

www.testarchiv.eu

Open Test Archive

Repositorium für Open-Access-Tests

Verfahrensdokumentation:

SIRS/CCSERRS

Die klinischen Caregiver-Child Socioemotional and Relationship Rating Scales von McCall - deutsche Version

Müller, J. M. (2024)

Müller, J. M. (2024). SIRS/CCSERRS. Die klinischen Caregiver-Child Socioemotional and Relationship Rating Scales von McCall - deutsche Version [Verfahrensdokumentation, Beobachtungsantwortbogen, Rating- und Auswertungshinweise]. In Leibniz-Institut für Psychologie (ZPID) (Hrsg.), Open Test Archive. Trier: ZPID.
<https://doi.org/10.23668/psycharchives.15230>

Alle Informationen und Materialien zu dem Verfahren finden Sie im Testarchiv unter:
<https://www.testarchiv.eu/de/test/9008727>

Leibniz-Institut für Psychologie (ZPID)

Universitätsring 15
54296 Trier

www.leibniz-psychology.org

Inhaltsverzeichnis

1. Testkonzept.....	2
1.1 Theoretischer Hintergrund.....	2
1.2 Testaufbau	5
1.3 Auswertungsmodus	5
1.4 Auswertungshilfen.....	10
1.5 Auswertungszeit	10
1.6 Itembeispiele.....	10
1.7 Items	10
2. Durchführung	11
2.1 Testformen.....	11
2.2 Altersbereiche	11
2.3 Durchführungszeit	11
2.4 Material.....	12
2.5 Instruktion	12
2.6 Durchführungsvoraussetzungen	12
3. Testkonstruktion	14
4. Gütekriterien.....	16
4.1 Objektivität.....	16
4.2 Reliabilität	19
4.3 Validität.....	28
4.4 Normierung	34
5. Anwendungsmöglichkeiten.....	34
6. Kurzfassung	35
Diagnostische Zielsetzung.....	35
Aufbau	35
Grundlagen und Konstruktion.....	35
Empirische Prüfung und Gütekriterien.....	36
7. Bewertung.....	36
8. Literatur	37

1. Testkonzept

1.1 Theoretischer Hintergrund

Die Erfassung von Beziehungs- und Interaktionsaspekten spielt im klinischen Rahmen der Erfassung, Erklärung und Prognose kindlicher und elterlicher Belastungsfaktoren neben der Prüfung theoretischer Überlegungen auch in der Versorgung nach Einschätzung von Ärzten, Psychologen und Pädagogen eine wichtige Rolle (Ergebnisse einer Befragung; Müller, Hoffmann & Janssen, 2023). Dabei ist es zunächst nachrangig, ob der Beziehungs- und Interaktionsaspekt als Ursache für eine aktuelle Belastung bei Mutter oder Kind, als Folge oder als Risikofaktor bzw. Zielgröße für eine geplante Intervention verstanden wird. Die Erfassung des Beziehungs- und Interaktionsaspektes über Beobachtungsverfahren ist jedoch in der Regel aufwändig (Müller et al., 2023) und eine Best Practice hat sich bislang nicht etablieren können, so dass Anwender wie Forscher je nach Fragestellung vor der Frage nach dem passendsten Instrument stehen, da die Antwort neben der Fragestellung insbesondere vom Alter der Kinder abhängt.

Die Caregiver-Child Socioemotional and Relationship Rating Scale (CCSERRS, im Weiteren auch als ‚SIRS‘ bezeichnet) wurde 2010 von McCall, Groark und Fish veröffentlicht und dient der Erfassung verschiedener Beziehungs- und Interaktionsaspekte zwischen einem erwachsenen Versorger und einem hiervon abhängigen Kind. Neben der oben zitierten Erstveröffentlichung des SIRS in 2010 wurde der SIRS in weiteren Studien u. a. als Evaluationsinstrument eingesetzt (siehe McCall et al., 2010b, Hawk et al., 2018).

Die Entwicklung des SIRS basierte nach McCall auf dem Bedarf und dem gleichzeitigen Mangel an geeigneten Beobachtungsinstrumenten im Kontext der (emotionalen) Betreuung von Waisenkindern in entsprechenden Institutionen. Wichtige Merkmale des Instruments sollten seine universelle Anwendbarkeit in der Praxis sein, wobei mit wenigen Items wichtige Konstrukte im Kontext der Beziehung zwischen Bezugsperson und zu betreuendem Kind ökonomisch erhoben werden sollten. Das SIRS war der Versuch von McCall, grundlegende Aspekte der sozial-emotionalen Interaktion zwischen Pflegeperson und Kind ohne Einschränkungen des Beobachtungssettings und auf der Basis einer kurzen Beobachtungszeit abzubilden. Obwohl die ersten Entwicklungsschritte im Kontext von Pflegekindern erfolgten, sollte das Instrument auch für Eltern und ihre leiblichen Kinder sowie außerhalb des Pflegekinderkontextes anwendbar sein. McCall wollte das Beobachtungsinstrument SIRS sowohl Forschern als auch Praktikern frei zugänglich machen, um Erfahrungen und Weiterentwicklungen zu ermöglichen, was bei anderen Verfahren aufgrund von Urheber- oder Verlagsrechten leider nicht möglich ist.

Für Leser, die sich auch außerhalb des SIRS breit informieren möchten, geben Lotzin et al. (2015) einen umfassenden Überblick und eine Bewertung englischsprachiger Instrumente für den Altersbereich 0-3 Jahre, wobei die Autoren bereits hier auf die häufig fehlenden Studien zur Inhaltsvalidität hinweisen. Für den deutschsprachigen Raum stellt Jacob (2022) eine Auswahl verschiedener Beobachtungsverfahren für unterschiedliche Altersgruppen vor. Eine Auflistung deutschsprachiger Instrumente findet sich bei Müller, Hoffmann und Wonner (2018). Für die Anwender ist es oft schwierig, zu beurteilen, was das jeweilige Instrument im Unterschied zu anderen Beobachtungsinstrumenten misst. Wir ordnen daher im Folgenden das mit dem SIRS erfasste elterliche Erziehungsverhalten in den Rahmen des umfassenden Artikels

von Ellen Skinner et al. (2005) ein, der eine sehr gute inhaltliche Orientierung bietet (siehe dazu ausführlich auf Seite 6; Exkurs Skinner). Diese Suche nach einem wissenschaftlich fundierten und ökonomischen Beobachtungsinstrument wird derzeit durch fehlende psychometrische Testbeschreibungen und Vergleichsstudien erschwert: Es ist primär vor allem unklar, was die einzelnen Beobachtungsinstrumente messen und wie sie sich inhaltlich voneinander abgrenzen. Diese Problematik besteht seit Jahrzehnten und hat sich mit der zunehmenden Fülle an Instrumenten noch verschärft. Ellen Skinner et al. (2005) listen in ihrem umfassenden Review zu diesem Thema eine beeindruckende Anzahl englischsprachiger Bezeichnungen/Skalenbezeichnungen auf. Dies ist auf die immer wieder problematische Vorgehensweise zurückzuführen, ein Instrument auf der Basis einer einmaligen Erprobung mit einer relativ kleinen Stichprobe ohne psychometrische Analyse und ohne expliziten Bezug zu definierten Konstrukten zu entwickeln, so dass unklar bleibt, was die Skalen letztlich messen und wie trennscharf die Skalen eines Instruments innerhalb eines Instruments sind und sich mit Inhalten aus anderen Instrumenten überlappen.

Ein weiteres Problem bei der Suche und Auswahl eines Beobachtungsinstrumentes besteht darin, dass diese oft nicht frei zugänglich sind. Es wurde bereits erwähnt, dass Beobachtungsinstrumente von verschiedenen Professionen genutzt werden, welche allerdings sehr unterschiedliche Grundvoraussetzungen für die Bewertung von Erhebungsinstrumenten besitzen, weshalb wir im Folgenden zumindest in groben Zügen die grundlegenden Kenntnisse vermitteln möchten.

Die Vorteile eines wissenschaftlich fundierten Instruments liegen vor allem darin, dass die Auswahl der Beobachtungsinhalte und deren Zuordnung zu Konstrukten inhaltlich wie statistisch begründet werden. Inhaltlich müssten die Skalen innerhalb einer Theorie oder einer diagnostischen Fragestellung nachvollziehbar eingebettet werden. Erst anschließend kann empirisch im Rahmen einer psychometrischen Analyse das resultierende Instrument evaluiert und beschrieben werden in Hinblick auf seine Genauigkeit und Aussagekraft (Reliabilität; Validität). Eine psychometrische Analyse berichtet über Aspekte, welche eine korrekte Interpretation und Anwendung abzusichern helfen. Dort wird beispielsweise auf den Aspekt der Content Validity (Inhaltsvalidität) hingewiesen, also den Grad, inwieweit die Stichprobe von Items, Aufgaben oder Fragen eines Tests repräsentativ für eine definierte Grundgesamtheit oder einen Inhaltsbereich ist. Das zu messende Konstrukt muss deshalb ausführlich beschrieben sein. An dieser Stelle sei in einer leicht überarbeiteten Übersetzung von DeepL vom 01.12.2023 aus Mahoney et al. (1996) seine Kritik am häufigen Vorgehen zitiert: „Im Allgemeinen sind die meisten der Forschungsskalen, die derzeit zur Bewertung der Eltern-Kind-Interaktion verwendet werden, nicht entwickelt worden, um einen vordefinierten Bereich der elterlichen Effektivität widerzuspiegeln. Vielmehr sind sie ein Sammelsurium von Merkmalen und Konstrukten, die sich aus einer von zwei Quellen entwickelt haben. Erstens: Sie entstammen aus theoretischen Artikeln als potenzielle Einflüsse auf verschiedene Aspekte der Entwicklung von Kindern. Zweitens wurden sie in früheren Untersuchungen als elterliche Charakteristika mit einer Vielzahl von kindlichen Ergebnissen in Verbindung gebracht (zitiert in Mahoney et al., 1996)“. Beide Quellen führen zu einer eklektischen Sammlung von Konstrukten und Items, mit der Folge, dass bei einem neuen Beobachtungsinstrument unklar bleibt, wie der Zusammenhang zwischen dem Beobachtungsmerkmal (einem Item) und dem Konstrukt theoretisch hergeleitet wird. Zum anderen bleibt unklar, inwieweit die Konstrukte untereinander geordnet sind, d. h. ob sie nicht sehr verwandte, d. h. inhaltlich ähnliche Aspekte

erfassen. Dies würde nicht nur dem Aspekt der Ökonomie widersprechen, sondern insbesondere eine eindeutige Zuordnung der Items zu einer Skala verhindern. Aus Sicht der Testkonstruktion ist aber genau dies zu gewährleisten bzw. das Ziel einer Testkonstruktion. Die Zuordnung von Items zu Skalen ist eben nicht beliebig und gelingt bei verschiedenen Instrumenten auch nicht immer vergleichbar gut. Empirisch sollten Items, die zu einem bestimmten Konstrukt (z. B. Responsivität) gezählt werden, höher mit dem entsprechenden Skalenwert korrelieren als Items zu alternativen Konstrukten (z. B. Intrusivität). Wird dies nicht berücksichtigt, korrelieren die Konstrukte so hoch miteinander, dass sie empirisch redundant werden und der Aufwand nicht den gewünschten Erkenntnisgewinn bringt. Dieser Aspekt kann auf Itemebene z. B. im Rahmen einer Faktorenanalyse durch die Ladung des Items auf dem Konstrukt beschrieben werden, die stark vereinfacht der Korrelation eines Beobachtungsaspektes mit dem Skalenwert entspricht. Während der Testentwicklung könnten zur Verbesserung der endgültigen Skalen entsprechende Items ausgeschlossen werden (sog. Phase der Itemauswahl), was bei vielen Instrumenten nicht geschehen ist und hier auch nicht Teil der Analysen sein wird, da wir nur eine Testübersetzung und -auswertung vornehmen. Neben der Eindeutigkeit der Skala verbessert diese Phase in der Regel auch die Homogenität der Skala (Cronbachs Alpha) und damit die Reliabilität der Skala. Dementsprechend können mit solchen Skalen präzisere Messungen durchgeführt werden, die dann die Festlegung von Cut-off-Werten ermöglichen, die eine möglichst trennscharfe Unterscheidung zwischen gesunden und behandlungs- oder beratungsbedürftigen Personen erlauben. Generell gibt es in der Phase der Itemauswahl eine Vielzahl von Optimierungsstrategien und Kriterien, die bisher bei der Testentwicklung im Bereich der Beobachtungsverfahren von Eltern-Kind-Interaktionen nicht genutzt wurden. Ein Kriterium wäre beispielsweise eine hohe Beurteilerübereinstimmung, so dass Beobachtungsaspekte, die nicht von zwei Beurteilern gleichermaßen zuverlässig eingeschätzt werden können, nicht in die endgültige Version des Instruments aufgenommen werden.

Erschwerend kommt neben der Vielzahl der Instrumente die unterschiedliche Benennung der Skaleninhalte hinzu, die eine Vergleichbarkeit und Abgrenzung verschiedener Instrumente nahezu unmöglich macht. Im Folgenden wird daher versucht, die Skalen von McCall inhaltlich zu klassifizieren, d. h. inwieweit sie im Rahmen bekannter Konstrukte interpretierbar sind. Die faktorielle Validität eines Verfahrens kann bei der Testentwicklung durch exploratorische oder konfirmatorische Faktorenanalysen untersucht werden, wobei nur die Antworten auf die Items für die Analysen herangezogen werden, ohne dass externe Kriterien eine Rolle spielen, weshalb dieser Aspekt auch als interne Validität bezeichnet wird. Eine zentrale Frage bei diesen Analysen ist, wie viele Faktoren bzw. Skalen ein Beobachtungsinstrument empirisch unterscheiden kann. Gleichzeitig können die einzelnen Indikatoren (psychometrische Items) auch hinsichtlich einer ökonomischen Erfassung (ausreichende Anzahl von Items pro Dimension) beurteilt werden. Erst auf dieser Grundlage kann dann eine externe Validierung an anderen Instrumenten oder einem Kriterium (z. B. mütterliche Belastung und kindliche klinische Symptomatik) überprüft werden. Ein weiterer wichtiger Zusatzaspekt bei Beobachtungsinstrumenten ist die Beschreibung der Interrater-Reliabilität, die beschreibt, ob zwei voneinander unabhängige Beobachter (Rater) zu vergleichbaren Ergebnissen kommen. Diese wird der Beurteilungsobjektivität zugeordnet, obwohl häufig der Fehlschluss gezogen wird, dass in der Praxis Einzelbeurteilungen reliabel wären. Die Interraterreliabilität in Forschungsstudien beschreibt neben der Skalenhomogenität einen Teil der Reliabilität und häufig werden die aggregierten Beurteilungen von zwei Beurteilern für Itemanalysen

verwendet. In der Praxis liegt jedoch häufig nur ein Urteil vor und es ist wahrscheinlich, dass ein Einzelurteil nicht die gleiche Messgenauigkeit aufweist wie die aggregierten Urteile zweier Beurteiler aus Forschungsstudien. Dies ist auch bei den hier vorgestellten Ergebnissen zu berücksichtigen.

1.2 Testaufbau

Jedes SIRS-Item wird auf einer 4-Punkte-Skala (0 = nie, 1 = selten, 2 = häufig, 3 = immer) beantwortet, welche die Häufigkeit reflektiert, mit der das jeweilige Verhalten (positiv oder negativ) aufgetreten ist, relativ zu den potentiellen Gelegenheiten, bei denen es hätte auftreten können. Die insgesamt 18 „Items“ werden den folgenden Bereichen zugeordnet:

- (1) Anteilnahme der Bezugsperson (Item 1-4)
- (2) Bezugsperson/Kind-gerichtetes Verhalten (Item 5-6)
- (3) Verhaltenskontrolle (Item 7-10)
- (4) Erregungszustand (Affekt) der Bezugsperson (Item 11-13)
- (5) Kontakterwartung und Responsivität des Kindes (Item 14-15)
- (6) Child Affect (Gefühle des Kindes) (Item 16-17)
- (7) Beziehung zum Kind (Item 18)

Jedes „Item“ erfasst im Sinne von McCall ein eigenständiges Konstrukt, welches über dieses Item global eingeschätzt wird. Zum besseren Verständnis gibt McCall eine verhaltensnahe Beschreibung (Operationalisierung). Die Bewertung erfordert aber in jedem Fall eine Integrations- und Bewertungsleistung des Beurteilers (geht über reines „Zählen“ von Verhaltensweisen hinaus).

