

**Hans-Günther Roßbach**

**Analyse von Meßinstrumenten  
zur Erfassung von Qualitätsmerkmalen frühkindlicher  
Betreuungs- und Erziehungsumwelten**



**Institut für sozialwissenschaftliche Forschung e.V.**

**Roßbach, Hans-Günther:**

Analyse von Meßinstrumenten zur Erfassung von Qualitätsmerkmalen  
frühkindlicher Betreuungs- und Erziehungsumwelten.

Münster: Institut für sozialwissenschaftliche Forschung e.V., 1993

ISBN 3-926083-06-9

(Sozialwissenschaftliche Forschungsdokumentationen ; 4)

ISSN 0937-5066

**Sozialwissenschaftliche Forschungsdokumentationen**

ISBN 3-926083-06-9

©Institut für sozialwissenschaftliche Forschung e.V., Münster  
Steinfurter Str. 555, D-4400 Münster, F.R.G.

Alle Rechte vorbehalten

Printed in Germany

# *Inhaltsverzeichnis*

	<i>Seite</i>
0. Vorwort . . . . .	1
1. Bedeutung von Meßinstrumenten zur Erfassung der Qualität frühkindlicher Erziehungs- und Betreuungsumwelten . . . . .	3
1.1 Screeninginstrumente zur Erfassung der Qualität frühkindlicher Erziehungs- und Betreuungsumwelten . . . . .	3
1.2 Die Situation in den USA . . . . .	5
1.3 Zum Qualitätsbegriff . . . . .	8
1.4 Funktionen der Screeninginstrumente . . . . .	11
2. Überblick über existierende Screeninginstrumente . . . . .	15
2.1 Instrumente für institutionelle Gruppenbetreuungen . . . . .	17
2.1.1 Early Childhood Classroom Observation . . . . .	17
2.1.2 Classroom Practices Inventory . . . . .	25
2.1.3 CDA Assessment Observation Instrument . . . . .	29
2.1.4 Assessment Profile for Early Childhood Programs . . . . .	36
2.1.5 Early Childhood Environment Rating Scale . . . . .	43
2.1.6 Infant/Toddler Environment Rating Scale . . . . .	54
2.1.7 Preschool Classroom Implementation Rating Instrument . . . . .	59
2.1.8 Child Care Facility Schedule . . . . .	62
2.2 Instrumente für Tagespflege . . . . .	65
2.2.1 Assessment Profile for Family Day Care . . . . .	65
2.2.2 Family Day Care Rating Scale . . . . .	71
2.2.3 Family Day Home Observation Instrument . . . . .	78
2.2.4 Evaluating Home Based Day Care . . . . .	82
2.3 Instrumente für häusliche Umwelten . . . . .	86
2.3.1 Home Observation for Measurement of the Environment . . . . .	86
3. Zusammenfassung und Vergleich der Instrumente . . . . .	93
Literatur . . . . .	115

## ***Vorwort***

Der vorliegende Arbeitsbericht zur *Analyse von Meßinstrumenten zur Erfassung von Qualitätsmerkmalen frühkindlicher Betreuungs- und Erziehungsumwelten* wurde ermöglicht durch eine Reisebeihilfe der Deutschen Forschungsgemeinschaft für einen sechswöchigen Forschungsaufenthalt in den USA. Bei dieser Gelegenheit sei dem Frank Porter Graham Child Development Center der University of North Carolina at Chapel Hill für die guten Arbeitsmöglichkeiten und die Unterstützung bei der meistens sehr schwierigen Beschaffung der Forschungsinstrumente sowie der darauf bezogenen Literatur gedankt. Ein besonderer Dank gilt Dr. Thelma Harms und Dr. Richard M. Clifford, die vielfältige Anregungen gaben und aufgrund derer Kontakte in manchen Fällen sonst nicht zugängliche Informationen beschafft werden konnten.



# **1. Bedeutung von Meßinstrumenten zur Erfassung der Qualität frühkindlicher Erziehungs- und Betreuungsumwelten**

## **1.1 Screeninginstrumente zur Erfassung der Qualität frühkindlicher Erziehungs- und Betreuungsumwelten**

In den letzten beiden Jahrzehnten haben sich in der Bundesrepublik Deutschland gravierende Veränderungen in den Familienstrukturen und den Orientierungen von Frauen ergeben, die die Rahmenbedingungen für das Aufziehen und die Betreuung von Kindern nachhaltig verändern (vgl. ausf. Tietze, Roßbach & Jakobs, in Vorbereitung; Tietze & Roßbach, 1991). Dies hat zu einer erheblichen Zunahme von nicht-kernfamilialen Betreuungen von Kindern vor dem Schuleintritt geführt, wobei besonders institutionelle Gruppenbetreuungen zugenommen haben. Wurden 1960 (alte Bundesländer) weniger als ein Drittel (28,1 %) der Kinder im Alter von drei Jahren bis zum Schuleintritt in einem Kindergarten betreut, so waren es 1989 bereits gut zwei Drittel (67,7 %). Auch auf dem Gebiet der ehemaligen DDR ist die Versorgungsquote in diesem Zeitraum von 46,1 auf 95,1 % gestiegen und erreichte 1989 damit fast eine Vollversorgung (vgl. Roßbach & Tietze, 1991, S. 557).<sup>1</sup> Ähnliche Entwicklungen zeigen sich auch in anderen europäischen Ländern (vgl. Moss, 1988, 1990). Anders als bei den älteren Kindern - und auch anders als in mehreren anderen Ländern - stellt sich der Besuch einer institutionellen Gruppenbetreuung bei den jüngeren Kindern (unter drei Jahren) dar. Nach einer eigenen Erhebung wurden 1988 nur 3,2 % der Kinder unter drei Jahren in der Bundesrepublik (alte Bundesländer) institutionell, d.h. in einer Krippe oder in einem Kindergarten, betreut (vgl. Tietze & Roßbach, 1991). Vergleicht man dieses Datum mit den entsprechenden Quoten in Dänemark (48 %), Belgien (20 %) oder Frankreich (20 %) (vgl. Moss, 1990), so läßt sich vermuten, daß in der Zukunft auch in der Bundesrepublik eine Zunahme der institutionellen Gruppenbetreuungen für Kinder dieser Altersgruppe zu erwarten ist - dies nicht zuletzt auch deshalb, weil die ehemalige DDR in der Versorgung mit Krippenplätze für Kinder dieser Altersgruppe (1989: 55,6 %, vgl. Tietze & Roßbach, 1991) im europäischen Rahmen an der Spitze lag.

Neben der institutionellen Gruppenbetreuung macht eine Mehrheit der Kinder noch

---

<sup>1</sup> Diese Angaben für das Gebiet der alten Bundesrepublik weichen von den offiziellen Angaben ab, wonach 1960 32,8 % und 1989 79,0 % der Kinder einen Kindergarten besuchten (Bundesminister für Bildung und Wissenschaft, 1990). Dort werden aber die Plätze in Kindergärten nur auf die Kinder von drei Altersjahrgängen (Kinder im Alter von drei bis unter sechs Jahren) bezogen. Da aber auch jüngere Kinder und vor allem ein guter halber Jahrgang der sechsjährigen Kinder einen Kindergarten besuchen, stellen diese Angaben eine Überschätzung der tatsächlichen Platzangebots dar. Um diese Überschätzung zu vermeiden, werden bei den Angaben im Text die Kindergartenplätze nicht auf drei, sondern auf dreieinhalb Altersjahrgänge bezogen.

weitere nicht-kernfamiliale Betreuungserfahrungen: Gut die Hälfte aller Kinder wird durch Personen des sozialen Netzwerkes, d.h. durch die Großeltern, durch andere Verwandte oder durch Freunde/Nachbarn, betreut (Kinder unter drei Jahren: 53,7 %; Kinder von drei bis unter sechs Jahren: 54,9 %); etwa jedes dreizehnte bis vierzehnte Kind erfährt eine Betreuung durch bezahlte Personen, d.h. durch eine Kinderfrau, eine Tagesmutter oder einen Babysitter (Kinder unter drei Jahren: 6,9 %; Kinder von drei bis unter sechs Jahren 7,9 %; vgl. Tietze & Roßbach, 1991, S. 561).<sup>2</sup> In der Öffentlichkeit wird hier den Tagesmutterbetreuungen eine besondere Aufmerksamkeit entgegengebracht (vgl. z.B. Bundesministerium für Jugend, Familie und Gesundheit, 1980).

Nicht zuletzt in Zusammenhang mit der Zunahme der Erwerbstätigkeit von Müttern mit kleinen Kindern - und der damit gegebenen Notwendigkeit der Betreuung der Kinder während der Abwesenheit ihrer Mütter - konzentriert sich die gegenwärtige öffentliche Diskussion primär auf den Aspekt einer quantitativen Ausweitung öffentlicher Hilfen zur Betreuung von kleinen Kindern (vgl. z.B. die Diskussion um das Recht auf einen Kindergartenplatz). Fragen, die sich auf eine Sicherung der pädagogischen Qualität der verschiedenen Betreuungsformen richten, werden seltener gestellt. Dementsprechend liegen kaum gesicherte Informationen darüber vor, wie die konkrete Betreuungspraxis in den verschiedenen Betreuungsformen aussieht, d.h. welche Qualitätsmerkmale - wobei der Qualitätsbegriff hier zunächst in einem beschreibenden Sinne gebraucht wird - die Umwelten aufweisen, in denen Kinder im vorschulischen Alter betreut und erzogen werden. Wenn auch zu Zeiten der Bildungsreform dem Vorschulbereich eine besondere Aufmerksamkeit gewidmet wurde und in vielen Projekten Reformmaßnahmen, speziell Curricula, entwickelt und erprobt wurden, so läßt sich dennoch in wissenschaftlichen Studien eine Vernachlässigung der Erziehungsrealität mit ihren Bedingungen konstatieren (vgl. Fried, Roßbach, Tietze & Wolf, 1992).

Als Konsequenz dieser Vernachlässigung von qualitativen Merkmalen frühkindlicher Erziehungs- und Betreuungsformen gibt es in der Bundesrepublik kaum erprobte und wissenschaftlich abgesicherte Instrumente, mit denen die konkreten Bedingungen in frühkindlichen Erziehungs- und Betreuungsformen in einer Weise erhoben werden

---

<sup>2</sup> Im vorliegenden Bericht wird der Begriff "Tagesmutter" oder "Tagesmutterbetreuung" immer als Kürzel für die Betreuung durch eine nicht mit dem Kind verwandte Person in einer Tagespflegestelle, d.h. in der Wohnung dieser Betreuungsperson, verwandt. Im amerikanischen Sprachgebrauch entspricht dies den Begriffen "Family Day Care" und "Family Day Care Provider".

Da in der Regel die Betreuer in frühkindlichen Erziehungs- und Betreuungsformen weiblich sind, wird im folgenden auch immer die weibliche Form benutzt (z.B. Tagesmutter, Betreuerin, Erzieherin, Gruppenleiterin).

können, die meßtheoretischen Standards entspricht. Empirische Untersuchungen stehen damit immer vor der Aufgabe, dafür geeignete Meßinstrumente in einem langwierigen Prozeß selbst entwickeln oder nach in anderen Ländern erprobten Vorlagen adaptieren zu müssen. Dies gilt zum einen für Instrumente, die sich Spezialaspekten der frühkindlichen Erziehungs- und Betreuungsumwelten zuwenden, wie z.B. der Beobachtung von Kind-Kind-Interaktionen, Kind-Betreuer-Interaktionen und Kind-Spielzeug-Interaktionen oder der gezielten Erfassung von räumlich-materiellen oder architektonischen Merkmalen einer Betreuungsform. Zum anderen trifft dies vor allem aber auf Instrumente zu, die weniger isolierte Aspekte erfassen wollen, als vielmehr darauf abzielen, einen Gesamteindruck der Qualität einer frühkindlichen Erziehungs- und Betreuungsform zu geben und in diesem Sinne als Screeninginstrumente einsetzbar sind.

Wenn auch Screeninginstrumente Einzelaspekte von Betreuungsformen weniger differenziert erfassen können als dafür entwickelte Spezialinstrumente - und deshalb der Einsatz eines Screeninginstruments mit einem "Informationsverlust" verbunden ist -, so dürften doch die Vorteile einer routinemäßigen und zeitökonomischen Erfassung des Gesamteindrucks der Qualität einer Betreuungsform solchen Screeninginstrumenten einen breiten Anwendungsbereich geben: Surveystudien, die an größeren Stichproben von Erziehungs- und Betreuungsformen durchgeführt werden, um einen Überblick über die Qualität der Betreuungsformen für die nachwachsende Generation zu geben, sind zwangsweise auf solche zeitökonomisch und routinemäßig einsetzbaren Screeninginstrumente angewiesen. Ein Einsatz scheint aber auch in Studien, die sich Spezialaspekten von Betreuungsformen zuwenden, geraten, um zu vermeiden, daß durch die Fokussierung auf Einzelaspekte das Gesamtbild einer Betreuungsform verloren geht. Schließlich dürften solche Screeninginstrumente in verschiedenen Funktionen in Erzieheraus- und Fortbildung und in Projekten, die auf eine direkte Praxisverbesserung abzielen, einsetzbar sein (vgl. hierzu weiter unten). Nach Wissen des Berichterstatters ist aber ein solches Screeninginstrument zur Globalerfassung der Qualität frühkindlicher Erziehungs- und Betreuungsformen gegenwärtig in der Bundesrepublik nicht vorhanden.

## ***1.2 Die Situation in den USA***

In den USA - teilweise auch in Kanada - stellt sich aufgrund des anders gelagerten gesellschaftlichen Kontextes im frühkindlichen Betreuungs- und Erziehungsbereich (z.B. weniger staatliche Regulierungen, andere Trägerstrukturen, stärkere Kommerzialisierung) die Situation anders dar. Befürchtungen über negative Auswirkungen einer nicht-kernfamilialen Betreuung in der frühen Kindheit haben dazu geführt, daß schon seit gut 15 Jahren vermehrte wissenschaftliche Anstrengungen auf die Untersuchung der Qualität von frühkindlichen Betreuungs- und Erziehungsumwelten



gerichtet wurden (vgl. auch Doherty, 1991). In diesem Kontext wurden verschiedene Screeninginstrumente zur Einschätzung der Qualität solcher Umwelten entwickelt. Von besonderer Bedeutung im vorliegenden Zusammenhang sind auch verschiedene groß angelegte Praxisprogramme, in denen - angesichts des relativ niedrigen Niveaus staatlicher Regulierungen, die für die Errichtung und den Unterhalt einer Betreuungsform erfüllt werden müssen, und angesichts der relativ geringen Anforderungen an die formale Ausbildung der Betreuerinnen in Einrichtungen - versucht wurde und wird, bedeutsame Impulse zur Verbesserung der frühkindlichen Betreuungs- und Erziehungsformen zu geben. Im Rahmen dieser Programme - die zum Teil von Berufsorganisationen des Fachpersonals initiiert wurden - wurden ebenfalls mehrere Instrumente zur globalen Einschätzung der Qualität der jeweiligen Betreuungsformen entwickelt und breit angewandt. Zu nennen sind hier vor allem die folgenden drei Programme:

- Schon 1975 wurde das *CDA (Child Development Associates) National Credentialing Program* begonnen.<sup>3</sup> Zunächst konnte nur Fachpersonal, das in institutioneller Gruppenbetreuung für Kinder im Alter von drei bis fünf Jahren arbeitet, an diesem Programm teilnehmen, ab Mitte der 80er Jahre wurde das Programm auf Fachpersonal in Gruppenbetreuungen für jüngere Kinder und auf Betreuerinnen in Tagespflegestellen ausgeweitet. Die Teilnahme an diesem Programm ist freiwillig. In ihm werden die Kompetenzen einer Betreuerin im Hinblick auf einen Basissatz von Kompetenzen (CDA Competencies) evaluiert. Falls eine Betreuerin diese Kompetenzen zeigt, wird ihr als Anerkennung ein Zeugnis (CDA Credential) für ihre guten Leistungen vergeben. Da dieses Programm nicht nur diese Evaluationskomponente aufweist, sondern in weiten Teilen informelle - zum Teil auch formelle - Fortbildungskomponenten sowie direkte Maßnahmen zur Praxisverbesserung enthält, dürfte hiermit ein bedeutendes pädagogisches Programm zur Professionalisierung des Fachpersonals in frühkindlichen Erziehungs- und Betreuungsformen vorliegen (vgl. z.B. Council for Early Childhood Professional Recognition, 1990, 1991a). Gegenwärtig wird der formale Ablauf des Programms revidiert, wobei gleichzeitig die Fortbildungskomponente verstärkt und formalisiert wird (Competence, 1991,

---

<sup>3</sup> Zum Unterschied zwischen Accreditation und Credentialing:

- Bei der *Akkreditierung* wird einer Einrichtung eine Anerkennung für gute Qualität ausgesprochen.
- Beim *Credentialing* erhält dagegen eine Betreuerin/Erzieherin ein Zeugnis für gute Leistungen.

Während bei Einrichtungen diese Unterscheidung als bedeutsam erscheint - eine Erzieherin mit "credential" kann die Einrichtung wechseln, ohne ihr "credential" zu verlieren -, dürfte sie bei Tagespflegestellen, bei denen in der Regel von einer Einheit zwischen der Tagespflegestelle und der Tagesmutter ausgegangen werden kann, weniger wichtig sein und eher als künstlich erscheinen.

1992).

- Seit 1984 führt die *NAEYC (National Association for the Education of Young Children)* auf freiwilliger Basis ihr *Accreditation Program* durch. Einrichtungen des frühkindlichen Bereichs, die bestimmten von der NAEYC gesetzten Qualitätsstandards genügen, wird dabei eine entsprechende Anerkennung ausgesprochen. Auch hier geht es nicht nur um die Evaluation der Übereinstimmung der Qualitätsmerkmale einer Einrichtung mit den NAEYC Akkreditierungskriterien, vielmehr besteht für eine Einrichtung die erste und längste Phase der Teilnahme an diesem Akkreditierungsprogramm darin, Maßnahmen zur Verbesserung der Qualität der eigenen Einrichtung so einzuleiten, daß die Akkreditierungskriterien erfüllt werden können (vgl. National Association for the Education of Young Children, 1991a).
- 1988 hat *NAFDC (National Association for Family Day Care)* ein vergleichbares *Accreditation Program* für Tagespflegestellen ins Leben gerufen. Auch hier ist die Teilnahme freiwillig. Tagespflegestellen, die bestimmten Qualitätskriterien genügen, wird dann die Akkreditierung ausgesprochen. Ein Teil dieses Programs besteht ebenfalls aus einer Fortbildungskomponente (vgl. Sibley & Abbott-Shim, 1989; National Association for Family Day Care, ohne Jahr).

Zur Evaluation der Qualität einer Betreuungsform bzw. der Arbeit einer Betreuerin werden in den drei genannten Programmen formale Beobachtungsinstrumente - im Sinne von Screeninginstrumenten - eingesetzt. Typischerweise werden die Instrumente dabei zunächst im Rahmen einer Selbstevaluation von den Betreuerinnen ausgefüllt, wobei dieser Prozeß einer systematischen Selbstevaluation zugleich auch als Methode einer Praxisverbesserung zu betrachten ist. Anschließend werden dann in der Regel Beobachtungen durch andere Personen anhand des gleichen Instruments benutzt, um die Selbstevaluation zu verifizieren bzw. um zu einer objektiven Evaluation zu kommen. Da die Vergabe oder Vorenthaltung eines "Qualitätssiegels" für die betroffenen Einrichtungen bzw. Betreuerinnen im Hinblick auf ihre "Werbekraft" von größerer Bedeutung ist, stellt ein solcher instrumenteller Zugriff zugleich auch besondere Anforderungen an die Meßinstrumente, vor allem an ihre Objektivität.

Im Hinblick auf diesen Arbeitsbericht bleibt damit festzuhalten, daß in den USA eine Vielzahl von Screeninginstrumenten für eine globale Einschätzung der Qualität frühkindlicher Betreuungs- und Erziehungsformen existiert. Allerdings ist eine unreflektierte Übernahme von Instrumenten für die Situation in der Bundesrepublik oder eine einfache Übersetzung ins Deutsche aus verschiedenen Gründen nicht möglich. Ein Grund hierfür ist zunächst einmal, daß sich die Situation der frühkindlichen

Erziehung und Betreuung in der Bundesrepublik von jener in den USA in mehreren Aspekten unterscheidet. Darüber hinaus weisen die verschiedenen Instrumente oftmals zwei Schwächen auf:

- In vielen Fällen wurden die Instrumente ad hoc für jeweils spezifische Anwendungszwecke entwickelt. Oftmals fehlen damit systematische Begründungen der in die Instrumente aufgenommen Items. Obwohl auf einer abstrakteren Ebene durchaus Modellvorstellungen zur Strukturierung von Umwelten auf verschiedenen Ebenen vorliegen - vgl. z.B. die von Bronfenbrenner entwickelten Modellvorstellungen (vgl. Bronfenbrenner, 1981; Bronfenbrenner & Crouter, 1983) -, fehlen weitgehend "empirienähere" Modelle, die verschiedene Dimensionen einer frühkindlichen Erziehungs- und Betreuungsumwelt ausdifferenzieren und so die Entwicklung konkreter Meßinstrumente anleiten können (vgl. auch Tietze, 1985; Robbich, Clifford & Harms, 1991).
- In mehreren Fällen sind die Instrumente selbst nicht veröffentlicht und liegen nur im Rahmen von "grauer" Literatur vor. Oftmals fehlen Angaben zu den technischen Gütekriterien bzw. sind diese nicht veröffentlicht. Die Zusammenstellung einer Übersicht über solche Screeninginstrumente wird dadurch zum Teil erheblich erschwert. Wenige Informationen liegen zudem über die faktorielle Struktur der Instrumente vor, d.h. über die Dimensionen, die empirisch von den Instrumenten erfaßt werden (vgl. Robbich et al., 1991).

Im Hauptteil dieses Arbeitsberichtes wird ein systematischer Überblick über Screeninginstrumente zur Erfassung von Qualitätsmerkmalen frühkindlicher Betreuungs- und Erziehungsformen gegeben, um diese Instrumente im deutschsprachigen Raum bekannt zu machen, ihre Qualität abzuschätzen und damit für verschiedene Anwendungskontexte Auswahlhilfen zu geben. Vorher wird - zum Abschluß dieser Einleitung - noch kurz auf den Qualitätsbegriff, der diesen Instrumenten zugrunde liegt, eingegangen. Ebenfalls werden kurz die verschiedenen Anwendungskontexte von solchen Screeninginstrumenten und die damit gegebenen Anforderungen an sie zusammengefaßt.

### **1.3 Zum Qualitätsbegriff**

Die Begriffe "Qualität" und "qualitative Merkmale" von frühkindlichen Betreuungs- und Erziehungsformen lassen sich zunächst als rein beschreibende Begriffe verstehen, um jene Merkmale zu kennzeichnen,

- die als wichtig für die Charakterisierung der Erziehungs- und Betreuungsformen angesehen werden,

- in denen sich die verschiedenen Formen (potentiell) unterscheiden und
- in denen bestimmte Ausprägungen mit bestimmten Entwicklungszuständen auf Seiten der Kinder in Verbindung stehen.

Die Screeninginstrumente enthalten - je nach Instrument unterschiedlich - eine Sammlung von solchen "qualitativen Merkmalen". Allerdings sind diese Sammlungen nicht normativ neutral und dienen somit nicht nur beschreibenden Zwecken. Ihnen gehen vielmehr normative Entscheidungen insofern voraus,

- als die jeweilige Auswahl keine Zufallsauswahl aus dem Gesamt der möglichen "qualitativen Merkmale" darstellt, sondern jeweils nur das betrachtet wird, was den Autoren als wichtig erscheint.
- als die Itempole der ausgewählten Items jenes widerspiegeln, was nach Ansicht der Autoren gut oder schlecht ist.

Insofern stellen die Screeninginstrumente Operationalisierungen einer bestimmten, inhaltlichen Vorstellung einer *guten* Qualität von frühkindlichen Betreuungs- und Erziehungsformen dar. Aus diesem Grunde spielen die Konzepte der Autoren über eine gute Qualität der Formen eine wichtige Rolle bei der Beurteilung der Instrumente. Gerade deshalb ist es bedauerlich, daß oftmals diese Konzepte nicht explizit entwickelt werden, sondern weitgehend implizit bleiben (vgl. weiter unten).

Vorstellungen über eine gute Qualität sind nicht statisch, sondern unterliegen einem historischen Wandel (vgl. Clifford, Harms, Pepper & Stuart, in Druck). Beeinflußt durch die Untersuchungen von Bowlby (1951) wurde in den 50er Jahren eine gute Qualität nicht-elterlicher Betreuungs- und Erziehungsformen definiert in Begriffen von Wärme und Pflege (ähnlich wie jener durch die Eltern selbst). In den 60er und 70er Jahren erfolgt eine Umdefinition. Der Schwerpunkt lag nun auf kognitiver Stimulierung, und als qualitativ gut wurde jene Betreuung betrachtet, die die kognitive Entwicklung der betreuten Kinder stimuliert. In den 80er und 90er Jahren hingegen wird eine breitere Definition einer guten Qualität verfolgt, die sich auf das "ganze" Kind richtet und neben kognitiver Stimulierung auch Gesundheits- und Sicherheitsaspekte sowie soziale und emotionale Unterstützung des Kindes umfaßt.

Darüber hinaus hängt der Qualitätsbegriff von den Funktionen ab, die die frühkindlichen Erziehungs- und Betreuungsformen erfüllen sollen (vgl. Ruopp & Travers 1982, zit. nach Doherty, 1991, S. 2f.): Bezieht sich deren Funktion hauptsächlich darauf, mütterliche Erwerbstätigkeit zu stimulieren, so läßt sich Qualität definieren als die Quantität vorhandener Plätze, auch wenn diese die Kinder nur

aufbewahren. Betrachtet man Kinderbetreuung als einen umfassenderen, über die Ermöglichung von mütterlicher Erwerbstätigkeit hinausgehenden Service für Eltern, so könnte sich Qualität beziehen auf eine Betreuung, die nahe der elterlichen Wohnung lokalisiert ist, die zuverlässig und finanziell erschwinglich ist und die zeitlich flexibel genutzt werden kann. Ein völlig anderes Bild einer qualitativ guten Betreuung zeigt sich aber dann, wenn die Funktion einer Betreuung definiert wird als Unterstützung der physischen, emotionalen, sozialen und intellektuellen Entwicklung eines Kindes, seines diesbezüglichen Wohlbefindens und der Eltern in ihrer Elternrolle. Eine solche Definition schließt nicht aus, daß die Betreuung auch Bedürfnisse der Eltern im Hinblick auf ihre eigenen beruflichen oder anderen Verpflichtungen berücksichtigt.

