufig durcheinander.

CAPI10. Schläge, die bei einem Kind nur blaue Flecken verursachen, sind okay.

CAPI11. Ich versuche immer nach meinem Kind zu sehen, wenn es schreit.

CAPI12. Ich handle manchmal ohne zu denken.

CAPI13. Man kann sich nicht auf andere verlassen.

CAPI14. Ich bin ein glücklicher Mensch.

CAPI15. Ich mache gerne etwas zusammen mit meiner Familie.

CAPI16. Teenager-Mädchen müssen geschützt werden.

CAPI17. Ich bin oft innerlich ärgerlich.

CAPI18. Manchmal fühle ich mich ganz allein in dieser Welt.

CAPI19. In einem Zuhause sollte immer alles an seinem Platz sein.

CAPI20. Manchmal mache ich mir Sorgen, dass ich nicht alle Bedürfnisse eines Kindes befriedigen kann.

CAPI21. Messer sind gefährlich für Kinder.

CAPI22. Ich fühle mich häufig abgelehnt.

CAPI23. Ich bin oft innerlich einsam.

CAPI24. Kleine Jungs sollten niemals Mädchenspiele lernen.

CAPI25. Ich fühle mich häufig sehr frustriert.

- CAP126. Kinder sollten immer gehorchen.
- CAP127. Ich mag alle Kinder.
- CAP128. Manchmal habe ich Angst, dass ich die Kontrolle über mich verliere.
- CAP129. Manchmal wünsche ich mir, dass mein Vater mich mehr geliebt hätte.
- CAP130. Ich habe ein Kind, das ungeschickt ist.
- CAP131. Ich weiß, was die richtige und falsche Art ist zu handeln.
- CAP132. Meine Telefonnummer ist nirgendwo aufgeführt.
- CAP133. Die Geburt eines Kindes führt normalerweise zu Problemen in einer Ehe.
- CAP134. Ich bin immer ein guter Mensch.
- CAP135. Ich mache mir nie Sorgen um meine Gesundheit.
- CAP136. Ich mache mir manchmal Sorgen, dass ich nicht genug zu essen haben werde.
- CAP137. Ich wollte nie jemand anderem wehtun.
- CAP138. Ich habe immer Pech.
- CAP139. Ich bin normalerweise ein stiller Mensch.
- CAP140. Kinder sind wie die Pest.
- CAP141. Die Dinge in meinem Leben gingen normalerweise gegen mich.
- CAP142. Ein Baby jedes Mal hochzunehmen wenn es schreit, verwöhnt es.
- CAP143. Ich bin manchmal sehr still.
- CAP144. Ich verliere manchmal die Geduld.
- CAP145. Ich habe ein Kind, das böse ist.
- CAP146. Manchmal denke ich zuerst an mich selbst.
- CAP147. Ich fühle mich manchmal wertlos.
- CAP148. Meine Eltern haben sich nicht wirklich um mich gekümmert.
- CAP149. Ich bin manchmal sehr traurig.
- CAP150. Kinder sind wirklich kleine Erwachsene.
- CAP151. Ich habe ein Kind, das Sachen kaputt macht.
- CAP152. Ich mache mir oft Sorgen.
- CAP153. Es ist in Ordnung, ein Kind für eine Weile in schmutzigen Windeln zu lassen.
- CAP154. Ein Kind sollte nie widersprechen.
- CAP155. Manchmal ist mein Verhalten kindisch.

- CAPI56. Ich rege mich oft schnell auf.
- CAPI57. Manchmal habe ich schlechte Gedanken.
- CAPI58. Jeder muss zuerst an sich selbst denken.
- CAPI59. Ein Kind das schreit wird niemals glücklich sein.
- CAPI60. Ich habe nie eine andere Person gehasst.
- CAPI61. Kinder sollten nicht Schwimmen lernen.
- CAPI62. Ich tue immer das, was richtig ist.
- CAPI63. Ich bin oft beunruhigt.
- CAPI64. Ich habe ein Kind, das oft krank ist.
- CAPI65. Manchmal mag ich die Art, in der ich handle, nicht.
- CAPI66. Manchmal gelingt es mir nicht, alle meine Versprechen zu halten.
- CAPI67. Man hat mir viel Schmerz zugefügt.
- CAPI68. Kinder sollten sauber bleiben.
- CAPI69. Ich habe ein Kind, das oft in Schwierigkeiten gerät.
- CAPI70. Ich werde nie wütend auf andere.
- CAPI71. Mit anderen komme ich immer zurecht.
- CAPI72. Ich denke oft darüber nach, was ich zu tun habe.
- CAPI73. Ich finde es schwer, mich zu entspannen.
- CAPI74. Heutzutage weiß man nicht so recht, auf wen man zählen kann.
- CAPI75. Mein Leben ist glücklich.
- CAPI76. Ich habe eine körperliche Behinderung.
- CAPI77. Kinder sollten Spielkleidung und gute Kleidung haben.
- CAPI78. Andere Leute verstehen nicht, wie ich mich fühle.
- CAPI79. Ein Fünfjähriger, der sein Bett einnässt, ist schlecht.
- CAPI80. Kinder sollten still sein und zuhören.
- CAPI81. Ich habe einige enge Freunde in meiner Nachbarschaft.
- CAPI82. Die Schule ist ein erster Linie für die Erziehung des Kindes verantwortlich.
- CAPI83. Meine Familie streite viel.
- CAPI84. Ich habe Kopfschmerzen.
- CAPI85. Als Kind wurde ich misshandelt.