1.3 Auswertungsmodus

Vor der Auswertung müssen die Items 1, 2, 5, 7-9, 11, 14, und 16 vor einer Summenbildung rekodiert werden, da diese Items unerwünschte Verhaltensweisen repräsentieren. Die Summenbildung wird für die Anwendung empfohlen und der Cut-off (siehe unter „Testkonstruktion“, Tabellen 7-8) verwendet diese Art der Skalierung. Für die psychometrische Analyse und Auswertung wird jedoch der Mittelwert der Antworten verwendet. Dies hat den Vorteil, dass die Lage der Antworten auf der Antwortskala im Hinblick auf Boden- und Deckeneffekte leichter interpretiert werden kann und für alle Skalen der gleiche Wertebereich (von 0 bis 3) verwendet wird. Dies erleichtert die Interpretation für bestimmte Auswertungen.

Jedes Item wird also auf einer vierstufigen Antwortskala von 0 bis 3 kodiert. McCall gruppiert in seiner Originalpublikation die Items sowohl inhaltlich als auch psychometrisch, so dass sich drei **Auswertungsmöglichkeiten der Scorebildung A, B und C** ergeben.

A) McCall (2010a, ebenso Groark et al., 2011) bildet über alle 18 Items – nach entsprechender Umkodierung (siehe dazu unter „Testkonstruktion“) von gespiegelten Fragen - einen Total Score aus dem Durchschnitt der einzelnen Itemantworten. In McCall 2010b wurde der Summenscore gebildet. Wir benennen diese Auswertungsvariante mit SIRS_McCall_01.

B) McCall bildet 2010 im Rahmen einer explorativen Faktorenanalyse drei Skalen, wobei er in diesem Zusammenhang darauf hinweist, dass diese Skalenbildung vorläufig ist.

B1: SIRS_McCall_02: Caregiver–Child Mutual Positive Engagement (1, 2, 3, 4, 10, 12, 13, 14, 15, 17, 18)

B2: SIRS_McCall_03: Caregiver Punitiveness (9, 11, 16)

B3: SIRS_McCall_04: Child–Caregiver Directed Interaction (5, 6, 7, 8)

C) McCall bildet zuletzt im Rahmen einer inhaltlichen Gruppierung der Items sieben Skalen (Itemnummer in Klammern, siehe auch „Testaufbau“).

C1: SIRS_McCall_05: Verfügbarkeit und Responsivität (1, 2, 3, 4)

C2: SIRS_McCall_06: Bezugsperson/Kind gerichtetes Verhalten (5, 6)

C3: SIRS_McCall_07: Verhaltenskontrolle des Kindes“ (7, 8, 9, 10)

C4: SIRS_McCall_08: Emotionen der Bezugsperson gegenüber dem Kind (11, 12, 13)

C5: SIRS_McCall_09: Kontakterwartung und Responsivität des Kindes an der Bezugsperson (14, 15)

C6: SIRS_McCall_10: Emotionen des Kindes gegenüber der Bezugsperson (16, 17)

C7: SIRS_McCall_11: Beziehung des Kindes zur Bezugsperson (18)

Die von McCall erstellte eklektische Itemsammlung deckt verschiedene Aspekte einer Bezugsperson/Eltern-Kind-Interaktion ab. Die hierauf gebildeten elf Skalen auf insgesamt 18 Einzelratings folgen jedoch keinem expliziten und kohärenten theoretischen Modell. Zur Einordnung und inhaltlichen Orientierung wird im Folgenden auf die Arbeit von Skinner et al. (2005) verwiesen, die auf der Basis einer sehr umfangreichen Literaturrecherche einen Überblick über wesentliche Konstrukte erarbeitet hat (vgl. TEXTBOX 1). Auf Grundlage ihrer eigenen Forschungsergebnisse benennt sie sechs jeweils unipolaren Skalen des Erziehungsverhalten.

Exkurs Skinner: An dieser Stelle ist es wichtig, zu verstehen, dass Skinner diese sechs Dimensionen empirisch aus den drei von ihr identifizierten bipolaren Skalen ableitet, die in der Literatur als Warmth/Acceptance vs. Rejection, Firm Control vs. Lax Control und Autonomy vs. Psychological Control bezeichnet werden. Eine bipolare Skala geht davon aus, dass sich die Enden (d. h. niedrige und hohe Werte) inhaltlich ausschließen und nur eine inhaltliche Ausprägung messen. Dementsprechend kann z. B. eine Mutter, die viel Verhalten im Bereich Warmth/Acceptance zeigt, kein Verhalten im Bereich Rejection zeigen. Gleiches gilt für die anderen Skalen. Es ist jedoch nicht möglich, mit einer solchen Skala ein alternierendes Verhalten zu kodieren, d. h. ein Schwanken zwischen Acceptance und Rejection. Die Aufteilung einer bipolaren Skala in jeweils zwei unipolare Skalen hat in der Anwendung verschiedene praktische und theoretische Vorteile. Beispielsweise können nun beim Elternverhalten sowohl vermehrt akzeptierendes als auch häufig ablehnendes Verhalten getrennt kodiert werden und somit positives und negatives Elternverhalten über einen jeweils eigenen Score abgebildet werden. Damit ergibt sich z. B. die neue Möglichkeit, wenn ein Elternteil weder das eine noch das andere Verhalten zeigt, dies als ‚unterinvolviertes Verhalten‘ zu interpretieren. Eine separate Skala wäre dann nicht mehr notwendig. Entsprechend kann auch inkonsistentes, ambivalentes oder wechselndes Verhalten über diese Skala in der Auswertung kodiert werden. Das elterliche Verhalten könnte in diesem Fall aus Sicht des Kindes als weniger vorhersehbar wahrgenommen werden. Insbesondere diese dynamische Darstellung von Beziehungsstressoren scheint in verschiedenen Studien stärker mit der Ausprägung von

Belastungserleben zusammenzuhängen als beispielsweise ein rein negatives (dafür aber vorhersagbares) Verhalten (Bushman et al., 2009; Janßen et al., 2019).

TEXTBOX 1. Sechs Konstrukte nach dem Modell von Ellen Skinner (2005, S. 186, Figure 1) und Begriffe, welchen diesen Bereichen zugeordnet sind (*Nachdruck der Tabelle mit freundlicher Genehmigung des Wiley-Verlages*)

Dimension	Definition	Related Constructs	
Warmth	Expression of love, affection, caring, and enjoyment. Appreciation, emotional availability. "Love: Positive evaluation, sharing, expression of affection, emotional support" (Schaefer, 1965, p. 415).	Approving Acceptance Love Support Supportive control	Positive involvement Closeness Connection Child-centered
Rejection	Active dislike, aversion, and hostility. Harsh, over-reactive, irritable, critical, disapproving. "Hostility: Irritability, negative evaluation, rejection" (Schaefer, 1965, p. 415).	Deprecating Hostility Harsh Disapproval Negativity Cold Derogation	Critical Over-reactivity Aversion Irritability Dislike Irritable explosive discipline
Structure	Provision of information about pathways to reach desired outcomes. Predictable, consistent. Clear expectations, firm maturity demands. "[E]xtent to which parents provide clear and consistent guidelines, expectations, and rules for child behavior" (Grolnick & Ryan, 1989, p. 144)	Demandingness- Firm control Behavioral control Contingent responsiveness Behavior contingency Directive Assertive control	Strictness Supervision Organization Regulation Rule-setting Regularity of routine Household organization
Chaos	Interferes with or obscures the pathways from means to ends Noncontingent, inconsistent, erratic, unpredictable, arbitrary, or, undependable.	Permissiveness Non-directive Lax control, Laxness Unpredictable Undependable Non-contingent	Erratic Casual Under-controlled Laissez faire Inconsistent discipline
Autonomy Support	Allow freedom of expression and action. Encourage child to attend to, accept, and value genuine preferences and opinions. "[D]egree to which parents value and use techniques which encourage independent problem-solving, choice, and participation in decisions" (Grolnick & Ryan, 1989, p. 144)	Psychological autonomy Freedom Responsiveness Democratic	Permissiveness Non-directive Autonomy-granting
Coercion	Restrictive overcontrolling intrusive autocratic style Strict obedience is demanded "[E]xternally dictating outcomes, and motivating ... through punitive disciplinary techniques, pressure, or controlling rewards" (Grolnick & Ryan, 1989, p. 144)	Arbitrary control Demandingness- Restrictiveness Autocratic Psychological control Controllingness Inflexible rigid discipline	Intrusive control Intrusive support Strict control Over-restrictive Over-controlling Power assertion Intrusiveness

Die unipolare Aufspaltung ermöglicht also im Anschluss mehr Beschreibungsmöglichkeit, als es eine bipolare Konzeption. Neben der impliziten Erfassung von Konstrukten durch unipolare Skalen ergeben sich auch therapeutisch neue Möglichkeiten der Messung. Beispielsweise kann therapeutisch ein gezielter Aufbau von positivem Elternverhaltensweisen erfasst werden, ohne dass dies implizit bedeuten würde, dass hierüber gleichzeitig eine Minimierung von negativen Verhaltensweisen erreicht worden wäre – wenngleich dies vermutlich ein therapeutisch erwünschtes Ziel ist.

Aufgrund der angestrebten klinischen Anwendung werden neben der Orientierung an den sechs bipolaren Konstrukten von Ellen Skinner im Folgenden die im klinischen Bereich gebräuchlicheren Konstrukte verwendet, um die inhaltliche Zuordnung weiterer klinischer Skalen zu verdeutlichen. Dies würde die Interpretierbarkeit der Items und Skalen von McCall erleichtern und thematische Lücken aufzeigen. Die Skalen von McCall in den Versionen B und C von McCall werden vom Autor inhaltlich zugeordnet, ebenso die Items des SIRS:

POSITIVES ELTERNVERHALTEN

I. Skinner: Involvement/Responsivität (Warmth)

Skalen:

SIRS_McCall_02: Caregiver–Child Mutual Positive Engagement

SIRS_McCall_05: Verfügbarkeit und Responsivität

SIRS_McCall_06: Bezugsperson/Kind gerichtetes Verhalten

SIRS_McCall_08: Emotionen der Bezugsperson gegenüber dem Kind

Items:

3. Mentale Verfügbarkeit/Empfänglichkeit der Bezugsperson in Bezug auf Kinder (+)

4. Bezugsperson reagiert auf das Kind (+)

12. Bezugsperson zeigt positiven Affekt und Zuneigung (+)

II. Skinner: Acceptance (Autonomy support)

Skalen: -

Items:

6. Bezugsperson lässt Kind führen (+)

III. Skinner: Structure (Support, Guiding)

Skalen: -

Items:

5. Bezugsperson führt Kind (+)

8. Verhaltens-Kontrolle/-Gehorsam/-Disziplin der Bezugsperson (+)

10. Unterstützung/Empathie/Anleitung der Bezugsperson (+)

NEGATIVES ELTERNVERHALTEN

IV. Skinner: Rejection [Hostility; low involvement; low engagement]

Skalen: SIRS_McCall_03: Caregiver Punitiveness

Items:

1. Distanziertheit der Bezugsperson (-)

2. Bezugsperson reagiert nicht (-)

9. Bezugsperson bestraft das Kind (-)

11. Negativer Affekt der Bezugsperson (-)

V. Skinner: Intrusivness (Coersion, Psychological control)

Skalen: -

Items:

7. Intrusivität der Bezugsperson (-)

VI. Skinner: Impaired Predictability (Chaos, Inconsistence)

Skalen: -

Items: -

Nicht zuordenbare Items: 13. *Lebhaftigkeit und Ausdruckskraft der Bezugsperson*. Begründung: Diese Beschreibung wird eher der elterlichen Persönlichkeit zugeordnet. Insgesamt erfassen die von McCall vorgeschlagenen elterlichen Skalen nur einen Teil der in der Literatur durch Skinner vorgeschlagenen Konstrukte.

Inhaltliche Bewertung und Verortung der Skalen von McCall zum Kindverhalten
Neben den elterlichen Verhaltensweisen sollen im Weiteren auch die kindlichen Verhaltensweisen und Tendenzen innerhalb eines allgemeineren Rahmens verortet werden. Die kindbezogenen SIRS Items lauten (mit vorgehender Itemnummer):

- 6. Bezugsperson lässt Kind führen (+)
- 14. Emotionale Distanziertheit des Kindes gegenüber der Bezugsperson (-)
- 15. Kindliche Responsivität und positive Erwartung in Bezug auf die Bezugsperson (+)
- 16. Negativer Affekt des Kindes gegenüber der Bezugsperson (-)
- 17. Positiver Affekt des Kindes gegenüber der Bezugsperson (+)
- 18. Beziehung des Kindes zur Bezugsperson (+)

Die beiden Items 5 und 6 weisen beide einen Bezug zum Therapie Konzept der Parent–Child Interaction Therapy (PCIT; Timmer et al., 2023) auf, weshalb diese beiden typischen Interaktionsmuster in enger Beziehung zum Elternverhalten stehen und Item 5 auch dort unter positivem Elternverhalten/3. Structure (Support, Guiding) aufgeführt ist. Eindeutig kindbezogenes Verhalten wird dagegen mit Item 6 erfasst. Üblicherweise wird das Verhalten des Kindes in der Interaktion mit der Mutter im klinischen Kontext jedoch im Rahmen der Bindungstheorie von Bowlby (1958) interpretiert, wobei u.E. alle fünf Items des SIRS zumindest proximal in diesem Kontext interpretiert werden können. Eine zweite Interpretation für die Items 15 und 16 ergibt sich im Kontext der Persönlichkeits- bzw. Temperamentsforschung, z. B. über die übergeordneten Dimensionen von Tellegen (1985) zum positiven und negativen Affekt (Tellegen, 1985; Watson et al., 1999). Detaillierte Facetten des Temperaments können mit Hilfe des Children’s Behavior Questionnaire von Rothbart et al. (2001; Rothbart et al., 1994; Prokasky et al., 2017) erfasst werden. Diese werden hier lediglich aufgeführt, um zu verdeutlichen, welche kindlichen Verhaltenstendenzen für die soziale Interaktion von Bedeutung sein können, im SIRS jedoch keine Entsprechung finden und somit auf Lücken hinweisen. Die Temperamentsbereiche nach Rothbart et al. (2001) sind:

Surgency: Shyness, Impulsivity, Activity Level, High-intensity Pleasure

Negative Affectivity: Soothability, Anger, Discomfort, Fear, Sadness

Effortful Control: Inhibitory Control, Low-intensity Pleasure, Attention Focusing, Perceptual Sensitivity

Die oben genannten Facetten des Temperaments bzw. des positiven und negativen kindlichen Affekts stellen grundlegende Verhaltenstendenzen dar, die auch im Rahmen eines Bindungssystems relevant werden (vgl. Seifer & Schiller, 1995; Belsky & Isabella, 1988). Entsprechend ist das kindliche Verhalten nicht nur als Folge des elterlichen Interaktionsverhaltens zu verstehen, sondern auch als Resultat allgemeiner Eigenschaften bzw. Voraussetzungen oder als Vulnerabilitätsfaktoren vor der Interaktion des Kindes, die sich auch außerhalb des Eltern-Kind-Systems in sozialen Interaktionen zeigen.

Neben den kindlichen Voraussetzungen sind bei der Interpretation des kindlichen Interaktionsverhaltens auch die elterlichen Voraussetzungen zu berücksichtigen (Clark et al., 2000) (insbesondere deren Persönlichkeit und sozial-emotionale Kompetenzen sowie deren Ziele) und beide zusammen in der ‚Passung‘ des kindlichen Temperaments mit der elterlichen Persönlichkeit (vgl. dazu ausführlich Achtergarde et al., 2015). Daraus folgt auch, dass ein Kind bei anderen Interaktionspartnern weniger ‚auffälliges‘ oder ‚schwieriges‘ Interaktionsverhalten zeigen kann oder möglicherweise sogar nur dort oder mit bestimmten Interaktionspartnern auffällig interagiert.

Abschließend sei darauf hingewiesen, dass die Aussagen der Eltern über ihr Kind bekanntermaßen relativ schwach mit den unabhängigen Einschätzungen z. B. von Erziehern oder Therapeuten korrelieren, wie dies in klinischen Symptomchecklisten seit Jahrzehnten beobachtet wird, und dass diese Urteilsunterschiede nicht allein situativ erklärt werden können, sondern auch ihren Ursprung in der verzerrten Wahrnehmung besonders belasteter Eltern haben können, was an anderer Stelle ausführlich beschrieben wurde (Müller, Achtergarde & Furniss, 2011, Müller & Furniss 2013, Müller, Romer & Achtergarde 2014).

Insgesamt erfassen die kindbezogenen Inhalte des SIRS den wichtigen Aspekt der Bindung sowie grundlegende temperamentsbezogene Aspekte, ohne jedoch aufgrund der begrenzten Anzahl von Items ein genaueres oder umfassenderes Bild der genannten Konstrukte zeichnen zu können. Für eine erste grobe Orientierung des kindlichen Verhaltens in der Interaktion mit der Bezugsperson scheint der SIRS wichtige Inhalte zu erfassen, wenngleich Lücken analysiert werden konnten.