Die im Hauptteil des Arbeitsberichts vorgestellten Screeninginstrumente schließen sich alle dieser breiten Definition der Funktion von frühkindlichen Erziehungs- und Betreuungsformen an und richten sich auf das "ganze" Kind im oben beschriebenen Sinne. Ein solcher breiter Qualitätsbegriff, der sich nicht nur auf im engeren bildungsbezogene Aspekte bezieht, sondern ebenfalls Aspekte der Pflege, Betreuung, Gesundheit usw. umfaßt, dürfte gerade auch für die bundesrepublikanische Diskussion um die Kindergartenerziehung wichtig sein. Oftmals läßt sich hier eine Einschränkung auf die Bildungsseite und eine Vernachlässigung von Betreuungs- und Pflegeaspekten (Einrichtung nur kurzzeitig geöffnet, keine Mittagsmahlzeiten, keine oder eine reduzierte Verantwortung für Gesundheitsaspekte usw.) erkennen (vgl. Tietze et al., in Vorbereitung).

Alle im folgenden vorgestellten Screeninginstrumente folgen auch insofern einem breiten Verständnis von Qualität, als sie verschiedene Bereiche von frühkindlichen Betreuungs- und Erziehungsformen ansprechen. Ruopp, Travers, Glantz & Coelen (1979) unterscheiden drei Bereiche, in denen sich die Qualität einer Betreuungs- und Erziehungsform zeigt:

- "Inputs", d.h. z.B. Ausbildung der Erzieherinnen, zahlenmäßige Erzieher-Kind-Relation, Raumgröße und Ausstattung
- Erfahrungen der Kinder, d.h. z.B. die Qualität der Erzieherinnen-Kind-Interaktionen, der tatsächliche Tagesablauf, das implementierte Curriculum
- "Outcomes", d.h. z.B. die Veränderungen in den sozialen, intellektuellen oder sprachlichen Fähigkeiten des Kindes.

Entsprechend dem Charakter eines Screeninginstruments sind die Möglichkeiten der Instrumente, den kindlichen Entwicklungsstand und seine Veränderungen über Zeit

zu erfassen, sehr eingeschränkt. Alle decken aber - mit unterschiedlichen Schwerpunktssetzungen - die ersten beiden Aspekte ab, d.h., sie umfassen sowohl Merkmale des so definierten Inputs in eine frühkindliche Erziehungs- und Betreuungsform als auch Merkmale der dort tatsächlich ablaufenden Prozesse.

#### **1.4 Funktionen der Screeninginstrumente**

Zusammenfassend können drei Einsatzbereiche solcher Screeninginstrumente unterschieden werden: Forschung, Aus- und Fortbildung, direkte Praxisverbesserung.

Empirische Untersuchungen zur Qualität von frühkindlichen Erziehungs- und Betreuungsformen und ihren Auswirkungen auf den kindlichen Entwicklungsstand - speziell solche Studien, die an größeren Stichproben durchgeführt wurden - lassen sich weitgehend zwei unterschiedlichen Ansätzen zuordnen (vgl. Harms & Clifford, in Druck). In einem Teil der Studien wurde die Qualität einer Betreuungsform über Input-Faktoren wie Gruppengröße, Ausbildung der Erzieherinnen oder zahlenmäßiges Erzieher-Kind-Verhältnis bzw. über eine Kombination solcher Faktoren definiert (vgl. z.B. die National Day Care Study, Ruopp et al., 1979; Howes & Olenick, 1986; Howes, 1990). Ein solcher Ansatz ist insofern begrenzt, da zwar solche Input-Faktoren bedeutende Bedingungen für tatsächliches Betreuungs- und Erziehungsgeschehen in den verschiedenen Formen darstellen, dieses aber nicht determinieren. Bei gleichen Input-Bedingungen sind durchaus unterschiedliches Betreuungs- und Erziehungsgeschehen innerhalb der Formen und damit unterschiedliche Auswirkungen auf die Kinder möglich. Andere Studien haben sich deshalb stärker dem tatsächlichen Betreuungs- und Erziehungsgeschehen in den Formen zugewandt. Forschungspraktisch haben sie dazu oftmals globalere Einschätzungen der Qualität einer frühkindlichen Erziehungs- und Betreuungsform benutzt, die versuchen, das Gesamt solcher prozessualen Aspekte zu bewerten. In einigen Studien wurden solche Globalbeurteilungen zusätzlich zu Input-Faktoren oder anderen Merkmalen einbezogen (vgl. z.B. die National Child Care Staffing Study, Whitebook, Howes & Phillips, 1990). Gerade im Rahmen von solchen Untersuchungen haben forschungsökonomisch anwendbare Screeninginstrumente in der Forschung ihren primären Einsatzbereich. Sie erlauben eine solche Globalbeurteilung der Qualität einer frühkindlichen Betreuungs- und Erziehungsform, die (neben Rahmenbedingungen, d.h. Inputs) ebenfalls prozessuale Aspekte berücksichtigt. Ein Einsatz scheint aber auch in Studien, die sich Spezialaspekten von Betreuungsformen zuwenden, geraten, um zu vermeiden, daß durch die Fokussierung auf Einzelaspekte das Gesamtbild einer Betreuungsform verloren geht.

Aus diesen Einsatzmöglichkeiten entstehen besondere Anforderung an die Qualität solcher Screeninginstrumente in der Forschung. Grundbedingung dürfte sein, daß das

Screeninginstrument informativ über die Qualität der untersuchten Formen berichtet und in der Lage ist, zwischen der Qualität der untersuchten Formen zu diskriminieren. Neben den klassischen Gütekriterien der Objektivität, Reliabilität und Validität dürften eine leichte Handhabbarkeit und leichte Trainierbarkeit der Beobachter sowie ein zeitökonomischer Einsatz eine besondere Rolle spielen. Die letzten Bedingungen dürften vor allem dann von Bedeutung sein, wenn das Screeninginstrument in Untersuchungen von Spezialaspekten frühkindlicher Betreuungs- und Erziehungsformen "nur so mitlaufen soll".

Screeninginstrumente - als Operationalisierungen von Qualität - können ebenfalls in Aus- und Fortbildung eingesetzt werden. Die Qualität einer Betreuungsform stellt nicht nur einen globalen Eindruck dar, den ein Betrachter von dieser Form gewinnt. Wenn auch die Einzelaspekte zusammenhängen, so lassen sich doch verschiedene Aspekte von Qualität unterscheiden, die u.U. einer differenzierteren Betrachtung bedürfen. Um eine solche differenzierte Betrachtung zu erleichtern, ist ein Raster erforderlich, das hilft, den Blick auf die spezifischen Komponenten zu lenken.<sup>4</sup> Die Auseinandersetzung mit einem der zur Diskussion stehenden Screeningsinstrumente bzw. seine Anwendung läßt sich dann als ein Training der spezifischen Wahrnehmung von Qualitätsaspekten betrachten. In diesem Sinne können die Screeninginstrumente fruchtbar in Aus- und Fortbildung von Betreuungspersonen des frühkindlichen Bereichs - nicht nur von Erzieherinnen in Einrichtungen - eingesetzt werden. Bei dieser Funktion treten die Bedeutung technischer Gütekriterien eines Instruments oder seine zeitökonomische Anwendung zugunsten seiner didaktisch-methodischen Möglichkeiten etwas in den Hintergrund.

Ein weiterer, im Hinblick auf eine Praxisverbesserung bedeutsamer Einsatz von Screeninginstrumenten dürfte im Rahmen von "Selbst"-Evaluationsprozessen gegeben sein.<sup>5</sup> Selbstevaluation kann als die Methode bezeichnet werden, bei der von den Praktikern selbst initiiert eine Betreuungsform kritisch durchleuchtet wird und nach Diskussion der Ergebnisse Veränderungen eingeleitet werden (Doherty, 1991, S. 75). Solche Selbstevaluationsprozesse bedürfen der Unterstützung von außen. Screeninginstrumente können dabei den Prozeß der Selbstevaluation anleiten und

---

<sup>4</sup> Das Gesamtbild der Qualität einer frühkindlichen Betreuungs- und Erziehungsform wird anschließend dann wieder aus den Einzelaspekten analytisch gebildet. Ein solches Vorgehen verhindert auch, daß ein erster Gesamteindruck durchschlägt und dann alle Aspekte - im Sinne eines Halo-Effekts - gleichermaßen beurteilt werden.

<sup>5</sup> Dieser Funktion lassen sich auch die Anwendungen von einigen der Screeninginstrumente in den USA im Rahmen der beschriebenen Akkreditierungs- bzw. "Credentialing" Programme sowie im Rahmen der Überprüfung staatlicher Regulierungen zuordnen (vgl. Clifford et al., in Druck; Harms & Clifford, in Druck).

systematisieren. Im Anschluß an die Datenerhebung können dann über Stärken und Schwächen einer Betreuungsform geredet und Veränderungsprozesse zur Verbesserung einer Betreuungsform eingeleitet werden. Selbstevaluation muß aber nicht zwingend implizieren, daß z.B. ausschließlich die jeweilige Erzieherin einer Kindergartengruppe ihre eigene Gruppe beobachtet und evaluiert. Denkbar sind auch Ergänzungen durch Beobachtungen von anderen Erzieherinnen oder Fachberater, wobei gerade Differenzen in den Wahrnehmungen stimulierend für Diskussion und anschließende Veränderungsprozesse sein dürften. Eine solche Selbstevaluation - die die Praktiker selbst zu Subjekten der Veränderungsprozesse macht - kann als Teil einer Professionalisierungsstrategie betrachtet werden. Allerdings stellen sich hier auch besondere Anforderungen an Screeninginstrumente, die in diesem Prozeß eingesetzt werden können. Zumindest fünf Anforderungen erscheinen als wichtig, damit ein Screeninginstrumente seine Rolle in Selbstevaluationsprozessen produktiv erfüllen kann (vgl. in einem ähnlichen Zusammenhang auch Roßbach, 1981):

- Das Screeninginstrument soll nicht nur einen globalen Blick auf die frühkindliche Betreuungs- und Erziehungsform liefern, vielmehr muß die Aufmerksamkeit auf spezifische Aspekte der Form gelenkt werden, die - zumindest im Prinzip - durch Aktionen der Betreuerin oder weiterer Verantwortlicher verändert werden können.
- Die Aspekte, die das Instrument anspricht, sollen zugleich auch von den Betreuungspersonen als bedeutsame Qualitätsaspekte betrachtet werden. Ansonsten dürften selbstinitiierte Veränderungsprozesse kaum erwartet werden können.<sup>6</sup>
- In vielen Fällen werden Menschen "blind" gegenüber ihrer eigenen Praxis. Damit sind sie überfordert, eine differenzierte Evaluation ihrer eigenen Arbeit zu leisten. Die Hauptanforderung an ein Screeninginstrument - besonders wenn die Betreuungspersonen ihre eigene Praxis selbst einschätzen - bezieht sich damit auf seine Fähigkeit, den Betreuern eine stärker objektive Bewertung der eigenen Arbeit zu ermöglichen. Das Screeninginstrument sollte ihnen helfen, Distanz zu der eigenen Praxis zu gewinnen und damit ihnen einen gewissermaßen neuen Blick auf ihre Praxis zu ermöglichen.
- Die Nützlichkeit eines solchen Screeninginstruments hängt vor allem auch von

---

<sup>6</sup> Dies impliziert nicht, daß im Einzelfall immer zwingend den in dem Screeninginstrument implizierten Bewertungen eines Aspektes als qualitativ gut oder qualitativ schlecht gefolgt werden muß. Andere Bewertungen werden dann aber explizit bewußt und diskutierbar. Gleichwohl sollten in der Mehrheit der Aspekte die impliziten Bewertungen des Instruments mit jenen der Betreuungspersonen übereinstimmen.



dem Zeitaufwand ab, der zu seiner Anwendung erforderlich ist. Die Nützlichkeit steigt zudem, wenn das Instrument leicht zu handhaben und der Aufwand für eine Einarbeitung gering ist.

- Schließlich muß das Screeninginstrument technische Gütekriterien wie Objektivität, Reliabilität und Validität erfüllen. Gerade der Prozeß einer Selbstevaluation benötigt technisch gute Hilfsmittel.

Diese Anforderungen an ein Screeninginstrument in Forschung, Aus- und Fortbildung sowie in Prozessen einer direkten Praxisveränderung müssen bei der in Kapitel 2 erfolgenden Vorstellung der verschiedenen Screeninginstrumente im Auge behalten werden.

## 2. Überblick über existierende Screeninginstrumente

Im folgenden wird ein Überblick über Screeninginstrumente zur Einschätzung der Qualität frühkindlicher Betreuungs- und Erziehungsformen gegeben. Dabei wird nicht der Anspruch auf Vollständigkeit erhoben. Auch mag der Berichterstatter Instrumente, die nach den folgenden Kriterien berücksichtigt hätten werden müssen, übersehen haben. Dieser Fehler liegt allein bei ihm. Andererseits muß berücksichtigt werden, daß die Instrumente oftmals nicht veröffentlicht vorliegen und somit nur im Rahmen von grauer Literatur - und damit erschwert - zugänglich sind. Gleichwohl wird davon ausgegangen, daß dieser Überblick weitgehend die Instrumente umfaßt, die derzeit in den USA zum Einsatz kommen. Im einzelnen werden aus dem Überblick Instrumente nach den folgenden Kriterien *ausgeschlossen*:

- *Instrumente, die sich auf Betreuungs- und Erziehungsformen für ältere Kinder beziehen*: Berücksichtigt werden damit nur Betreuungs- und Erziehungsformen für Kinder von der Geburt zum Schulbeginn; in einigen wenigen Fällen werden hier aufgrund der Besonderheiten des us-amerikanischen Betreuungssystems auch Hortformen - also Betreuung vor und/ oder nach dem täglichen Schulbesuch - mitberücksichtigt, soweit dort Kinder bis zum Alter von ca. acht Jahren betreut werden.
- *Ältere Instrumente, die vor 1980 entwickelt wurden und gegenwärtig praktisch kaum noch benutzt werden*.
- *Instrumente, die ausschließlich auf die Überprüfung der Voraussetzungen für eine Lizenzierung einer Betreuungsform gerichtet sind*: Als Beispiel hierfür mag das Child Development Program Evaluation System (CDPES) dienen, das in Pennsylvania seit 1978 benutzt wird und auch für Entwicklung anderer "Lizenzierungssysteme" herangezogen wurde (vgl. Fiene & Melnick, 1991). Das CDPES ist zudem weitgehend auf Gesundheits- und Sicherheitsaspekte beschränkt.
- *Instrumente, die nicht auf eine Gesamterfassung der Qualität einer frühkindlichen Erziehungs- und Betreuungsform bezogen, sondern auf ausgewählte Aspekte beschränkt sind*: Als Beispiele für nicht berücksichtigte Instrumente können herangezogen werden die Early Childhood Education Linkage System Scale, die sich im Sinne eines Selbstevaluations-Instruments ausschließlich auf Gesundheits- und Sicherheitsaspekte bezieht (vgl. Fiene & Melnick, 1991), oder die Early Childhood Physical Environment Scales, die das Ausmaß der räumlichen Abgrenzung von Aktivitätsbereichen in Einrichtungen - und damit auch nur einen Teilaspekt der physischen Umwelt im engeren - mißt (vgl.

Moore, 1986). Ebenfalls nicht berücksichtigt werden durch dieses Kriterium reine Verhaltens-Beobachtungssysteme, d.h. Instrumente, die sich überwiegend auf situativ beobachtbare Kind-Kind-Interaktionen, Betreuer-Kind-Interaktionen oder Kind-Spielzeug-Interaktionen beziehen. Als prominenteste Beispiele für nicht berücksichtigte Instrumente kann hier auf die Rahmen der National Day Care Study (Ruopp et al., 1979) eingesetzten Beobachtungssysteme verwiesen werden: zum einen auf das Prescott-SRI Child Observation System (Child Focus Instrument) (Stallings, Wilcox & Travers, 1976; Abt, 1980), das sich als Weiterentwicklung des Day Care Environment Inventory (Prescott, Kritchevsky & Jones, 1972) auf das Verhalten der Kinder konzentriert, zum anderen auf das SRI Preschool Observation System (Adult Focus Instrument) (Stallings et al., 1976; Abt, 1980), das - neben der Erfassung von einigen Rahmendaten z.B. zur Anzahl betreuter Kinder, Öffnungszeiten und einigen physischen Merkmalen wie Raumgröße, Ausstattung mit Spielmaterialien (Physical Environment Inventory) sowie der Aktivitäten und Gruppenbildungen (Classroom Snapshot) - sich hauptsächlich auf das Verhalten der Betreuerin bezieht.<sup>7</sup>

- *(einfache) Checklisten von Qualitätsmerkmalen:* Darunter werden hier reine Auflistungen von Qualitätsmerkmalen verstanden, die selbst nicht als "Meß"-Instrumente zu verstehen sind, sondern die (nur) dazu dienen, einen Maßstab für Veränderungsprozesse zu bilden, an deren Ende dann die Erreichung der Qualitätsstandards steht. Als Beispiel hierfür kann das im Rahmen von Head Start entwickelte und eingesetzte Self-Assessment/Validation Instrument (SAVI, 1979) dienen, in dem bis zu 350 Einzelaspekte/Standards angesprochen werden, d.h. deren Erfüllung festgestellt wird oder für deren Nicht-Erfüllung Gründe angegeben werden müssen.<sup>8</sup> Werden aber in diesem Prozeß "Meß"-Instrumente im engeren eingesetzt, die dazu dienen, in einem zeitökonomischen Einsatz Qualitätsmerkmale zu evaluieren, so sollen diese Instrumente berücksichtigt werden. Letzteres trifft z.B. auf die im Rahmen des NAEYC Akkreditierungsprogramms (National Association for the Education of Young Children, 1991a) eingesetzte Early Childhood Classroom Observation zu (National Academy of Early Childhood Programs, 1985; National Association for the Education of Young Children, 1991b, S. 15-40).

---

<sup>7</sup> Das SRI Observation System wurde - in einer etwas anderen Form - im Rahmen der Evaluation des Head Start Follow Through Planned Variation Project entwickelt und eingesetzt (vgl. Stallings, 1977).

<sup>8</sup> Zur Zeit ist daran gedacht, entsprechende Leitlinien speziell für Tagespflege aufzustellen. Allerdings liegen dazu noch keine veröffentlichte Unterlagen vor.

Positiv formuliert: Der folgende Überblick über Screeninginstrumente zur Qualitätseinschätzung enthält solche Instrumente, die sich auf Betreuungs- und Erziehungsformen für Kinder im vorschulischen Alter beziehen, die in einem engeren Sinne als "Meß"-Instrumente gekennzeichnet werden können, d.h. für einen routinemäßigen Einsatz geeignet sind und anhand der technischen Gütekriterien eines Meßinstrumentes beschrieben werden können, die für eine breitere, nicht nur auf ein bestimmtes Projekt beschränkte Anwendung geeignet sind, die eine Gesamteinschätzung der Qualität einer frühkindlichen Erziehungs- und Betreuungsform leisten und die relativ neue Instrumente darstellen bzw. derzeit in Gebrauch sind. Desweiteren soll ein gewisser Schwerpunkt auf Instrumenten liegen, die sich auf die Umwelten beziehen, in denen unmittelbar das Betreuungs- und Erziehungsgeschehen stattfindet. Am Beispiel von institutionellen Gruppenbetreuungen wäre dies die Gruppe, wie sie mit dem Gruppenraum bzw. Räumen, den in dieser Gruppe betreuten Kindern sowie den in dieser Gruppe arbeitenden Erzieherinnen gegeben ist.

Die Instrumente sind dabei entsprechend ihres Anwendungsbereichs gegliedert: Zuerst werden Instrumente vorgestellt, die sich auf institutionelle Gruppenbetreuung in Einrichtungen des frühkindlichen Bereichs beziehen. Anschließend werden Instrumente für eine Qualitätseinschätzung von Tagespflegestellen diskutiert. Den Abschluß bildet ein Instrument zur Bewertung der häuslichen Erziehungs- und Betreuungsumwelt eines Kindes. Die Hauptcharakteristika der verschiedenen Instrumente werden am Ende dieses Berichts in Tabellenform zusammengefaßt (vgl. Kapitel 3). Bei der Darstellung der Instrumente wird versucht, jeweils bei allen Instrumenten die gleichen Aspekte anzusprechen. Aufgrund der teilweise unzureichenden oder nur begrenzt zugänglichen Informationen zu verschiedenen Instrumenten wird dies jedoch nicht immer möglich sein.

## ***2.1 Instrumente für institutionelle Gruppenbetreuungen***

### ***2.1.1 Early Childhood Classroom Observation***

#### ***Entstehung und Anwendungskontext***

Die Early Childhood Classroom Observation (National Academy of Early Childhood Programs, 1985; National Association for the Education of Young Children, 1991b, S. 15-40) wurde als Teil des Akkreditierungsprogramms der NAEYC (National Association for the Education of Young Children, 1991a) entwickelt.

Zum besseren Verständnis des Anwendungskontextes eine kurze Skizze des Akkreditierungsprogramms: Das Akkreditierungsprogramm besteht in seiner Grundstruktur darin zu überprüfen,

ob eine Einrichtung des frühkindlichen Bereichs<sup>9</sup> einen Satz von Qualitätskriterien erfüllt. Falls dies festgestellt wird, wird der Einrichtung eine besondere Anerkennung, die Akkreditierung, verliehen. Die NAEYC-Kriterien für eine gute Qualität wurden zu Beginn der 80er Jahre in einem etwa dreijährigen Zeitraum entwickelt. Dazu wurden etwa 50 Evaluationsberichte und die Forschungsliteratur zu Effekten von Merkmalen frühkindlicher Betreuungsformen auf die Entwicklung der Kinder gesichtet. Zudem wurden in verschiedenen Wellen Experten des frühkindlichen Bereichs sowie das Fachpersonal selbst um kritische Stellungnahmen zu den Qualitätskriterien aufgefordert. In der Entwicklungsphase wurden z.B. 250 Experten aufgefordert, einen Satz von Kriterien (der sich kaum von den letztendlich ausgewählten Kriterien unterscheidet) auf einer vierstufigen Wichtigkeitsskala einzuschätzen. Bei einer Rücklaufquote von 78 % wurden alle Kriterien als "important" (3) oder "essential" (4) eingestuft (vgl. Bredekamp, 1989). Aufgrund der Erfahrungen in den ersten fünf Jahren des Akkreditierungsprogramms wurden dann die Kriterien revidiert. Die revidierten Kriterien liegen seit Juli 1991 vor, allerdings werden die alten Kriterien im Akkreditierungsprogramm noch bis September 1992 benutzt. Insgesamt werden die NAEYC-Kriterien in den USA als eine anerkannte Definition von guter Qualität einer institutionellen Gruppenbetreuung betrachtet (vgl. Harms & Clifford, in Druck).

Insgesamt liegen 91 Kriterien vor, die in 10 Komponenten (Dimensionen) gegliedert sind, wobei die Anzahl der Kriterien pro Komponente von 3 bis 22 variiert:

- A. Interactions among staff and children
- B. Curriculum
- C. Staff-parent interaction
- D. Staff qualifications and development
- E. Administration
- F. Staffing
- G. Physical environment
- H. Health and safety
- I. Nutrition and food service
- J. Evaluation

Das Akkreditierungsprogramm wird von der National Academy of Early Childhood Programs, einer Abteilung von NAEYC, verwaltet. Der Akkreditierungsprozeß läuft in vier Stufen ab:

- Im ersten Schritt führt die Einrichtung, d.h. die Mitarbeiter der Einrichtung, anhand der Materialien des Programms eine Selbststudie durch. Das Ziel dieser Phase besteht darin, Stärken und Schwächen der Arbeit der eigenen Einrichtung festzustellen und Veränderungen so zu initiieren, daß die NAEYC-Kriterien erfüllt werden. Diese Selbststudie wird durch verschiedene Instrumente angeleitet: die *Early Childhood Classroom Observation*, die von jeder Gruppenleiterin und von der Einrichtungsleiterin ausgefüllt werden muß, den *Administrator Report*, ein Arbeitsbuch, mit dem die Einrichtungsleiterin die administrativen Aspekte der Einrichtung beurteilen soll, den *Staff Questionnaire*, der von allen Betreuern, die mit Kindern arbeiten, ausgefüllt werden muß und der sich auf verschiedene Verwal-

---

<sup>9</sup> Um an dem Programm teilnehmen zu können, müssen in der Einrichtung mindestens zehn Kinder im Alter von null bis fünf Jahren ganztags oder halbtags oder Kinder im Schulalter (bis acht Jahre) vor und/oder nach der Schule von mindestens zwei Erwachsenen betreut werden.

tungsaspekte bzw. Aspekte, die das Personal direkt angehen, bezieht, und den *Parent Questionnaire*, mit dem allen Eltern die Gelegenheit gegeben werden soll, die Einrichtung, besonders die Qualität der Interaktionen zwischen Personal und Eltern, zu evaluieren. Eine solche Selbststudie kann oftmals bis zu einem Jahr dauern und dient zunächst dazu, die Aspekte, die schlecht abschnitten, zu verbessern. Wenn die Einrichtung mit ihren Verbesserungen zufrieden ist, kann sie die Akkreditierung beantragen und damit den nächsten Schritt einleiten - sie muß es aber nicht, sie kann auch diesen pädagogischen Prozeß der Selbststudie an dieser Stelle abbrechen.