- CAP186. Schläge sind die beste Bestrafung.
- CAP187. Ich mag es nicht von anderen berührt zu werden.
- CAP188. Menschen, die um Hilfe bitten, sind schwach.
- CAP189. Kinder sollten gewaschen werden, bevor sie ins Bett gehen.
- CAP190. Ich lache nicht sehr viel.
- CAP191. Ich habe einige enge Freunde.
- CAP192. Man sollte sich um seine eigenen Bedürfnisse kümmern.
- CAP193. Ich habe Ängste, von denen niemand etwas weiß.
- CAP194. Meine Familie hat Probleme zurechtzukommen.
- CAP195. Das Leben erscheint mir oft nutzlos.
- CAP196. Ein Kind sollte mit einem Jahr sauber sein.
- CAP197. Ein Kind in einer Schlammputze ist ein guter Anblick.
- CAP198. Die Leute verstehen mich nicht.
- CAP199. Ich fühle mich oft wertlos.
- CAP100. Andere Leute haben mein Leben unglücklich gemacht.
- CAP101. Ich bin immer ein freundlicher Mensch.
- CAP102. Manchmal weiß ich nicht, warum ich so handele, wie ich es tue.
- CAP103. Ich habe viele persönliche Probleme.
- CAP104. Ich habe ein Kind, das sich oft verletzt.
- CAP105. Ich fühle mich häufig aus der Fassung gebracht.
- CAP106. Die Leute nutzen mich manchmal aus.
- CAP107. Mein Leben ist gut.
- CAP108. Ein zuhause sollte makellos sein.
- CAP109. Ich lasse mich durch meine Probleme schnell aus der Fassung bringen.
- CAP110. Ich höre nie auf Klatsch.
- CAP111. Meine Eltern haben mich nicht verstanden.
- CAP112. Viele Dinge im Leben machen mich ärgerlich.
- CAP113. Mein Kind hat besondere Probleme.
- CAP114. Ich mag die meisten Kinder nicht.
- CAP115. Kinder sollte man sehen und nicht hören.

- CAP116. Die meisten Kinder sind gleich.
- CAP117. Es ist wichtig für Kinder zu lesen.
- CAP118. Ich bin häufig niedergeschlagen.
- CAP119. Kinder sollten gelegentlich an ihre Eltern denken.
- CAP120. Ich rege mich oft auf.
- CAP121. Die Leute kommen mit mir nicht zurecht.
- CAP122. Ein gutes Kind hält seine Spielsachen und seine Kleider sauber und ordentlich.
- CAP123. Kinder sollten ihre Eltern glücklich machen.
- CAP124. Es ist unnatürlich für ein Kind, manchmal zu widersprechen.
- CAP125. Ich bin nie unfair zu anderen.
- CAP126. Gelegentlich genieße ich es, mich um mein Kind nicht kümmern zu müssen.
- CAP127. Kinder sollten immer adrett sein.
- CAP128. Ich habe ein Kind, das langsam ist.
- CAP129. Eltern müssen Strafen benutzen, wenn sie das Verhalten eines Kindes kontrollieren wollen.
- CAP130. Kinder sollten nie Schwierigkeiten bereiten.
- CAP131. Normalerweise bestrafe ich mein Kind, wenn es schreit.
- CAP132. Ein Kind braucht sehr strenge Regeln.
- CAP133. Kinder sollten sich nie den Anordnungen ihrer Eltern widersetzen.
- CAP134. Ich fühle mich oft besser als andere.
- CAP135. Kinder gehen mir manchmal auf die Nerven.
- CAP136. Als Kind hatte ich oft Angst.
- CAP137. Kinder sollten immer still und höflich sein.
- CAP138. Ich rege mich oft auf und weiß nicht, warum.
- CAP139. Meine tägliche Arbeit regt mich oft auf.
- CAP140. Manchmal habe ich Angst, dass meine Kinder mich nicht lieben werden.
- CAP141. Ich habe ein gutes Sexualleben.
- CAP142. Ich habe Artikel und Bücher über Kindererziehung gelesen.
- CAP143. Ich fühle mich häufig sehr allein.
- CAP144. Man sollte Ärger nicht zeigen.
- CAP145. Ich fühle mich häufig allein.