1.4 Auswertungshilfen

Als Auswertungshilfen stehen für den deutschen Sprachraum unsere Mittelwertsvergleiche unserer klinischen und nicht-klinischen Stichprobe der Kinder- und Jugendpsychiatrie Münster in Tabelle 4 zu Verfügung, sowie die Änderung dieser bei veränderten Beobachtungszeiten. Ergänzend sind hierbei die Abhängigkeit der Skalen untereinander in Tabelle 5 zu beachten, sowie die Reliabilität und das Personenunterscheidungsvermögen in Tabelle 6 nebst Angaben zur Sensitivitäts- und Spezifität in Tabelle 7.

1.5 Auswertungszeit

Der Aufwand für die Einschätzung einer Interaktion beträgt ca. 10-20 Minuten.

1.6 Itembeispiele

4. Bezugsperson reagiert auf das Kind

1.7 Items

Anteilnahme der Bezugsperson

1. Distanziertheit der Bezugsperson
2. Bezugsperson reagiert nicht
3. Mentale Verfügbarkeit/Empfänglichkeit der Bezugsperson in Bezug auf Kinder
4. Bezugsperson reagiert auf das Kind

Bezugsperson/Kind-gerichtetes Verhalten

5. Bezugsperson führt Kind

6. Bezugsperson lässt Kind führen

Verhaltenskontrolle

7. Intrusivität der Bezugsperson

8. Verhaltens-Kontrolle/-Gehorsam/-Disziplin der Bezugsperson

9. Bezugsperson bestraft das Kind

10. Unterstützung/Empathie/Anleitung der Bezugsperson

Erregungszustand (Affekt) der Bezugsperson

11. Negativer Affekt der Bezugsperson

12. Bezugsperson zeigt positiven Affekt und Zuneigung

13. Lebhaftigkeit und Ausdruckskraft der Bezugsperson

Kontakterwartung und Responsivität des Kindes

14. Emotionale Nicht-Responsivität des Kindes

Nicht-Reagieren auf die Bezugsperson

15. Kindliche Responsivität/Positive Erwartung

Child Affect (Gefühle des Kindes)

16. Negative Gefühle des Kindes

17. Positive Gefühle des Kindes

Beziehung zum Kind

18. Beziehung des Kindes zur Bezugsperson

2. Durchführung

2.1 Testformen

Es existiert die Version aus der Originalpublikation von McCall. Parallelförmige oder Testformen für spezielle Populationen liegen nicht vor. Die Vergleichbarkeit mit fremdsprachigen Versionen wurde für die kindbezogenen Items bereits von Esins et al. (2017) evaluiert. Eine spanische Übersetzung liegt ebenfalls vor und kann direkt bei McCall bezogen werden.

2.2 Altersbereiche

Ähnlich wie im Original wird empfohlen, SIRS bei mindestens 18 Monate alten Kindern bis maximal 6 Jährigen anzuwenden.

2.3 Durchführungszeit

In der Originalpublikation von McCall werden keine Durchführungszeiten angegeben. McCall et al. (2010) schreiben dazu (Übersetzung von DeepL am 01.12.2021 auf Seite 207: „Das Instrument ist recht allgemein gehalten und kann verwendet werden, um das Verhalten der Betreuungsperson zu bewerten, z. B. in einem Freispielkontext für nur 5 Minuten oder in

Fütterungs-, Bade-/Anzieh- und Freispielsituationen für jeweils für 5 Minuten oder länger“. In der Anwendung des SIRS wurden die Interaktionen von Groark et al. (2011) in zwei 5-minütigen Episoden während jeder der drei Aktivitäten (Füttern, Baden/Anziehen und Freispiel) für insgesamt maximal 30 Minuten bewertet, wobei die Bewertungen aus zwei Episoden gemittelt wurden. Das Fehlen einer standardisierten Beobachtungszeit für die Anwendung des SIRS hat den Nachteil, dass direkte Vergleiche der Ergebnisse zwischen verschiedenen Kontexten und Beobachtungsverfahren möglicherweise eingeschränkt sind. Wir empfehlen eine Beobachtungszeit von ca. 20 Minuten, da dies eine Einordnung der Ergebnisse mit denen unserer Studie ermöglicht.

2.4 Material

Es existiert außerhalb der Originalpublikation u. W. kein Manual von McCall oder Antwortbogen. Die vorliegende Publikation soll zumindest für den deutschsprachigen Raum ein Angebot für interessierte klinische Anwender schaffen.

Das Beobachtungsinstrument in der Publikation von McCall findet sich zusammen mit einer Beschreibung der Items und wie sie zu bewerten sind auf der Webseite zu Download (Beobachtungsantwortbogen und Rating- und Auswertungshinweise).

Die vorliegende Verfahrensdokumentation dient als Manual.

2.5 Instruktion

Es liegt in den Publikationen von McCall oder Groark keine Instruktion für die Interaktionspartner vor. Wir empfehlen die in unserer Datenerhebung erfolgte Instruktion für die Dyade ‚Eine gute Zeit haben‘, da sie zum einen bewusst keine Vorgaben für die Gestaltung oder den Inhalt einer Interaktion macht, sodass nicht eine Kooperation der Interaktionspartner im Sinne der Erfüllung einer externen Anforderung gemessen wird, jedoch ein Hinweis existiert, dass beide Interaktionspartner insoweit aufeinander eingehen bzw. interagieren müssen, dass beide wiederholt eine Interaktion miteinander in der Zukunft eingehen möchten. Es sei erwähnt, dass im Rahmen unserer Datenerhebung eine Aufklärung im Rahmen der Verwendungsweise des Videomaterials und diesbezüglicher Auswertungen sowie des Datenschutzes erfolgt. Derartige Aufklärung müssen unseres Erachtens immer dokumentiert werden, da sie potentiell den Interaktionsverlauf beeinflussen können und entsprechend für die Interpretation bekannt und damit dokumentiert sein müssen.

2.6 Durchführungsvoraussetzungen

In beiden Veröffentlichungen unter der Beteiligung von McCall erfolgte die Anwendung für die Betreuer-Kind-Beziehung in amerikanischen und russischen Waisenhäusern, wobei deren Betreuer in der Regel sehr viele Kindern zugeordnet (8 bis 23 Kinder) waren, was bei der Beurteilung der SIRS-Skalen Mittelwerte im Vergleich mit den unseren Mittelwerten von Eltern und deren Kindern zu berücksichtigen ist. Wir empfehlen eine Interaktionsbeobachtung in einem Raum mit altersgerechtem Spielmaterial, in dem keine anderen Personen anwesend sind, da dies eine Einordnung der Ergebnisse mit denen unserer Untersuchung ermöglicht.

Grundsätzlich sollten Eltern über die Zielsetzung bzw. Aufgabe der Interaktionsbeobachtung im Rahmen des diagnostischen Prozesses oder eines Monitorings im Laufe einer Intervention für Therapieentscheidung oder zur internen Qualitätssicherung aufgeklärt werden. Ebenso sollte

die Aufklärung über die Art der Dokumentation bzw. die zeitlich begrenzte Archivierung mit Löschdatum des ggf. anfallenden Videomaterials und dessen Verbleib und Zugriffsmöglichkeiten (Datenschutz) erfolgen. Wir empfehlen, diese und weitere Durchführungsvoraussetzungen einer Interaktionsbeobachtung und Auswertung standardisiert zu dokumentieren, welche in nachfolgender Textbox 2 schematisch von uns (Müller et al., 2023) beschrieben wurde. In dieser wird auch die Ausbildung bzw. Qualifikation des Diagnostikers und weitere Bedingungen der Testsituation dokumentiert.

TEXTBOX 2: Standardisierte Beschreibung einer Eltern-Kind-Interaktions-Beziehungsdiagnostik (mit freundlicher Genehmigung des Verlages Vandenhoeck & Ruprecht) aus der Publikation von Müller, Hoffmann & Janssen (2023, S. 45)

- 1) Diagnostiker/in:
 - a. Primäre Ausbildung (Psychologie, Medizin, Pädagogik)
 - b. gegebenenfalls sekundäre Ausbildung (Systemische Ausbildung, Kognitiv-Verhaltenstherapeutisch, Tiefenpsychologisch/psychoanalytisch, in Ausbildung)
- 2) Diagnostische Fragestellungen/Anlass bezüglich Kind, Bezugsperson, Dyade
- 3) Beobachtungsinhalte mit Beispielen
 - a. Bezugsperson: Erziehungsverhalten (Involviertheit, Responsivität, feindseliges Verhalten, Intrusivität), Sensitivität, Bonding, Emotionale und kommunikative Kompetenzen, Affekt, Persönlichkeit, Stress, Psychopathologie
 - b. Kind: Bindungsverhalten, Affekt, Temperament, Regulation, sozial-verbale Kompetenzen, Involviertheit, Responsivität, Bedürfnisse, Psychopathologie (externalisierendes/internalisierendes Verhalten)
 - c. Dyade: Kooperation, Reziprozität, shared attention, Synchronizität, Match and Repair, Konfliktlösung, führendes Verhalten, Machtverhalten
- 4) Beobachtungsplanung
 - a. Aufklärung sowie Instruktion vor der Interaktion
 - b. Auswahl von Ort/Räumlichkeiten (Zuhause, Klinik, Spielzimmer, Besprechungsraum, Draußen)
 - c. Auswahl des Settings (z. B. freies Spiel, Eltern- oder Kind geführte Interaktion, während des Elterngesprächs, Alltagssituation)
 - d. Wahl der Beobachtungsdokumentation (Protokollbogen, Itemsammlung, Memo, Freitext)
- 5) Umsetzung
 - a. Anzahl und Beschreibung der Settings, Beobachtungsbeginn und Ende, Vertrautheit
 - b. Anwesende Personen während der Eltern-Kind-Interaktion
 - c. Teilnehmend (Passive Rolle vs. Interaktionspartner), nicht-teilnehmende Beobachtung
 - d. Video- und/oder Audioaufzeichnung
- 6) Auswertung
 - a. Qualitativ: Freitext, episodische Beschreibung, Deutung
 - b. Quantitativ: Skalenwerte, Kategoriale Einordnung
- 7) Befund bzw. Beantwortung der Fragestellung und Einordnung im Rahmen einer übergeordneten Fragestellung (z. B. Diagnosestellung; Therapieziele)

3. Testkonstruktion

Das MKI-Projekt steht als Abkürzung für Mutter-Kind-Interaktion, wobei inhaltlich die Eltern-Kind-Interaktion gemeint ist, und zielt auf die Entwicklung, Überprüfung und Verbesserung von Beobachtungsinstrumenten zur Erfassung verschiedener Aspekte der Eltern-Kind-Beziehung ab. Zu Beginn unserer Forschungsarbeiten haben wir PIRGAS (Zero To Three; 2005) aus dem DC:0-3R erprobt, das jedoch zahlreiche inhaltliche und methodische Einschränkungen aufweist (Müller et al., 2013) und nur über ein einziges globales Item erfasst wird, weshalb keine Fehlerkompensation erfolgt, die nur mit mehreren Indikatoren bzw. Items zu einem Konstrukt erreicht werden kann. Dies gilt in vergleichbarer Weise für die Relationship Problem Checklist (RPCL), die zudem nur binär kodierbar ist (siehe auch Baans et al., 2023). Insgesamt fokussiert die DC:0-3R inhaltlich auf negatives Elternverhalten und es fehlen im Manual klare Hinweise, in welchem Setting (Wie viele Personen dürfen anwesend sein?) oder mit welchen Vorgaben (Dürfen Eltern und ihr Kind ohne Vorgaben ‚frei spielen‘ oder sollen sie Aufgaben des Testleiters umsetzen?) Verhaltensbeobachtungen instruiert werden sollten.

Im Rahmen des MKI-Projekts erschien es uns zielführend, weitere englischsprachige Instrumente zu übersetzen und für den deutschsprachigen Raum verfügbar zu machen. Im Forschungskontext spielt die Etablierung bzw. Verbreitung eines Verfahrens eine wichtige Rolle, um vergleichende Studien sowie die Bündelung empirischer Erfahrungen mit den oft recht überschaubaren Stichprobengrößen in einer konvergenten Weiterentwicklung zu ermöglichen.

Grundlage der Entwicklung des SIRS von McCall war seine Sichtung verfügbarer Instrumente, beispielsweise das Home Observation for Measurement of the Environment (HOME, Bradley & Caldwell, 1995; Caldwell & Bradley, 1984), welches jedoch einen zu starken inhaltlichen Fokus auf den Umgebungskontext beinhaltet, sowie methodische Mängel durch ein binäres Antwortformat aufwies, welche zudem für Gruppensettings mit mehreren anwesenden Personen konzipiert waren. Ähnliches gilt für das Inventory and the Early Childhood Environmental Rating Scale, Revised edition (ECERS), die Infant/Toddler Childhood Environmental Rating Scale (ITERS) oder die Family Day Care Rating Scale (FDCRS; Harms & Clifford, 1989; Harms, Clifford & Cryer, 1998; Harms, Cryer & Clifford, 1990).

Das SIRS von McCall war eines der ganz wenigen Beobachtungsinstrumente, das zumindest für die Testentwicklung bereits eine explorative Faktorenanalyse zur Untersuchung der psychometrischen Struktur nutzte. Die Anwendungsbreite des SIRS sowie seine einfache Anwendbarkeit machten die Skalen im Kontext einer Familientagesklinik für Kinder im Alter von 0 bis 6 Jahren interessant, zudem erfasst er in Ergänzung zum Play-Pab (Wagner et al., 2015; Manual und Kodierbogen siehe Wagner et al., 2017) auch explizit das Verhalten von Kindern. Ziel dieser Veröffentlichung ist es, das von uns übersetzte und damit deutschsprachige Beobachtungsinstrument SIRS von McCall sowohl Forschern als auch Praktikern frei zugänglich zu machen, um Erfahrungen und Weiterentwicklungen zu ermöglichen und es anhand der oben genannten Kriterien umfassender als bisher zu beschreiben.

In diesem Papier berichten wir über unsere Erfahrungen und versuchen, eine Einschätzung über mögliche Einsatzmöglichkeiten und Grenzen des SIRS zu geben. Für die Bewertung eines Erhebungsinstruments ist eine inhaltliche Verortung im Lichte von Übersichtsarbeiten wie der von Skinner (2005) relevant, insbesondere die strukturelle Validität mit dimensionaler Analyse über Faktorenanalysen oder modellprüfende Verfahren einschließlich einer Beschreibung und Bewertung der Skaleneigenschaften. Schließlich wäre für die praktische Anwendung die

Normierung als Interpretationshilfe für die einzelnen Testergebnisse von Bedeutung, die jedoch unseres Wissens bislang für kein Beobachtungsverfahren im klinischen Kontext der Eltern-Kind-Interaktion vorliegt und hier auch nicht geleistet werden kann.

Übersetzung

Die Itembeschreibungen wurden aus der Originalpublikation 2017 von Sandra Achtergarde, Thomas Meier, Jörg M. Müller und Olena Skorozenina mehrfach ins Deutsche übersetzt und auf Übersetzungsfehler rückübersetzt. Abweichungen wurden inhaltlich diskutiert und die aus unserer Sicht beste Variante ausgewählt.

Itembeschreibung, Antwortskala, Instruktion im Manual

Insgesamt enthält der SIRS 18 ‚Constructs to be rated‘ (siehe Textbox 3).

TEXTBOX 3: Die 18 ‚Constructs to be rated‘ respektive Items des SIRS von McCall (2010)

Elternbereich

1. Distanziertheit der Bezugsperson (-)
2. Bezugsperson reagiert nicht (-)
3. Mentale Verfügbarkeit/Empfänglichkeit der Bezugsperson in Bezug auf Kinder (+)
4. Bezugsperson reagiert auf das Kind (+)
5. Bezugsperson führt Kind (+)
6. Bezugsperson lässt Kind führen (+)
7. Intrusivität der Bezugsperson (-)
8. Verhaltens-Kontrolle/-Gehorsam/-Disziplin der Bezugsperson (+)
9. Bezugsperson bestraft das Kind (-)
10. Unterstützung/Empathie/Anleitung der Bezugsperson (+)
11. Negativer Affekt der Bezugsperson (-)
12. Bezugsperson zeigt positiven Affekt und Zuneigung (+)
13. Lebhaftigkeit und Ausdruckskraft der Bezugsperson (+)

Kindbereich

14. Emotionale Nicht-Responsivität des Kindes (-)
15. Kindliche Responsivität/Positive Erwartung (+)
16. Negative Gefühle des Kindes (-)
17. Positive Gefühle des Kindes (+)
18. Beziehung des Kindes zur Bezugsperson (+)

Hierbei ist anzumerken, dass McCall vom üblichen Sprachgebrauch abweicht, da latente Konstrukte prinzipiell nicht beobachtbar sind und daher auch nicht global beurteilt werden können, sondern über mehrere Indikatoren erfasst werden müssen. Fachlich wird daher im Folgenden von 18 Items bzw. Indikatoren gesprochen, die im Manual anhand von Verhaltensbeispielen näher beschrieben werden. In der Textbox ist in Klammern angegeben, ob das Item einen negativ bewertenden Inhalt beschreibt und dazu führt, dass die Antworten zu diesem Item vor einer Auswertung umgepolt werden müssen.