- Im zweiten Schritt wird ein Bericht über die Selbststudie an die National Academy of Early Childhood Programs erstellt. Dieser Bericht - die *Program Description* - erfolgt in einer formalisierten Weise. Auch wenn die Arbeit einer Einrichtung nicht mit allem NAEYC-Kriterien übereinstimmt, kann eine Akkreditierung beantragt werden; es müssen dann aber die Nicht-Übereinstimmungen kommentiert werden.
- Im dritten Schritt besuchen je nach Größe der Einrichtung ein oder mehrere trainierte Gutachter (validator) einen oder mehrere Tage die Einrichtung und überprüfen, ob die Program Description akkurat die Realität in der Einrichtung widerspiegelt. Der Gutachter selbst trifft keine Entscheidungen über die Akkreditierung; er berichtet nur an die Academy. Der Gutachter besichtigt die gesamte Einrichtung, beobachtet mit der Early Childhood Classroom Observation in den Gruppen (bei kleinen Einrichtungen in allen Gruppen, bei größeren werden einige Gruppen zufällig ausgewählt), überprüft schriftliche Unterlagen und geht mit der Einrichtungsleiterin die Program Description durch.
- Aufgrund des Berichts der Gutachter und der schriftlichen Unterlagen entscheidet dann eine Kommission aus drei Experten des frühkindlichen Bereichs über die Akkreditierung. Die Entscheidung über eine Akkreditierung erfolgt dabei nicht schematisch nach einem quantitativen Kriterium (z.B. x % der NAEYC-Qualitätskriterien müssen erfüllt werden), vielmehr basieren die Akkreditierungsentscheidungen auf "professional judgement". Dies hängt damit zusammen, daß die einzelnen Kriterien unterschiedlich wichtig sind und sie deshalb nicht schematisch zusammengefaßt werden können. So werden einige Kriterien als wichtiger als andere angesehen. Da aber dieses "Wichtigkeitssystem" nirgendwo explizit festgeschrieben ist, gibt es eine größere Subjektivität in den Entscheidungen der Kommissionen. Andererseits aber haben die ersten Jahre des Akkreditierungsprogramms nach Bredekamp (1989) gezeigt, daß es eine hohe Konsistenz zwischen den verschiedenen Kommissionen gibt. Wird eine Akkreditierung ausgesprochen, so ist sie für drei Jahre gültig, wobei aber jährliche Berichte abgeliefert werden müssen. Nach drei Jahren muß der Prozeß mit einer Selbststudie neu begonnen werden.

Die Early Childhood Classroom Observation wendet sich dem Geschehen in einzelnen Gruppen zu, in denen Kinder im vorschulischen Alter (null bis fünf Jahre) halbtags oder ganztags bzw. Schulkinder (bis zu acht Jahren) vor und/oder nach der Schule betreut werden. Sie enthält aus dem Gesamtsatz der NAEYC-Kriterien nur

die Kriterien, die in einer Gruppe beobachtet werden können.<sup>10</sup> Bei einigen Items wurde die Formulierung etwas verändert, um ein Kriterium besser beobachtbar zu machen. Ebenso wurden einige Kriterien der Gesamtliste in mehrere Teilkriterien aufgegliedert. Dennoch sind die Items in der Early Childhood Classroom Observation sehr eng mit den NAEYC-Qualitätskriterien verbunden und dürften von daher auch ihre (Experten-) Validität beziehen.

Im Akkreditierungsprozeß wird die Early Childhood Classroom Observation zum einen von nicht-trainierten Laien (Gruppenerzieherinnen, Einrichtungsleiterinnen), zum anderen von trainierten Experten (Gutachtern/validators) angewandt. Über das Training der Gutachter liegen keine Informationen vor. Der Einsatz durch Laien in der Selbststudie setzt aber voraus, daß das Instrument aus sich selbst heraus verständlich ist und somit kein spezielles Training benötigt.

### *Aufbau des Instruments*

Die Early Childhood Classroom Observation besteht aus 68 Items/Kriterien in fünf der insgesamt zehn Dimensionen der NAEYC-Kriterien:<sup>11</sup>

- A. Interactions among staff and children, 15 Items
- B. Curriculum, 23 Items
- G. Physical environment, 11 Items
- H. Health and safety, 18 Items
- I. Nutrition and food service, 1 Item

Items aus den Dimensionen C. Staff-parent interaction, D. Staff qualification and development, E. Administration, F. Staffing und J. Evaluation sind nicht enthalten, da sie nicht direkt beobachtbar sind bzw. sich nicht auf die einzelnen Gruppen, sondern auf Aspekte der Einrichtung insgesamt beziehen. Im Rahmen des Akkreditierungsprogramms werden diese Aspekte durch weitere Instrumente abgedeckt, auf die hier aber nicht eingegangen werden soll.

Die Items/Kriterien sind numeriert und werden als kurze (numerierte) Statements

---

<sup>10</sup> Die Numerierung der Items (mit Buchstaben und Zahlen) in der Early Childhood Classroom Observation entspricht der Numerierung in der Gesamtliste der NAEYC-Kriterien.

<sup>11</sup> Im folgenden wird die Version von 1985 beschrieben (National Academy of Early Childhood Programs, 1985). Die revidierte Version, die ab September 1992 eingesetzt wird, unterscheidet sich allerdings nur unwesentlich von dieser älteren Version. Ein Item zum Umgang mit Medien wurde zusätzlich aufgenommen, so daß das Instrument nun aus 69 Items besteht. Daten zur Anwendung dieser neueren Version liegen noch nicht vor.

vorgegeben, wobei auf einer dreistufigen Skala eingeschätzt werden muß, inwieweit das Statement zutrifft. Die drei Stufen sind:

- 1 - not met: Es gibt wenige Hinweise darauf, daß dieses Kriterium in der Gruppe erfüllt ist.
- 2 - partially met: Es gibt einige Hinweise darauf, daß dieses Kriterium in der Gruppe erfüllt ist.
- 3 - fully met: Es gibt klare Hinweise darauf, daß dieses Kriterium in der Gruppe erfüllt ist.

Bei vielen Items stehen zusätzlich noch sogenannte Indikatoren (important points to look for); im Prozeß des Beobachtens muß angekreuzt werden, wenn diese Indikatoren vorhanden sind. Anhand der dreistufigen Ratingsskalen werden aber nicht diese Indikatoren, sondern nur die numerierten Items bewertet. Für eine Einschätzung eines Kriteriums als "3" müssen alle Indikatoren vorhanden sein. Auch wenn alle Indikatoren vorhanden sind, muß aber nicht zwingend eine "3" vergeben werden. Bei einer Vergabe einer "1" oder "2" muß ein Kommentar gegeben werden. Zusätzlich stehen bei einigen Kriterien noch erläuternde Beispiele. Wenn in einigen Fällen Items sich auf spezifische Altersgruppen beziehen, die aber in der zu beobachtenden Gruppe nicht vorhanden sind, kann "not applicable" angekreuzt werden.

Die Items/Kriterien brauchen nicht in der Reihenfolge bearbeitet zu werden, in der sie im Instrument stehen. Wenn es in einer Gruppe mehr als eine Erzieherin gibt, dann sollen sich die Ratings auf alle Erwachsenen in der Gruppe beziehen. Als Orientierungsregel wird formuliert: "Rate criteria on the quality of the children's experience" (National Academy of Early Childhood Programs, 1985, S. 3).

Im folgenden wird jeweils ein Beispiel aus der Dimension A. Interactions among staff and children und aus der Dimension G. Physical environment gegeben (anhand der dreistufigen Ratingskala werden nur die Skalenwerte, die unmittelbar neben der Numerierung stehen, eingeschätzt, nicht die jeweils darunter stehenden Indikatoren):



A-1. Staff interact frequently with children showing affection and respect.

- 0 Staff interact nonverbally by smiling, touching, holding.
- 0 Staff talk with individual children during routines (arriving/departing, eating, and activities).

G-2. Space is arranged to accommodate children individually, in small groups, and in large group.

- 0 There are clear pathways for children to move from one area to another without disturbing activities.
- 0 Areas are organized for easy supervision by staff.

Die Early Childhood Classroom Observation besteht zwar insgesamt nur aus 68 Items, die auf der dreistufigen Ratingskala bewertet werden müssen. Andererseits müssen die Beobachter aber auf eine viel größere Anzahl von Aspekten achten, deren Beschaffenheit bei dem Rating der Items berücksichtigt werden muß. So gibt es z.B. 80 dieser Indikatoren (important points to look for), wobei diese Indikatoren aber nur bei 22 der 68 Items stehen.<sup>12</sup> Insofern sind die einzustufenden Items recht komplex. Es geht nicht nur um ein einfaches "Aspekt ist vorhanden" versus "Aspekte ist nicht vorhanden". Vielmehr müssen komplexere Informationsverarbeitungsprozesse bei den Beobachtern vorausgesetzt werden, in denen viele Einzelinformationen zu globaleren, zum Teil von den konkreten Einzelaspekten abstrahierenden Ratings zusammengefaßt werden.

Für die Zwecke dieses Berichtes sollen drei Niveaus unterschieden werden, die sich darauf beziehen, inwieweit bei der Bearbeitung der Items von beobachteten Einzelaspekten abstrahiert werden muß.

- *Abstraktionsniveau 1 oder geringes Abstraktionsniveau:* Die Items sind so formuliert, daß keine oder nur geringe Abstraktionen von beobachtbaren Einzelaspekten erforderlich werden. Die Items sind mehr im Sinne von Checklisten formuliert; an die Fähigkeiten des Beobachters, verschiedene Einzelaspekte zu einem Gesamturteil zusammenzufassen, werden keine oder nur geringe Anforderungen gestellt.
- *Abstraktionsniveau 2 oder mittleres Abstraktionsniveau:* Die Items sind so formuliert, daß zwar der Beobachter bei der Bearbeitung der Items von Einzelaspekten abstrahieren und zu einer zusammenfassenden Beurteilung kom-

---

<sup>12</sup> Die Early Childhood Classroom Observation von 1991 enthält sogar 110 solcher "important points to look for" (National Association for the Education of Young Children, 1991b, S. 15-40).

men muß. Jedoch sind diese zusammenfassenden Beurteilungen noch relativ nahe an Einzelaspekten orientiert, oder es werden nur zusammenfassende Urteile im Hinblick auf wenige Einzelaspekte erforderlich.

- *Abstraktionsniveau 3 oder höheres Abstraktionsniveau:* Die Items sind komplex, sie verlangen vom Beobachter eine zusammenfassende Bewertung von mehreren Einzelaspekten. An die Verarbeitungskapazitäten des Beobachters werden erhebliche Anforderungen gestellt.

Im Sinne dieser Arbeitseinteilung können die Items der Early Childhood Classroom Observation dem höheren Abstraktionsniveau 3 zugeordnet werden.

Wenn eine Gruppenleiterin ihre eigene Gruppe einschätzt, so benötigt sie ca. 1 1/2 Stunden, um die Early Childhood Classroom Observation auszufüllen, wobei dies dann eher auf einem Erinnern an jenes, was die Erzieherin gewöhnlich tut, als auf Beobachtung beruht. Die Einschätzung soll immer das reflektieren, was typischerweise geschieht, und nicht Wünsche oder zukünftige Pläne. Die Einrichtungsleiterin und der Gutachter (validator) führen demgegenüber eine Beobachtung durch, für die 1 bis 2 Stunden benötigt werden. Wenn während der Beobachtung keine oder zuwenig Informationen zur Beurteilung eines Items anfallen, so soll anschließend die Gruppenleiterin befragt werden; das entsprechende Item muß dann mit einem "R" für report gekennzeichnet werden.

Additive Zusammenfassungen der Ergebnisse in den fünf erfaßten Dimensionen oder für das Instrument insgesamt sind im Rahmen des Akkreditierungsprozesses nicht geplant. Ebenfalls gibt es kein quantitatives Kriterium dafür, ab wann z.B. von einer guten Qualität der beobachteten Gruppe gesprochen werden kann. Ähnlich wie bei der Akkreditierungsentscheidung insgesamt (siehe weiter oben) wird davon ausgegangen, daß die einzelnen Items nicht gleichgewichtet sind und somit eine rein additive Zusammenfassung weniger aussagekräftig ist.

### **Gütekriterien**

Berücksichtigt man den relativ breiten Gebrauch der Early Childhood Classroom Observation, so überrascht, daß nur relativ wenig Informationen zu den technischen Gütekriterien dieses Instruments vorliegen. In einer Study außerhalb des Akkreditierungsprogramms (n = 5 Gruppen) berichten Holloway and Reichhart-Erickson eine Interrater-Reliabilität von .89 (vgl. Harms & Clifford, in Druck; persönliche Mitteilung von S. Bredekamp an die Autoren). Mithin scheint eine relativ hohe Objektivität in der Anwendung der Early Childhood Classroom Observation gegeben zu sein, was wohl auch auf die relativ einfache dreistufige Ratingsskala, auf der die Kriterien

zu bewerten sind, zurückgeführt werden kann.

Umfangreichere Informationen zu den technischen Kriterien liegen nur aus einem Feldtest mit einer Vorläuferversion mit noch 75 Items vor, der vor dem Beginn des Akkreditierungsprogramms durchgeführt wurde und der zu Modifikationen an dem Instrument - u.a. zu der Reduktion auf 68 Items - führte (vgl. Bredekamp, 1986). In insgesamt 31 Einrichtungen wurden 131 Gruppen (unterschiedliche Altersgruppen) von den Gruppenleiterinnen, den Einrichtungsleiterinnen sowie von außenstehenden Gutachtern (validators) anhand des Instruments eingeschätzt. Für die Auswertungen wurde alle Einschätzungen zusammengenommen, so daß Werte von 387 "Gruppen" in die Analysen eingingen.<sup>13</sup> Die interne Konsistenz (Cronbach's Alpha) des gesamten Instruments - als Maß für die Reliabilität - beträgt .70 (standardisiertes Alpha .86). Hierbei muß allerdings die hohe Anzahl von 75 Items berücksichtigt werden. Insgesamt fallen die Item-Total-Korrelationen (alle Items additiv zusammengefaßt) recht unterschiedlich und insgesamt eher niedrig aus. Die höchsten Korrelationen um .50 finden sich bei Items, die curriculare Aktivitäten ansprechen, ansonsten liegen die Korrelationen in der Regel um .20 bis .40, in vielen Fällen aber niedriger.

Zur Überprüfung der Konstrukt-Validität, genauer: der faktoriellen Validität, wurde eine Hauptkomponentenanalyse mit anschließender VARIMAX-Rotation durchgeführt. Als Ergebnis zeigten sich drei Faktoren mit Eigenwerten größer als 1, die zusammen 48 % der Varianz aufklären.<sup>14</sup> Die veröffentlichte Ladungsmatrix enthält allerdings jeweils nur die höchste Ladung eines Items auf einem Faktor, so daß komplexe Ladungen nicht berichtet werden und somit die Bewertung dieser Faktorenanalyse erschwert ist. Berichtet werden 26 Items mit Ladungen größer .35. Die drei Faktoren werden wie folgt benannt:

- preschool curriculum (24,3 % aufgeklärte Varianz)
- positive interaction (15 %)
- balanced schedule of activity (8,4 %)

Aufgrund der Ergebnisse dieses Feldtestes wurde die Early Childhood Classroom

---

<sup>13</sup> Technisch gesprochen sind damit die einzelnen "Gruppenwerte" nicht unabhängig, vielmehr ist in der Regel eine Gruppe dreimal in dem Datensatz enthalten, einmal eingeschätzt durch die Gruppenleiterin, dann durch die Einrichtungsleiterin und schließlich durch den außenstehenden Gutachter.

<sup>14</sup> Die Faktorenanalyse wurde wiederum mit allen Beobachtungen durchgeführt. Getrennte Analysen für die Beobachtungen der Gruppenleiterinnen, der Einrichtungsleiterinnen und der außenstehenden Gutachter bestätigten aber die Ergebnisse der gemeinsamen Faktorenanalyse mit allen Einschätzungen (Bredekamp, 1986, S. 113-115).

Observation modifiziert. Obwohl Bredekamp (1986) auf die Notwendigkeit verweist, eine weitere Reliabilitäts- und Validitätsstudie mit der letztlich eingesetzten Early Childhood Classroom Observation an einer größeren Stichprobe durchzuführen, wurde eine solche Studie bisher nicht durchgeführt (persönliche Mitteilung, S. Bredekamp, 7. April 1992).

### **2.1.2 Classroom Practices Inventory** **Entstehung und Anwendungskontext**

Das CPI (Classroom Practices Inventory) wurde in dem Projekt "Academic Environments in Early Childhood: Challenge or Pressure" entwickelt (vgl. Hyson, Hirsh-Pasek & Rescorla, 1990). Das Projekt untersucht den Zusammenhang zwischen elterlichen Einstellungen hinsichtlich akademischer Orientierungen in der Erziehung in der frühen Kindheit, der elterlichen Auswahl einer Einrichtung für ihr Kind und verschiedener Aspekten des kindlichen Verhaltens. Eine Forschungsfrage war u.a., ob Eltern mit hohen akademischen Wertorientierungen auch Einrichtungen für ihr Kind auswählen, die ihrerseits eine hohe akademische Orientierung aufweisen. Um das Ausmaß der akademischen Orientierung einer Einrichtung bzw. einer Gruppe zu messen, wurde das CPI entwickelt. Das CPI versteht sich dabei als eine Operationalisierung der NAEYC Guidelines for Developmentally Appropriate Practices für Einrichtungen bzw. Gruppen mit Kindern im Alter von vier und fünf Jahren (Bredekamp, 1987).

Die NAEYC Guidelines for Developmentally Appropriate Practices für Kinder dieser Altersgruppe bestehen aus 46 Statements, die in 23 Paaren von Statements angeordnet sind. Jeweils ein Statement eines Paares beschreibt eine angemessene Praxis, während das zweite dann die korrespondierende unangemessene Praxis enthält. Die NAEYC Guidelines for Developmentally Appropriate Practices sind entstanden bzw. verstehen sich als Gegenposition zu einem mehr formalen, akademischen und "lehrerorientierten" Ansatz in der frühkindlichen Erziehung, wobei eine solche akademische Orientierung als entwicklungsmäßig unangemessen betrachtet und kritisiert wird. Die NAEYC Guidelines sind aber nicht unumstritten, auch unter Experten des frühkindlichen Bereichs selbst (vgl. Hyson et al., 1990, S. 488). Sie stellen *einen* Ansatz zur pädagogischen Arbeit in Einrichtungen des frühkindlichen Bereichs dar, der zudem mehr auf den Prozeß des Erziehungsgeschehens und weniger auf curriculare Inhalte bezogen ist. Allerdings - dabei dürfte die Kennzeichnung des Ansatzes als "entwicklungsmäßig angemessen" wohl eine Rolle spielen - werden in der Diskussion oftmals dieser Ansatz verabsolutiert und damit schnell alternative Ansätze als "entwicklungsmäßig unangemessen" diskreditiert.

Die NAEYC Guidelines konnten nicht direkt als Items für das CPI übernommen werden. Die konkreten Items wurden u.a. ausgewählt bzw. umformuliert nach den folgenden Regeln:

- Das Item soll aufgrund von direkten Beobachtungen während eines mehr-

stündigen Besuchs einschätzbar sein.

- Das Item soll sich direkt auf die Diskussion über akademische Orientierung der Arbeit vs. stärkere Spielorientierung beziehen.

Neben Items zur akademischen Orientierung enthält das CPI auch einige Items zum sozialen Klima. Soziales Klima und akademische Orientierung in einer Gruppe können unabhängige Aspekte sein. Aus diesem Grunde wurden einige Items aus den NAEYC-Akkreditierungskriterien übernommen, um die affektive Komponente der pädagogischen Arbeit mit zu erfassen.

Das CPI beabsichtigt nicht, ein generelles Maß für die Qualität einer Einrichtung bzw. einer Gruppe wie z.B. die Early Childhood Classroom Observation zu sein. Vielmehr soll mit dem Ausmaß der akademischen Orientierung ein eingeschränkterer Aspekt der pädagogischen Arbeit beobachtet und eingeschätzt werden. Andererseits jedoch sprechen die Items des CPI die pädagogische Arbeit in Gruppen relativ breit an, so daß das CPI hier in diesem Überblick berücksichtigt wird.

Die vorliegende Version des CPI wird von trainierten, außenstehenden Beobachtern angewandt; über den Umfang des erforderlichen Trainings gibt es keine genauen Angaben.

### ***Aufbau des Instruments***

Das CPI besteht aus zwei Teilen mit zusammen 26 Items:

- Die ersten 20 Items stehen unter der Überschrift "Program Focus" (bzw. Program/Activity Focus). Zehn Items sind dabei so formuliert, daß eine Zustimmung zu ihnen eine "angemessene" Praxis indiziert, während die anderen zehn Items so formuliert sind, daß eine Zustimmung eine "unangemessene" Praxis indiziert. Die Items sind gemischt angeordnet.
- Sechs weitere Items stehen unter der Überschrift "Emotional Climate"; vier Items sind positiv und zwei Items negativ formuliert.

Die Items sollen aufgrund von Beobachtungen auf einer fünfstufigen Ratingskala eingeschätzt werden, wobei zu jedem Item noch zusätzliche Kommentare gegeben werden können:

- 1 - not at all like this classroom; der angesprochene Aspekt trifft auf die beobachtete Gruppe nicht zu.

... bis ..

- 5 - very much like this classroom; der angesprochene Aspekt trifft völlig auf die beobachtete Gruppe zu.

Zwei Beispiele für Items auf dem Bereich "Program Focus" (das erste repräsentiert eine "angemessene" Praxis, das zweite eine "unangemessene"):

1. Children select their own activities from a variety of learning areas the teacher prepares, including dramatic play, blocks, science, math, games and puzzles, books, recording, art, and music.
10. Reading and writing instruction emphasizes direct teaching of letter recognition, reciting the alphabet, coloring within the lines, and being instructed in the correct formation of letter.

Die 26 Items müssen als relativ komplexe Items charakterisiert werden. Sie verlangen in der Regel nicht (nur) eine Einschätzung von diskreten Einzelaspekten, vielmehr müssen in dem Rating mittels der fünfstufigen Skala verschiedene Einzelaspekte zusammengefaßt werden. Dementsprechend können die Items dem Abstraktionsniveau 3 (höheres Abstraktionsniveau) zugeordnet werden.

Die Items sollen unmittelbar *nach* dem Besuch einer Gruppe ausgefüllt werden. Für die Beobachtungen in einer Gruppe werden etwa 2 1/2 Stunden benötigt.

Die Ergebnisse zu den einzelnen Items werden additiv zusammengefaßt, wobei die im Sinne des Konzepts der entwicklungsangemessenen Praxis negativ formulierten Items vorher umgepolt werden müssen, so daß hohe Skalenwerte immer eine angemessene Praxis indizieren. Es gibt eine Gesamtskala und vier Subskalen: appropriate program focus items (zehn Items, die im Sinne des Konzepts der entwicklungsangemessenen Praxis positiv formuliert sind), inappropriate program focus items (zehn Items, die in diesem Sinne negativ formuliert sind), total program focus items (alle 20 Items aus dem Bereich Program Focus) und emotionale climate (alle sechs Items aus diesem Bereich). Durch die Berechnung von Skalenmittelwerten kann die ursprüngliche fünfstufige Skala zur Beurteilung der Qualität einer Gruppe herangezogen werden.

### **Gütekriterien**

In zehn Einrichtungen wurden die Gruppen (die Anzahl der Gruppen wird nicht berichtet) jeweils unabhängig von zwei Beobachtern eingeschätzt. Die Meßwertreihen der beiden Beobachtungen korrelieren mit .86. Im Durchschnitt stimmten die

Beobachter in 64 % der Antworten exakt überein. Betrachtet man noch Abweichungen von einem Skaleneinheit auf der fünfstufigen Skala als Übereinstimmung, so beträgt die durchschnittliche Übereinstimmung 98 %. Dem CPI kann somit eine gute Objektivität zuerkannt werden.

Das CPI insgesamt sowie die einzelnen Subskalen weisen eine sehr gute Reliabilität im Sinne der internen Konsistenz auf (Cronbach's Alpha, bei 207 Gruppen aus 58 Einrichtungen): Für die Gesamtskala aus 26 Items ergibt sich .96; für die Skalen appropriate program focus items .92, inappropriate program focus items .93, total program focus items .96 und emotional climate .88. Zugleich ergeben sich sehr hohe Korrelationen der Skalen untereinander (größer/gleich .75) sowie zwischen den Skalen und dem Gesamtwert (größer/gleich .90), so daß die Skalen hochgradig miteinander zusammenhängende Aspekte der pädagogischen Arbeit in Gruppen messen.

Eine Faktorenanalyse zur Aufdeckung der faktoriellen Validität ( $n = 207$  Gruppen; VARIMAX Rotation) ergab vier Faktoren mit Eigenwerten größer 1. Die vier Faktoren erklären zusammen 69 % der gesamten Varianz der 26 Items und reproduzieren größtenteils die unterlegte Struktur: Ein Faktor umfaßt die Items zum sozialen Klima, ein anderer die Items aus der Skala inappropriate program focus items. Die zehn Items aus der Skala appropriate program focus items verteilen sich auf die beiden anderen Faktoren. Allerdings ist diese Faktorenanalyse nur schwierig zu beurteilen, da nicht die volle Ladungsmatrix, sondern nur jeweils für ein Item die höchste Ladung auf einem Faktor berichtet wird. Die offensichtlich vorhandene komplexe Struktur mit Items mit höheren Ladungen auf mehreren Faktoren wird damit verdeckt (vgl. Hyson et al., 1990, S. 482f.). Zudem berichten die Autoren offensichtlich fälschlicherweise nicht die auf die einzelnen Faktoren entfallenden Varianzanteile *nach* der Rotation, sondern jene *vor* der Rotation.