-
- CAPI146. Ich sage manchmal schlimme Wörter.
- CAPI147. Im Moment bin ich total verliebt.
- CAPI148. Meine Familie hat viele Probleme.
- CAPI149. Ich tue nie etwas, was schlecht für meine Gesundheit ist.
- CAPI150. Ich bin immer glücklich mit dem, was ich habe.
- CAPI151. Andere Leute haben mein Leben schwer gemacht.
- CAPI152. Ich lache fast jeden Tag.
- CAPI153. Ich mache mir manchmal Sorgen, dass meine Bedürfnisse nicht befriedigt werden.
- CAPI154. Ich fühle mich oft ängstlich.
- CAPI155. Ich handele manchmal dumm.
- CAPI156. Man sollte seine Angelegenheiten für sich behalten.
- CAPI157. Ich erhebe nie meine Stimme im Ärger.
- CAPI158. Als Kind wurde ich von meinen Eltern mies behandelt.
- CAPI159. Manchmal denke ich zuerst an mich und dann an andere.
- CAPI160. Ich sage immer die Wahrheit.

2.2 Zugehörigkeit der Items zu den Skalen

Tabelle G.3: Zuordnung der CAPI-Items zu den Skalen

Abuse Scale (77 Items)
3, 5, 7, 9, 13, 14, 17, 18, 19, 22, 23, 24, 25, 26, 28, 29, 32, 26, 28, 29, 41, 45, 47, 49, 52, 54, 56, 63, 67, 68, 69, 73, 74, 75, 76, 77, 78, 80, 81, 83, 84, 90, 93, 94, 95, 98, 99, 100, 102, 103, 105, 107, 108, 109, 111, 112, 113, 115, 118, 120, 122, 127, 128, 129, 130, 132, 134, 138, 141, 143, 145, 147, 148, 151, 152, 153, 154
Lie Scale (18 Items)
12, 34, 35, 44, 46, 57, 62, 66, 70, 106, 110, 146, 149, 150, 155, 157, 159, 160
Random Response Scale (18 Items)
1, 11, 16, 27, 31, 33, 43, 53, 58, 59, 60, 61, 65, 72, 89, 114, 116, 119
Inconsistency Scale (20 Itempaare)
3-76, 4-6, 5-9, 38-41, 44-70, 52-63, 58-72, 62-65, 75-118, 79-98, 83-94, 85-158, 87-141, 90-152, 95-107, 100-151, 105-120, 122-127, 124-133, 143-145
Distress scale (36 Items)
5, 9, 17, 18, 22, 23, 25, 28, 29, 36, 41, 47, 49, 52, 56, 63, 73, 78, 84, 93, 95, 98, 99, 102, 103, 105, 109, 111, 112, 118, 120, 138, 143, 145, 153, 154

Tabelle G.3 (Fortsetzung)

Rigidity scale (14 Items)
7, 19, 24, 26, 32, 54, 68, 80, 108, 115, 122, 127, 130, 132
Unhappiness scale (11 Items)
14, 38, 75, 77, 81, 90, 107, 134, 141, 147, 152
Problems with child and self scale (6 Items)
3, 45, 69, 76, 113, 128
Problems with family scale (4 Items)
39, 83, 94, 148
Problems from others scale (6 Items)
13, 67, 74, 100, 129, 151

3. Items und Skalen des EBSK

3.1 Wortlaut der Items des EBSK

EBSK1. Ich habe ein Kind, das oft in Schwierigkeiten gerät.

EBSK2. Ich fühle mich oft ängstlich.

EBSK3. Ich habe ein Kind, das langsam ist.

EBSK4. Ich werde nie wütend auf andere.

EBSK5. Ein Kind sollte nie widersprechen.

EBSK6. Ein Kind, das schreit, wird niemals glücklich sein.

EBSK7. Heutzutage weiß man nicht so recht, auf wen man sich verlassen kann.

EBSK8. Man kann sich nicht auf andere verlassen.

EBSK9. Ich habe ein gutes Sexualleben.

EBSK10. Ich fühle mich häufig sehr aus der Fassung gebracht.

EBSK11. Ich lache fast jeden Tag.

EBSK12. Andere Leute haben mein Leben schwer gemacht.