Jedes Item wird durch eine verhaltensnahe Beschreibung näher beschrieben und die Bewertung erfordert in jedem Fall eine Integrations- und Bewertungsleistung des Beurteilers und geht somit über ein einfaches „Zählen“ von Verhaltensweisen hinaus. Grundsätzlich wird jedes ‚Item‘ auf einer 4-Punkte-Skala (0 = nie, 1 = selten, 2 = häufig, 3 = immer) bewertet, die die relative Häufigkeit widerspiegelt, mit der das jeweilige Verhalten unter den potentiellen Gelegenheiten, bei denen es hätte auftreten können, aufgetreten ist. Die Antwortskala basiert somit nicht auf einer konkreten Auszählung beobachteter Verhaltensweisen, sondern entspricht einer kontextbezogenen Einschätzung. Dieser Ansatz hat sowohl Vor- als auch Nachteile. So kann der SIRS beispielsweise auch dann gut eingesetzt werden, wenn der Beobachtungszeitraum variiert. Das Handbuch gibt auch nur grobe Empfehlungen für das Beobachtungssetting, d. h. wer ist anwesend, wo findet die Beobachtung statt, welche Instruktionen wurden den beobachteten Teilnehmern vorher gegeben. Der Nachteil ist, dass diese Anpassung der Beurteilungsleistung möglicherweise zu Lasten der Vergleichbarkeit, der Genauigkeit und damit der Aussagekraft des SIRS gehen könnte. Zukünftige Studien müssen jedoch noch untersuchen, welchen Einfluss die Anzahl, das Setting sowie die Dauer der Beobachtung auf die Genauigkeit und Aussagekraft haben (siehe Murphy & Hall, 2021).

Psychometrische Bewertung der Originalskalen

Nachdem unter „Auswertungsmodus“ eine inhaltliche Bewertung der eltern- und kindbezogenen Items und Skalenbezeichnungen vorgenommen wurde, soll nun die psychometrische Bewertung ergänzt werden. Zusammenfassend lässt sich feststellen, dass die Items vereinzelt Schwächen in Bezug auf Deckeneffekte aufweisen, die auch in der Originalpublikation von McCall zu finden sind. Die insgesamt 10 Skalen in drei Auswertungsvarianten können je nach Fragestellung eingesetzt werden. Der Summenscore hat den Vorteil, dass er aufgrund der hohen Itemanzahl die höchste Messgenauigkeit aller Auswertungsvarianten aufweist. Dies kann helfen, Verbesserungseffekte bei kleineren Stichproben nachzuweisen, ohne jedoch einzelne Aspekte der Interaktion abbilden zu können. Der Summenscore weist nach unseren Beobachtungen eine hohe Sensitivität und eine befriedigende Spezifität auf. Diese Vorteile gehen zu Lasten der Einzelaspekte des SIRS, das mit den drei Skalen der Variante B vielleicht den besten Kompromiss zwischen Genauigkeit für die getrennten Inhalte bietet. Variante C mit den sechs Skalen plus dem Einzelitem 18 liefert die inhaltlich detaillierteste Beschreibung, allerdings ist die Anzahl der Items mit teilweise nur zwei Items so gering, dass die Messgenauigkeit zu sehr darunter leidet. Im Vergleich der Auswertungsvarianten mittels einer Strukturgleichung schneidet die dritte Variante C mit sechs Skalen am besten ab, zeigt aber auch, dass die Skalen bzw. die Modelle insgesamt einer Überarbeitung bedürfen. Die Beurteilerübereinstimmung ist zufriedenstellend und die Cut-off-Werte können diagnostische Entscheidungen unterstützen.

4. Gütekriterien

4.1 Objektivität

Stichproben

Zur psychometrischen Überprüfung des SIRS wurden die Ratings von 86 Dyaden aus der Inanspruchnahmepopulation der Familientagesklinik für Säuglinge, Klein- und Vorschulkinder der Klinik für Kinder- und Jugendpsychiatrie, -psychosomatik, und -psychotherapie (Furniss et

al., 2013; Postert et al., 2014) des Universitätsklinikums Münster sowie 30 nicht-klinische Fälle umfasst, womit insgesamt $N = 116$ den Analysen zugrunde lagen. Von den teilnehmenden Erwachsenen waren 99 % weiblich, das Alter der Eltern lag zwischen 18 und 56 Jahren. Das Durchschnittsalter der teilnehmenden Kinder betrug $M = 4.61$ Jahre ($SD = 1.61$; 61.21 % Jungen). Die häufigste klinische Diagnose war „Sonstige emotionale Störung des Kindesalters“ (F93.8).

Prozedere und Beobachtungssettings

Nach Aufklärung und Einholung der Einwilligungserklärung wurden die Eltern und deren Kind in einem kind- bzw. altersgerechten Spielzimmer für mindestens 15 Minuten und maximal 20 Minuten mittels einer fest installierten Videokamera in ihrer Interaktion aufgenommen. Die Interaktion wurde mit Beginn der Interaktion im Raum beurteilt. Grundsätzlich handelte es sich bei unserem Setting um eine Freispielsituation von Elternteil und Kind, bei der sich nur die Eltern im Raum aufhielten.

Ratertraining

Die insgesamt vier Rater (je zwei für unterschiedlichen Beobachtungsdauern von 10 und 20 Minuten) durchliefen vor der Beurteilung ein kurzes Training, wobei jeder Rater dieselben fünf Fälle beurteilte und Abweichungen zu anderen Ratern besprochen wurde. Dies diente vor allem dazu, sich mit dem Bewertungsinstrument vertraut zu machen. Dieses Kurztraining bildet die Situation ab, in der Gutachter in der Praxis kaum Zeit für ein instrumentenspezifisches Training haben und kommt dem ‚learning by doing‘ in der Versorgungspraxis nahe. Unsere Umfrage (Müller et al., 2023) belegt, dass instrumentenspezifische Trainings eher die Ausnahme als die Regel sind. Zudem ist derzeit aufgrund fehlender empirischer Forschung unklar, ob Trainings im Bereich der Eltern-Kind-Interaktionsbeobachtung zu einer erhofften Steigerung der Reliabilität und Validität führen und welche Rolle diesbezüglich die Beobachtungsdauer, Setting sowie urteilerspezifische Einflüsse spielen.

Deskriptive univariate Itemanalyse

In Tabelle 1 und Tabelle 2 sind die Mittelwerte und Standardabweichungen für die SIRS-Items jeweils für eine Beobachtungszeit von 10 und 15 bis 20 Minuten für beide Rater und ihre aggregierten Urteile zusammen mit der Korrelation zwischen den Ratern A und B dargestellt.

Beobachtungsdauer

Alle Analysen und Interpretation hinsichtlich des Einflusses der Beobachtungsdauer auf die Ergebnisse im SIRS sind Teil der Doktorarbeit von Frau Annkathrin Angresius (2024).

Die Mittelwerte der SIRS Items in Tabelle 1 zeigt, dass beide Rater für die 10 Minuten Beobachtungszeit für 7 von 18 Items (1, 2, 7, 8, 9, 11 und 16) überwiegend positive (≥ 2.75 ; Deckeneffekt) oder negative ($\leq .25$) Bewertung abgeben, was für die 15-20-minütigen Beobachtungszeit in Tabelle 2 deutlich seltener auftritt. In der Originalpublikation von McCall 2010 fallen 9 von 18 Items ebenso durch Deckeneffekt auf und in der Interventionserhebung immerhin noch 6. Deckeneffekt sind auf Itemebene kein grundsätzliches Problem für die darauf basierenden Skalenwerte, allerdings wäre eine stärkere Streuung der Itemschwierigkeiten sicherlich wünschenswert, zudem sie keine idealen Voraussetzungen für bi- und multivariaten Analysen darstellen. Für die Bewertung der Eignung eines Items fehlt noch die Bewertung der

Interraterreliabilität, die Item-Skala Korrelation sowie innerhalb einer Faktorenanalyse bzw. einer Strukturprüfung, wie eindeutig die Items jeweils einer Skala zugeordnet werden können.

Insgesamt schneiden die Items hinsichtlich der Interraterreliabilität zwischen den Ratern im SIRS gut ab, wobei Item 5 sowohl für die kürzere als auch für die längere Beobachtungszeit nicht zufriedenstellend genau beurteilt werden kann. Hinsichtlich der Interrater-Korrelationen unterscheiden sich die Ergebnisse auf Itemebene überraschenderweise nur wenig, d. h. eine Verlängerung der Beobachtungsdauer führt nur zu einer geringen Verbesserung der Reliabilität.

Tabelle 1

Mittelwert, Standardabweichung und die Korrelation zwischen den Ratern A und B für 10 Minuten Beobachtungsdauer für die 18 Items des SIRS von McCall für N = 116 gemischt klinische und nicht-klinische Dyaden (Angresius, 2024)

Item	Rater A		Rater B		Gemittelte Urteile A+B		Corr (A, B)
	M	SD	M	SD	M	SD	r
BP Distanziertheit	2.86	0.35	2.81	0.51	2.83	0.37	.49
BP reagiert nicht	2.84	0.39	2.79	0.47	2.81	0.38	.53
BP Mentale Verfügbarkeit	2.66	0.55	2.67	0.55	2.66	0.48	.54
BP reagiert auf das Kind	2.66	0.49	2.68	0.56	2.67	0.44	.41
BP führt Kind	2.16	0.75	2.45	0.71	2.30	0.51	-.02
BP lässt Kind führen	2.33	0.60	2.49	0.81	2.41	0.60	.44
BP Intrusivität	2.93	0.28	2.70	0.54	2.81	0.36	.48
BP Verhaltens-Kontrolle	2.69	0.49	2.76	0.53	2.73	0.43	.45
BP bestraft das Kind	2.97	0.15	2.96	0.23	2.97	0.15	.17
BP Unterstützung	2.39	0.62	2.40	0.75	2.39	0.61	.55
BP Negativer Affekt	2.85	0.40	2.84	0.43	2.84	0.39	.77
BP positiven Affekt	2.34	0.58	1.89	0.77	2.11	0.58	.47
BP Lebhaftigkeit	2.15	0.64	1.63	0.74	1.89	0.61	.53
Kind Responsivität	2.91	0.27	2.47	0.64	2.69	0.38	.30
Kind Positive Erwartung	2.46	0.55	2.37	0.69	2.41	0.55	.25
Kind Negative Gefühle	2.90	0.31	2.87	0.39	2.88	0.31	.30
Kind Positive Gefühle	2.31	0.56	2.07	0.74	2.19	0.56	.50
Kind BP Beziehung	2.22	0.53	2.32	0.75	2.27	0.55	.45
Arithmetic Mean	2.60	0.52	2.51	0.64	2.55	0.50	.45

Anmerkungen. BP = Bezugsperson; Scoring auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer. Items 1, 2, 5, 7-9, 11, 14, und 16 stellen Fragen nach negativen Verhaltensweisen und wurde gespiegelt, sodass für alle Items gilt, dass hohe Werte erwünschtes Verhalten darstellt.

Tabelle 2

Mittelwert, Standardabweichung und die Korrelation zwischen den Ratern A und B für 15-20 Minuten Beobachtungsdauer für die 18 Items des SIRS von McCall für N = 116 gemischt klinische und nicht-klinische Dyaden

Item	Rater A		Rater B		Gemittelte Urteile A+B		Corr (A, B)
	M	SD	M	SD	M	SD	r
1. BP Distanziertheit	2.63	0.73	2.57	0.78	2.60	0.64	.42
2. BP reagiert nicht	2.47	0.86	2.41	0.86	2.44	0.74	.48
3. BP Mentale Verfügbarkeit	2.53	0.69	2.31	0.89	2.42	0.69	.49
4. BP reagiert auf das Kind	2.47	0.79	2.30	0.94	2.38	0.73	.43
5. BP führt Kind	1.47	0.82	1.94	1.02	1.70	0.73	.23
6. BP lässt Kind führen	1.93	0.83	2.33	0.87	2.13	0.74	.51
7. BP Intrusivität	2.64	0.76	2.75	0.53	2.69	0.56	.52
8. BP Verhaltens-Kontrolle	2.16	0.89	2.64	0.64	2.40	0.66	.45
9. BP bestraft das Kind	2.78	0.55	2.93	0.33	2.85	0.39	.59
10. BP Unterstützung	2.17	0.97	2.04	1.06	2.11	0.90	.54
11. BP Negativer Affekt	2.62	0.78	2.75	0.52	2.69	0.57	.53
12. BP positiven Affekt	2.27	0.80	1.78	0.85	2.02	0.73	.55
13. BP Lebhaftigkeit	2.08	0.77	1.61	0.85	1.85	0.68	.41
14. Kind Responsivität	2.46	0.87	2.24	0.89	2.35	0.76	.48
15. Kind Positive Erwartung	2.31	0.70	2.24	0.85	2.27	0.66	.46
16. Kind Negative Gefühle	2.75	0.58	2.75	0.51	2.75	0.49	.59
17. Kind Positive Gefühle	2.19	0.65	1.91	0.87	2.05	0.66	.50
18. Kind BP Beziehung	2.19	0.64	2.27	0.83	2.23	0.62	.39
Arithmetic Mean	2.45	0.72	2.46	0.74	2.46	0.63	.48

Anmerkungen. Scoring auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer. Items 1, 2, 5, 7-9, 11, 14, und 16 stellen Fragen nach negativen Verhaltensweisen und wurde gespiegelt, sodass für alle item gilt, dass hohe Werte positives Verhalten darstellt. Farblich unterliegt sind Mittelwerte, welche einen deutlichen Deckeneffekt ($M \geq 2.75$; $SD \leq .25$) anzeigen.

4.2 Reliabilität

Die Überprüfung der Zuordnung von Items zu Skalen nach McCall mittels Strukturgleichungsmodellen des SIRS findet auf der Basis aggregierter Urteile statt, da zu diesem Zeitpunkt vor allem die korrekte Zuordnung von Items zu Skalen interessiert und Unterschiede zwischen den Ratern in ihrem Effekt ausgeblendet werden sollen. Für den praktischen Kontext wäre im Weiteren wichtig darauf hinzuweisen, dass die folgenden Ergebnisse entsprechend nicht auf einzelne Rater übertragbar sind, da die aggregierten Urteile deutlich stabilere und reliablere Urteile darstellen.

Deskriptive bivariate Itemanalyse

Bevor im Folgenden die Messstruktur mittels Faktorenanalysen bzw. Strukturgleichungen untersucht wird, sollen die Grundlagen dieser Analysen in Form von bivariaten Korrelationen dargestellt werden. Zur Orientierung sei vorab gesagt, dass die Items innerhalb einer McCall-Skala grundsätzlich relativ hoch korrelieren sollten, während die Korrelation eines Items mit einer anderen Skala deutlich niedriger sein sollte. Da die Konstrukte erwartungsgemäß miteinander korrelieren, wären für konvergente Korrelationen ca. $r > .60$ günstig, während divergente Korrelationen ca. $.40$ nicht überschreiten sollten.

Die Korrelationen zwischen den Items in Tabelle 3 zeigen eine sehr große Streubreite von $r = .00$ bis $r = .83$ für die 10 Minuten Beobachtungen und von $r = .14$ bis $r = .91$ für die 15-20-minütigen Beobachtungen. Im Schnitt korrelieren bei 15-20 Minuten Urteilen die Items mit $r = .54$ deutlich höher als bei 10 Minuten ($r = .42$). Beide Varianten bieten eine gute Grundlage für die Bildung mehrerer Skalen, deren deskriptive Eigenschaften im Folgenden beschrieben werden.