An den Gruppen aus 10 Einrichtungen wurden verschiedene weitere Aspekte einer Konstrukt-Validität überprüft. Einschränkend muß aber festgehalten werden, daß die Einrichtungen alle Halbtageseinrichtungen waren, die von Kindern aus der mittleren und der oberen Mittelschicht besucht wurden: Die akademischen Orientierungen von Müttern (hohe Werte entsprechen einer höheren Erwartung und einer positiven Einstellung zu frühem formalem Lernen und direkter Unterweisung der Kinder durch Erwachsene) korrelieren mit  $-.41$  mit den CPI-Werten der Einrichtungen, die ihre Kinder besuchen. Mütter mit hohen akademischen Erwartungen wählen demnach eher Einrichtungen mit niedrigeren CPI-Werten und damit einer höheren akademischen Orientierung aus. Es gibt eine leicht positive Korrelation ( $r = .33$ ) derart, daß Kinder aus Gruppen mit hohen CPI-Werten höher in ihrer Kreativität eingeschätzt werden. Entwicklungsgemäße Umwelten - nach der NAEYC-Definition - haben

demnach eher kreativere Kinder. Diese Beziehung bleibt auch dann bestehen, wenn - als Maß für den familialen Hintergrund - die mütterlichen akademischen Orientierungen kontrolliert werden. Kinder aus Gruppen mit höheren CPI-Werten zeigen sich auch weniger ängstlich in Testsituationen ( $r = .23$ ; Angst gemessen auf einer siebenstufigen Skala, wobei höhere Werte Gelassenheit und Vertrauen in Testsituationen indizieren). Diese Beziehung bleibt bestehen, wenn wiederum die mütterlichen akademischen Orientierungen kontrolliert werden ( $r = .26$ ). Betrachtet man nur die Kinder von Müttern mit einer hohen akademischen Orientierung ( $n = 38$ ), so zeigt sich auch hier, daß Kinder aus Einrichtungen mit einem entwicklungsgemäßen Programm (nach der NAEYC-Definition) als weniger ängstlich in Testsituationen eingeschätzt werden. Zusätzlich hatten diese Kinder positivere Einstellungen zum (späteren) Schulbesuch.

Die berichteten Ergebnisse verweisen auf sinnvoll interpretierbare Beziehungen der CPI-Werte zu anderen Konstrukten und bestätigen damit eine Konstrukt-Validität. Allerdings bedürfen diese Beziehungen zu ihrer Bestätigung nach Ansicht der Autoren noch weiterer Untersuchungen an größeren Stichproben und vor allem an Stichproben von Einrichtungen mit einem anderem sozialen Einzugsbereich. Ebenfalls darf aus den berichteten Beziehungen nicht vorschnell auf kausale Beziehungen zwischen dem Ausmaß der entwicklungsgemäß angemessenen Arbeit in einer Gruppe (gemessen über das CPI) und den verschiedenen anderen Maßen geschlossen werden, da die Kinder und Familien nicht zufällig auf die verschiedenen Einrichtungen aufgeteilt wurden, die Kontrolle von anderen möglichen Ursachen (Kontrolle von mütterlichen Einstellungen, teilweise Alter des Kindes) eher gering ausfällt und auch keine längsschnittlichen Daten vorliegen.

### ***2.1.3 CDA Assessment Observation Instrument/Advisor Report Form Entstehung und Anwendungskontext***

Das CDA Assessment Observation Instrument, auch Advisor Report Form oder Field Advisor Formal Observation genannt (Council for Early Childhood Professional Recognition, 1991b, 1992; Phillips, 1991a, S. 111-150), wurde im Rahmen des CDA (Child Development Associates) National Credentialing Program entwickelt (vgl. z.B. Council for Early Childhood Professional Recognition 1990, 1991a; Competence, 1991, 1992). *Dieses Instrument kann sowohl in institutionellen Gruppenbetreuungen als auch bei Tagespflegestellen eingesetzt werden.*

Zum besseren Verständnis des Anwendungskontextes wiederum eine kurze Skizze des CDA National Credentialing Programs, wobei die Beschreibung dadurch erschwert wird, daß derzeit der Ablauf des Programms umgestellt wird: Ein Child Development Associate (CDA) "... is a person who is able to meet the specific needs of children and who, with parents or other adults, work to nurture children's physical, social, emotional, and intellectual growth in a child development



framework. The CDA Credential is awarded to child care providers and home visitors who have demonstrated their skills in working with young children and their families by successfully completing the CDA assessment process." (Council of Early Childhood Professional Recognition, 1990, S.2) Das CDA Credential können beantragen (beim Antrag muß sich für einen Bereich entschieden werden):

- Betreuerinnen in Einrichtungen für die Arbeit mit Kindern unter drei Jahren oder für die Arbeit mit Kindern von drei bis fünf Jahren
- Betreuerinnen in Tagespflegestellen, wobei sie in dieser Tagespflegestelle mindestens zwei nicht verwandte Kinder im Alter von fünf Jahren oder jünger betreuen müssen
- Mitarbeiterinnen in pädagogischen Programmen, in denen sie regelmäßig Familien zu Hause besuchen, um dort mit fünfjährigen oder jüngeren Kindern zu arbeiten und die Eltern in Erziehungs- und Betreuungsfragen zu unterstützen.

Das CDA Credential ist ein Zeugnis für qualitativ gute Arbeit, das an die Betreuerin selbst vergeben wird - und nicht z.B. an die Einrichtung, in der sie arbeitet. Die Qualität der Arbeit wird mit Bezug auf die CDA Competency Standards bewertet. Für die verschiedenen Betreuungsformen haben die CDA Competency Standards die gleiche Struktur. Sie sind gegliedert in sechs globale Kompetenzziele (*Competency Goals*), d.h. in sechs allgemeine Aussagen über Ziele des Erzieherverhaltens. Diese sechs allgemeinen Ziele sind insgesamt in 13 Funktionsbereichen (*Functional Areas*) ausdifferenziert. Diese Funktionsbereiche beschreiben die Hauptaufgaben oder Hauptfunktionen, die eine Betreuerin erfüllen muß, um die Kompetenzziele zu erreichen. Kompetenzziele und Funktionsbereiche sind für alle Betreuungsformen gleich (in Klammer die Funktionsbereiche):

- I. to establish and maintain a safe, healthy, learning environment (1. Safe, 2. Health, 3. Learning Environment)
- II. to advance physical and intellectual competence (4. Physical, 5. Cognitive, 6. Communication, 7. Creative)
- III. to support social and emotional development and provide positive guidance (8. Self, 9. Social, 10. Guidance)
- IV. to establish positive and productive relationships with families (11. Families)
- V. to ensure a well-run, purposeful program responsive to participant needs (12. Program Management)
- VI. to maintain a commitment to professionalism (13. Professionalism)

Die CDA Competency Standards bilden die wohl umfassendste Beschreibung der Qualität einer frühkindlichen Betreuungs- und Erziehungsform (vgl. Modigliani, 1991). Viele der anderen Instrumente beziehen sich bei der Begründung der Auswahl ihrer Items auf diese Standards (siehe weiter unten).

Das CDA National Credentialing Program wird von dem Council for Early Childhood Professional Recognition in Washington, DC, verwaltet. In der derzeit noch gültigen Form läuft es in fünf Stufen ab; die dafür benötigte Zeit kann bis zu einem Jahr umfassen:

1. *Application*: Aufgrund eines Antrags an das Council erhält die Betreuerin alle benötigten

## Unterlagen und Instrumente.

2. *Team Formation/Information Collection:* Die Informationssammlung über die Qualität der Arbeit wird von einem lokalen Vierer-Team durchgeführt, dem angehören: die Kandidatin selbst (d.h. die beantragende Betreuerin), ein Berater (advisor), ein Elternvertreter und ein CDA Repräsentant. Den Berater und den Elternvertreter muß die Kandidatin selbst aussuchen, wobei bestimmte Voraussetzungen erfüllt sein müssen. Die Kandidatin stellt über ihre eigene Arbeit ein sogenanntes Portfolio zusammen. Der Elternvertreter verteilt Fragebögen an die Eltern der von der Kandidatin betreuten Kinder und beobachtet auch die Kandidatin bei ihrer Arbeit mittels einer halbstandardisierten Beobachtung. Die zentrale Rolle in diesem Prozeß der Informationssammlung nimmt aber der Berater ein. Mit der Advisor Report Form beobachtet er dreimal - in mindestens dreiwöchigen Abständen - die Kandidatin bei der Arbeit. Unmittelbar nach den Beobachtungen wird der Kandidatin vom Berater ein Feedback über ihre Leistungen gegeben, so daß diese in der Folge ihre Arbeit verbessern kann. Die Beobachtungen durch den Berater haben somit nicht rein evaluative Funktionen, sondern vor allem die Funktion, Verbesserungen anzuregen. Am Ende dieser Phase beobachtet dann auch der CDA Repräsentant die Kandidatin bei der Arbeit und befragt sie über ihre Kenntnisse und ihre Arbeit in den 13 Funktionsbereichen.
3. *Team Meeting:* Anschließend trifft sich das lokale Vierer-Team, sichtet alle vorliegenden Informationen und spricht eine Empfehlung über die Vergabe des Credentials aus. Alle vier Mitglieder des Teams - also auch die Kandidatin selbst - haben dabei gleiches Stimmrecht.
4. *Credential Award:* Nach diesem Treffen werden alle Unterlagen an das Council geschickt, das dann, wenn das Vierer-Team sich positiv entschieden hat, alle Unterlagen vollständig sind und die Prozedur formgerecht abgelaufen ist, der Kandidatin das Credential zuschickt.
5. *Renewal:* Das CDA Credential ist zunächst drei Jahre gültig, danach kann es für jeweils fünf Jahre umfassende Perioden erneuert werden. Bei einer Erneuerung wird eine verkürzte Prozedur verfolgt.

Zur Zeit wird das CDA National Credentialing Program revidiert, insbesondere wird die Aus- bzw. Fortbildungskomponente verstärkt. Dabei gibt es nun zwei Wege, um das Zeugnis für qualitativ gute Arbeit zu erhalten:

1. *CDA Professional Preparation Program - CDA P:* Dies ist eine völlig neue dreiphasige Prozedur, um das Credential zu erlangen (Competence, 1991). In einer ersten Phase (Field Work) steht die Arbeit in einem Praxisfeld im Vordergrund. Daneben muß die Kandidatin - angeleitet durch einen Berater (field advisor) - ein Textbuch durcharbeiten (acht Themenbereiche, die auf die CDA Competency Goals bezogen sind) (Phillips, 1991b). In der zweiten Phase (Coursework) nimmt die Kandidatin an einer Serie von Seminaren (z.B. Wochenendseminare, Sommerkurse) teil, in denen verschiedene Themen unterrichtet werden, die auch dem Textbuch zugrunde liegen. Am Ende dieser Phase steht ein schriftlicher Test zur Wissenskontrolle. In der dritten Phase (Final Evaluation) muß die Kandidatin in der Praxis - unter Anleitung durch den Berater - ein zusammenhängendes Set von praktischen Übungen durchführen. Ebenfalls werden Fragebögen an die Eltern der von der Kandidatin betreuten Kinder verteilt. Ein CDA Repräsentant überprüft zum Abschluß die Leistungen

der Kandidatin, u.a. in einem halbstandardisierten mündlichen Interview. Das CDA P<sub>3</sub> läuft seit September 1991.

Im Rahmen dieser Prozedur werden nur noch von dem Berater (field advisor) formale Beobachtungen der Leistungen der Betreuerin durchgeführt. Als Instrument wird die Field Advisor Formal Observation - auch CDA Assessment Observation Instrument genannt - eingesetzt. Am Ende der ersten Phase (Field Work) werden die Ergebnisse der Beobachtung direkt nach der Beobachtung mit der Kandidatin ausführlich diskutiert, um ihr Gelegenheit zur Verbesserung ihrer Arbeit zu geben. In der dritten Phase (Final Evaluation) führt der Berater nur eine zweite Beobachtung durch und schickt den ausgefüllten Bogen, ohne mit der Kandidatin darüber zu diskutieren, an das Council zurück.<sup>15</sup> Es kann hier aber als problematisch angesehen werden, daß nur noch der Berater - der zugleich zu größeren Teilen für das Training der Kandidatin verantwortlich ist - eine formale Beobachtung der Leistungen der Kandidatin durchführt.

2. *CDA Direct Assessment*: Dieser Weg zu dem CDA Credential entspricht weitgehend der früheren, oben beschriebenen Prozedur (Competence, 1992). Die entscheidenden Unterschiede bestehen darin, daß nun ein höherer formaler Schulabschluß (High School) vorausgesetzt wird (was vorher nicht der Fall war) und daß nun die Kandidatin in den letzten fünf Jahren 120 Stunden formale Fortbildung in acht Themenbereichen (die auf die CDA Competency Goals bezogen sind) nachweisen muß. Wie bei CDA P<sub>3</sub> geschieht damit auch eine engere Verknüpfung des an den CDA Competency Goals orientierten Evaluationsprozesses mit einem direkten Training in diesen Competency Goals. Weiterhin gibt es nicht mehr die lokalen Vierer-Teams. Ansonsten werden die gleichen Instrumente zur Überprüfung der Leistungen der Kandidatin eingesetzt wie beim CDA P<sub>3</sub> (siehe oben). Die Prozedur der Direct Assessment ist am 2. Juni 1992 in Kraft getreten.

Auch in der Direct Assessment werden damit die Leistungen der Kandidatin in der Praxis nur noch von einem Berater (Advisor) beobachtet. Nach den vorliegenden Unterlagen ist zudem unklar, ob der Berater nur *eine* Beobachtung durchführt, die direkt ohne Diskussion mit der Kandidatin an das Council geschickt wird, oder ob er wie in CDA P<sub>3</sub> zweimal die Kandidatin bei der Arbeit beobachtet. Bei nur einer Beobachtung ohne Feedback an die Kandidatin würde die bisher vorhandene pädagogische Funktion aufgegeben.

Zur Zeit liegen verschiedene Versionen des Beobachtungsinstrumentes vor. Im folgenden wird nur die derzeit benutzte Form - Advisor Report Form (Council for Early Childhood Professional Recognition, 1991b) - diskutiert. Das im CDA P<sub>3</sub> und in der Direct Assessment benutzte Instrument unterscheidet sich aber nur dadurch,

---

<sup>15</sup> Gegenüber der alten Prozedur des CDA Programms wird damit die pädagogische Funktion dieser formalen Beobachtung eingeschränkt, da die Kandidatin nur noch einmal und nicht mehr zweimal ein Feedback zu ihrer Arbeit erhält.

daß es nicht mehr für drei, sondern für weniger Beobachtungen konzipiert ist.<sup>16</sup> Die Advisor Report Form wird sowohl bei Betreuerinnen in Einrichtungen (Betreuerinnen in Gruppen mit Kinder unter drei Jahren und in Gruppen mit Kindern von drei bis fünf Jahren) als auch bei Betreuerinnen in Tagespflegestellen angewandt. Die Berater (advisor) werden für ihre Beobachtungsaufgaben nicht speziell trainiert, so daß das Instrument aus sich selbst heraus verständlich sein muß. In der alten CDA-Prozedur beobachtet auch ein CDA Repräsentant mit Hilfe dieses Instruments die Arbeit der Kandidatin. Hierzu gibt es u.a. einen Trainingsfilm. Weitere Informationen über das Training der CDA Repräsentanten liegen dem Berichtersteller jedoch nicht vor.

Die zu beobachtenden Items sind eng auf die CDA Competency Standards in den 13 Funktionbereichen bezogen bzw. können als Umsetzungen dieser Standards in Items eines Beobachtungsinstruments angesehen werden. Aus den Funktionsbereichen 12 (Program Management) und 13 (Professionalism) wurden keine Items aufgenommen, da diese beiden Bereiche nicht direkt während eines Besuches beobachtbar sind.

### *Aufbau des Instruments*

Die Advisor Report Form besteht aus insgesamt 43 relativ komplexen Items aus den ersten 11 Funktionsbereichen der CDA Competency Standards bzw. den ersten vier Competency Goals (vgl. weiter oben):

- I. to establish and maintain a safe, healthy, learning environment:
  1. Safe, 3 Items
  2. Health, 5 Items
  3. Learning Environment, 4 Items
  
- II. to advance physical and intellectual competence:
  4. Physical, 4 Items
  5. Cognitive, 3 Items
  6. Communication, 5 Items
  7. Creative, 5 Items
  8. Self, 3 Items
  
- III. to support social and emotional development and provide positive guidance:

---

<sup>16</sup> Die derzeit benutzte Version der Advisor Report Form ist so aufgebaut, daß der gleiche Bogen bei allen drei Beobachtungen des Beraters benutzt werden muß. Soweit aus den dem Berichtersteller zugänglichen Unterlagen ersichtlich ist, sind die neuen Versionen nur noch für eine Beobachtung konzipiert.

- 9. Social, 4 Items
- 10. Guidance, 3 Items

IV. to establish positive and productive relationships with families:

- 11. Families, 4 Items

Die 43 Items sind wie folgt zu bearbeiten:

- Die einzuschätzenden Items sind numeriert und werden als kurze Statements vorgegeben. Diese Items müssen auf einer dreistufigen Skala im Hinblick auf ihr Vorhandensein eingeschätzt werden:
    - 1 = rarely; der durch das Item angesprochene Aspekt konnte nicht und nur selten beobachtet werden
    - 2 = sometimes; der durch das Item angesprochene Aspekt konnte einige Male beobachtet werden
    - 3 = mostly; der durch das Item angesprochene Aspekt konnte meistens beobachtet werden.
- Nur die 43 numerierten Items werden mit dieser dreistufigen Skala eingeschätzt.
- Zu jedem Item gibt noch ein oder mehrere sogenannte "indicator statements". Diese Indikatoren beschreiben spezifische Aspekte der durch das Item angesprochenen Kompetenz. Bei den drei Besuchen soll angekreuzt werden, ob diese Aspekte vorhanden sind oder nicht (für jeden Besuch gibt es eine Ankreuzmöglichkeit). Zusätzlich stehen bei den meisten Indikatoren noch Beispiele.
  - Erst nach Vollendung der dritten Beobachtung müssen die 43 numerierten Items anhand der dreistufigen Skala eingeschätzt werden. Wenn alle zu einem Item gehörenden Indikatoren angekreuzt wurden, ist eine "3" wahrscheinlich, sie muß aber nicht zwingend vergeben werden. Für Ratings von "1" oder "2" müssen Begründungen angegeben werden. Konnte ein Item nicht aufgrund von Beobachtungen eingeschätzt werden, sondern wurden dazu Befragungen erforderlich, muß ein "I" in die Kommentarzeile geschrieben werden.

Die Items brauchen nicht unbedingt in der Reihenfolge bearbeitet werden, in der sie in dem Instrument stehen. In jedem Funktionsbereich ist noch Platz für weitere -

vom Berater einzutragende - Items vorgesehen, die die Leistungen der Kandidatin zusätzlich beschreiben können. Nach jeder Beobachtung soll der Kandidatin ein Feedback über ihre Leistungen gegeben werden. In jedem Funktionsbereich ist zudem Raum für die Eintragung von Verbesserungsvorschlägen (können beim ersten, zweiten, dritten Besuch oder immer gemacht werden). In jedem Funktionsbereich muß mindestens eine Empfehlung eingetragen werden.<sup>17</sup>

Wie bereits erwähnt, sind aus den Funktionsbereichen 12 (Program Management) und 13 (Professionalism) keine Items in dem Instrument enthalten, da diese beiden Bereiche nicht direkt während der Besuche beobachtbar sind. Es gibt aber jeweils eine Checkliste, die aufgrund eines Interviews mit der Kandidatin ausgefüllt werden muß. In jedem der beiden Funktionsbereiche soll der Berater drei Kommentare zur Arbeit der Kandidatin in diesem Bereich machen sowie eine Empfehlung zur Verbesserung der Arbeit geben.

Zwei Beispiele für Items (das erste aus dem Funktionsbereich 2. Health, das zweite aus dem Funktionsbereich 7. Creative); anhand der dreistufigen Skala werden nur die nummerierten Items eingeschätzt, nicht die jeweils darunter stehenden "indicator statements":

- 2.1 General hygiene practices are implemented consistently to cut down spread of infectious disease.

--- Toys and equipment are kept clean.  
--- Tissues are used once and discarded in closed plastic-lined container.  
--- Correct handwashing procedure used by adult before food, after each diapering/toileting, and whenever needed.

Ex. liquid soap, running water, and individual paper towels used.

- 7.4 Dramatic play experiences with a variety of age-appropriate props are available.

--- Dramatic play available daily.  
--- Creativity in dramatic play is encouraged.  
Ex. children make up own roles, children use props in their own way.  
--- Variety of props allows children to copy familiar behavior.  
Ex. dolls; blankets; mirrors; hats; pocketbooks; small people/animal toys; kitchen toys; dress-ups with large buttons; snaps, zippers  
--- New props and ideas made available regularly to add variety to play.

---

<sup>17</sup> Das Instrument für CDA P<sub>3</sub> sieht vor, daß jeweils mindestens zwei Empfehlungen gegeben werden müssen.

Vergleichbar mit den Items der Early Childhood Classroom Observation (vgl. Kap. 2.1.1) werden an den Beobachter mit diesen 43 komplexen Items relativ hohe Anforderungen gestellt. Er muß eine Vielzahl von Einzelinformationen berücksichtigen - insgesamt gibt es bei den 43 Items zu den ersten 11 Funktionsbereichen 130 "indicator statements" - und bei seinen Einschätzungen der 43 Items auf der dreistufigen Skala zu einem zusammenfassenden Urteil verarbeiten. Die Items sind damit auf einem höheren Abstraktionsniveau angesiedelt.

Für jede Beobachtung benötigt der Berater 2 1/2 bis 3 Stunden. Zusätzlich muß er noch Zeit für die Gespräche mit der Kandidatin über die Beobachtungen einplanen. Der Beobachter soll sich während der Beobachtung weitgehend zurückhalten und z.B. nicht mit den Kinder sprechen, um das Geschehen nicht zu beeinflussen. Zusammenfassungen der Ergebnisse der einzelnen Items zu Skalen sind in diesem Prozeß nicht vorgesehen.

### **Gütekriterien**

Untersuchungen zu Gütekriterien wurden bisher noch nicht durchgeführt (vgl. Clifford et al., in Druck, persönliche Mitteilung von C.B. Phillips an die Autoren).<sup>18</sup> Inhaltliche Validität wird dem Instrument aufgrund seiner engen Verbindung zu den CDA Competency Standards zugesprochen (vgl. Harms & Clifford, in Druck).

### **2.1.4 Assessment Profile for Early Childhood Programs**

#### **Entstehung und Anwendungskontext**

Das Assessment Profile (Abbott-Shim & Sibley, 1987a, 1987b) richtet sich nicht auf die Qualität einzelner Gruppen in Einrichtungen, sondern stellt ein umfassendes System dar, um die Qualität einer ganzen Einrichtung zu bewerten. Das Profile repräsentiert einen Satz von Standards für die Arbeit in frühkindlichen Einrichtungen, der nach Ansicht der Autoren umfassend ist und alle Dimensionen einer solchen Einrichtung anspricht. Ursprünglich wurde das Instrument von M. Abbott-Shim und A. Sibley zur Evaluation der Effektivität eines Fortbildungsprogramms entwickelt. Während des Entwicklungsprozesses wurde das Profile von verschiedenen Experten im frühkindlichen Bereich kritisch durchgesehen. Vor der Veröffentlichung wurde in einem fünfjährigen Zeitraum ein Feldtest in 90 Einrichtungen in Atlanta durchgeführt. Eine weitergehende Begründung der Auswahl der Items liegt dem Bericht-

---

<sup>18</sup> Modigliani (1991, S. 21) spricht mit Bezug auf das gesamte CDA Programm davon, daß es vor seinem Beginn umfangreich einem Feldtest unterzogen worden wäre, um eine befriedigende Reliabilität und Validität zu erreichen. Leider fehlen dazu aber sämtliche Angaben.

erstatte nicht vor. Die Autoren weisen noch darauf hin, daß es - im Sinne einer inhaltlichen Validität - eine hohe Übereinstimmung zwischen dem Itemsatz des Profils einerseits und andererseits den Akkreditierungskriterien der NAEYC, den CDA Competency Standards sowie dem Itemsatz der ECERS (Early Childhood Environment Rating Scale; Harms & Clifford, 1980, siehe 2.1.5) gibt (vgl. Quality Assist, ohne Jahr).

Das Assessment Profile stellt ein umfassendes Instrument für den Einsatz in Selbstevaluationsstudien von Einrichtungen dar. Selbstevaluation wird dabei als der Prozeß betrachtet, der das Bewußtsein der Praktiker über das vergrößert, was täglich in ihrer Einrichtung passiert. Eine erfolgreiche Selbstevaluation, die auch zu Verbesserungen in der Praxis führt, setzt voraus, daß alle Beteiligten bereit sind, sich selbst zu beobachten, eigene Handlungen in Frage zu stellen und Veränderungen einzuleiten. Eine solche Selbstevaluation muß objektiv und gründlich durchgeführt werden. In diesem Prozeß der Selbstevaluation ist das Profile ein strukturierter Führer der Beobachtungen (structured observation guide).

Das Assessment Profile ist in vier sogenannte Komponenten aufgeteilt. Eine Komponente erfaßt - unter der Überschrift "Administration" - alle Aspekte einer Einrichtung, die nicht gruppenspezifisch sind. Die anderen drei Komponenten beziehen sich auf die Qualität der Gruppen, wobei hier drei Altersgruppen unterschieden werden.<sup>19</sup>

- Infant - Gruppen, in denen Kinder im Alter von bis zu zwei Jahren betreut werden
- Preschool - Gruppen, in denen Kinder im Alter von zwei bis fünf Jahren betreut werden
- School-age - Gruppen, in denen Schulkinder im Alter von fünf bis zehn Jahren vor und/oder nach dem Schulbesuch betreut werden.

Auch wenn das Assessment Profile im Rahmen einer Selbstevaluation eingesetzt wird, so bedeutet dies nicht, daß eine Gruppenleiterin ihre eigene Gruppe einschätzt und dadurch - aufgrund von potentiellen Befangenheiten - u.U. pädagogische Chancen verloren gehen. Für die Anwendung des Profils in den einzelnen Gruppen sollen zwei Beobachter ausgewählt werden. Dies können z.B. zwei Eltern oder

---

<sup>19</sup> Bei den einzelnen Items sprechen zwar verschiedene Items auf der Einrichtungsebene ähnliche Aspekte an wie Items auf der Gruppenebene, insgesamt aber sind die Items jeweils spezifisch entweder auf die Einrichtung insgesamt oder auf gruppenspezifische Merkmale zugeschnitten.



andere Personen außerhalb der Einrichtung (community resource people) sein. Die beiden Beobachter führen dann ein etwa dreitägiges (Selbst-) Training in der Anwendung des Profiles durch. Dieses Training besteht darin, daß sie z.B. nach einer sorgfältigen Durchsicht und Diskussion des Instruments unabhängig voneinander Gruppen in anderen Einrichtungen beobachten und so lange ihre Einschätzungen vergleichen, bis sie mindestens eine Übereinstimmung von 85 % erreicht haben. Erst danach sollen sie mit den Beobachtungen in den Gruppen der zu evaluierenden Einrichtung beginnen. Ebenfalls zwei Personen - dies können die gleichen wie bei den Gruppen sein, aber auch andere - bearbeiten den Einrichtungsteil. Dabei füllen beide zunächst unabhängig voneinander das Instrument aus, diskutieren dann ihre Einschätzungen und berichten zusätzlich zu ihren Einzelantworten eine "gemeinsame, abgestimmte" Antwort.