EBSK13. Ich bin häufig niedergeschlagen.

EBSK14. Andere Leute verstehen nicht, wie ich mich fühle.

EBSK15. Ich bin oft innerlich ärgerlich.

EBSK16. Mein Leben ist glücklich.

EBSK17. Die Dinge in meinem Leben gingen normalerweise gegen mich.

EBSK18. Meine Familie hat Probleme zurechtzukommen.

EBSK19. Man erwartet zu viel von mir.

EBSK20. Die Leute verstehen mich nicht.

EBSK21. Kinder sollten nicht Schwimmen lernen.

EBSK22. Die meisten Kinder sind gleich.

EBSK23. Viele Dinge im Leben machen mich ärgerlich.

EBSK24. Andere Leute haben mein Leben unglücklich gemacht.

EBSK25. Ich rege mich oft auf und weiß nicht, warum.

EBSK26. Kinder sollten still sein und zuhören.

EBSK27. Ich habe eine körperliche Behinderung.

EBSK28. Ich fühle mich häufig abgelehnt.

EBSK29. Als Kind wurde ich von meinen Eltern mies behandelt.

EBSK30. Kinder sollten sauber sein.

EBSK31. Ich habe Ängste, von denen niemand etwas weiß.

EBSK32. Ich fühle mich häufig allein.

EBSK33. Ich fühle mich häufig sehr frustriert.

EBSK34. Ein Kind braucht sehr strenge Regeln.

EBSK35. Ich rege mich oft auf.

EBSK36. Als Kind wurde ich misshandelt.

EBSK37. Ich versuche immer nach meinem Kind zu sehen, wenn es schreit.

EBSK38. Ich habe viele persönliche Probleme.

EBSK39. Mein Leben ist gut.

EBSK40. Ich mache mir nie Sorgen um meine Gesundheit.

EBSK41. Kinder sollten Spielkleidung und gute Kleidung haben.

EBSK42. Ich bin ein verwirrter Mensch.

EBSK43. Meine Familie streitet viel.

EBSK44. Man hat mir viel Schmerz zugefügt.

EBSK45. Kinder sollten immer gehorchen.

EBSK46. Ich finde es schwer, mich zu entspannen.

EBSK47. Ich habe immer Pech.

EBSK48. Ich bin oft innerlich einsam.

EBSK49. Ich fühle mich häufig sehr allein.

EBSK50. Ich tue immer das was richtig ist.

EBSK51. Andere tun mir nie leid.

EBSK52. Ich tue nie etwas, was schlecht für meine Gesundheit ist.

EBSK53. Ich mag die meisten Kinder nicht.

EBSK54. Mein Kind hat besondere Probleme.

EBSK55. Ich habe einige enge Freunde in meiner Nachbarschaft.

EBSK56. Manchmal fühle ich mich ganz allein in dieser Welt.

EBSK57. Ich bin normalerweise ein stiller Mensch.

EBSK58. Ich fühle mich oft wertlos.

EBSK59. Ich erhebe nie meine Stimme im Ärger.

EBSK60. Kinder sollten gelegentlich an ihre Eltern denken.

EBSK61. Das Leben erscheint mir oft nutzlos.

EBSK62. Ich lache nicht sehr viel.

EBSK63. Es ist in Ordnung ein Kind für eine Weile in schmutzigen Windeln zu lassen.

2.2 Zugehörigkeit der Items zu den Skalen

Tabelle G.4: Zuordnung der EBSK-Items zu den Skalen

Belastungsskala (47 Items)
1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 25, 26, 27, 28, 30, 31, 33, 34, 35, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 54, 55, 56, 57, 58, 61, 62
Lügenskala (5 Items)
4, 40, 50, 52, 59
Zufalls-Antwort-Skala (8 Items)
6, 21, 22, 37, 51, 53, 60, 63
Inkonsistenzskala (6 Itempaare)
11-62, 12-24, 17-47, 29-36, 32-49, 39-61

H. Danksagung

Mein besonderer Dank gilt allen, die mich bei der Erstellung dieser Dissertation unterstützt haben: Prof. Dr. Günther Deegener dafür, dass er mir als Doktorvater stets mit konstruktivem Rat hilfreich zur Seite gestanden hat. Den weiteren Koautoren des EBSK-Manuals, Prof. Dr. Gottfried Spangler und Dr. Wilhelm Körner, für die produktive Zusammenarbeit. Dr. Kurt Becker und Marion Spengler für das Gegenlesen des Manuskripts. Anna Schneider dafür, dass sie die kleinen und großen Unannehmlichkeiten des Alltags von mir ferngehalten hat. Allen Probanden dafür, dass sie an der Untersuchung teilgenommen haben.