Tabelle 3

Korrelation zwischen den SIRS-Items für 10 (oberes Dreieck) und 15-20 Minuten (unteres Dreieck)

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	.	.63	.80	.66	.25	.66	.30	.46	.28	.56	.48	.56	.42	.44	.34	.10	.36	.52
2	.85	.	.74	.71	.31	.69	.40	.68	.51	.67	.70	.59	.47	.43	.29	.26	.30	.42
3	.78	.86	.	.83	.34	.80	.46	.61	.47	.76	.68	.69	.57	.37	.39	.27	.36	.60
4	.80	.91	.88	.	.29	.76	.40	.56	.38	.73	.62	.63	.51	.40	.41	.40	.38	.65
5	.52	.64	.68	.61	.	.40	.34	.48	.26	.07	.37	.17	.15	.44	.00	.16	.07	.17
6	.71	.76	.82	.75	.83	.	.49	.52	.35	.74	.58	.59	.47	.43	.57	.41	.52	.72
7	.55	.66	.70	.63	.61	.64	.	.53	.58	.50	.69	.36	.22	.19	.22	.39	.15	.36
8	.59	.65	.68	.63	.64	.68	.71	.	.62	.50	.75	.45	.37	.42	.16	.22	.13	.34
9	.56	.60	.58	.57	.47	.55	.74	.75	.	.40	.72	.37	.32	.21	.13	.08	.10	.26
10	.71	.81	.87	.84	.61	.76	.62	.62	.55	.	.60	.67	.58	.20	.48	.32	.42	.62
11	.60	.69	.71	.68	.60	.68	.81	.81	.82	.71	.	.53	.44	.31	.22	.32	.20	.39
12	.68	.75	.73	.80	.45	.63	.47	.50	.45	.76	.59	.	.67	.14	.34	.21	.39	.55
13	.54	.49	.53	.58	.14	.41	.20	.30	.26	.52	.31	.73	.	.12	.24	.16	.33	.37
14	.41	.46	.49	.40	.44	.55	.33	.35	.28	.43	.32	.29	.28	.	.38	.18	.25	.47
15	.53	.49	.54	.51	.48	.65	.39	.40	.33	.54	.40	.40	.36	.80	.	.26	.72	.66
16	.23	.37	.34	.39	.37	.40	.32	.39	.37	.30	.25	.27	.19	.38	.35	.	.23	.33
17	.58	.56	.55	.58	.40	.60	.32	.36	.38	.50	.36	.55	.49	.74	.74	.43	.	.58
18	.59	.57	.65	.59	.54	.70	.50	.57	.52	.56	.54	.51	.46	.74	.81	.42	.74	.

Deskriptive univariate Beschreibung der Skalen von McCall in den Auswertungsvarianten A-C

In Tabelle 4 sind die Mittelwerte der SIRS-Skalen für alle drei Auswertungsvarianten A, B und C sowie die Standardabweichungen pro Rater und deren Aggregation mit der Interrater-Korrelation für 10 Minuten und in Tabelle 5 für 15 bis 20 Minuten Beobachtungszeit dargestellt,

wobei eine Unterteilung der Gesamtstichprobe in klinische und nicht-klinische Gruppen noch einmal ergänzt wird. Betrachtet man zunächst allein die Mittelwerte der Skalen im Range der Antwortenkodierung, so zeigen alle Dyaden im Durchschnitt insgesamt ein recht positives Elternverhalten. Wie erwartet zeigt die klinische Stichprobe durchgängig geringere Scores ungeachtet der Skala, sodass die Dyade insgesamt in der klinischen Stichprobe (Total score = 2.19) negativer bewertet werden als jene der nicht-klinischen Stichprobe (Total score = 2.75).

Tabelle 4

Mittelwerte und Standardabweichungen der Skalen des SIRS von McCall je Rater und deren Aggregat sowie die Interrater- Korrelation für 10 Minuten Beobachtungsdauer einer deutschen Stichprobe für die Auswertungsvariante A, B, C

		Rater A		Rater B		Rater A+B		
		M	SD	M	SD	M	SD	r _{A,B}
Total Sample N = 116								
1	Auswertungsvariante A Total score	2.59	0.30	2.51	0.39	2.55	0.32	.65
Auswertungsvariante B								
2	Caregiver Positive Engagement	2.53	0.36	2.37	0.46	2.45	0.37	.63
3	Caregiver Punitiveness	2.91	0.21	2.89	0.27	2.90	0.22	.73
4	Child-Caregiver directed interaction	2.53	0.37	2.60	0.49	2.56	0.37	.45
Auswertungsvariante C								
5	Anteilnahme der BP	2.75	0.38	2.74	0.44	2.74	0.37	.64
6	BP /Kind-interagierendes Verhalten	2.24	0.51	2.47	0.69	2.36	0.46	.18
7	Verhaltenskontrolle der BP	2.75	0.31	2.71	0.38	2.73	0.31	.63
8	Erregungszustand (Affekt) der BP	2.44	0.45	2.12	0.52	2.28	0.44	.65
9	Kontakterwartung und Responsivität des Kindes	2.68	0.37	2.42	0.54	2.55	0.39	.49
10	Gefühle des Kindes	2.60	0.36	2.47	0.44	2.54	0.35	.56
11	Beziehung zum Kind	2.22	0.53	2.32	0.75	2.27	0.55	.45
- Tabelle wird fortgeführt -								

		- Tabelle fortgeführt -					
		Rater A		Rater B		Rater A+B	
		M	SD	M	SD	M	SD
Clinical Sample N = 86							
1	<i>Auswertungsvariante A</i> Total score	2.54	0.32	2.43	0.41	2.48	0.33
<i>Auswertungsvariante B</i>							
2	Caregiver Positive Engagement	2.45	0.37	2.28	0.48	2.36	0.38
3	Caregiver Punitiveness	2.88	0.23	2.85	0.30	2.87	0.25
4	Child-Caregiver directed interaction	2.52	0.41	2.52	0.52	2.52	0.40
<i>Auswertungsvariante C</i>							
5	Anteilnahme der BP	2.70	0.42	2.67	0.49	2.68	0.41
6	BP /Kind-interagierendes Verhalten	2.27	0.56	2.38	0.74	2.33	0.50
7	Verhaltenskontrolle der BP	2.69	0.33	2.64	0.42	2.66	0.34
8	Erregungszustand (Affekt) der BP	2.37	0.47	2.09	0.54	2.23	0.46
9	Kontakterwartung und Responsivität des Kindes	2.61	0.39	2.33	0.53	2.47	0.40
10	Gefühle des Kindes	2.53	0.36	2.37	0.46	2.45	0.36
11	Beziehung zum Kind	2.15	0.55	2.15	0.77	2.15	0.55
Control Sample N = 30							
1	<i>Auswertungsvariante A</i> Total score	2.74	0.16	2.73	0.20	2.74	0.16
<i>Auswertungsvariante B</i>							
2	Caregiver Positive Engagement	2.76	0.20	2.62	0.25	2.69	0.21
3	Caregiver Punitiveness	2.98	0.06	2.99	0.03	2.99	0.03
4	Child-Caregiver directed interaction	2.54	0.24	2.82	0.27	2.68	0.21
<i>Auswertungsvariante C</i>							
5	Anteilnahme der BP	2.90	0.16	2.93	0.13	2.92	0.12
6	BP /Kind-interagierendes Verhalten	2.16	0.36	2.73	0.44	2.44	0.32
7	Verhaltenskontrolle der BP	2.91	0.14	2.90	0.13	2.91	0.11
8	Erregungszustand (Affekt) der BP	2.66	0.32	2.21	0.44	2.44	0.36
9	Kontakterwartung und Responsivität des Kindes	2.89	0.16	2.69	0.48	2.79	0.26
10	Gefühle des Kindes	2.82	0.23	2.74	0.22	2.78	0.18
11	Beziehung zum Kind	2.43	0.41	2.82	0.38	2.63	0.36

Anmerkungen. BP = Bezugsperson; Scoring auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer. Alle Scores zeigen erwünschtes Verhalten mit höheren Scores.

Tabelle 5

Mittelwerte und Standardabweichungen der Skalen des SIRS von McCall je Rater und deren Aggregat sowie die Interrater- Korrelation für 15-20 Minuten Beobachtungsdauer einer deutschen Stichprobe für die Auswertungsvarianten A, B, C

		Rater A		Rater B		Rater A+B		r _{A,B}
		M	SD	M	SD	M	SD	
Total Sample N = 116								
1	Auswertungsvariante A Total score	2.34	0.54	2.32	0.58	2.33	0.51	.65
Auswertungsvariante B								
2	Caregiver Positive Engagement	2.34	0.57	2.15	0.71	2.25	0.57	.61
3	Caregiver Punitiveness	2.72	0.53	2.81	0.32	2.76	0.39	.65
4	Child-Caregiver directed interaction	2.05	0.69	2.41	0.64	2.23	0.59	.57
Auswertungsvariante C								
5	Anteilnahme der BP	2.52	0.71	2.40	0.79	2.46	0.66	.54
6	BP /Kind-interagierendes Verhalten	1.70	0.75	2.14	0.90	1.92	0.70	.44
7	Verhaltenskontrolle der BP	2.44	0.68	2.59	0.51	2.51	0.54	.66
8	Erregungszustand (Affekt) der BP	2.32	0.63	2.05	0.60	2.19	0.56	.63
9	Kontakterwartung und Responsivität des Kindes	2.39	0.73	2.24	0.78	2.31	0.67	.57
10	Gefühle des Kindes	2.47	0.52	2.33	0.55	2.40	0.49	.66
11	Beziehung zum Kind	2.19	0.64	2.27	0.83	2.23	0.62	.39
Clinical Sample N = 86								
1	Auswertungsvariante A Total score	2.22	0.57	2.17	0.60	2.19	0.52	
Auswertungsvariante B								
2	Caregiver Positive Engagement	2.20	0.59	1.97	0.73	2.09	0.57	
3	Caregiver Punitiveness	2.63	0.59	2.75	0.35	2.69	0.43	
4	Child-Caregiver directed interaction	1.94	0.76	2.28	0.68	2.11	0.63	
Auswertungsvariante C								
5	Anteilnahme der BP	2.39	0.77	2.20	0.83	2.30	0.69	
6	BP /Kind-interagierendes Verhalten	1.62	0.83	1.94	0.95	1.78	0.75	
7	Verhaltenskontrolle der BP	2.29	0.73	2.48	0.55	2.39	0.57	
8	Erregungszustand (Affekt) der BP	2.20	0.67	1.95	0.63	2.07	0.58	
9	Kontakterwartung und Responsivität des Kindes	2.23	0.79	2.07	0.82	2.15	0.70	
10	Gefühle des Kindes	2.37	0.56	2.19	0.57	2.28	0.51	
11	Beziehung zum Kind	2.12	0.70	2.07	0.87	2.09	0.64	
- Tabelle wird fortgeführt -								

- Tabelle wird fortgeführt -

		Rater A		Rater B		Rater A+B		r _{A,B}
		M	SD	M	SD	M	SD	
Control Sample N = 30								
1	Auswertungsvariante A Total score	2.70	0.16	2.22	0.57	2.75	0.17	
Auswertungsvariante B								
2	Caregiver Positive Engagement	2.74	0.19	2.20	0.59	2.67	0.22	
3	Caregiver Punitiveness	2.97	0.07	2.63	0.59	2.99	0.04	
4	Child-Caregiver directed interaction	2.37	0.23	1.94	0.76	2.80	0.23	
Auswertungsvariante C								
5	Anteilnahme der BP	2.91	0.16	2.39	0.77	2.95	0.10	
6	BP /Kind-interagierendes Verhalten	1.94	0.31	1.62	0.83	2.71	0.38	
7	Verhaltenskontrolle der BP	2.84	0.17	2.29	0.73	2.89	0.12	
8	Erregungszustand (Affekt) der BP	2.69	0.29	2.20	0.67	2.33	0.40	
9	Kontakterwartung und Responsivität des Kindes	2.84	0.18	2.23	0.79	2.73	0.36	
10	Gefühle des Kindes	2.75	0.19	2.37	0.56	2.73	0.17	
11	Beziehung zum Kind	2.39	0.38	2.12	0.70	2.83	0.31	

Anmerkungen. BP = Bezugsperson; Scoring auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer. Alle Scores zeigen erwünschtes Verhalten mit höheren Scores.

Graphischer Vergleich zwischen klinischen und nicht-klinische Verteilungen der Scores auf den Skalen von McCall in den Auswertungsvarianten A-C

Einen besseren Eindruck als die Mittelwerte vermittelt ein Histogramm in ABBILDUNG 1. Hier ist deutlich zu erkennen, dass die Kontrollstichprobe überwiegend positives Elternverhalten zeigt, während die Streuung in der klinischen Stichprobe erheblich vergrößert ist und fast die gesamte Bandbreite der Antwortskala nutzt. Die resultierende Effektgröße (standardisierte Darstellung von Gruppenunterschieden mit Cohen d) ist unten gesondert für alle Skalen beschrieben.

Die Skalen unterscheiden sich hinsichtlich der Darstellung von Unterschieden recht erheblich, wie aus den folgenden ABBILDUNG 2-3 ersichtlich ist. Beispielsweise streuen die Wert in der Skala BP/Kind-interagierendes Verhalten sehr stark, während fast alle Dyaden auf der Skala *Caregiver Punitiveness* recht positiv abschneiden. Entsprechend zeigen die Eltern unserer Stichprobe kaum strafendes Verhalten.

ABBILDUNG 1. Verteilung des Total score in der Auswertungsvariante A des SIRS von McCall bei für 15-20 Minuten Beobachtungsdauer für die gemittelten Urteile auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer.

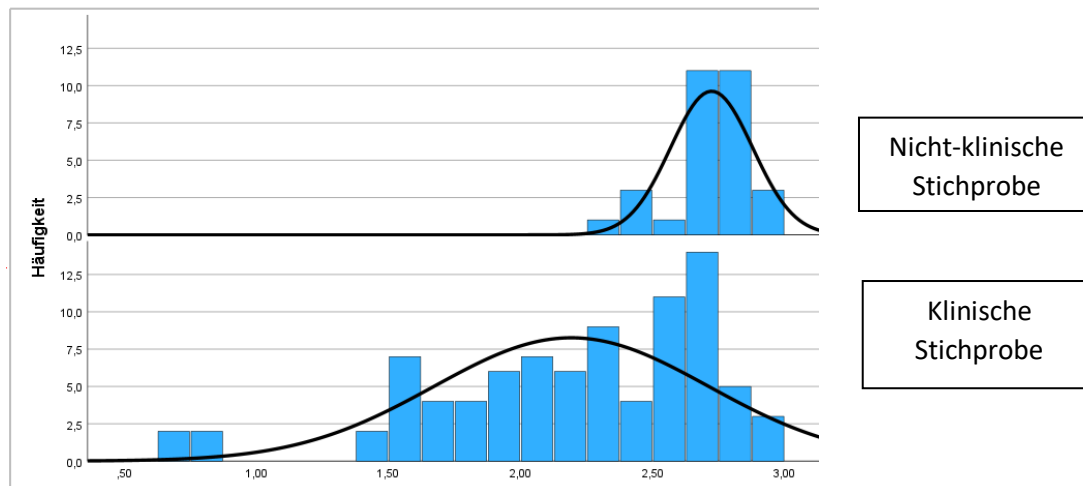


ABBILDUNG 2. Verteilung für die gemittelten Urteile auf vierstufiger Antwortskala (0 = nie, 1 = selten, 2 = häufig, 3 = immer) des SIRS von McCall bei für 15-20 Minuten Beobachtungsdauer in der Auswertungsvariante B für die drei Skalen Caregiver Positive Engagement (B1), Caregiver Punitiveness (B2) und Child-Caregiver directed interaction (B3).