Das Assessment Profile wurde auch in verschiedenen Forschungsstudien eingesetzt (vgl. Quality Assist, ohne Jahr). Hierzu wurden die Beobachter sowohl durch eine theoretische Einführung in das Instrument als auch durch zwei halbtägige Probebeobachtungen trainiert. Das Training dauert insgesamt ca. 2 Tage. Nach diesem Training werden Beobachterübereinstimmungen von 85 % bis 100 % berichtet.

### *Aufbau des Instruments*

Die Materialien zum Assessment Profile sind in zwei kleinen Heften zusammengefaßt. Ein Heft enthält neben dem Manual den Teil zur Bewertung der Einrichtung insgesamt, das andere Heft enthält die Bögen für die Bewertung der Gruppen (getrennt nach den drei Altersstufen). Innerhalb von jeder Komponente sind die Instrumententeile strukturgleich aufgebaut:

- Innerhalb jeder *Komponente* werden verschiedene *Dimensionen* unterschieden, z.B. für Preschool Gruppen die sechs Dimensionen: safety & health, learning environment, scheduling, curriculum, interacting, individualizing.
- Jede Dimension enthält verschiedene *Standards*, d.h. allgemeine Qualitätsbeschreibungen. In allen Dimensionen und Komponenten zusammen gib es 106 Standards. Die Anzahl der Standards variiert von Dimension zu Dimension. (Anzahl Standards, Einrichtung insgesamt: 24; Preschool Gruppe: 30; Infant Gruppe: 27; School-age Gruppe: 25.)
- Jeder Standard wird durch spezifische *Kriterien* operationalisiert. Die Anzahl der Kriterien variiert von Standard zu Standard. *Diese Kriterien sind die eigentlich zu beobachtenden und einzuschätzenden Items.*

Die vier Instrumententeile enthalten zusammen 524 Kriterien,<sup>20</sup> die sich auf die Komponenten und Dimensionen wie folgt verteilen:

- Einrichtung:
  - + Physical Facilities, 31 Items
  - + Food Service, 18 Items
  - + Program Management, 28 Items
  - + Personnel, 42 Items
  - + Program Development, 26 ItemsInsgesamt also 5 Dimensionen mit 145 Items.
  
- Preschool Gruppe:
  - + Safety & Health, 24 Items
  - + Learning Environment, 18 Items
  - + Scheduling, 23 Items
  - + Curriculum, 28 Items
  - + Interacting, 32 Items
  - + Individualizing, 22 ItemsInsgesamt also 6 Dimensionen mit 147 Items.
  
- Infant Gruppe:
  - + Safety & Health, 41 Items
  - + Nutrition, 11 Items
  - + Learning Environment, 19 Items
  - + Interacting, 31 Items
  - + Individualizing, 27 ItemsInsgesamt also 5 Dimensionen mit 129 Items.
  
- School-age Gruppe:
  - + Safety & Health, 16 Items
  - + Learning Environment, 15 Items
  - + Curriculum, 26 Items
  - + Interacting, 29 Items
  - + Scheduling, 17 ItemsInsgesamt also 5 Dimensionen mit 103 Items.

Die Items/Kriterien sind als einfache kurze Statements formuliert, die mit "ja" oder

---

<sup>20</sup> Eigentlich sind es nur 523 Kriterien, da die Dimension Interacting bei Infant Gruppen nicht 31, sondern nur 30 Items enthält. Da aber auf den Auswertungsblättern diese Dimension aus Versehen mit 31 Items vermerkt wurde, wird immer diese Itemzahl gezählt.

"nein" beantwortet werden *müssen*. Eine andere Antwortmöglichkeit hat der Beobachter nicht, d.h., er muß sich entscheiden: "Ja" bedeutet, daß ein Kriterium positiv erfüllt ist; "nein" bedeutet, daß der Beobachter den entsprechende Aspekt nicht oder nicht konsistent beobachtet hat oder daß der Aspekt bei einer Befragung als "kommt nicht vor" charakterisiert wurde. Zu Beantwortung der Items können drei Informationsquellen herangezogen werden: Beobachtung (*observation*), Durchsicht von Dokumenten (*document review*) und Befragungen von Betreuerinnen bzw. der Einrichtungsleiterin (*report*). In der Regel soll auf Beobachtungen zurückgegriffen werden. Neben jedem Item ist angegeben, aufgrund welcher Informationsquelle es beantwortet werden soll (*o*, *d* und *r*). Alle mit *o* gekennzeichneten Items sollen während der Beobachtung selbst beantwortet werden.

Im folgenden sind zwei Beispiele für Items aus dem Instrument für Preschool-Gruppen wiedergegeben. Zu zwei Standards (siehe weiter oben) werden dabei alle einzuschätzenden Kriterien/ Items aufgelistet; das erste Beispiel stammt aus der Dimension Learning Environment, das zweite aus der Dimension Interacting:

- Classroom reflects child as an individual.
  - o, r* 1. A quiet activity area exists in the room where one or two children may choose to be alone.
  - o* 2. Individual cubbies or containers are present.
  - o* 3. Cubbies or containers are labelled with a picture and/or name.
  - o* 4. Children's work is displayed at the child's eye level.
  
- Teacher is responsive to the children.
  - o* 1. Child is allowed to speak to the teacher without interruption.
  - o* 2. Teacher talks with the child at the child's eye level.
  - o* 3. Child's statements are acknowledged with a verbal response or a physical gesture.
  - o* 4. Child's emotions are acknowledged with a verbal response or a physical gesture.
  - o* 5. Teacher verbally acknowledges each child's arrival and departure.

Im Vergleich zu der Early Childhood Classroom Observation (vgl. 2.1.1) und dem CDA Assessment Observation Instrument (vgl. 2.1.3) erscheint die Itemzahl in dem Assessment Profile relativ hoch. Hierbei ist allerdings zu berücksichtigen, daß im Assessment Profile die Items, d.h. die einzuschätzenden Kriterien, viel stärker den Charakter von diskreten Einzelaspekten haben. Mithin werden auch beim Beobachter weniger Informationsverarbeitungsprozesse benötigt als bei den beiden anderen Instrumenten. Entsprechend der oben eingeführten Unterscheidung liegen die Items hier auf einem mittleren Abstraktionsniveau.

Die in der Early Childhood Classroom Observation und im CDA Assessment Instrument einzuschätzenden Items entsprechen in einigen Hinsichten im Assessment Profile mehr den Standards.<sup>21</sup> Diese drei Instrumente würden sich somit hier durch das Abstraktionsniveau unterscheiden, auf dem die eigentlichen Ratings stattfinden:

- In der Early Childhood Classroom Observation und im CDA Assessment Instrument werden Qualitätsaspekte auf einem eher höheren Abstraktionsniveau eingeschätzt. In dem Prozeß des Ratings selbst werden Einzelinformationen durch den Beobachter auf diesem Niveau zusammengefaßt.
- Im Assessment Profile hingegen sind die Einschätzungen eher an Einzelaspekten orientiert und verlangen nur Abstraktionen auf einem mittleren Niveau. Eine Zusammenfassung der gewonnenen Informationen auf einem höheren Niveau - der Ebene der Standards - muß dann analytisch erfolgen.

Für die gesamte Selbstevaluationsstudie wird ein Zeitaufwand von etwa einer Woche angegeben. Die Beobachtungsbögen für die Gruppen sollen während der Beobachtungszeit in den Gruppen ausgefüllt werden. In einem Faltblatt (Quality Assist, ohne Jahr) geben die Autoren an, daß während eines Vormittags ein bis vier Gruppen beobachtet werden können, wobei die Beobachter in 15- bis 20-minütigen Intervallen rotierend die Gruppen wechseln.

In einer solchen Selbstevaluation gehen die ausgefüllten Instrumente nicht an eine außenstehende Instanz, die dann die Ergebnisse aufbereitet und interpretiert. Vielmehr müssen die Ergebnisse von der Einrichtung selbst aufbereitet und interpretiert werden. Dazu werden im Assessment Profile verschiedene Hilfsmittel mitgeliefert. Allgemein lassen sich die Ergebnisse auf drei Ebenen betrachten:

- Die spezifischste Ebene ist die der einzelnen Items/Kriterien. Allerdings dürften sich hier auch Schwierigkeiten ergeben, die gegebene Informationsfülle angemessen zu verarbeiten.
- Eine zweite Ebene bezieht sich auf die einzelnen Dimensionen und die in ihnen enthaltenen Standards. Auf einem Auswertungsblatt (Dimension Score sheet) werden dazu - getrennt für die Komponenten - für jeden Standard in jeder Dimension sowie für jede Dimension insgesamt die Anzahl der mit "ja" beantworteten Kriterien eingetragen. Daneben stehen dann jeweils die maximal erreichbaren Punktzahlen pro Standard und Dimension, so daß durch einen Vergleich zwischen maximaler und erreichter Punktzahl die Leistungen in

---

<sup>21</sup> Im Instrumententeil für Preschool Gruppen gibt es im Profile (bei 147 Items) 30 Standards. Diese Anzahl entspricht schon eher der Anzahl der in den beiden anderen Instrumenten einzuschätzenden 68 bzw. 43 Aspekten.

jedem Standard und in jeder Dimension beurteilt werden können. Ab wann eine Differenz zwischen erreichter Punktzahl und maximaler Punktzahl auf notwendige Verbesserungsmaßnahmen hinweist, bleibt allerdings der Interpretation der Beteiligten überlassen.

- Die allgemeinste Ebene der Auswertung bieten die sogenannten Assessment Profiles. Für jede beobachtete Gruppe sowie für die Einrichtung insgesamt gibt es ein Auswertungsblatt, in dem die erreichten Punktzahlen in jeder Dimension - transformiert auf eine Prozentskala - eingetragen werden. Fallen die Eintragungen in den Bereich von 90 - 100 %, so wird von einer exzellenten Leistung (exception program performance) gesprochen. Der Bereich von 70 bis 89 % indiziert ein durchschnittliches Qualitätsniveau, und Werte unter 69 % indizieren Bereiche, die der Verbesserung bedürfen. Zusätzlich gibt es noch drei Summary Assessment Profiles, in die jeweils die Durchschnitte für Preschool, Infant und School-age Gruppen eingetragen werden können.

### *Gütekriterien*

Zur Bewertung der Gütekriterien liegt dem Berichterstatter nur ein Faltblatt vor (Quality Assist, ohne Jahr). Wie bereits erwähnt, berichten die Autoren, daß nach dem zweitägigen Training bei einem Einsatz in Forschungsstudien Beobachterübereinstimmungen von 85 % bis 100 % erreicht werden können. Sie verweisen auf verschiedene neuere Studien, in denen die Beobachterübereinstimmungen bei 90 % und höher liegen. Veröffentlichte Ergebnisse aus diesen Studien liegen dem Berichterstatter noch nicht vor. Die Autoren berichten von internen Konsistenzen (Cronbach's Alpha) von über .90 (.90 bei 39 Infant Gruppen, .97 bei 28 Toddler Gruppen, .96 bei 46 Preschool Gruppen und .90 bei 40 Einrichtungsbögen); hierbei muß aber berücksichtigt werden, daß die einzelnen Teile jeweils aus deutlich über 100 Items bestehen. Die Halbierungsreliabilitäten variieren von .73 (Infant Gruppen) bis .94 (Preschool Gruppen).

In 135 Gruppen wurde von trainierten Beobachtern sowohl das Assessment Profile als auch die ECERS (Early Childhood Environment Rating Scale; Harms & Clifford, 1980, vgl. 2.1.5) bzw. die ITERS (Infant/Toddler Environment Rating Scale; vgl. Harms, Cryer & Clifford, 1990, vgl. 2.1.6) angewandt. Die relativ hohen Korrelationen (.49 bei 39 Infant Gruppen, .66 bei 28 Toddler Gruppen und .64 bei 46 Preschool Gruppen) verweisen auf eine gute (kriteriumsbezogene) Validität. Wie bereits erwähnt wurde die inhaltliche Validität durch einen Vergleich mit den NAEYC Akkreditierungskriterien, den CDA Competencies und den Items der ECERS überprüft, wobei sich nach Angaben der Autoren eine hohe Übereinstim-

mung ergab.

Zur Zeit befindet sich eine Forschungsversion des Assessment Profiles (verkürzte Itemzahl) in Druck (persönliche Mitteilung, M. Abbott-Shim, 9. April 1992).

### **2.1.5 Early Childhood Environment Rating Scale**

#### **Entstehung und Anwendungskontext**

Die ECERS (Early Childhood Environment Rating Scale) wurde von T. Harms und R.M. Clifford (1980) am Frank Porter Graham Child Development Center der University of North Carolina at Chapel Hill entwickelt und wird derzeit breit in Forschungsarbeiten in den USA benutzt. Es gibt Übersetzungen des Instruments in verschiedene Sprachen, u.a. ins Deutsche (Tietze & Roßbach, 1986; vgl. auch Hagen & Roßbach, 1987; Roßbach, 1990). Die ECERS ist zwar für alle institutionellen Gruppenbetreuungen im frühkindlichen Bereich geeignet, jedoch ist sie am angemessensten für Gruppen mit Kinder über 2 1/2 Jahren (vgl. Clifford et al., 1989). Für Kinder unter drei Jahren liegt eine spezielle Version vor - ITERS (Infant/Toddler Environment Rating Scale) (Harms et al., 1990) -, die bei Gruppenbetreuung für Kinder dieses Alters vorgezogen werden sollte (siehe 2.1.6).

Die ECERS wurde für den Einsatz in Forschung und Erzieherinnenfortbildung entwickelt. Sie kann sowohl von außenstehenden Beobachtern als auch im Rahmen von Selbstevaluationsprozessen von Erzieherinnen für ihre eigenen Gruppen angewandt werden. Die Anwendung der ECERS verlangt ein Training, wofür verschiedene Materialien entwickelt wurden.

Die ECERS will einen Überblick über die Qualität der Arbeit in Gruppen in Einrichtungen des frühkindlichen Bereichs geben. Umwelt wird dabei explizit breit definiert, indem sie neben materiellen Aspekten auch Merkmale der Interaktionen umfaßt. Die ECERS besteht aus 37 Items, die nach folgenden Leitfragen ausgewählt wurden (Harms & Clifford, 1986):

- Was benötigen Kinder in jeder Gruppenbetreuung, um entwicklungsangemessene Fortschritte im physischen, sozialen, kognitiven und emotionalen Bereich zu machen?
- Wie sollen die Erzieherinnen die Gruppenumwelt strukturieren, damit die Entwicklungsbedürfnisse der Kinder erfüllt werden können?
- Was benötigen die Erzieherinnen für sich selbst, um ihre Arbeit professionell ausüben zu können?

Der Auswahl und der Formulierung der Items der ECERS unterliegen dabei fünf grundlegende pädagogische Orientierungen:

1. Kinder sind aktive Lerner; sie lernen durch ihre Aktivitäten, durch das, was sie tun, hören, erfahren und denken.
2. Kinder lernen (auch) durch die Interaktionen mit ihren Eltern, Erzieherinnen und anderen Erwachsenen wie auch durch die Interaktionen mit anderen Kindern. Verbale und non-verbale Interaktionen mit Erwachsenen sind eine wichtige Quelle für kindliche Lernprozesse. Eine wichtige Aufgabe für die Erzieherinnen besteht deshalb darin, Interaktionen und einen produktiven Umgang mit Spielen sowie mit Routinetätigkeiten vorab zu planen.
3. Eine räumlich-materiale Umwelt, die so organisiert ist, daß Kinder maximal unabhängig und erfolgreich sein können, erlaubt den Kindern mehr Zeit für produktive Interaktionen, Diskussionen und Freude.
4. Kinder benötigen emotionale Wärme, räumliche Möglichkeiten, die ihrem Bedürfnis nach Sanftheit/Kuscheligkeit entgegenkommen, und vorhersagbare Routinen, um sich sicher und geschützt zu fühlen.
5. Eine "gute" Umwelt für Kinder sollte auch die Bedürfnisse der Erwachsenen, die in ihnen als Betreuer arbeiten, erfüllen.

### *Aufbau des Instruments*

Die ECERS besteht aus 37 Ratingskalen, die in sieben Bereiche oder Subskalen gegliedert sind:

- 1. Personal care routines, 5 Items
- 2. Furnishings and display for children, 5 Items
- 3. Language-reasoning experiences, 4 Items
- 4. Fine and gross motor activities, 6 Items
- 5. Creative activities, 7 Items
- 6. Social development, 6 Items
- 7. Adult needs, 4 Items

Jedes Item muß auf einer siebenstufigen Skala eingeschätzt werden, wobei - für jedes Item gesondert und speziell auf das jeweilige Item abgestimmt - die Skalenstufen 1 (inadequate), 3 (minimal), 4 (good) und 7 (excellent) ausführlich beschrieben werden:

- Die Beschreibung unter "1" bezieht sich auf eine Situation in der Gruppe, in der der durch das Item angesprochene Aspekt völlig unzureichend vorhanden oder gestaltet ist.
- Die Beschreibung unter "3" bezieht sich auf eine Situation, in der zu einem Aspekt ein Minimum von Material, Raum oder Supervision gegeben ist.
- Die Beschreibung unter "5" repräsentiert eine Situation, in der zu einem Aspekt adäquates Material in adäquaten räumlichen Möglichkeiten gegeben ist und in der die Kinder unter Supervision der Erzieherinnen mit dem Material interagieren.
- Die Beschreibung unter "7" bezieht sich auf eine Situation, in der alle Bedingungen aus den darunter liegenden Stufen erfüllt sind *und* die Kinder zu unabhängiger Auseinandersetzung mit dem durch das Item angesprochenen Aspekt aufgefordert werden sowie individuell auf einzelne Kinder abgestimmte Planungen der Erzieherinnen erkennbar sind.

Entsprechend den jeweiligen Beschreibungen der Skalenpunkte sollen die 37 Items eingeschätzt werden. Die Skalenpunkte 2, 4 und 6 sollen dann vergeben werden, wenn alle Bedingungen für die Vergabe einer niedrigeren Stufe, aber nicht alle der höheren Stufe erfüllt sind. Vereinfachend läßt sich festhalten: Niedrigere Skalenpunkte beziehen sich in der Regel auf Vorhandensein von Material oder Raum, für höhere Skalenpunkte wird immer zusätzlich ein auf den jeweiligen Aspekt bezogenes produktives Handeln der Erzieherinnen gefordert. Bei einigen Items sind noch erläuternde Beispiele angegeben.

Die 37 Skalen werden von einem trainierten Beobachter während eines etwa zwei-stündigen Besuchs einer Gruppe bearbeitet, wobei keine strikte Reihenfolge vorgegeben ist, in der die Skalen bearbeitet werden müssen. Einige Skalen können ausgefüllt werden, wenn noch keine Kinder da sind, für andere müssen Interaktionen der Kinder beobachtet werden, während eine letzte Gruppe von Skalen Befragungen der Erzieher verlangt. Es kann immer nur eine Gruppe gleichzeitig bearbeitet werden. Die Einschätzungen der 37 Qualitätsaspekte sollen immer auf der gegenwärtigen Situation, wie sie beobachtet oder von der Erzieherin berichtet wird, beruhen, nicht auf zukünftigen Plänen oder Veränderungswünschen.

Im folgenden zwei Beispiele für Items der ECERS aus der deutschen Erprobungsversion (Tietze & Robbach, 1986); das erste Beispiel stammt aus dem Bereich 3. Language-reasoning activities, das zweite aus dem Bereich 5. Creative activities:



## 11. Sprachverstehen

(Material: Bücher, Schallplatten, Bilderlotto und andere Bilderspiele, Pinnwand usw.)

1

Wenige Materialien vorhanden, geringer Gebrauch zur Förderung des Sprachverstehens (z.B. keine feste tägliche Erzählstunde)

3

Einige Materialien vorhanden, aber entweder nicht allgemein zugänglich (geschlossener Schrank) oder nicht regelmäßig für die Sprachentwicklung genutzt.

5

Viele Materialien vorhanden, die nach freier Wahl oder Anleitung angewendet werden können. Täglich mindestens eine geplante Aktivität (z.B. vorlesen, Geschichten erzählen, Bildgeschichten usw.)

7

Alle Bedingungen aus Einschätzung 5 müssen erfüllt sein, zusätzlich: Erzieherin ist den ganzen Tag über ein gutes Sprachvorbild (z.B. gibt klare Anweisungen, benutzt treffende Worte bei Beschreibungen). Erzieherin plant zusätzliche Aktivitäten für Problemkinder.

## 25. Rollenspiele

1

Keine spezielle Ausstattung für Verkleidungs- oder Rollenspiele.

3

Requisiten für Rollenspiele unterstützen hauptsächlich Vater-Mutter-Kind-Rollen bzw. Rollen in häuslichen Situationen. Keine oder nur gewisse Ausstattung für Rollenspiele, die Verkehr, Arbeit oder Abenteuer einschließen.

5

Vielfältige Requisiten für Rollenspiele einschließlich für die Themen Verkehr, Arbeit, Abenteuer und Phantasie. Platz im Gruppenraum und außerhalb des Gruppenraums (entweder im Freien, in einem Mehrzweckraum oder in einer Gymnastikhalle) ermöglicht ein lebhaftes Spiel.

7

Alle Bedingungen aus Einschätzung 5 müssen erfüllt sein, zusätzlich: Bilder, Geschichten und Ausflüge werden dazu benutzt, um die Rollenspiele zu bereichern.

Die Beispiele zeigen deutlich, daß die hier verwendeten Ratingskalen relativ komplexe Instrumente darstellen, die hohe Anforderungen an die Verarbeitungskapazitäten der Beobachter stellen. Es werden keine Einzelmerkmale eingeschätzt, vielmehr müssen für ein Rating viele Einzelaspekte beobachtet und dann zu einem einheitlichen Urteil zusammengefaßt werden. Die Items sind damit dem Abstraktionsniveau 3 (höheres Abstraktionsniveau) zuzuordnen.

Die Ergebnisse zu den 37 Ratingskalen werden sowohl insgesamt über alle Items als auch bezogen auf die sieben Subskalen additiv zusammengefaßt. Durch die Bildung von Durchschnittswerten für das gesamte Instrument bzw. für die einzelnen Subskalen kann die ursprüngliche siebenstufige Skala zur zusammenfassenden Beurteilung der Qualität herangezogen werden. Zur Auswertung auf Subskalenebene gibt es ein spezielles Auswertungsblatt (classroom profile), in dem die pro Subskala erreichten Punktzahlen graphisch veranschaulicht werden. Um zu berücksichtigen, daß es pro Subskala unterschiedlich viele Items gibt, sind die Punkteskalen dabei pro Subskala so gestaucht bzw. in die Länge gezerrt, daß alle Subskalen auf eine einheitlich Skala abgebildet werden.

### **Gütekriterien**

Die ECERS wurde und wird zur Zeit in vielen Studien in den USA und Kanada als Instrument für eine globale Qualitätseinschätzung eingesetzt. Im folgenden kann nicht auf Analysen zu den Gütekriterien in allen Studien eingegangen werden, vielmehr werden nur einige der Hauptergebnisse berichtet.

Harms & Clifford (1980, 1983) berichten von verschiedenen Studien, in denen jeweils zwei Beobachter unabhängig voneinander dieselben Gruppen mit der ECERS einschätzen. Dabei ergaben sich Rangkorrelationen der beiden Einschätzungen von .90 (n = 22 Gruppen), .79 (n = 18) und .88 (n = 25). Für die ersten sechs Subskalen berichtet McCartney (1984) für eine Studie in Bermuda eine Interrater-Reliabilität von .82 (n = 8). Ähnlich hohe Werte werden auch für die National Child Care Staffing Study berichtet (vgl. Whitebook et al., 1990; Howes, Phillips & Whitebook, in Druck).<sup>22</sup>

Die Test-Retest-Reliabilität beträgt .96 (Harms & Clifford, 1983). Dabei wurden 31 Gruppen im Abstand von zwei Wochen zweimal eingeschätzt. An einer Stichprobe von 25 Gruppen wurde die interne Konsistenz überprüft (Cronbach's Alpha). Sie beträgt für die Gesamtskala .83, für die sieben Subskalen .32 bis .79 (Harms & Clifford, 1980). Für eine deutsche Erprobungsfassung ergaben sich bei 43 Gruppen Alphakoeffizienten von .93 für die Gesamtskala und .43 bis .76 für die Subskalen (vgl. Hagen & Roßbach, 1987; Roßbach, 1990). McCartney (1984) weist daraufhin, daß die ersten sechs Subskalen in ihrer Studie (n = 8) sehr hoch miteinander korrelieren (.60 bis .92). Insgesamt kann der ECERS eine gute Reliabilität zugesprochen

---

<sup>22</sup> Allerdings werden hier nicht Beobachterübereinstimmungen speziell für die ECERS berichtet. Vielmehr wird auf alle in der Studie benutzten Instrumente verwiesen, wobei sich Beobachterübereinstimmungen über 85 % ergaben (Whitebook et al., 1990, S. 21; Howes et al., in Druck).

werden, obwohl die internen Konsistenzen bei einigen Subskalen niedrig ausfallen.<sup>23</sup>

Die inhaltliche Validität der ECERS wurde überprüft, indem sieben Experten des frühkindlichen Bereichs die Wichtigkeit der einzelnen Items sowie der Gesamtskala einschätzen sollten. Von den insgesamt 266 Ratings der Experten lauteten 78 % "very important" (höchster Skalenwert), nur 1 % der Ratings lautete "low importance". Im Sinne einer kriteriumsorientierten Validität schätzten Experten - denen die ECERS unbekannt war - die Qualität von 18 ihnen gut bekannte Gruppen auf einer siebenstufigen Skala von 1 "low" bis 7 "high" ein. Diese Einschätzungen wurden dann verglichen mit unabhängig davon gewonnenen ECERS-Werten. Es ergab sich eine Rangkorrelation von .74 (Harms & Clifford, 1983). In der Bermuda Studie (McCartney, 1984) zeigten sich weiterhin moderate Korrelationen der ECERS-Werte mit einem anderen Instrument zur Einschätzung der frühkindlichen Gruppenumwelt, dem Day Care Environment Inventory (Prescott et al., 1972).