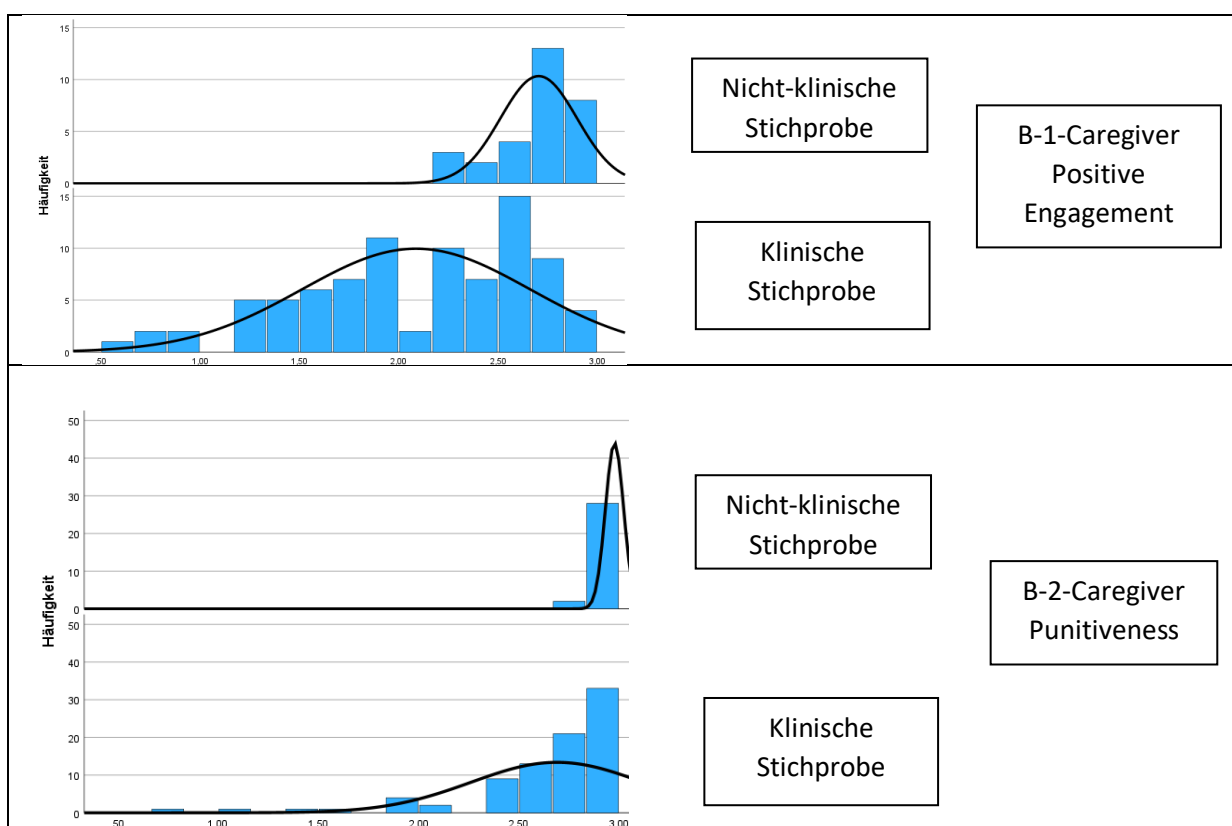
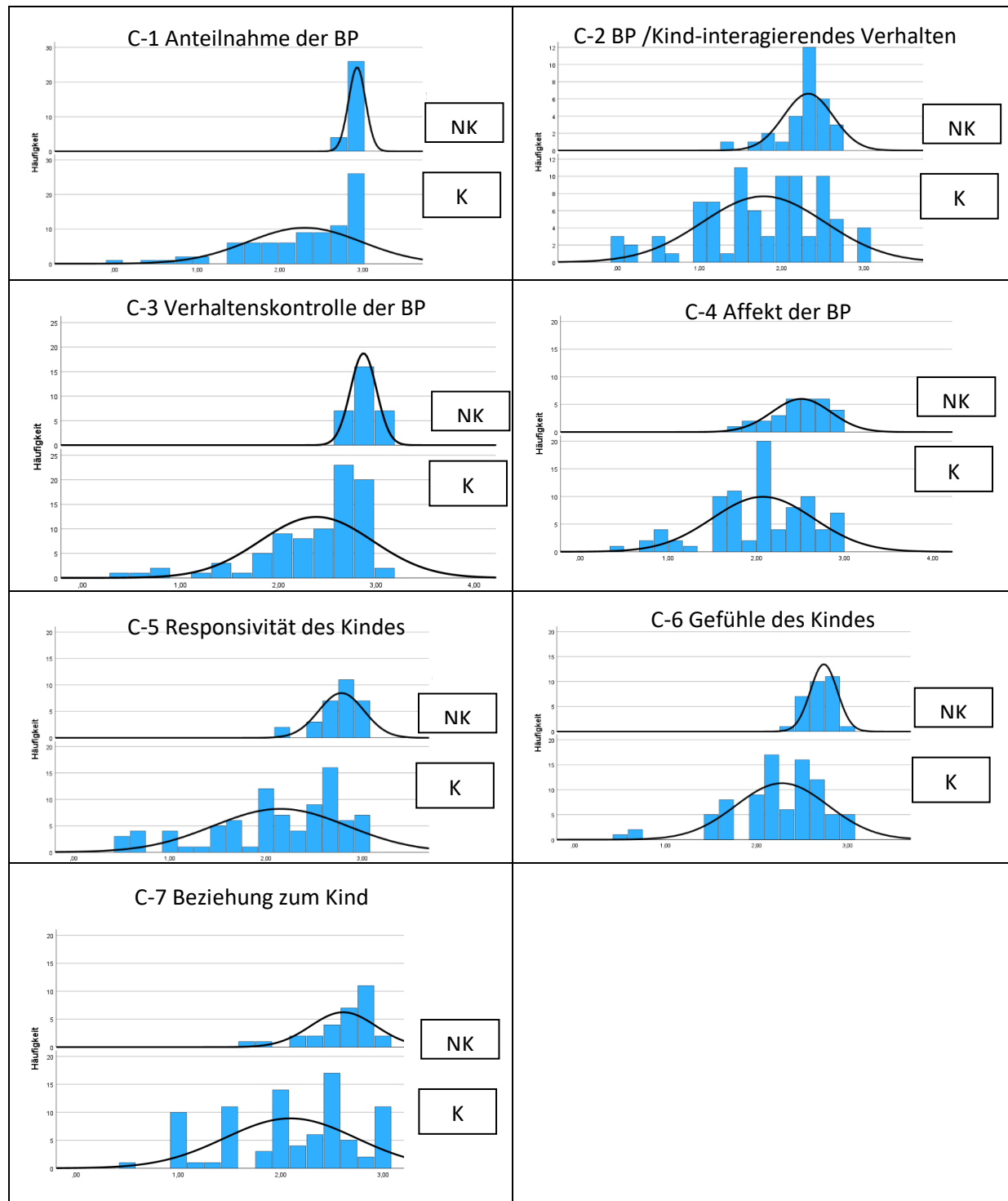


ABBILDUNG 3. Verteilung der für die gemittelten Urteile auf vierstufiger Antwortskala (0 = nie, 1 = selten, 2 = häufig, 3 = immer) des SIRS von McCall bei für 15-20 Minuten Beobachtungsdauer für die klinische (K) und nicht-klinische (NK) Stichprobe in der Auswertungsvariante C.



Beobachtungsdauer und Scores auf den Skalen von McCall in den Auswertungsvariante A-C

Wie bei den Items zeigt sich der Einfluss der Beobachtungsdauer nun auch kumuliert in den reduzierten Mittelwerten der Skalen (siehe Tabelle 6). Dies bedeutet, dass mit zunehmender Beobachtungsdauer auch mehr unerwünschtes Elternverhalten erkannt wird und somit die Skalen in der längeren Beobachtung besser zwischen den Eltern differenzieren können. Am Beispiel des Summenscores sinkt der Mittelwert im Wertebereich 0-3 von 2.55 für die 10-minütige Beobachtung auf 2.33 für die 15-20-minütige Beobachtung. Betrachtet man dieses Ergebnis nochmals getrennt für die klinische und die nicht-klinische Gruppe, so fällt auf, dass insbesondere der Mittelwert in der klinischen Gruppe von 2.48 auf 2.19 sinkt, während in der nicht-klinischen Gruppe kein Einfluss der Beobachtungsdauer festzustellen ist (2.75 vs. 2.74). Die Beobachtungsdauer ist also insbesondere für eine klinische Stichprobe von größerer Relevanz. Dies zeigt sich wiederum in der stark erhöhten Standardabweichung, die von 0.33 auf 0.52 ansteigt. Auch hier sind die einzelnen Skalen unterschiedlich stark betroffen, was darauf hindeutet, dass unterschiedliche Inhalte auch unterschiedlich lange Beobachtungszeiten erfordern könnten (Angresius, 2024).

Tabelle 6

SIRS Skalen Interrater-Korrelation für 15-20 Minuten Beobachtungsdauer auf der Basis aggregierter Urteile zweiter Rater für N = 116 Dyaden für die Auswertungsvarianten A, B, C; Cohen's D für Mittelwertsunterschiede zwischen der klinischen (n = 86) und nicht-klinischen (n = 30) Stichprobe, die Probability of Distinct Test Scores nach Müller (2006a, b) sowie die kritische Differenz als Mindestabstand für einen signifikanten Unterschied zwischen zwei Testwerten (Angresius, 2024)

	Items	Mean	SD	Korr Rater A-B	d	Cronbach s Alpha	PDTS	Kritische Differenz
Auswertungsvariante A Total score	18	41.93	9.18	.65	1.23	.95	58.91	5.69
Auswertungsvariante B								
Caregiver Positive Engagement	11	24.71	6.29	.61	1.15	.95	61.83	3.90
Caregiver Punitiveness	3	8.29	1.17	.65	0.81	.71	19.33	1.74
Child-Caregiver directed interaction	4	8.93	2.35	.57	1.24	.89	45.23	2.16
Auswertungsvariante C								
Anteilnahme der BP	4	19.68	5.25	.54	1.09	.96	59.85	2.91
BP /Kind-interagierendes Verhalten	2	7.68	2.80	.44	1.38	.91	53.82	2.33
Verhaltenskontrolle der BP	4	20.09	4.33	.66	1.01	.85	36.36	4.65
Erregungszustand (Affekt) der BP	3	13.11	3.34	.63	0.48	.79	36.57	4.24
Kontakterwartung und Responsivität des Kindes	2	9.25	2.69	.57	0.92	.89	46.61	2.47
Gefühle des Kindes	2	9.59	1.94	.66	1.00	.58	17.68	3.49
Beziehung zum Kind	1	4.46	1.24	.39	1.29	-	-	-

Anmerkungen. BP = Bezugsperson; Scoring auf vierstufiger Antwortskala von 0 = nie, 1 = selten, 2 = häufig, 3 = immer; d = Cohen D für den Vergleich clinical vs. non-clinical sample.

Exkurs PDTs: Der deskriptive Testkennwert PDTs (Müller, 2006a, 2006b) gibt die durchschnittliche Wahrscheinlichkeit der Unterscheidung zwischen zwei zufällig ausgewählten Testergebnissen (Probability of Distinct Test Scores) einer Stichprobe an. Der Wert variiert zwischen 0 und 100 %, wobei bessere Tests einen höheren Wert bzw. eine höhere Wahrscheinlichkeit der Unterscheidung zwischen Testergebnissen erreichen. Der PDTs verwendet die kritischen Differenzen, um zu entscheiden, ob sich zwei Testergebnisse signifikant voneinander unterscheiden. Es wird ein vollständiger Vergleich aller möglichen Testpaare auf der Grundlage der Testergebnisse einer Stichprobe durchgeführt. Der Kennwert ist besser als Cronbachs Alpha geeignet, Einschränkungen in der Anwendung einer Skala darzustellen, wenn die Verteilung der Testwerte zu häufig ähnlichen Testergebnissen führt. Beispielsweise ist der PDTs-Wert niedriger, wenn die Verteilung der Testergebnisse Deckeneffekte aufweist. Die folgenden sprachlichen Bezeichnungen für die Bewertung eines PDTs-Scores für Skalen sind der Veröffentlichung von Müller et al. (2010) entnommen:

< 30 % = "sehr schlecht"
30 %-45 % = "schlecht"
45 %-60 % = "mäßig"
60 %-75 % = "gut"
75 %-90 % = "sehr gut"
> 90 % = "ausgezeichnet"

4.3 Validität

Cut-off Score mit Sensitivitäts- und Spezifitätswerten

Die Trennschärfe eines Testverfahrens zwischen klinisch auffälligen und unauffälligen Personen kann mit Hilfe von ROC-Kurven (Receiver Operating Characteristic) beschrieben werden (siehe ABBILDUNG 4 für den Summenscore, der verschiedene Einzelwerte als Cut-off-Werte mit den resultierenden Sensitivitäts- und Spezifitätswerten grafisch darstellt).

Exkurs Sensitivitäts- und Spezifitätswerte: Wenn eine klinisch auffällige Dyade durch ein Instrument/Skala zuverlässig identifiziert wird, ist die Sensitivität des Instruments/Skala hoch. In der Screening-Eingangsdiagnostik ist es wichtig, keinen klinisch relevanten Fall zu übersehen, daher ist die Sensitivität in diesem Kontext wichtiger als die Spezifität, die die Zuverlässigkeit angibt, dass eine durch einen SIRS-Score als auffällig identifizierte Dyade tatsächlich klinisch auffällig ist. Die falsch-positiv-Rate (positiv hier im Sinne von klinisch auffällig) wird hierbei als 1-Spezifität angegeben und diagnostisch sind möglichst geringe falsch-positive wünschenswert, da eine Eltern-Kind-Dyade beispielsweise auf Basis dieser Informationsgrundlage einen Therapieplatz oder eine Beratung erhalten könnte, obwohl sie diese nicht benötigen. Liegt ein SIRS-Wert über dem Cut-off, so weist der Score auf eine eher positive Eltern-Kind-Interaktion hin. Klinische Auffälligkeiten sind hier entsprechend seltener zu erwarten. Umgekehrt zeigen Werte gleich oder unter dem Cut-off an, dass die Wahrscheinlichkeit für klinische Auffälligkeiten oder ein Risiko hierfür ansteigt. Eine ROC Analyse zeigt in graphischer Form für jeden denkbaren Cut-off Wert auf einer Skala die jeweils resultierenden Sensitivitäts- und 1-Spezifitätswerte an. Je weiter diese Linien von der Diagonalen entfernt liegt, desto besser ist

das Instrument. Die Fläche unter der Kurve wird als AUC (Area-under-Curve) bezeichnet und entspricht grob der Reliabilität.

Zur Festlegung eines Cut-off auf Basis einer ROC Analyse (siehe ABBILDUNG 12) und in Hinblick auf einen Kompromiss zwischen hoher Sensitivität und Spezifität dient die Überlegung, dass innerhalb eines klinischen Samples zumindest grob auf der Basis von N = 981 Patienten aus vier klinischen Inanspruchnahme (siehe Baans et al., 2023, Tabelle 1) über den PIRGAS ca. 38 % als auffällig betrachtet werden. Wir setzen die Spezifität auf 80 %. In den nachfolgenden Tabellen 7 und 8 berichten wir unter diesen Annahmen einen entsprechender Cut-off für den Gesamtscore und Caregiver Positive Engagement, welcher bei Gleichheit oder Unterschreitung als Indikator für eine diagnostische Entscheidung verwendet werden kann. Sollte ein Anwender eine abweichende diagnostische Präferenz festlegen, kann aus den nachfolgenden Tabellen der für diese Fragestellung optimale Wert abgelesen werden.

ABBILDUNG 4. ROC für den SIRS Total Score für die klinische und nicht-klinische Gruppe (N = 116).

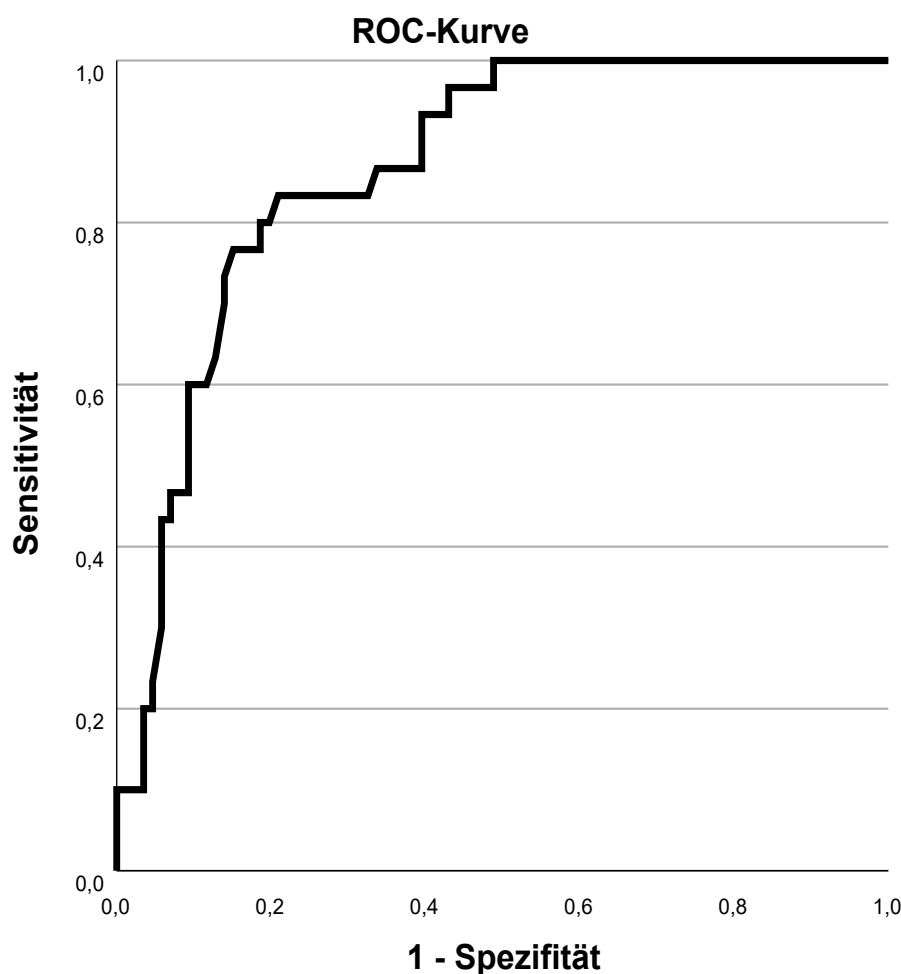


Tabelle 7

Cut-off-Score für den **Total Score** in der Auswertungsvariante A als Summenwert der Ratings mit Sensitivität und Spezifitätswerten (Angresius, 2024)

Summenwert	Sensitivität	1 - Spezifität	Summenwert	Sensitivität	1 - Spezifität
12	1.00	1.00	38	1.00	0.61
14	1.00	0.98	39	1.00	0.56
15	1.00	0.97	40	1.00	0.54
21	1.00	0.95	41	0.97	0.49
26	1.00	0.94	42	0.97	0.44
27	1.00	0.92	43	0.97	0.43
28	1.00	0.87	44	0.93	0.41
29	1.00	0.85	45	0.87	0.38
30	1.00	0.84	46	0.87	0.36
31	1.00	0.81	47	0.83	0.34
32	1.00	0.78	48	0.77	0.15
33	1.00	0.77	49	0.50	0.09
34	1.00	0.73	50	0.37	0.06
35	1.00	0.70	51	0.17	0.04
36	1.00	0.67	52	0.07	0.00
37	1.00	0.64	53	0.03	0.00

Tabelle 8

Cut-off Score in der Auswertungsvariante B für **Caregiver Positive Engagement** mit Sensitivität und Spezifitätswerten (Angresius, 2024)

Summenwert	Sensitivität	1 - Spezifität	Summenwert	Sensitivität	1 - Spezifität
6	1.00	1.00	20	1.00	0.67
8	1.00	0.99	21	1.00	0.62
9	1.00	0.98	22	1.00	0.55
10	1.00	0.97	23	1.00	0.54
11	1.00	0.95	24	1.00	0.49
12	1.00	0.94	25	0.90	0.44
13	1.00	0.92	26	0.87	0.40
14	1.00	0.90	27	0.83	0.34
15	1.00	0.88	28	0.83	0.23
16	1.00	0.83	29	0.70	0.15
17	1.00	0.79	30	0.47	0.13
18	1.00	0.76	31	0.27	0.05
19	1.00	0.72	32	0.03	0.00

Korrelation zwischen den Skalen des SIRS von McCall

Nach der psychometrischen Betrachtung der einzelnen Skalen interessiert im Weiteren wie unabhängig oder auch redundant die Skalen des SIRS sind, da sich hieraus vorab zumindest grob ableiten lässt, wieviel ‚neue‘ Information in einer Skala im Vergleich zu den verbleibenden gewonnen wird und ob die Einschätzung durch Urteiler auf aller Skalen notwendig ist.

Einschränkend muss vorangestellt werden, dass die Bestimmung der inkrementellen Validität in Bezug auf ein Außenkriterium hier noch nicht berichtet wird und weiteren Analysen und Publikationen vorbehalten bleibt. In Tabelle 9 sind die Skalenkorrelationen aller Skalen nach McCall der Auswertungsvarianten A, B, C auf der Basis der aggregierten Urteile für 10 (oberhalb der Diagonalen) und 15-20 Minuten (unterhalb der Diagonalen) Beobachtungsdauer dargestellt.

Tabelle 9

Skalenkorrelation aller Skalen nach McCall der Auswertungsvarianten A, B, C auf der Basis aggregierter Urteile für 10 (oberhalb der Diagonalen) und 15-20 Minuten (unterhalb der Diagonalen) Beobachtungsdauer auf der Basis von Rohwerten

SIRS Skalen der Auswertungsvarianten A (01), B (02-04) und C (05-11)	01	02	03	04	05	06	07	08	09	10	11
01 Auswertungsvariante A Total Score	-	.97	.75	.85	.91	.79	.87	.84	.67	.63	.76
Auswertungsvariante B											
02 Caregiver Positive Engagement	.97	-	.62	.70	.90	.68	.77	.83	.70	.67	.80
03 Caregiver Punitiveness	.81	.69	-	.74	.65	.58	.81	.66	.35	.50	.44
04 Child-Caregiver directed interaction	.89	.78	.82	-	.75	.93	.84	.63	.48	.38	.54
Auswertungsvariante C											
05 Anteilnahme der BP	.94	.93	.70	.81	-	.71	.78	.76	.51	.44	.62
06 BP /Kind-interagierendes Verhalten	.85	.76	.68	.94	.76	-	.64	.55	.51	.39	.56
07 Verhaltenskontrolle der BP	.90	.81	.88	.91	.85	.77	-	.75	.41	.40	.55
08 Erregungszustand (Affekt) der BP	.85	.86	.67	.66	.82	.59	.77	-	.33	.42	.52
09 Kontakterwartung und Responsivität des Kindes	.71	.75	.45	.54	.53	.58	.48	.43	-	.62	.69
10 Gefühle des Kindes	.73	.74	.66	.55	.59	.56	.52	.52	.71	-	.61
11 Beziehung zum Kind	.80	.80	.61	.66	.64	.65	.63	.60	.81	.71	-

Die Skaleninterkorrelationen in der ersten Zeile und auch darunter weisen insgesamt relativ hohe Korrelationen aller Skalenvarianten auf, was zunächst die Bildung eines Gesamtscores nahelegt. Je mehr Skalen gebildet werden, desto mehr treten der Auswertungsvariante C deutlich unabhängige Skalen auf, was neben der Eigenständigkeit bzw. dem eigenen Informationsgehalt auch an der abnehmenden Reliabilität durch kürzere Skalen liegen kann. Für die Frage nach dem besten Kompromiss zwischen unabhängigen Skalen und Datenpassung können schließlich die Model-Fit-Maße aus den Strukturgleichungen Hinweise liefern.

Strukturprüfung der Skalenbildung nach McCall

Zur Überprüfung der drei von McCall vorgeschlagenen Auswertungsvarianten und die hierüber definierten Messstrukturen wurden entsprechende Strukturgleichungsmodelle in ihren Parametern geschätzt und Model-Fit-Maße ermittelt, die einen Hinweis darauf geben, ob die vorgeschlagene Skalenbildung zu den Items ‚passt‘. Zur Beschreibung der Varianz, die durch die Modelle aufgeklärt werden soll, verwenden wir den Baseline Model Chi-Square = 2046.18 und einen Baseline Model Chi-Square DF von 153, woraus sich ein Baseline Model-Fit Chi-Square/DF=13.37 ergibt. Dieser Ausgangswert soll durch die im Folgenden getesteten Modelle möglichst in Richtung eines Wertes von unter 2 reduziert werden (Schermelleh-Engel et al., 2003), wobei die Model-Fit-Indizes nicht absolut interpretiert werden dürfen, da ihre Sensitivität gegenüber Rahmenbedingungen wie Stichprobenumfang und modellinternen Zusammenhängen sowie der Datenvorverarbeitung für die spätere Interpretation eine Rolle spielt, auf die im Folgenden näher eingegangen wird.

Exkurs: Zwei Varianten der Standardisierung von Raterurteilen je nach Fragestellung

Wie zuvor erwähnt beruhen die folgenden Modelltestungen auf aggregierten Werten, wobei allerdings zwei Varianten zunächst diskutiert werden sollen. Am einfachsten und vielleicht üblichsten ist eine simple Aggregation der Raterurteile zu einem gemittelten Urteil, welches dann die Datengrundlage für die Prüfung einer Strukturgleichung bildet. Dieses Vorgehen würde allerdings den Rater mit der größeren Antwortvarianz stärker gewichten, weshalb die Antwort z-Standardisiert werden können, was die 2. Variante darstellt. Bei einer z-Standardisierung wird jede Antwort relativ zu den verbleibenden Antworten eines Urteilers neu skaliert, sodass anschließend jeder Rater numerisch nur Urteile hat, welchen einen Mittelwert von Null und eine Standardabweichung von Eins aufweisen. Unterschiede zwischen den Ratern hinsichtlich deren Strenge bzw. Milde oder deren Tendenz zu extremen Urteilen werden dabei ausgeglichen. Uns interessieren primär an dieser Stelle, ob die Items sich zu Skalen bündeln lassen und nicht der Einfluss der Rater, welche nach einer Festlegung der Messstruktur erst dann die Möglichkeit bietet die Einflüsse der Rater zu untersuchen. Bei unserem Vorgehen werden also gleich zwei Varianzquellen ausgeschaltet, welche in der Praxis eine Rolle spielen können: zum ersten die Mittelwertsunterschiede zwischen den Ratern und Unterschiede in deren Varianzen. Für ein transparentes Vorgehen werden in Tabelle 10 deshalb beide Varianten und die assoziierten Model-Fit Werte.

Tabelle 10

Model-Fit Maße der von McCall vorgeschlagenen Item-Skala Zuordnung für die drei Auswertungsvarianten A, B und C für 15-20 Minuten Beobachtungsdauer, sowohl für die Aggregationsvariante der Rohscores beider Rater und für die Variante, dass die Einzelantworten der Rater vor deren Aggregation z-standardisiert wurden

	Skalen	RMSEA	Chi-Square χ	χ df	χ / df	Pr > χ	AGFI	Bentler CFI
Raw Scores								
Auswertungsvariante A	1	.21	827.26	135	6.13	<.0001	.39	.69
Auswertungsvariante B	3	.19	690.77	132	5.23	<.0001	.46	.75
Auswertungsvariante C	6	.17	435.00	104	4.18	<.0001	.51	.84
Z-Scores								
Auswertungsvariante A	1	.20	736.01	135	5.45	<.0001	.41	.68
Auswertungsvariante B	3	.18	631.10	132	4.78	<.0001	.48	.74
Auswertungsvariante C	6	.14	350.22	104	3.37	<.0001	.56	.86

Die Modell-fit Indices in Tabelle 10 zeigen an, dass alle drei Auswertungsvarianten A-C von McCall nicht hinreichend zu den Daten passen, allerdings scheint die Auswertungsvariante C mit den sechs Skalen im Vergleich deutlich besser abzuschneiden und zwar sowohl bei den Roh- wie auch bei den z-standardisierten Werten. Die Beschreibung der konvergenten Itemladungen für die Auswertungsvariante C von McCall mit insgesamt 6 Skalen für die Items 1-17 sind in Tabelle 11 dargestellt.

In Tabelle 11 ist ersichtlich, dass alle Pfade signifikant sind, was für die Passung des Modells spricht. Allerdings zeigt sich ein Schätzproblem für das 6. Item, das den nominalen Rang von 1 leicht überschreitet. Insgesamt laden die Items durchschnittlich hoch, übersteigen aber nicht immer die Korrelation zwischen den Skalen, was vermutlich ein Grund für die unbefriedigende Modellanpassung ist. Die Varianzaufklärung liegt bei der sechsfaktoriellen Lösung auf dem sehr hohen Niveau von 88.6 %, bei der dreifaktoriellen Lösung bei 77.6 % und bei der einfaktoriellen Lösung immer noch bei 59.2 %.

Tabelle 11

Konvergente (a priori den Skalen zugeordnete) Itemladungen des SIRS auf den jeweiligen Skalen für die Auswertungsvariante C von McCall auf der Basis von über beide Rater aggregierten Rohwerten

SIRS Item	Ladung	Signifikanz	Skala
1. Distanziertheit der BP (-)	.82	(**)	F1
2. BP reagiert nicht (-)	.88	(**)	F1
3. Mentale Verfügbarkeit/Empfänglichkeit der BP (+)	.93	(**)	F1
4. BP reagiert auf das Kind (+)	.91	(**)	F1
5. BP gerichtetes Verhalten (+)	.75	(**)	F2
6. Kind gerichtetes Verhalten (+)	1.02	(**)	F2
7. Intrusivität der BP (-)	.79	(**)	F3
8. Verhaltens-Kontrolle/-Gehorsam/-Disziplin der BP	.78	(**)	F3
9. BP bestraft das Kind (-)	.80	(**)	F3
10. Unterstützung/Empathie/Anleitung der BP (+)	.86	(**)	F3
11. Negativer Affekt der BP (-)	.91	(**)	F4
12. BP zeigt positiven Affekt und Zuneigung (+)	.64	(**)	F4
13. Lebhaftigkeit und Ausdruckskraft der BP	.37	(**)	F4
14. Emotionale Nicht-Responsivität des Kindes	.79	(**)	F5
15. Kindliche Responsivität/Positive Erwartung	.93	(**)	F5
16. Negative Gefühle des Kindes	.43	(**)	F6
17. Positive Gefühle des Kindes	.73	(**)	F6

Anmerkung. BP = Bezugsperson.

4.4 Normierung

Wie bei allen Beobachtungsinstrumenten fehlt auch beim SIRS eine repräsentative Normierung. Als Interpretationshilfe bzw. Referenzwerte dienen die Cut-off-Werte (siehe unter „Testkonstruktion“, Tabellen 7-8).

5. Anwendungsmöglichkeiten

Bei der Wahl des Instruments stellt sich auch die Frage nach der Anwendbarkeit im eigenen Kontext. Einen ersten Überblick geben Mahoney et al. (1996), deren Liste wir im Folgenden ergänzen. Zu den Kontextbedingungen gehört z. B. der Einfluss der Instruktion der Dyade (d. h. was man ihnen vor der Interaktion sagt, warum und wozu die Beobachtung notwendig ist, ob man die Evaluationsziele bekannt gibt usw.). Dazu gehört auch die Angabe des Beobachtungsortes, z. B. ob die Beobachtung im häuslichen Kontext oder eher in der ungewohnten Umgebung einer Klinik stattfindet, was zu einer erhöhten anfänglichen Unsicherheit der Dyade führen kann. Ebenso kann die Anwesenheit eines Beobachters die Interaktion beeinflussen, z. B. indem das Kind Kontakt mit dem Beobachter aufnimmt, anstatt zu interagieren. Die Liste wird noch erweitert durch die Frage, ob eine simultane Auswertung während der Interaktion stattfindet oder ob die Auswertung auf der Basis der Videoaufzeichnung erfolgt, was dem Beobachter mehr Zeit und Ruhe sowie die Möglichkeit gibt, eine Szene wiederholt unter verschiedenen Aspekten zu beurteilen. Schließlich ist zu erwähnen, dass auch die Dauer der Beobachtung und die verschiedenen Phasen im Prozess

einer längeren Interaktion einen Einfluss auf die inhaltliche Bewertung und deren Genauigkeit sowie auf das Training der Beurteiler haben können.

Das Beobachtungsinstrument SIRS von McCall kann im Forschungskontext und in der klinischen Diagnostik der Eltern-Kind-Interaktion eingesetzt werden. McCall gibt einen Altersbereich von mindestens 18 Monaten bis ca. 6 Jahren an. Für die Auswertungsvarianten aus der Originalpublikation von McCall gibt es keine vorgegebene oder durchschnittliche Durchführungszeit. Wir empfehlen eine Beobachtungszeit von ca. 20 Minuten sowie eine Interaktionsbeobachtung in einem Raum mit altersgerechtem Spielmaterial, bei der keine weiteren Personen anwesend sind, da dies eine Einordnung der Ergebnisse mit denen unserer Untersuchung ermöglicht.

Voraussetzung für die Anwendung des SIRS ist zum einen die Kenntnis der in diesem Bereich wesentlichen Konstrukte nach Ellen Skinner, die positives von negativem Elternverhalten unterscheiden, wobei das SIRS nicht alle wesentlichen Bereiche erfasst. Die kindbezogenen Skalen können sowohl im Kontext der Bindungstheorie als auch im Rahmen von Temperamentsmerkmalen interpretiert werden. Eine weitere Voraussetzung für die Interpretation der Skalen ist psychometrisches Vorwissen, da die Messgenauigkeit der Skalen zwischen den verschiedenen Auswertungsvarianten stark variiert. Der diagnostische und therapeutische Nutzen kann nur eingeschränkt beurteilt werden, da eine breite empirische Erprobung der Skalen noch aussteht. Die Ergebnisse von McCall und auch die unseren weisen jedoch auf ein Potenzial der Skalen hin, das je nach klinischer, psychologischer und pädagogischer Anwendungsfrage relevant sein kann. Die Anwendung sollte im Rahmen einer standardisierten Beschreibung der Interaktionsbeobachtung nach Müller et al. (2023) erfolgen.

6. Kurzfassung

Diagnostische Zielsetzung:

Die Caregiver-Child Socioemotional and Relationship Rating Scale (CCSERRS, nach McCall und im Folgenden als ‚SIRS‘ bezeichnet) wurde 2010a von McCall, Groark und Fish erstmals entwickelt und ins Deutsche adaptiert und richtet sich an Ärzte, Psychologen und Pädagogen, die eine Eltern-Kind-Interaktionsbeobachtung planen und sich über das SIRS als mögliches Beobachtungsinstrument informieren möchten.

Aufbau:

Das SIRS erfasst in 18 teils globalen Einschätzungen die folgenden Bereiche/Skalen: 1. Interesse der Bezugsperson; 2. von der Bezugsperson an das Kind gerichtetes Verhalten; 3. Verhaltenskontrolle der Bezugsperson; 4. Erregungszustand (Affekt) der Bezugsperson; 5. Kontakterwartung und Responsivität des Kindes; 6. Gefühle des Kindes; 7. Beziehung des Kindes zur Bezugsperson.