Wie bereits erwähnt, wurde die ECERS in einer Vielzahl von Studien eingesetzt. Im Sinne einer Konstruktvalidität zeigen sich dabei inhaltlich sinnvoll zu interpretierende Beziehungen zu anderen Konstrukten. Hierzu ohne Anspruch auf Vollständigkeit einige Beispiele: In der bereits mehrfach angesprochen Studie in Bermuda (McCartney, 1984; Phillips, McCartney & Scarr, 1987) wurden u.a. in einem hierarchischen Regressionsansatz die ECERS-Einschätzungen von 8 Einrichtungen bzw. Gruppen in Beziehung gesetzt mit Maßen der Sprachentwicklung von Kindern (Peabody Picture Vocabulary Test, Preschool Language Assessment Instrument, Adaptive Language Inventory und eine standardisierte Kommunikationsaufgabe) sowie der sozialer Entwicklung (gemessen über das Classroom Behavior Inventory und den Preschool Behavior Questionnaire, wobei diese Instrumente sowohl von den Erzieherinnen als auch von den Eltern ausgefüllt wurden). Analyseebene sind jeweils die Kinder, wobei die Fallzahlen je nach Kriteriumsmaß von 46 bis 156 variieren. In dem hierarchischen Ansatz werden jeweils die ECERS-Werte als letzter Prädiktor aufgenommen, vorher werden berücksichtigt: Alter des Kindes, Aspekte des familialen Hintergrunds (gemessen über zwei Variablen zur Wertorientierung) sowie die Betreuungserfahrungen des Kindes (Alter des Eintritts in institutionelle Gruppenbetreuung, Anzahl der bisherigen Betreuungsstunden in der jeweiligen Einrichtung). Durch dieses Vorgehen werden somit die Effekte der Qualität der Gruppen gemessen über die ECERS konservativ eingeschätzt; die gemeinsame erklärende Varianz mit den anderen Prädiktoren wird diesen zugeschlagen. Durchgängig zeigen die ECERS-

---

<sup>23</sup> Bei der deutschen Erprobungsstudie (n = 43 Gruppen) ergaben sich einige negative Item-Total-Korrelationen. Zusätzlich hatten einige Items höhere Korrelationen mit anderen Skalen als mit der, der sie theoretisch zugeordnet sind (vgl. Roßbach, 1990).

Werte eigenständige Erklärungskraft, wobei ihre zusätzliche Erklärungskraft bei diesem Prädiktorensatz - gemessen über den  $R^2$ -Zuwachs - bei einigen Kriterien, speziell der sozialen Entwicklung, erheblich ist und bis zu .39 reicht. Zusätzlich ergeben sich zum Teil hohe Korrelationen zwischen dem ECERS Gesamtwert und verschiedenen Indikatoren der verbalen Interaktion in den Gruppen (vgl. McCartney, 1984, S. 255).

In der National Day Care Staffing Study wurde sowohl die ECERS als auch die ITERS (Infant/ Toddler Environment Rating Scale; Harms et al., 1990; vgl. 2.1.6) als Maß für eine globale Qualitätseinschätzung der untersuchten Gruppen benutzt (Whitebook et al., 1990). Die ECERS wurde in insgesamt 313 Gruppen eingesetzt. Bei den Analysen wurde jedoch nicht der ECERS-Gesamtwert benutzt, sondern zwei aufgrund von Faktorenanalysen gewonnene Skalen (zu den Faktorenanalysen vgl. weiter unten). Die beiden Faktoren stehen in vielfältigen und inhaltlich sinnvoll zu interpretierenden Beziehungen zu anderen Konstrukten z.B. der Lernumwelt in den Gruppen (z.B. Gruppengröße, zahlenmäßiges Erzieher-Kind-Verhältnis), zu Merkmalen der Erzieherinnen (z.B. Ausbildung, Bezahlung, Zufriedenheit mit der Tätigkeit) wie auch zu Merkmalen des kindlichen Entwicklungsstandes (z.B. gemessen über den Peabody Picture Vocabulary Test, das Adaptive Language Inventory sowie verschiedenen Maßen der Sozialentwicklung). In einer Folgeuntersuchung (Howes et al., in Druck) wurden diese beiden Faktoren auch in ein Pfadmodell integriert. Dabei zeigt sich, daß die über die ECERS gemessene Qualität sinkt bei steigender Gruppengröße und sich verschlechterndem zahlenmäßigen Erzieher-Kind-Verhältnis (Pfadkoeffizienten von -.25 und -.28). Bei höherer Qualität steigt die Sicherheit der Kinder im Umgang mit den Erzieherinnen sowie die soziale Orientierung der Kinder (Pfadkoeffizienten z.B. von .24 und .31), so daß sich insgesamt indirekte kausale Effekte auf die soziale Kompetenz der Kinder im Umgang mit Peers von .16 bis .20 ergeben.

Die ECERS wurde ebenfalls im Victoria Day Care Research Project in Kanada - zusammen mit einer parallelen Form für Tagespflege<sup>24</sup> - eingesetzt (vgl. Goelman & Pence, 1987, 1988). Bezogen auf kindliche Aktivitäten (gemessen über die Child Observation Form) zeigen sich aber nur Beziehungen zu "information activities", d.h. zu Aktivitäten, die nicht zwingend im Umkreis von Instruktionen auftreten, in denen aber spezifische Informationen von den Erzieherinnen an die Kinder gegeben werden (n = 25 Gruppen): Je höher die über die ECERS erfaßte Qualität, desto umfangreicher solche Aktivitäten (Korrelation mit dem ECERS-Gesamtwert .27, mit den Subskalen 3. Language-reasoning experiences .33, 5. Creative activities .35 und

---

<sup>24</sup> Es handelt sich hier um eine Vorform der FDCRS (Family Day Care Rating Scale) von Harms & Clifford (1989); vgl. Kapitel 2.2.2.

6. Social development .28). Es zeigen sich aber keine signifikante Beziehungen zum kindlichen Entwicklungsstand (gemessen z.B. über den Peabody Picture Vocabulary Test, den Expressive One-Word Picture Vocabulary Test). Die Autoren erklären dies damit, daß trotz der Variation in der Qualität gemessen über die ECERS die Gruppen relativ homogen waren, so daß die ECERS-Werte keine prädiktive Kraft haben (zu Effekten bei Tagespflege vgl. Kapitel 2.2.2.).

In einer kleinen Studie zur Erprobung einer deutschsprachigen Version der ECERS in der Bundesrepublik (n = 43 Gruppen aus 18 Einrichtungen) zeigten sich niedrige, aber plausible Beziehungen zwischen drei faktorenanalytisch gewonnenen Skalen sowie dem Gesamtwert (vgl. weiter unten) einerseits und externen Kriterien andererseits (Hagen & Roßbach, 1987; Roßbach, 1990): So steigt z.B. die Qualität, gemessen über die ECERS in Einrichtungen, in denen das Ausbildungsniveau der Erzieherinnen höher liegt (Korrelation mit dem ECERS-Gesamtwert von .29). Allerdings muß bei dieser Studie berücksichtigt werden, daß hier nur eine erste deutsche Erprobungsfassung zum Einsatz kam und damit weitere Analysen und gegebenenfalls Modifikationen der Erprobungsversion erforderlich werden.

Zur faktoriellen Struktur der ECERS liegen dem Berichtersteller - gemessen an dem breiten Einsatz der ECERS - nur relativ wenige Informationen vor. Verglichen aber mit den anderen Screeninginstrumenten zur Qualitätseinschätzung frühkindlicher Betreuung- und Erziehungsformen ist die faktorielle Struktur der ECERS noch am umfangreichsten untersucht worden. In der Erprobungsstudie der deutschsprachigen Version der ECERS wurde mit den Daten aus 43 Gruppen in 18 Einrichtungen eine Faktorenanalyse gerechnet (Hauptkomponentenanalyse mit anschließender VARIMAX - Rotation) (vgl. Hagen & Roßbach, 1987; Roßbach 1990). Berücksichtigt wurden 31 Items der ECERS.<sup>25</sup> Drei Faktoren wurden extrahiert, die zusammen 53 % der gesamten Varianz erklären. Die Faktoren wurden wie folgt benannt (in Klammern jeweils der durch den Faktor erklärte Varianzanteil): Räumlich-materiale Ausstattung (23,4 %), Atmosphäre/pädagogische Anleitung durch Erzieherinnen (15,0 %) und Engagement der Erzieherinnen bei Zusatzaktivitäten (14, 6%).<sup>26</sup> Viele Items haben aber eine komplexe Ladungsstruktur, d.h. höhere Ladungen nicht nur auf einem Faktor. Die der ECERS zugrunde liegende Differenzierung in sieben

---

<sup>25</sup> Eine Faktorenanalyse mit 31 Items und nur 43 Fällen muß mit Vorsicht betrachtet werden. Deshalb wurden verschiedene Teilanalysen mit zufällig ausgewählten Itemuntergruppen gerechnet. Die Ergebnisse wiesen jedoch weitgehend eine vergleichbare Struktur auf wie die Faktorenanalyse mit allen 31 Items.

<sup>26</sup> Der dritte Faktor bezieht sich auf das Engagement der Erzieherinnen bei solchen Aktivitäten, die in den deutschen Kindergärten eher eine untergeordnete Rolle spielen wie z.B. Körperpflege oder Betreuung bei Mahlzeiten.

Subskalen konnte nicht reproduziert werden.

In Hong Kong wurde eine Faktorenanalyse mit Daten aus 53 Gruppen und einem leicht modifizierten Itemsatz (Ausschluß von vier Items, drei neue Items) gerechnet (Tse, persönliche Mitteilung; vgl. Roßbach, 1990). Fünf Faktoren wurden identifiziert; die 67,1 % der Varianz aufklären (die Aufteilung der Varianz auf die einzelnen Faktoren ist dem Berichtersteller nicht bekannt): 1. Provision for the learning activities of children, 2. Engagement and activities of children, 3. Space, material, equipment, 4. Supervision and rapport, 5. Provisions for needs of adults and children. Diese Faktorenstruktur ist zwar einerseits der Situation in den Einrichtungen in Hong Kong angemessen, reproduziert andererseits aber ebenfalls nur teilweise die sieben ECERS-Subskalen: Der Faktor 3 "Space, material, equipment" ist vergleichbar mit der ECERS-Subskala 4 "Fine and gross motor activities", der Faktor 5 "Provisions for needs of adults and children" entspricht der ECERS-Subskala 7 "Adult needs".

Zur weiteren Aufdeckung der Faktorenstruktur wurden parallele Faktorenanalysen (Modell gemeinsamer Faktoren, quadrierte multiple Korrelationen als Anfangsschätzungen für die Kommunalitäten, VARIMAX-Rotation) berechnet (Roßbach et al., 1991). Für die Bundesrepublik wurden Daten aus 68 Gruppen und für die USA Daten aus 185 Gruppen benutzt. Die Daten wurden jeweils aus verschiedenen Studien genommen. Aufgrund von fehlenden Werten in dem USA-Datensatz wurden nur die Items der ersten sechs Subskalen herangezogen; zusätzlich wurden in beiden Datensätzen ein Item (Item 33. Provisions for exceptional children) sowie nur in der Bundesrepublik zwei weitere Items (3. Nap/rest, 31. Cultural awareness) ausgeschlossen. Für den USA-Datensatz wurden vier Faktoren extrahiert, die 82,9 % der gemeinsamen und 40,8 % der gesamten Varianz aufklären: 1. Teachers' stimulation of child development (16,3 % der gesamten Varianz), 2. Availability and arrangement of materials and space as prerequisites for learning (11,8 %), 3. Fulfillment of routine care needs of children (8,2 %), 4. Provisions for motor development (4,6 %). Mehrere Items haben höhere Ladungen auf verschiedenen Faktoren. Bildet man entsprechend dieser vier Faktoren vier additive Skalen aus den Markiervariablen, so korrelieren diese Skalen moderat miteinander (.43 bis .68), wobei allerdings die ersten beiden Skalen mit einer Korrelation von .82 gut zwei Drittel der Varianz gemeinsam haben. Obwohl der erste Faktor mehr auf die Erzieherinnen und der zweite mehr auf räumlich-materiale Aspekte gerichtet sind, können beide als zwei Seiten eines höheren Faktors betrachtet werden: der Fähigkeiten von Erzieherinnen sowohl ihren direkten Umgang mit dem Kind als auch räumlich-materiale Aspekte so zugestalten, daß die kindliche Entwicklung stimuliert wird. Auch wenn diese vier Faktoren nicht die theoretische Struktur der ECERS reproduzieren, so liegen mit ihnen doch interpretierbare Dimensionen frühkindlicher Betreuungs- und Erzie-

lungsumwelten vor.<sup>27</sup>

Für den deutschen Datensatz wurden drei Faktoren extrahiert, die zusammen 65 % der gesamten gemeinsamen und 40,3 % der gesamten Varianz aufklären: 1. Allgemeine räumlich-materiale Bedingungen für Lernen, Grobmotorik und Pflege (17,4 % der gesamten Varianz), 2. Atmosphäre/pädagogische Anleitung durch Erzieherinnen (15,4 %), 3. Raum und Mobiliar für Entspannung und Rückzug (7,5 %).<sup>28</sup> Entsprechend dieser Faktoren wurden additive Skalen aus den - zum Teil komplexen - Markiertvariablen gebildet. Die Skalen zu den ersten beiden Faktoren korrelieren mit .74 relativ hoch miteinander, die Korrelationen mit der dritten Skala sind jedoch niedrig (.36 und .19). Wiederum können damit die ersten beiden Faktoren als beeinflusst von einem höher geordneten Faktor angesehen werden. Die jeweils ersten beiden Faktoren in den USA und in der Bundesrepublik sind vergleichbar. Der US-Faktor 2 "Availability and arrangement of materials and space as prerequisites for learning" und der deutsche Faktor 1 "Allgemeine räumlich-materiale Bedingungen für Lernen, Grobmotorik und Pflege" beziehen sich beide auf die räumlich-materiale Ausstattung, wobei allerdings der deutsche Faktor mehr die gesamten räumlich-materialen Rahmenbedingungen indiziert, während sich der amerikanische Faktor stärker auf Bedingungen für Lernprozesse richtet. Der US-Faktor 1 "Teachers' stimulation of child development" und der deutsche Faktor 2 "Atmosphäre/pädagogische Anleitung durch Erzieherinnen" sind insofern ähnlich, als beide Faktoren sich direkter auf die Erzieherinnen richten. Ein leichter Unterschied besteht jedoch insofern, als der amerikanische Faktor sich wiederum mehr auf die Stimulierung der kindlichen Entwicklung bezieht, während der deutsche Faktor - zumindest teilweise - mehr auf die allgemeine Atmosphäre und stärker auf traditionelle Inhalte der Kindergartenpädagogik gerichtet ist. Es wird davon ausgegangen, daß diese Unterschiede nicht das Ergebnis von methodologischen Problemen darstellen, sondern daß

---

<sup>27</sup> Die Subskalen der ECERS differenzieren verschiedene inhaltliche Bereiche einer frühkindlichen institutionellen Erziehungs- und Betreuungsumwelt. Die Subskala 5 "Creative activities" enthält alle Items, die sich auf die Stimulation der Kreativität des Kindes beziehen - sei es durch direkte Erzieherinnen-Kind-Interaktionen, sei es durch die Bereitstellung entsprechenden Materials. Die extrahierten Faktoren 3 und 4 folgen dem gleichen Organisationsprinzip, indem sie sich auf Pflegeaspekte sowie auf die motorische Entwicklung beziehen. Anders aber die ersten beiden Faktoren: Hier wird - gewissermaßen quer zu den inhaltlichen, curricularen Bereichen - unterschieden zwischen einer aktiven, direkten Stimulierung durch die Erzieherinnen und einer Stimulierung durch ein Arrangement von räumlich-materialen Bedingungen. Dementsprechend verteilen sich dann auch z.B. die Items der ECERS-Subskala 5 "Creative activities" neu auf die beiden Faktoren.

<sup>28</sup> Die weitgehende Übereinstimmung mit der weiter oben berichteten Faktorenanalyse von 43 deutschen Gruppen beruht auch darauf, daß diese 43 Gruppen knapp zwei Drittel der 68 hier analysierten Gruppen ausmachen.

sie unterschiedliche Akzentuierungen institutioneller Gruppenbetreuungen in den beiden Ländern sensibel widerspiegeln.

Wie bereits weiter oben erwähnt, wurde in der National Child Care Staffing Study die ECERS in 313 Gruppen eingesetzt. Diese Daten wurden einer Faktorenanalyse unterzogen (Maximum Likelihood Methode, schiefwinklige Rotation), bei der zwei Faktoren extrahiert wurden (vgl. Whitebook et al., 1990; Howes et al., in Druck): Der erste Faktor - bezeichnet als "Appropriate caregiving" - wird markiert durch Items zur Supervision, zur Atmosphäre und Sprachförderung. Er erklärt 52 % der gemeinsamen Varianz. Der zweite Faktor - "Developmentally appropriate activity" - enthält Items zur Materialausstattung, (Zeit-) Planung und den Aktivitäten in der Gruppe; er erklärt 48 % der gemeinsamen Varianz. Die Faktoren korrelieren mit .78 miteinander.<sup>29</sup> Für die Zwecke der National Child Care Staffing Study wurden additive Skalen aus den Markiertvariablen mit Ladungen von .50 und größer gebildet. Weitere technische Informationen zu der Faktorenanalyse (z.B. zu den Varianzanteilen der beiden Faktoren an der gesamten Varianz) sind in dem Abschlußbericht nicht enthalten bzw. liegen dem Berichtersteller nicht vor. Obwohl aufgrund des unterschiedlichen methodischen Zugriffs ein Vergleich dieser Faktorenanalyse mit den weiter oben beschriebenen (Roßbach et al., 1991) nur bedingt möglich ist, zeigen sich Ähnlichkeiten: Der Faktor "Appropriate caregiving" ist teilweise vergleichbar mit dem US-Faktor 1 "Teachers' stimulation of child development" und dem deutschen Faktor 2 "Atmosphäre/ pädagogische Anleitung durch Erzieherinnen"; der Faktor "Developmentally appropriate activity" ist - zumindest teilweise - vergleichbar mit dem US-Faktor 2 "Availability and arrangement of materials and space as prerequisites for learning" und dem deutschen Faktor 1 "Allgemeine räumlich-materiale Bedingungen für Lernen, Grobmotorik und Pflege".

Zusammenfassend läßt sich festhalten, daß zur ECERS eine Vielzahl von Informationen zu den Gütekriterien vorliegt. Die verschiedenen Studien weisen auf eine gute Objektivität, Reliabilität und Validität der ECERS hin, so daß hier von einem gut erprobten und technisch abgesicherten Instrument gesprochen werden kann. Dies schließt nicht aus, daß noch weitere Analysen zu den Gütekriterien notwendig werden (z.B. zur internen Konsistenz der Subskalen). Dies gilt auch für die Aufdeckung der faktoriellen Struktur der ECERS. Die verschiedenen Faktorenanalysen reproduzieren zwar nicht die unterlegte theoretische Gestalt der sieben Subskalen, gleichwohl aber zeigt sich zwischen den verschiedenen Faktorenanalysen eine gewisse Stabilität zumindest von zwei Hauptfaktoren - für die Weiterarbeit ein

---

<sup>29</sup> Es ist unklar, ob dies die Korrelation der beiden Faktoren als Ergebnis der schiefwinkligen Rotation ist oder die Korrelation der entsprechend der Faktorenanalysen gebildeten additiven Skalen.



erfreuliches Ergebnis.

### **2.1.6 Infant/Toddler Environment Rating Scale**

#### **Entstehung und Anwendungskontext**

Die ITERS (Infant/Toddler Environment Rating Scale) ist eine von T. Harms, D. Cryer und R. M. Clifford vom Frank Porter Graham Child Development Center der University of North Carolina at Chapel Hill entwickelte Adaptation der ECERS für Gruppenbetreuungen mit Kinder im Alter von unter 30 Monaten (Harms et al., 1990). Theoretischer Hintergrund und allgemeiner Aufbau dieses Instruments für eine Globaleinschätzung der Qualität in solchen Gruppen entsprechen der ECERS, so daß hier auf eine ausführlichere Beschreibung verzichtet werden kann (siehe weiter oben). Im Rahmen der ITERS werden "infants" als Kinder im Alter von unter 12 Monaten, "toddlers" als Kinder im Alter von 12 bis 30 Monaten definiert. Für die Einarbeitung in die ITERS gibt es verschiedene Trainingsmaterialien, deren Veröffentlichung unmittelbar bevorsteht.

Die Definition von Qualität, die den 35 Items der ITERS zugrunde liegt, ist weitgehend konsistent mit den NAEYC Akkreditierungskriterien (National Association for the Education of Young Children, 1991a) und den CDA Competency Standards für die Arbeit mit Kindern dieser Altersgruppe (Council for Early Childhood Professional Recognition, 1991c). 97 % bzw. 91 % der von den Items der ITERS angesprochenen Qualitätsaspekte sind ebenfalls in diesen Sammlungen von Qualitätskriterien enthalten. Rund 80 % der in den NAEYC Akkreditierungskriterien angesprochenen Aspekte werden auch von der ITERS angesprochen (vgl. Clifford et al., 1989; siehe auch weiter unten). Im Prinzip wird jeder Funktionsbereich in den CDA Competency Standards durch Items der ITERS angesprochen (Harms et al., 1990, S. 1f.).

#### **Aufbau des Instruments**

Die 35 Items sind in sieben Kategorien oder Bereiche aufgliedert:

- 1. Furnishings and display for children, 5 Items
- 2. Personal care routines, 9 Items
- 3. Listening and talking, 2 Items
- 4. Learning activities, 8 Items
- 5. Interaction, 3 Items
- 6. Program structure, 4 Items
- 7. Adult needs, 4 Items

Zur Einschätzung der Items werden wiederum wie bei der ECERS siebenstufige Ratingskalen benutzt, wobei die Skalenstufen 1 (inadequate), 3 (minimal), 5 (good) und 7 (excellent) jeweils ausführlich beschrieben werden. "Inadequate" meint dabei für Kinder dieser Altersstufe eine Betreuung, die noch nicht einmal kustodiale Betreuungsbedürfnisse des Kindes erfüllt, "Minimal" beschreibt eine Betreuungssituation, die kustodiale Bedürfnisse erfüllt und zugleich einige grundlegende Entwicklungsbedürfnisse des Kindes befriedigt. "Good" bezieht sich auf eine Betreuung, die die grundlegenden Entwicklungsbedürfnisse des Kindes erfüllt. "Excellent" schließlich beschreibt eine individualisierte Betreuung von hoher Qualität. Die Skalenstufen 1 und 3 beziehen sich in der Regel auf die Bereitstellung von grundlegenden räumlich-materialen Möglichkeiten sowie auf Gesundheits- und Sicherheitsvorsorge, die Stufen 5 und 7 verlangen positive Interaktionen, Planung und individualisierte Betreuung zusätzlich zu guten räumlich-materialen Bedingungen.

Im Vergleich zur ECERS bestehen in der ITERS die Beschreibungen der Skalenstufe 1 stärker aus einer Auflistung von negativen Aspekten. Falls *einer* von diesen negativen Aspekten in der Betreuungsform beobachtet wird, führt dies unweigerlich zu einer negativen Beurteilung (also einer 1). Ein höherer Wert kann somit nur dann erreicht werden, wenn keiner dieser negativen Aspekte beobachtet wird. Um einen Wert von 3 oder höher zu erhalten, müssen immer *alle* positiven Merkmale der jeweiligen Skalenstufe plus *alle* positiven Merkmale der darunter liegenden Skalenstufen gegeben sein.

Werden die Beobachtungen von einem außenstehenden Beobachter durchgeführt, so benötigt er etwa zwei Stunden für die Bearbeitung der ITERS. Die 35 Skalen brauchen nicht strikt in der Reihenfolge bearbeitet werden, wie sie im Instrument stehen. Alle Items müssen während des Besuches eingeschätzt werden. In der Regel beruhen die Einschätzungen auf Beobachtungen, bei einigen Items werden Befragungen der Erzieherinnen erforderlich. Es kann immer nur eine Gruppe gleichzeitig bearbeitet werden. Die Einschätzungen der 35 Qualitätsaspekte sollen immer auf der gegenwärtigen Situation, wie sie beobachtet oder von der Erzieherin berichtet wird, beruhen, nicht auf zukünftigen Plänen oder Veränderungswünschen.

Zwei Beispiele für Items der ITERS (das erste Item stammt aus dem Bereich 2. Personal care routines, das zweite aus dem Bereich 4. Learning activities):

## 6. Greeting/Departing

1

- Greeting is often neglected; departure not prepared for.
- Parents discouraged from entering area used for child's care.
- Parents do not have direct contact with caregiver.

3

- Caregiver usually greets child and parent and acknowledges departure.
- Parents allowed to enter area used for child's care.
- Parents and staff share information related to child's health and safety (Ex. special diets, accident reports).

5

- Caregiver greets each child and parent warmly and provides pleasant organized departure (Ex. conversation on arrival; clothes ready for departure).
- Parents bring child into caregiving area as part of daily routine.
- Separation problems handled sensitively.
- Written record of infant's daily feeding, diapering, and naps available for parents to see.

7

- Staff use greeting and departure as information-sharing time with parents.
- Staff give parents specific information about how the day went (Ex. play activities, mood, new skills).

## 23. Sand and water play

1

- Sand play not available for children 24 months and older.
- Water play not available for children 12 months or older.

3

- Some sand or water play provided outdoors or indoors at least once every 2 weeks yearround.
- Close supervision of sand/water play.
- Some toys used for sand/water play.

5

- Sand or water play at least once a week.
- Variety of toys for sand/water play.

7

- Sand or water play at least three times a week.
- Caregiver talks with children about textures, wet/ dryness; describes children's activities.