Grundlagen und Konstruktion:

Das SIRS von McCall et al. (2010) wurde ins Deutsche übersetzt und anhand einer klinischen Inanspruchnahmepopulation einer familienpsychiatrischen Tagesklinik (N = 86) sowie einer nicht-klinischen Kontrollstichprobe (N = 30) auf der Basis von 15-20-minütigen

videographierten Eltern-Kind-Interaktionen im Freispiel auf der Basis aggregierter Urteile zweier verblindeter Rater inhaltlich und psychometrisch evaluiert.

Mit uni-, bi- und multivariaten Verfahren wurden die psychometrischen Eigenschaften des SIRS auf Item- und Skalenebene u. a. mit Strukturgleichungsmodellen untersucht. Die Strukturprüfungen zeigen für drei Auswertungsvarianten ein gemischtes Ergebnis mit Hinweisen zur weiteren Überarbeitung der Items und Skalen.

Empirische Prüfung und Gütekriterien:

Reliabilität: Unsere Analysen zeigen eine Interrater-Reliabilität je nach Skala, eine Korrelation der Beurteiler zwischen $r = .39$ und $r = .65$ und die Homogenitäten der Skalen liegen je nach Skala zwischen $.58$ und $.95$, wobei die Skalen untereinander teilweise recht hoch korrelieren, was auf Redundanzen bzw. eine verminderte Unabhängigkeit der Skalen hindeutet.

Validität: Die Aussagekraft der Skalen wurde im Rahmenmodell von Ellen Skinner an dieser Stelle beschrieben und inhaltlich erfasst der SIRS in seinen Auswertungsvariante etablierte Konstrukte, wenngleich konvergente Studien zu inhaltsähnlichen Instrumenten noch ausstehen. In unserer Studie konnte der SIRS zumindest mit den übergeordneten Skalen eine hinreichende Trennung einer klinischen von einer nicht-klinischen Vergleichsstichprobe belegen.

Normen: Standards sind nicht verfügbar. Die Stichprobe umfasst jedoch eine klinische und eine nichtklinische Stichprobe, auf deren Grundlage zumindest vorläufige Grenzwerte vorgeschlagen werden.

7. Bewertung

Das Beobachtungsinstrument SIRS (CCSERRS; Caregiver-Child Socioemotional and Relationship Rating Scale von McCall et al., 2010) erfasst einen Teil der in der Literatur (vgl. Skinner et al., 2005) als wesentlich erachteten Dimensionen elterlichen Verhaltens, wenngleich die bipolare Konzeption mit der grundsätzlichen Unterscheidung von positivem und negativem Elternverhalten vom SIRS in seinen Skalen nicht vorgenommen wird. Das SIRS erfasst jedoch neben dem Elternverhalten auch das Verhalten des Kindes, das im Rahmen der Bindungstheorie sowie übergeordneter Temperamentsfaktoren interpretiert werden kann.

Der SIRS von McCall wurde im Kontext eines Forschungsprojektes neben weiteren Verfahren mit aufgenommen, da dieser ein international eingeführtes Instrument darstellt, welche auch Ergebnisse einer Faktorenanalyse während der Skalenentwicklung berücksichtigte, und vor allem neben dem häufig erfassten Elternverhalten auch kind-bezogene Skalen beinhaltet. Die Items des SIRS erfassen verschiedene wesentliche Aspekte des Interaktionsverhaltens, es fehlt ihm aber eine übergeordnete theoretische Einbindung, was die Interpretation der Skalen erschwert. Wir wählten deshalb als Bezugsrahmen das Rahmenmodell von Ellen Skinner, welches auch auf inhaltliche Lücken des SIRS hinweist. Der SIRS erfasst ein breiteres Spektrum relevanter Inhalte zum elterlichen Verhalten für die Bereiche 1. Involvement/Responsivität (Warmth), marginal den Bereich 2. Acceptance (Autonomy support), 3. Structure (Support, Guiding); 4. Rejection [Hostility; low involvement; low engagement]; marginal den Bereich 5. Intrusivness (Coersion, Psychological control) jedoch nicht den Bereich Impaired Predictability (Chaos, Inconsistence). Die Itemsammlung umfasst auch das Kindverhalten, welche im Rahmen

der Bindungstheorie von Bowlby (1985) interpretiert werden kann, und zumindest grob Temperamentsfaktoren nach von Tellegen (1985) hinsichtlich des positiven und negativen Affekts (Tellegen, 1985; Watson et al., 1999) erfasst.

Vorteilhaft ist die leichte Erlernbarkeit des Beobachtungsinstrumentes für die Anwendung sowie die einfache Auswertung, wenngleich drei Auswertungsvarianten existieren. Die Anwender sollten hierbei die psychometrischen Grenzen von Skalen beachten, wenn diese nur auf wenigen Items basieren und die Skalen teils inhaltlich nachvollziehbar hoch miteinander korrelieren. Vorteilhaft konnte der SIRS zeigen, dass er in der Lage ist, eine klinische von einer nicht-klinischen Stichprobe mit Cohens $d = 1.23$ zu unterscheiden. Gleichwohl bedürfen die Skalen einer zukünftigen Überarbeitung und weiteren Validierung und Normierung.

8. Literatur

- Achtergarde, S., Postert, C., Wessing, I., Romer, G. & Müller, J. M. (2015). Parenting and child mental health: Influences of parent personality, child temperament, and their interaction. *The Family Journal*, 23(2), 167–179.
<https://doi.org/10.1177/1066480714564316>
- Angresius, A. (2024). Einfluss der Beobachtungsdauer für die klinische Eltern-Kind Interaktionsbeobachtung mit den Instrumenten SIRS, PlayPab und EAS. Unveröffentlichtes Manuskript.
- Baans, N. E. U., Janßen, M., & Müller, J. M. (2024). Parent-Child Relationship Measures and Pre-Post Treatment Changes for a Clinical Preschool Sample Using DC:0-3R.
<https://doi.org/10.21203/rs.3.rs-3480635/v1>
- Belsky, J., & Isabella, R. A. (1988). Maternal, Infant, and Social-Contextual Determinants of Attachment Security. In J. Belsky & T. M. Nezworski (Eds.), *Clinical Implications of Attachment* (pp. 41-94). Lawrence Erlbaum Associates.
- Bowlby, J. (1985). The nature of the child's tie to his mother. *International Journal of Psycho-Analysis*, 39, 1–23.
- Bradley, R. H., & Caldwell, B. M. (1995). Caregiving and the regulation of child growth and development: Describing proximal aspects of caregiving systems. *Developmental Review*, 15, 38–85. <https://doi.org/10.1006/drev.1995.1002>
- Bushman, B. B., & Holt-Lunstad, J. (2009). Understanding social relationship maintenance among friends: Why we don't end those frustrating friendships. *Journal of Social and Clinical Psychology*, 28(6), 749–778. <https://doi.org/10.1521/jscp.2009.28.6.749>
- Caldwell, B. M., & Bradley, R. H. (1984). Home Observation for Measurement of the Environment. *American journal of mental deficiency*, 84(3), 235–244.
- Clark, L., Kochanska, G., & Ready, R. (2000). Mothers' personality and its interaction with child temperament as predictors of parenting behavior. *Journal of Personality and Social Psychology*, 79(2), 274-285. <https://doi.org/10.1037/0022-3514.79.2.274>
- Esins, S., Müller, J. M., Romer, G., Wagner, K., & Achtergarde, S. (2017). Klinische Validierung der Beurteilungsskala Caregiver-Child Socioemotional and Relationship Rating Scale für kindliches Interaktionsverhalten im Kleinkind- und Vorschulalter. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 66(3), 209–223.
<https://doi.org/10.13109/prkk.2017.66.3.209>

- Furniss, T., Müller, J. M., Achtergarde, S., Wessing, I., Averbek-Holocher, M., & Postert, C. (2013). Implementing psychiatric day treatment for infants, toddlers, preschoolers and their families: A study from a clinical and organizational perspective. *International Journal of Mental Health Systems*, 7(1), 12. <https://doi.org/10.1186/1752-4458-7-12>
- Groark, C. J., McCall, R. B., Fish, L., & Whole Child International Evaluation Team (2011). Characteristics of environments, caregivers, and children in three Central American orphanages. *Infant Mental Health Journal*, 32(2), 232–250. <https://doi.org/10.1002/imhj.20292>
- Harms, T., & Clifford, R. M. (1989). *The Family Day Care Rating Scale*. Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environmental Rating Scale* (Rev. ed.). Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (1990). *Infant/Toddler Childhood Environmental Rating Scale*. Teachers College Press.
- Hawk, B. N., McCall, R. B., Groark, C. J., Muhamedrahimov, R. J., Palmov, O. I., & Nikiforova, N. V. (2018). Caregiver sensitivity and consistency and children's prior family experience as contexts for early development within institutions. *Infant Mental Health Journal*, 39(4), 432–448. <https://doi.org/10.1002/imhj.21721>
- Jacob, A. (2022). Interaktionsbeobachtung von Eltern und Kind. Methoden - Indikation - Anwendung. Ein Praxisbuch (3. Aufl.). Kohlhammer Verlag.
- Janßen, M., Holodyski, M., Müller, J. M., Reinersmann, A., & Romer, G. (2019). Impaired predictability: enhanced fluctuations in the parenting behaviour of mothers of pre-school children with clinical diagnoses across three different play tasks. *European Child & Adolescent Psychiatry*, 28(12), 1645–1658. <https://doi.org/10.1007/s00787-019-01330>
- Keren, M., Feldman, R., & Tyano, S. (2003). A five-year Israeli experience with the DC: 0-3 Classification system. *Infant Ment Health Journal*. 24(4), 337–348. <https://doi.org/10.1002/imhj.10060>
- Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., & Ramsauer, B. (2015). Observational tools for measuring parent–infant interaction: A systematic review. *Clinical Child and Family Psychology Review*, 18(2), 99–132. <https://doi.org/10.1007/s10567-015-0180-z>
- Mahoney, G., Spiker, D., & Boyce, G. (1996). Clinical assessments of parent–child interaction: Are professionals ready to implement this practice? *Topics in Early Childhood Special Education*, 16(1), 26–50. <https://doi.org/10.1177/027112149601600105>
- McCall, R., Groark, C., & Fish, L. (2010a). A Caregiver-child Socioemotional and Relationship Rating Scale. *Infant Mental Health Journal*, 31(2), 201-219.
- McCall, R., Groark, C., Fish, L., Harkins, D., Serrano, G., & Gordon, K. (2010b). A socioemotional intervention in a Latin American orphanage. *Infant Mental Health Journal*, 31(5), 521–542. <https://doi.org/10.1002/imhj.20270>
- Murphy, N. A., & Hall, J. A. (2021). Capturing Behavior in Small Doses: A Review of Comparative Research in Evaluating Thin Slices for Behavioral Measurement. *Frontiers in Psychology*, 12, 667326. <https://doi.org/10.3389/fpsyg.2021.667326>
- Müller, J. M. (2006a). Probability of obtaining two statistically different test scores as a test index. *Educational & Psychological Measurement*, 66(4), 601–611. <https://doi.org/10.1177/0013164405284034>

- Müller, J. M. (2006b). SAS macros to compute the test index PDS. *Applied Psychological Measurement*, 30(4), 345–346. <https://doi.org/10.1177/0146621606286207>
- Müller, J. M., Achtergarde, S., & Furniss, T. (2011). The influence of maternal psychopathology on ratings of child psychiatric symptoms: An SEM analysis on cross-informant agreement. *European Child & Adolescent Psychiatry*, 20(5), 241–252. <https://doi.org/10.1007/s00787-011-0168-2>
- Müller, J. M., & Furniss, T. (2013). Correction of distortions in distressed mothers' ratings of their preschool children's psychopathology. *Psychiatry Research*, 210(1), 294–301. <https://doi.org/10.1016/j.psychres.2013.03.025>
- Müller, J. M., Romer, G., & Achtergarde, S. (2014). Correction of distortions in distressed mothers' ratings of their preschool children's Internalizing and Externalizing Scale score. *Psychiatry Research*, 215(1), 170–175. <https://doi.org/10.1016/j.psychres.2013.10.035>
- Müller, J. M., Achtergarde, S., Frantzmahn, H., Steinberg, K., Skorozhenina, O., Beyer, T., Füniss, T., & Postert, C. (2013). Inter-rater reliability and aspects of validity of the Parent-Infant Relationship Global Assessment Scale (PIR-GAS). *Child & Adolescent Psychiatry & Mental Health*, 7(17). <https://doi.org/10.1186/1753-2000-7-17>
- Müller, J. M., Hoffmann, V. A., & Janssen, M. (2023). Diagnostik der Eltern-Kind-Interaktion und -Beziehung: Ergebnisse einer multiprofessionellen und tätigkeitsspezifischen Umfrage und Vorschläge für eine standardisierte Dokumentation [Diagnostic of parent-child-interaction and their relationship: Results from a multiprofessional and Task Specific Survey]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 72(1), 23–49. <https://doi.org/10.13109/prkk.2023.72.1.23>
- Müller, J. M., Hoffmann, V. A., & Wonner, L. I. (2018, June 11). Systematisches Review zu den inhaltlichen Aspekten bei der Beurteilung einer Eltern- bzw. Mutter-Kind-Interaktion auf der Basis deutschsprachiger Verfahren - mit Appendix 1-4. Leibniz Institut für Psychologische Information und Dokumentation (ZPID). <https://doi.org/10.23668/psycharchives.855>
- Müller, J. M., Postert, C., Beyer, T., Furniss, T., & Achtergarde, S. (2010). Comparison of eleven short versions of the Symptom Checklist 90-Revised (SCL-90-R) for use in the assessment of general psychopathology. *Journal of Psychopathology & Behavioral Assessment*, 32(2), 246–254. <https://doi.org/10.1007/s10862-009-9141-5>
- Postert, C., Achtergarde, S., Wessing, I., Romer, G., Füniss, T., Averbek-Holocher, M., & Müller, J. M. (2014). Multiprofessionelle Intervallbehandlung psychisch kranker Kinder im Vorschulalter und ihrer Eltern in einer Familientagesklinik. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 63(10), 812–830. <https://doi.org/10.13109/prkk.2014.63.10.812>
- Prokasky, A., Rudasill, K., Molfese, V. J., Putnam, S., Gartstein, M., & Rothbart, M. (2017). Identifying child temperament types using cluster analysis in three samples. *Journal of Research in Personality*, 67, 190–201. <https://doi.org/10.1016/j.jrp.2016.10.008>
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. *Child Development*, 72(5), 1394–1408. <https://doi.org/10.1111/1467-8624.00355>
- Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology*, 103(1), 55–66. <https://doi.org/10.1037/0021-843X.103.1.55>

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, 8(2), 23–74. <https://doi.org/10.23668/psycharchives.12784>
- Seifer, R., & Schiller, M. (1995). The Role of Parenting Sensitivity, Infant Temperament, and Dyadic Interaction in Attachment Theory and Assessment. *Monographs of the Society for Research in Child Development*, 60 (2/3), 146–174. <https://doi.org/10.1111/j.1540-5834.1995.tb00209.x>
- Skinner, E., Johnson, S., & Snyder, T. (2005). Six Dimensions of Parenting: A Motivational Model. *Parenting: Science and Practice*, 5(2), 175–235. https://doi.org/10.1207/s15327922par0502_3
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Lawrence Erlbaum Associates.
- Timmer, S. G., Hawk, B., Usacheva, M., Armendariz, L., Boys, D. K., & Urquiza, A. J. (2023). The long and the short of it: A comparison of the effectiveness of Parent–Child Care (PC–CARE) and Parent–Child Interaction Therapy (PCIT). *Child Psychiatry and Human Development*, 54(1), 255–265. <https://doi.org/10.1007/s10578-021-01257-9>
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76(5), 820–838. <https://doi.org/10.1037/0022-3514.76.5.820>
- Wagner, K., Müller, J. M., Esins, S., Romer, G. & Achtergarde, S. (2015). Play-PAB-Verfahren und seine Validierung an einer vorschulpsychiatrischen Inanspruchnahmepopulation. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 64(9), 690-705. <https://doi.org/10.13109/prkk.2015.64.9.690>
- Wagner, K., Müller, J. M., Esins, S., Romer, G. & Achtergarde, S. (2017). Play-PAB. PLAY-Parenting Assessment Battery - deutsche modifizierte Fassung des Lab-PAB [Verfahrensdokumentation, Kodierbogen und Manual]. In Leibniz-Institut für Psychologie (ZPID) (Hrsg.), Open Test Archive. Trier: ZPID. <https://doi.org/10.23668/psycharchives.4592>
- Zero To Three/National Center for Infants, Toddlers and Families (2005). Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: DC: 0-3R (Rev. ed.). Zero To Three.