Die Beispiele zeigen wiederum, daß - wie bei der ECERS - die hier verwendeten Ratingskalen relativ komplexe Instrumente darstellen, die hohe Anforderungen an die Verarbeitungskapazitäten der Beobachter stellen. Er muß für jedes Rating viele Einzelaspekte beobachten und diese dann zu einem einheitlichen Urteil zusammenfassen. Damit wird für die Items ein höheres Abstraktionsniveau verlangt.

Die Ergebnisse zu den 35 Ratingsskalen können sowohl insgesamt über alle Items als auch bezogen auf die sieben Subskalen additiv zusammengefaßt werden. Durch die Bildung von Durchschnittswerten für das gesamte Instrument bzw. für die einzelnen Subskalen kann die ursprüngliche siebenstufige Skala zur zusammenfassenden Beurteilung der Qualität herangezogen werden. Ein spezielles Auswertungsblatt für die Darstellung der Ergebnisse auf Skalenebene ist im Unterschied zur ECERS nicht mehr vorhanden.

### **Gütekriterien**

Vor der Publizierung der ITERS haben die Autoren verschiedene Studien zur Analyse der Reliabilität und Validität des Instruments durchgeführt (vgl. Harms et al., 1990, S. 2; Clifford et al., 1989). An einer Stichprobe von 30 Gruppen mit Kindern unter 30 Monaten aus 30 verschiedenen Einrichtungen wurde die Interrater-Reliabilität - also das Ausmaß der Objektivität des Instruments - überprüft. Dazu wurden die 30 Gruppen unabhängig von jeweils zwei Beobachtern eingeschätzt.<sup>30</sup> Bei 29 der 35 Items weichen in mehr als 70 % der beobachteten Gruppen die Ratings nicht mehr als um einen Skaleneinheit auf der siebenstufigen Skala voneinander ab. Die Rangkorrelation der beiden Meßwertreihen zu den 30 Gruppen beträgt .84 für den additiven Gesamtwert; die niedrigste Rangkorrelation auf Subskalenebene beträgt .58 (Furnishing and display for children).<sup>31</sup> Die interne Konsistenz (Cronbach's Alpha) beträgt für die Gesamtskala .83. Bei den Subskalen variiert Alpha von .26 bis .79, wobei Alpha aber nur für zwei Skalen (Program structure und Adult needs) unter .70 liegt. Berücksichtigt man, daß einige Subskalen nur aus sehr wenigen Items bestehen, so dürfte die interne Konsistenz insgesamt als zufriedenstellend betrachtet werden. 18 der 30 Gruppen wurden nach drei bis vier Wochen ein zweites Mal eingeschätzt. Die Test-Retest-Reliabilität gemessen über die Rangkorrelation der Meßwertreihen zum ersten und zum zweiten Zeitpunkt beträgt für die Gesamtskala .79; die Subskalen variieren zwischen .58 (Program structure) und .82 (Interaction). Die Autoren erwarten ein höheres Ausmaß an Übereinstimmung, wenn zwischen den beiden Einschätzungen weniger Zeit liegt. Insgesamt kann

---

<sup>30</sup> Die Beobachtungen der Gruppen fanden unabhängig voneinander statt, aufgrund von Zeitproblemen mußten jedoch beide Beobachter zusammen die Erzieherinnen befragen, um Informationen zu Aspekten zu bekommen, die sie nicht beobachtet hatten. Allerdings war den beiden Beobachtern untersagt, ihre Ratings vor Beendigung der gesamten Prozedur miteinander zu diskutieren.

<sup>31</sup> In der National Child Care Staffing Study - in der die ITERS für Gruppen mit jüngeren Vorschulkindern und die ECERS für Gruppen mit älteren Vorschulkindern eingesetzt wurde - wird bezogen auf alle eingesetzten Instrumente von Beobachterübereinstimmungen von über 85 % berichtet (vgl. Whitebook et al., 1990, S. 21).

von einer zufriedenstellenden Objektivität und Reliabilität der ITERS ausgegangen werden.

Für eine Analyse der kriteriumsbezogenen Validität wurden zwei Experten, die keinerlei Erfahrungen mit der ITERS hatten und dieses Instrument auch vorher noch nicht gesehen hatten, gebeten, 12 Gruppen in die Kategorien "niedrige" und "hohe" Qualität einzuteilen. Diese Gruppen waren vorher aufgrund von ITERS-Werten so ausgewählt worden, daß jeder Experte drei Gruppen mit niedriger und drei Gruppen mit hoher Qualität zu beurteilen hatte. In 10 der 12 möglichen Fälle (83 %) stimmten die Einschätzungen der Experten mit den auf den ITERS-Werten basierenden Einteilungen überein. In den beiden Fällen der Nicht-Übereinstimmung lag einige Zeit zwischen den ITERS-Einschätzungen und den Experteneinschätzungen, wobei sich in dieser Zeit - wie Recherchen ergaben - die Qualität der Arbeit in den beiden Gruppen verändert hatte. Zwei weitere Studien wandten sich der inhaltlichen Validität der ITERS zu: Ein auf Itembasis durchgeführter Vergleich der ITERS mit sieben anderen, vergleichbaren Instrumenten ergab, daß im Durchschnitt 82 % der Items der ITERS auch in den anderen Instrumenten enthalten sind sowie daß 75 % der von den anderen Instrumenten angesprochenen Aspekte auch in der ITERS berücksichtigt werden. Fünf anerkannte Experten des frühkindlichen Bereichs wurde gebeten, auf einer fünfstufigen Wichtigkeitsskala (hohe Skalenwerte = wichtig) die 35 Items der ITERS zu bewerten. Im Durchschnitt wurden 86 % der Items als sehr wichtig oder extrem wichtig eingestuft. Im Durchschnitt ergab sich eine Einschätzung von 4,3 auf der fünfstufigen Wichtigkeitsskala. Insgesamt lassen diese Ergebnisse in bezug auf die hier angesprochenen Validitätsaspekte auf eine gute Validität des ITERS schließen.

Im Anschluß an diese Studien zur Reliabilität und Validität wurden vor der Veröffentlichung noch geringfügige Veränderungen an der ITERS vorgenommen, um die Reliabilität und Validität zu erhöhen.

Wie weiter oben bereits erwähnt, wurde die ITERS in der National Child Care Staffing Study zur Qualitätseinschätzung von Gruppen mit jüngeren Kindern ("infants" und "toddlers"; n = insgesamt 330 Gruppen) benutzt (vgl. Whitebook et al., 1990; Howes et al., in Druck). Eine Faktorenanalyse (Maximum Likelihood Methode, schiefwinklige Rotation) ergab zwei Faktoren: Appropriate caregiving (56 % der gemeinsamen Varianz) und Developmentally appropriate activity (44 %). Die beiden Faktoren bzw. entsprechend gebildete Skalen korrelieren hoch miteinander (für infant-Gruppen .82, für toddler-Gruppen .81) (zu den Faktorenanalysen vgl. ausführlicher die Ausführungen zu der entsprechenden Faktorenanalyse für die ECERS in Kap. 2.1.5; alle Ausführungen gelten entsprechend). Wie bei der ECERS stehen die beiden Faktoren in vielfältigen und inhaltlich sinnvoll zu interpretierenden Bezie-

hungen zu anderen Konstrukten z.B. der Lernumwelt in den Gruppen (z.B. Gruppengröße, zahlenmäßiges Erzieher-Kind-Verhältnis), zu Merkmalen der Erzieherinnen (z.B. Ausbildung, Bezahlung, Zufriedenheit mit der Tätigkeit) wie auch zu Merkmalen des kindlichen Entwicklungsstandes (z.B. gemessen über den Peabody Picture Vocabulary Test, das Adaptive Language Inventory sowie verschiedenen Maßen der Sozialentwicklung). In einer Folgeuntersuchung (Howes et al., in Druck) wurden diese beiden Faktoren - parallel zum Vorgehen für Gruppen mit älteren Kinder - auch in ein Pfadmodell integriert (vgl. auch hierzu ausführlicher Kap. 2.1.5). Somit kann der ITERS eine Konstruktvalidität zugesprochen werden.

### ***2.1.7 Preschool Classroom Implementation Rating Instrument***

#### ***Entstehung und Anwendungskontext***

Das PCI-H/S (Preschool Classroom Implementation Rating Instrument: High/Scope) wurde im Rahmen der Evaluation eines landesweiten Projekts in South Carolina entwickelt (vgl. Frede, Miller, Eads, Lee, MacNeill & Barton, 1991). Der Haupteinsatzbereich dieses Instruments besteht in der Überprüfung der allgemeinen Implementation des "High/Scope Curriculums - Young Children in Action" (Homann, Baret & Weikart, 1978). Das High/Scope Curriculum basiert auf der Entwicklungstheorie von Piaget, entspricht nach Angaben von Frede et al. (1991) aber zugleich den NAEYC Guidelines for Developmentally Appropriate Practices (vgl. Bredekamp, 1987; auch Kap. 2.1.2). Das PCI-H/S überprüft somit sowohl die Implementation des High/Scope Curriculums als auch zugleich, ob die pädagogische Arbeit den Kriterien für "entwicklungsangemessenes" Vorgehen entspricht. Das Instrument kann in Preschool Gruppen, d.h. hier in Gruppen mit Kinder im Alter von drei bis fünf Jahren, eingesetzt werden.

Von den insgesamt 54 Items sind acht Items spezifisch auf das High/Scope Curriculum abgestimmt. Ohne diese acht Items - der Teil ohne diese acht Items wird auch "cognitive/developmental subscale" genannt - sei auch ein allgemeinerer Einsatz in anderen Gruppen, die nicht diesem Curriculum folgen, möglich (Frede et al., 1991). Dennoch scheinen dem Berichtersteller auch andere Items zumindest implizit einem piagetorientierten Ansatz nahe zu stehen, so daß der Einsatz der PCI-H/S insgesamt einem piagetorientierten Qualitätsverständnis verbunden bleibt.

Der Einsatz der PCI-H/S setzt einen trainierten (Fremd-)Beobachter voraus, was wohl den Einsatz in Selbstevaluationsprozessen erschweren dürfte.

#### ***Aufbau des Instruments***

Das PCI-H/S besteht aus 54 Items, die in 12 Bereiche gegliedert sind:

- Room arrangement and materials, 8 Items
- Daily routine, 4 Items
- Planning time, 3 Items
- Work/free play time, 3 Items
- Clean-up time, 2 Items
- Recall time, 2 Items
- Small group time, 5 Items
- Outside time, 2 Items
- Circle time, 3 Items
- Teacher/child interaction, 15 Items
- Classroom management and organization, 5 Items
- Team evaluation and planning, 2 Items

Das PCI-H/S besteht aus einem Antwortbogen und einem Manual. Die einzelnen Items werden als kurze Statements formuliert, die anhand einer dreistufigen Skala eingeschätzt werden müssen:

- not evident: Die in dem Item angesprochen Materialien bzw. das dort beschriebene Verhalten ist nicht vorhanden.
- evident: Die in dem Item angesprochenen Materialien bzw. das dort beschriebene Verhalten ist prinzipiell vorhanden. Wann aber genau "evident" und nicht "optimal" vergeben werden soll, wird bei jedem einzelnen Item ausführlich im Manual beschrieben.
- optimal: Die in dem Item angesprochenen Materialien sind vollständig da; es gibt kaum Situation, wo ein in dem Item angesprochenes Verhalten nicht beobachtet werden kann.

Ferner gibt es noch die Antwortmöglichkeit "not observed", die dann angekreuzt werden soll, wenn ein Item nicht eingeschätzt werden kann, weil es während der Beobachtung keine Möglichkeit gab, das geforderte Verhalten zu zeigen. Bei jedem Item müssen für die Einschätzung Begründungen angegeben werden. Bei einigen Items kann dies bei einer Einstufung von "optimal" entfallen. Im Manual wird für jedes Item genau beschrieben, wann eine Einschätzung "not evident", "evident", "optimal" vergeben werden muß und welche Begründungen bzw. Erläuterungen angegeben werden müssen. Durch dieses Vorgehen wird vermutlich die Präzision der Einschätzungen gesteigert, andererseits werden dadurch erhöhte Anforderungen an den Beobachter gestellt, was dazu führt, daß ein umfangreicheres Training erforderlich wird (Frede u.a. 1991).

Zwei Beispiele für Items (das erste Beispiel stammt aus dem Bereich Room arrangement and materials, das zweite aus dem Bereich Small group time):

*Materials are easily accessible to children*

**Criteria:** At work time, children can reach materials and get them out of their container.

**Not evident:** Only the materials that teacher set out on tables are available to children OR those materials impede the use of other materials OR children ask teachers for materials that are not available or are too high to reach on 4 or more occasions.

**Evident:** The above is true for one or two areas of the room OR children ask for materials on 2 to 3 occasions.

**Optimal:** Children can reach or get materials on almost all occasions during work time.

**Justification:** List materials asked for, or describe situations where teacher-chosen materials impede use of others. No justification needed for optimal.

*Small group activities allow each child to make choices*

**Criteria:** Activities are structured so that children determine how to use materials. They do not follow teacher's directions.

**Not evident:** There is no small group time OR the activities are teacher-directed.

**Evident:** One adult allows children; the other does not.

**Optimal:** Both adults allow children to use materials in their own way.

**Justification:** Describe activity.

Obwohl hier zum Teil auf diskretes Verhalten von Erzieherinnen Bezug genommen wird, werden hier von dem Beobachter nicht Einzelmerkmale beurteilt. Auch hier muß er verschiedene Einzelaspekte zu einem zusammenfassenden Urteil bündeln. Insgesamt verlangen diese Items ein mittleres Abstraktionsniveau.

Im Manual wird die Empfehlung ausgesprochen, während der Beobachtung nur Notizen zu machen, und dann anschließend anhand der Ratingkriterien im Manual den Antwortbogen auszufüllen. Die Beantwortung der Mehrzahl der Items beruht auf Beobachtungen des Gruppengeschehens. Bei einigen wenigen Items - im Manual benannt - müssen die Betreuer interviewt werden. Der Beobachter soll sich in der



Gruppe sehr zurückhaltend verhalten; er soll Blickkontakte mit den Kinder vermeiden und ihnen auch möglichst keine Antworten geben. Es gibt keine Angaben dazu, wieviel Zeit für die Anwendung des PCI-H/S pro Gruppe benötigt wird.

Die Antworten auf die dreistufigen Skalen können zu additiven Gesamtskalen zusammengefaßt werden (vermutlich: not evident = 1, evident = 2, optimal = 3). Die auf diese Weise errechnete Punktzahl wird dann dividiert durch die Anzahl aller möglichen Punkte. Die Anzahl aller möglichen Punkte errechnet sich vermutlich nach: (54 - Anzahl nicht beobachteter Items) \* 3. Nähere Angaben fehlen in den dem Berichterstatter vorliegenden Unterlagen. Angaben, ab welchem Punktwert von einer qualitativ guten Betreuungsform gesprochen werden kann, liegen nicht vor.

### ***Gütekriterien***

Dem Berichterstatter liegen nur wenige Informationen zu den Gütekriterien der PCI-H/S vor. Die interne Konsistenz (Cronbach's Alpha) wird mit = .96 angegeben (vgl. Frede et al., 1991), was eine hohe Reliabilität in dieser Hinsicht indizieren würde. Desweiteren wird darauf verwiesen, daß die Skalenwerte des PCI-H/S die Werte eines Schulreifetests zu Beginn der ersten Klasse vorhersagen würden; weitere Angaben hierzu fehlen jedoch.

### ***2.1.8 Child Care Facility Schedule***

#### ***Entstehung und Anwendungskontext***

Dem Berichterstatter liegt zur Zeit nur ein kurzer Bericht über die CCFS (Child Care Facility Schedule) vor (vgl. Tsiantis et al., 1991). Die CCFS ist ein Instrument für eine Fremdbeobachtung von Einrichtungen des frühkindlichen Bereichs; die Autoren verweisen aber auch auf zukünftige Einsatzmöglichkeiten im Rahmen von Selbstevaluation, wozu allerdings noch Modifikationen erforderlich würden. Das Instrument zielt dabei eine Qualitätseinschätzung der gesamten Einrichtung an, nicht nur einzelner Gruppen. Das Instrument deckt die gesamte Altersspanne von Gruppenbetreuungen im Vorschulalter ab, d.h. Gruppen für Kinder im Alter von 0 bis 18 Monate, Gruppen für Kinder im Alter von 18 bis 36 Monaten sowie Gruppen für Kinder von drei bis sechs Jahren. Angaben zum Umfang des benötigten Trainings der Beobachter liegen nicht vor.<sup>32</sup>

Die CCFS wurde im Rahmen der World Health Organization entwickelt als eine

---

<sup>32</sup> In der Reliabilitätsstudie (vgl. weiter unten) bestand das Training darin, daß in Griechenland und in Nigeria jeweils zwei bzw. drei Beobachter fünf Einrichtungen mit der CCFS beurteilten und anschließend ihre Ratings diskutierten.

Prozedur zur Einschätzung der Qualität in frühkindlichen Einrichtungen, die in den verschiedensten Ländern eingesetzt werden kann. Als Ausgangspunkt für die Auswahl der Items wird eine Durchsicht bestehender Sammlungen von Qualitätskriterien genannt, besonders die NAEYC Akkreditierungskriterien (National Association for the Education of Young Children, 1991a). Die Kriterien wurden in einfache Statements umformuliert, wobei die in dem Item angesprochenen Aspekte entweder leicht während eines Besuches der Einrichtung beobachtet werden können müssen oder die Informationen durch eine kurze Befragung des Personal zu erhalten sind. Die Durchsicht ergab eine Sammlung von 80 Items. Nach einer Pilotstudie in Griechenland, in den Philippinen und in Nigeria wurden einige Items verändert und andere herausgenommen. 74 Items blieben dabei übrig.

### *Aufbau des Instruments*

Die 74 Items der CCFS sind in sieben Bereiche gegliedert.<sup>33</sup>

- Physical environment, 10 Items
- Health and safety, 11 Items
- Nutrition and food services, 4 Items
- Administration, 22 Items
- Staff family interaction, 11 Items
- Observable child behavior, 6 Items
- Curriculum, 10 Items.

Jedes Items muß auf einer dreistufigen Skala eingeschätzt werden:

- 0 = Die Bedingung/der durch das Item angesprochene Aspekt ist in der Einrichtung nicht vorhanden.
- 1 = Es ist fraglich, ob die Bedingung/der durch das Item angesprochene Aspekt in der Einrichtung vorhanden ist.
- 2 = Die Bedingung/der durch das Item angesprochene Aspekt ist eindeutig in der Einrichtung vorhanden.

Zusätzlich gibt es noch die Möglichkeit, ein Item als nicht anwendbar (not applicable) einzuschätzen. Die Informationsquellen für die Einschätzung der Items sind Beobachtung und Befragungen des Personals. Es gibt keine Angaben dazu, wieviel

---

<sup>33</sup> Im Anhang von Tsiantis et al. (1991) sind die Items wiedergegeben; allerdings ist unklar, auf welche Version sich die hier abgedruckte Itemsammlung von 80 Items bezieht.

Zeit für die Anwendung der CCFS benötigt wird.

Im folgenden jeweils drei Itembeispiele aus den Bereichen "Health and safety" und "Curriculum" (zitiert nach dem Anhang zu Tsiantis et al., 1991):

18. The centre assumes responsibility for safe arrival and departure of all children.
19. Children are carefully supervised, with at least two adults on the premises at all times.
21. All caregiving staff members have been trained in first aid.
71. The staff plan realistic curriculum goals for the children based on continuing assessment of individual needs and interests.
72. The daily schedule is planned to provide a variety of activities, including those that are: indoor/outdoor, quiet/active, individual/small group/large group.
75. Teaching/learning activities encourage language development and help children improve their ability to think, reason, question and experiment.

Wie die Beispiele zeigen, müssen die Beobachter bei der CCFS nicht nur Einzelmerkmale auf ihr Vorhandensein hin beurteilen. Vielmehr müssen bei einigen Items doch relativ komplexe Aspekte, die aus vielen Einzelmerkmalen bestehen, zusammenfassend durch den Beobachter beurteilt werden. Mithin werden an die Beobachter einige Anforderungen gestellt. Insgesamt lassen sich diese Items einem mittleren Abstraktionsniveau zuordnen. Für die Pilotstudien in Griechenland, in den Philippinen und in Nigeria stellen die Autoren aber fest, daß die Items leicht eingeschätzt werden konnten und die CCFS damit als ein leicht anzuwendendes Instrument angesehen werden kann. Für eine internationale Anwendung unter sehr unterschiedlichen Bedingungen des frühpädagogischen Bereichs dürfte eine solche leichte Handhabbarkeit eine der Hauptbedingungen für ein erfolgreich einsetzbares Instrument sein.

Der Artikel enthält keine Angaben dazu, ob und wie Skalen oder Gesamtwerte gebildet werden sollen. Eine solche Zusammenfassung der Informationen wäre aber wünschenswert, um zu zusammenfassenden Qualitätsbeurteilungen zu kommen. Additive Skalenbildungen liegen nahe.

### **Gütekriterien**

In Griechenland (Athen, 30 Einrichtungen) und Nigeria (Ibadan, 39 Einrichtungen) wurde die Interrater-Reliabilität überprüft. Jeweils zwei Beobachter haben dazu die CCFS unabhängig voneinander angewandt (Abstand der beiden Beobachtungen in Athen sechs bis zehn Tage; in Ibadan eine bis vier Wochen). Die Autoren berichten für Athen von bereichsspezifischen Korrelationen der beiden Meßwertreihen von .83 (Health and safety) bis .998 (Nutrition and food service). In Nigeria ergab sich eine Korrelation von .94 für das gesamte Instrument. Die Übereinstimmung der beiden Beobachtungen wurde auch ausführlich auf der Einzelitemebene überprüft, wobei sich ebenfalls eine hohe Übereinstimmung der Ratings auf der dreistufigen Skala zeigt. Insgesamt ergibt sich somit eine - überraschend - hohe Objektivität der CCFS. Ergebnisse zur Validität der CCFS werden nicht berichtet, es wird nur auf eine laufende Validitätsstudie in Griechenland verwiesen.

Als Ergebnis der Reliabilitätsstudie wurden einige Items verändert. Der neue Itemsatz besteht wieder aus 80 Items. Ebenfalls wird vorgeschlagen, statt der dreistufigen Skala nun eine vierstufige Skala (0 = Bedingung ist in der Einrichtung nicht vorhanden; 1 = Bedingung ist teilweise in der Einrichtung vorhanden; 2 = Bedingung ist adäquat in der Einrichtung vorhanden; 3 = Bedingung ist vollständig in der Einrichtung vorhanden) zu verwenden, da die dreistufige Skala nicht ausreichend die Nuancen zwischen den Einrichtungen erfassen könnte.

## **2.2 Instrumente für Tagespflege**

### **2.2.1 Assessment Profile for Family Day Care**

#### **Entstehung und Anwendungskontext**

Das Assessment Profile for Family Day Care (Sibley & Abbott-Shim, 1991) wurde für das Akkreditierungsprogramm von NAFDC (National Association for Family Day Care) entwickelt.

Zum besseren Verständnis eine kurze Skizze des NAFDC Akkreditierungsprogramms (vgl. Sibley & Abbott-Shim, 1989; National Association for Family Day Care, ohne Jahr): Die NAFDC Akkreditierung ist für Tagespflegestellen bzw. Tagespflegebetreuerinnen gedacht, die die jeweiligen lokalen Lizenzierungsbedingungen erfüllen und über diese Mindeststandards hinausgehend eine besondere Qualität ihrer Arbeit demonstrieren. Die Qualitätsstandards sind in einem Study Guide ausführlich beschrieben (Sibley & Abbott-Shim, 1989). Der Study Guide beschreibt in sieben Dimensionen, wie eine gute Tagespflegestelle aussehen sollte. Die sieben Dimensionen sind:

- Indoor safety
- Health
- Nutrition
- Indoor play environment
- Interacting

- Outdoor play environment
- Professional responsibilities

Die sieben Dimensionen werden narrativ beschrieben, wobei jeweils einige generelle Sätze bzw. Standards herausgehoben werden. Zu jedem von diesen generellen Standards werden dann einzelne Kriterien für eine gute Qualität genannt. Der Study Guide ist zunächst als eine Orientierungshilfe für die Tagespflegebetreuerinnen gedacht, um ihre eigene Qualität zu verbessern.

An dem Akkreditierungsprozeß sind drei Personen beteiligt, die unabhängig voneinander die Arbeit in der Tagespflegestelle bewerten: die Tagespflegebetreuerin selbst, ein Elternvertreter (parent validator), der selbst ein eigenes Kind in einer Tagespflegestelle hat oder hatte, wobei diese aber nicht die zu bewertende Tagespflegestelle sein darf, und ein Vertreter von NAFDC (NAFDC validator). Der Elternvertreter wird von der die Akkreditierung beantragenden Tagespflegebetreuerin selbst ausgewählt, wobei bestimmte Bedingungen erfüllt werden müssen. Nach Erhalt der Unterlagen von NAFDC hat die Tagespflegebetreuerin 90 Tage Zeit, um den Akkreditierungsprozeß abzuschließen.

In dem Akkreditierungsprozeß führt die Tagespflegebetreuerin zunächst eine Selbstevaluation durch. Dazu füllt sie das Assessment Profile for Family Day Care aus. Diese Selbstevaluation wird anschließend durch den Elternvertreter und den NAFDC Vertreter validiert, die ebenfalls das Assessment Profile for Family Day Care ausfüllen. Dabei geben sie aber nicht ihre Einschätzungen an die Tagespflegebetreuerin weiter. Vielmehr hinterlegen sie ihre Einschätzungen in versiegelten Umschlägen bei der Tagespflegebetreuerin, die diese dann mit weiteren Unterlagen an NAFDC zurückschickt. U.a. verteilt sie in diesem Prozeß auch Fragebögen an die Eltern, die sie wieder einsammeln muß und an NAFDC zurückschickt.

Die Unterlagen werden von NAFDC ausgewertet. Bei dieser Auswertung werden dann Durchschnittswerte aus den Werten der Assessment Profile der Tagespflegebetreuerin, dem Elternvertreter und dem NAFDC Vertreter gebildet. Diese Ergebnisse werden an die Tagespflegebetreuerin zurückgemeldet (vgl. dazu weiter unten). U.a. wird auch eine Liste der Kriterien zugeschickt, die entweder von der Tagespflegebetreuerin oder von einem der beiden anderen Beobachter mit "nein" (nicht vorhanden) bewertet wurden. Dies geschieht aber wohl nur für die Dimensionen, in denen im Durchschnitt der Bewertungen von der Tagespflegeperson, dem Elternvertreter und der NAFDC Vertreter 89 % oder weniger der Items mit "ja" (Aspekt ist vorhanden) bewertet wurden. Indem aber ansonsten nur Durchschnittswerte zurückgemeldet werden, wird u.U. eine pädagogische Chance verpaßt, denn gerade aus dem Vergleich und der Diskussion der Bewertungen durch die verschiedenen Personen dürfte die Tagespflegebetreuerin Hinweise für Verbesserungsmaßnahmen erhalten.

Bezogen auf die von NAFDC zurückgemeldeten Ergebnisse muß die Tagespflegebetreuerin anschließend einen zusammenfassenden Bericht schreiben. Dieser soll zum einen die Aspekte enthalten, die die Tagespflegebetreuerin aufgrund der Teilnahme am Akkreditierungsprozeß bereits verbessert hat oder beabsichtigt zu verbessern. Ebenfalls soll sie auf ihre Fortbildungspläne für das kommende Jahr eingehen. Wenn in irgendeiner Dimension nur ein Durchschnittswert von 89 % oder weniger (Durchschnitt der "Ja" Antworten aus den Bewertungen durch die Tagespflegebetreuerin selbst, dem Elternvertreter und dem NAFDC Vertreter) erreicht wurde, müssen spezielle Fragen zu dieser Dimension beantwortet werden, die dann von NAFDC auf dreistufigen Rating-

skalen (1 = unangemessene Antwort, 3 = befriedigende Antwort) bewertet werden.

Für die Vergabe der Akkreditierung betrachtet dann ein NAFDC Komitee drei Informationsquellen:

- die drei Assessment Profile. Eine Akkreditierung kann nur dann ausgesprochen werden, wenn im Durchschnitt über die drei Profile in jeder Dimension mindestens 80 % der Items mit "ja" (Aspekt vorhanden) (früher: 75 %) bewertet wurden.
- den zusammenfassenden Bericht der Tagespflegeperson, der dann anhand von dreistufigen Ratingsskalen beurteilt wird. Offensichtlich wird auch ein zusammenfassender Wert als Anteil der erreichten Punkte bezogen auf die Maximalzahl der zu erreichenden Punkte gebildet.
- die Elternfragebögen, deren Ergebnisse offensichtlich auch in einem Prozentwert zusammengefaßt werden. Dem Berichtersteller liegen keine Informationen vor, wie diese Prozentwertbildung geschieht.

Aus diesen drei Quellen wird dann ein zusammenfassender Prozentwert gebildet, wobei die Zusammenfassung gewichtet erfolgt: Das Ergebnis des Assessment Profiles wird mit 80 % gewichtet, die beiden anderen Teile jeweils nur mit 10 %. Eine Akkreditierung wird dann ausgesprochen, wenn der Gesamtwert 85 % oder besser beträgt (und - wie bereits erwähnt - in keiner Dimension im Assessment Profile weniger als 80 % der Aspekte mit "ja" bewertet wurden). Die Akkreditierung gilt dann für drei Jahre. Im Unterschied zum NAEYC Akkreditierungsprogramm (vgl. Kap. 2.1.1) und zum CDA National Credentialing Program (vgl. Kap. 2.1.3) wird hier somit der Akkreditierungsentscheidung ein quantitatives Kriterien zugrunde gelegt.

Die Items des Assessment Profile for Family Day Care sind eng auf die in dem Study Guide enthalten Qualitätsstandards und Kriterien bezogen; größtenteils sind sie als deren Operationalisierungen zu betrachten. Das Assessment Profile ist ausschließlich auf die Bewertung der Qualität von Tagespflegestellen hin ausgelegt. Dem Berichtersteller liegen keine Angaben zur Eingrenzung des Alters der in den Tagespflegestellen betreuten Kinder vor. Die Tagespflegebetreuerin und der Elternvertreter erhalten im Akkreditierungsprozeß kein Training im Umgang mit dem Assessment Profile, so daß besondere Anforderungen an eine leichte Handhabbarkeit des Instruments gestellt werden müssen. Die NAFDC Vertreter werden vermutlich in der Anwendung des Instruments trainiert; allerdings liegen keine Angaben über den Umfang des Trainings vor. Insgesamt aber scheint kein umfangreicheres Training in der Anwendung des Assessment Profiles erforderlich (vgl. auch Modigliani, 1991, S. 3).

### *Aufbau des Instruments*

Der Aufbau des Assessment Profiles for Family Day Care entspricht weitgehend dem Assessment Profile for Early Childhood Programs (vgl. Kap. 2.1.4), so daß die

Darstellung hier etwas gekürzt werden kann. In sieben Dimensionen werden zu 32 Standards insgesamt 277 Kriterien dargeboten, wobei sich 20 Kriterien speziell auf Infants (Kinder im Alter von 0 bis 18 Monate) beziehen und nur dann eingeschätzt werden dürfen, wenn Kindes dieses Alters in der Tagespflegestelle betreut werden.<sup>34</sup> Die 277 Kriterien sind die zu beobachtenden und einzuschätzenden Items. Die Verteilung auf die Dimensionen:

- Safety, 50 Items bezogen auf fünf Standards
- Health, 44 Items bezogen auf vier Standards
- Nutrition, 18 Items bezogen auf drei Standards
- Learning environment, 42 Items bezogen auf sieben Standards
- Interacting, 50 Items bezogen auf fünf Standards
- Outdoor environment, 25 Items bezogen auf drei Standards
- Professional responsibility, 48 Items bezogen auf fünf Standards

Neben der Beantwortung dieser Items muß stündlich die Anzahl der anwesenden Kinder und Erwachsenen sowie die Altersspanne der Kinder protokolliert werden.

Die Items/Kriterien sind als kurze Statements formuliert (weitere Erläuterungen zu den Kriterien finden sich in dem Study Guide, Sibley & Abbott-Shim, 1989). Die Beobachter müssen die Items mit "ja" oder "nein" beantworten, eine andere Möglichkeit haben sie nicht. Zur Beantwortung der Items können als Informationsquellen Beobachtungen, Durchsicht von Dokumenten und Befragungen der Tagespflegebetreuerin herangezogen werden. Neben jedem Item ist jeweils die benötigte Informationsquelle vermerkt ("o" für Beobachtung, "d" für Durchsicht von Dokumenten, "z" für Befragung der Tagespflegebetreuerin). Beobachtungen stehen als Informationsquellen im Vordergrund, nur bei einem Viertel der Items (70 Items) wird nicht auf Beobachtungen als Informationsquelle verwiesen. Alle Items sollen während des Besuchs der Tagespflegestelle beantwortet werden.

---

<sup>34</sup> Eine frühere Version bestand nur aus 189 Items in den sieben Dimensionen, wobei die Dimensionen leicht unterschiedlich benannt sind:

- Indoor safety, 34 Items
- Health, 27 Items
- Nutrition, 16 Items
- Indoor play environment, 22 Items
- Interacting, 49 Items
- Outdoor play environment, 14 Items
- Professional responsibilities, 27 Items.

Die dem Berichterstatter vorliegenden Auswertungsschemata (vgl. weiter unten) beziehen sich auf diese ältere Version.

Im folgenden werden als Beispiele zu zwei Standards alle Kriterien/Items aufgelistet. Der erste Standard stammt aus der Dimension Learning Environment, der zweite aus der Dimension Interacting:

- *A. The home is arranged to meet the developmental needs of children to explore independently.*
  - o 1. Specific shelves, drawers and/or containers which are accessible for children are designated for children's toys and plays.
  - o 2. Toys are available at child's reach without adult assistance or prohibition.
  - o 3. Toys are organized in a variety of containers and/or shelves.
  - o 4. Toys and materials are displayed in an orderly way.
  - r 5. Provider has a system to protect family personal space and possessions.
  
- *A. Provider is warm and nurturing with the children.*
  - o 1. Provider spontaneously expresses positive affection through physical gestures (such as smiles, hugs, pats, rubs a back, rocks, holds).
  - o 2. Provider is verbally spontaneous and positive with children (such as gives praise, encouragement, acknowledgement, conversation, cooing to infants).
  - o 3. Provider has a spontaneous sense of humor and engages children in laughter and smiling through verbal exchanges and/or playful games and activities.
  - o 4. Provider shares personal feelings and/or experiences as related to the activities and experiences of the day - OR all children are nonverbal.
  - o 5. Provider spends quiet, one-to-one time with children (such as holding, rocking, observing a mobile, looking at a picture book, assisting with a puzzle, rubbing child's back at nap time, conversation).

Wie aus den Beispielen ersichtlich (vgl. auch die entsprechenden Anmerkungen zum Assessment Profile for Early Childhood Programs), werden - relativ zu den bisher besprochenen Instrumenten - eher diskrete Einzelaspekte durch das Assessment Profile angesprochen. Mithin läßt sich von einem mittleren Abstraktionsniveau ausgehen. Damit werden auch weniger Anforderungen an die Informationsverarbeitungskapazitäten der Beobachter gestellt. Zum Teil sind die Items somit auch weniger anfällig für subjektive Interpretationen der Beobachter (vgl. Modigliani, 1991, S. 16). Die Ebene der Standards - die hier beim Assessment Profile analytisch gebildet werden muß - entspricht bei anderen Instrumenten der Ebene, auf der die Informationen selbst erhoben werden.

Der Elternvertreter und der NAFDC Vertreter benötigen etwa sechs Stunden für die



Bearbeitung des Profiles. Beide sollen sich dabei möglichst wenig am Geschehen beteiligen und die Rolle eines nicht-teilnehmenden Beobachters übernehmen.

Für die Aufbereitung der Beobachtungen gibt es - wie bei dem Assessment Profile for Early Childhood Programs - zwei Hilfsmittel:

- Auf einem Auswertungsblatt (Dimension Scores) werden für jeden Standard, für jede Dimension und für das Gesamtinstrument die Anzahl der mit "ja" beantworteten Kriterien aufgelistet. Daneben steht dann jeweils die maximal erreichbare Punktzahl (jeweilige Anzahl der Kriterien), so daß durch den Vergleich zwischen maximaler Punktzahl und erreichter Punktzahl die Leistungen in jedem Standard, in jeder Dimension und insgesamt beurteilt werden können.
- In ein weiteres Auswertungsblatt (Assessment Profile, Profile Graph) werden die in jeder Dimension erreichten Punktzahlen (= Anzahl der mit "ja" beantworteten Kriterien) eingetragen. Die jeweils in einer Dimension möglichen Punktzahlen sind dabei so auf eine gemeinsame Prozentskala abgetragen, daß die unterschiedlich umfangreichen Dimensionen direkt verglichen werden können.

Im Rahmen des NAFDC Akkreditierungsprogramms wird als Qualitätsniveau definiert, daß für eine Tagespflegestelle mindestens 80 % der Items (früher: 75 %) jeder Dimension mit "ja" beantwortet werden muß, damit eine ausgezeichnete Qualität und damit eine Akkreditierung ausgesprochen werden kann. Damit wird zwar einerseits ein eindeutiges quantitatives Kriterium definiert, zum anderen kritisiert aber z.B. Modigliani (1991, S. 16), daß durch dieses ungewichtete Vorgehen auch dann eine Akkreditierung ausgesprochen werden könnte, wenn einzelne, als unumgänglich betrachtete Kriterien nicht vorhanden sind.

### **Gütekriterien**

Zu den Gütekriterien des Assessment Profiles for Family Day Care liegen nur sehr wenige Informationen vor. Die Autoren berichten auf einem Faltblatt von einer durchschnittlichen Beobachterübereinstimmung von 92 % im Akkreditierungsprozeß (drei unabhängige Beobachtungen: Tagespflegebetreuerin, Elternvertreter, NAFDC Vertreter; an drei verschiedenen Tagen) (vgl. Quality Assist, ohne Jahr). Weitere Daten zur Objektivität und Reliabilität liegen nicht vor.

In einer Studie von 32 Tagespflegestellen fand Nelson (1989) - auf der Ebene der Gesamtwerte - Korrelationen von jeweils .62 zwischen dem Assessment Profile und

der FDCRS (Family Day Care Rating Scale, Harms & Clifford, 1989, vgl. Kap. 2.2.2) sowie zwischen dem Assessment Profile und dem Family Day Home Observation Instrument (Child Care Partnership of Dallas, 1987, vgl. Kap. 2.2.3). Aufgrund dieses Ergebnisses folgert Nelson, daß ihre Studie aufgrund der niedrigen Korrelation zur FDCRS - die gewissermaßen als Maßstab genommen wird - keine Konstruktvalidität des Assessment Profiles hätte nachweisen können. Näher scheint aber der Schluß zu liegen, daß das Assessment Profile und die FDCRS zum Teil unterschiedliche Qualitätsaspekte messen.

### **2.2.2 Family Day Care Rating Scale**

#### ***Entstehung und Anwendungskontext***

Die FDCRS (Family Day Care Rating Scale) ist eine von T. Harms und R.M. Clifford vom Frank Porter Graham Child Development Center der University of North Carolina at Chapel Hill entwickelte Adaptation der ECERS zur Einschätzung der Qualität von Tagespflegestellen (Harms & Clifford, 1989).<sup>35</sup> Theoretischer Hintergrund und Aufbau der 32 Items der FDCRS entsprechen der ECERS (vgl. Kap. 2.1.5). Das Instrument ist für Tagespflegestellen angemessen, in denen Kinder von der Geburt bis zum Schuleintritt betreut werden. Die FDCRS wurde entwickelt für Selbstevaluation durch die Tagespflegebetreuerin, für Supervision wie auch für den Einsatz in Forschung und Evaluation. Für die FDCRS existieren verschiedene Trainingsmaterialien, deren Veröffentlichung unmittelbar bevorsteht. Die FDCRS wird in den USA und Kanada breit angewandt und als bewährtes Instrument zur Beurteilung von Tagespflegestellen betrachtet (vgl. Modigliani, 1991, S. 17).

Die der FDCRS zugrunde liegenden Definition der Qualität einer Tagespflegestelle entspricht nach Angaben der Autoren den CDA Competency Standards für Tagespflegestellen (Council for Early Childhood Professional Recognition, 1991a). Tagespflege wird als eine professionelle Tätigkeit betrachtet ("small business"), die auch - entsprechend anderer "Heimarbeiten" - Anpassungen in der Wohnung erfordert. Alle Funktionsbereiche in den CDA Competency Standards werden durch Items der FDCRS angesprochen (vgl. Harms & Clifford, 1989, S. 1).

#### ***Aufbau des Instruments***

Die FDCRS besteht aus 32 Items, wobei drei Items unterteilt sind in Abhängigkeit

---

<sup>35</sup> Tagespflege wird hier definiert als Betreuung eines Kindes (speziell Kinder unter drei Jahren) in der Wohnung der Tagespflegebetreuerin in einer kleinen Gruppe von Kindern, wobei meistens eigene Kinder der Betreuerin mit betreut werden. In diesem Sinne ist auch Tagespflege in den USA eine Form von *Gruppenbetreuung*.

vom Alter der betreuten Kinder (Kinder unter zwei und Kinder über zwei Jahren). Zusätzlich gibt es acht ergänzende Items, die nur dann bearbeitet werden sollen, wenn behinderte Kinder (exceptional children) in der Tagespflegestelle betreut werden. Die 32 Items sind in sechs Subskalen organisiert:

- 1. Space and furnishings for care and learning, 6 Items
- 2. Basic care, 7 Items
- 3. Language and reasoning, 4 Items
- 4. Learning activities, 9 Items
- 5. Social development, 3 Items
- 6. Adult needs, 3 Items

Zur Einschätzung der Items werden wiederum wie bei der ECERS und der ITERS (vgl. 2.1.5 und 2.1.6) siebenstufige Ratingskalen benutzt, wobei die Skalenstufen 1 (inadequate; erfüllt nicht einmal kustodiale Betreuungsbedürfnisse des Kindes), 3 (minimal; erfüllt kustodiale Betreuungsbedürfnisse und teilweise Entwicklungsbedürfnisse des Kindes), 5 (good; erfüllt die Entwicklungsbedürfnisse des Kindes) und 7 (excellent; individualisierte Betreuung von hoher Qualität) jeweils ausführlich beschrieben werden. Die Skalenstufen "inadequate" und "minimal" konzentrieren sich auf die Bereitstellung von grundlegenden Materialien und auf Vorsichtsmaßnahmen im Hinblick auf die Gesundheit und Sicherheit der Kinder, die Skalenstufen "good" und "excellent" verlangen zusätzlich positive Interaktionen, Planung und individualisierte Betreuung.

Wie bei der ITERS bestehen die Beschreibungen der Skalenstufen 1 aus einer Auflistung von negativen Aspekten, wobei das Vorhandensein eines von diesen negativen Aspekten unweigerlich zur Vergabe der Qualitätsstufe "inadequate" führt. Die Skalenstufen 3, 5 und 7 können nur dann vergeben werden, wenn alle beschriebenen positiven Merkmale der Skalenstufe selbst sowie die der darunter liegenden Stufen erfüllt sind. Die Stufen 2, 4 und 6 müssen vergeben werden, wenn zwar alle positiven Merkmale einer tieferen Stufe erfüllt sind, nicht aber alle der jeweils höheren.

Ein außenstehender Beobachter benötigt etwa zwei Stunden für die Beobachtungen in der Tagespflegestelle. Hierzu soll eine Zeit ausgewählt werden, in der die betreuten Kinder wach und aktiv sind. Die Items brauchen nicht in der abgedruckten Reihenfolge bearbeitet werden. Alle Items müssen während des Besuchs des Beobachters in der Tagespflegestelle eingeschätzt werden. Als Informationsquellen für die Einschätzung der Items dienen in erster Linie Beobachtungen. In einigen Fällen müssen auch schriftliche Materialien durchgesehen und - falls erforderlich - die Betreuer befragt werden. Die Einschätzungen sollen immer auf der gegenwärtigen

Situation beruhen, nicht auf zukünftigen Plänen.

Im folgenden werden zwei Beispiele für Items der FDCRS wiedergegeben, das erste Item stammt aus dem Bereich 1. Space and furnishings for care and learning, das zweite aus dem Bereich 5. Social development:

### 3. Child-related display

1

- No child-related pictures, mobiles or children's artwork put up for children to look at.

3

- Some children's artwork displayed.

- Some store-bought or adult-made pictures put up especially for children to look at (Ex. nursery rhymes, ABC's, numbers, holidays).

5

- Much children's work displayed: at least two items per child enrolled.

- Some children's work down low on child's eye level.

7

- Many items of interest to children on child's eye level or where children are held up to see (Ex. in feeding area, near cribs, crawling and play areas).

- Display changed at least monthly to match children's activities and interest.

### 27. Tone

1

- Caregiver and children seem tense, voices sound angry, children cry often.

- Physical contact used mainly for control (Ex. hurrying children along, punishing).

3

- Physical contact mainly for routine care of children, little display of affection.

- Caregiver does not often smile at, talk to, or listen to children.

- May have favorite child who get most of the attention.

5

- Caregiver uses physical contact to show affection to all children (Ex. gentle holding, hugging, pat on the head).

- Caregiver and children seem relaxed, voices cheerful, a lot of smiling.

7

- Caregiver and children show respect for and kindness to one another.

- Caregiver praises children for being kind and helpful to others.

Die hier verwendeten siebenstufigen Ratingskalen stellen relativ komplexe Instrumente dar. Der Beobachter muß eine Vielzahl von Einzelinformationen verarbeiten und in eine zusammenfassende Einschätzung integrieren. Damit werden hohe Anforderungen an seine Verarbeitungskapazitäten gestellt. Die Items verlangen damit ein höheres Abstraktionsniveau.

Die Ergebnisse zu den 32 Items können additiv zu einem Gesamtwert der Qualität einer Tagespflegestelle zusammengefaßt werden. Indem ein Durchschnitt der Einschätzungen gebildet wird, kann die ursprüngliche siebenstufige Skala zur Interpretation der Qualität herangezogen werden. Eine Zusammenfassung ist ebenfalls auf der Ebene der sechs Subskalen möglich.

### *Gütekriterien*

Die FDCRS wurde in verschiedenen Untersuchungen zur Tagespflege in den USA und in Kanada eingesetzt, so daß Informationen zu den Gütekriterien aus verschiedenen Quellen vorliegen. In einigen der Studien wurde eine Vorform der FDCRS, die DCHERS (Day Care Home Environment Rating Scale, Harms, Clifford & Padan-Belkin, 1982), benutzt. Im folgenden kann nicht über alle Studien berichtet werden, vielmehr muß sich auf eine Auswahl beschränkt werden (vgl. auch Clifford et al., in Druck).

Harms & Clifford (1989) berichten für eine Vorform eine Interrater-Reliabilität von .86 ( $n = 19$ ; jeweils zwei Beobachter); einige Items hatten jedoch niedrigere Übereinstimmungen und wurden anschließend revidiert. Für die revidierte Form ergab sich in einer Studie von Jones & Meisel (1987) eine durchschnittliche Beobachterübereinstimmung von 80 % (zwei Beobachter, nur vier Tagespflegestellen). Ähnlich hohe Übereinstimmungen bei jeweils zwei Beobachtern berichten Nelson (1989) für 32 Tagespflegestellen (85 % Beobachterübereinstimmung) und Howes & Stewart (1987) für 55 Tagespflegestellen (über 90 %). Pence & Goelman (1991) verweisen ohne weitere Angaben auf drei unabhängige Tests, die Interrater-Reliabilitäten von .89, .79 und .84 ergaben. Insgesamt kann damit von einer hohen Objektivität der FDCRS ausgegangen werden.

Howes & Stewart (1987) haben an der Stichprobe von 55 Tagespflegestellen die interne Konsistenz (Cronbach's Alpha) der FDCRS auf der Ebene der Subskalen überprüft. Die Alpha-Werte reichen von .70 (Subskala 6. Adult needs) bis .93 (Subskala 4. Learning activities). Clifford et al., (in Druck) und Pence & Goelman (1991) weisen auf weitere Studien hin, die relativ hohe Alpha-Werte ergaben (über .90 für den Gesamtwert). Der FDCRS kann somit - im Hinblick auf die interne Konsistenz - eine gute Reliabilität zugesprochen werden. Informationen zur Test-Retest-Reliabilität liegen dem Berichtersteller nicht vor.

Die Autoren der FDCRS sehen eine inhaltliche Validität des Instruments zum einen dadurch gesichert, daß alle Funktionsbereiche der CDA Competency Standards (Council for Early Childhood Professional Recognition, 1991a) auch durch Items der FDCRS angesprochen werden (vgl. Harms & Clifford, 1989). Zum anderen sei die

FDCRS eine Adaptation der ECERS für Tagespflege. Die der ECERS über eine Expertenbefragung zugesprochene inhaltliche Validität (vgl. Kap. 2.1.5) würde damit indirekt auch auf eine inhaltliche Validität der FDCRS verweisen (vgl. Clifford et al., in Druck).

Verschiedene Studien weisen darauf hin, daß die über die FDCRS gemessene Qualität mit anderen Messung der Qualität einer Tagespflegestelle positiv korreliert. Nelson (1989) berichtet bei 32 Tagespflegestellen Korrelationen der FDCRS von .82 mit dem Family Day Home Observation Instrument (Child Care Partnership of Dallas, 1987; vgl. Kap. 2.2.3) und von .62 mit dem Assessment Profile for Family Day Care (Sibley & Abbott-Shim, 1991; vgl. Kap. 2.2.1). Pence & Goelman (1991) haben in 46 Tagespflegestellen sowohl die FDCRS als auch HOME (Home Observation for Measurement of the Environment; Caldwell & Bradley, vgl. Kap. 2.3.1) eingesetzt. Die moderate Korrelation von .34 nahmen sie zum Anlaß, aus beiden Meßwerten einen kombinierten Qualitätswert zu bilden (vgl. weiter unten). In einer Studie mit 161 lizenzierten Tagespflegestellen haben Stuart & Pepper (1988) die Qualität gemessen über die DCHERS (Vorform der FDCRS) verglichen mit einer globalen Qualitätseinschätzung der Tagespflegestellen (Einschätzung der Tagespflegestelle im Vergleich zu allen anderen, dem Rater bekannten Stellen auf einer 100 mm langen graphischen Skala; ein Wert von 80 bedeutet z.B., daß die eingeschätzte Tagespflegestelle eine bessere Qualität aufweist als 80 % der anderen Tagespflegestellen). Beide Qualitätseinschätzungen korrelieren mit .64. Allerdings wurden beide Einschätzungen durch denselben Beobachter erhoben, so daß diese Korrelation überhöht sein kann. Insgesamt jedoch kann von einer kriteriumsorientierten Validität der FDCRS ausgegangen werden.

Wie bereits erwähnt, wurde und wird die FDCRS in verschiedenen Studien in den USA und Kanada eingesetzt. Im Sinne einer Konstruktvalidität zeigen sich dabei vielfältige und inhaltlich interpretierbare Beziehungen zu verschiedenen anderen Konstrukten. An dieser Stelle kann nur auf einige dieser Studien eingegangen werden. Howes & Stuart (1987) untersuchten 55 Kinder (Durchschnittsalter 20 Monate), die in 55 verschiedenen Tagespflegestellen betreut wurden. Zur Konstruktion eines Qualitätsindikators für die Tagespflegestellen wurde eine Hauptkomponentenanalyse gerechnet, in die neben den sechs Subskalen der FDCRS drei weitere Variablen eingingen: Anzahl der betreuten Kinder, Anzahl der betreuenden Erwachsenen sowie zahlenmäßiges Betreuer-Kind-Verhältnis. Die erste Hauptkomponente erklärt 68 % der gesamten Varianz. Deshalb wurde nur eine additive Skala gebildet, wobei die Einzelvariablen mit ihren Ladungen auf der ersten Hauptkomponente gewichtet wurden. Dieser Qualitätsindikator korreliert signifikant mit zwei komplexen, faktorenanalytisch gebildeten Indikatoren zu familialen Merkmalen ("Nurturing and supported":  $r = .42$  für Mädchen und  $.36$  für Jungen; "Restricted

and stressed":  $r = -.41$  für Mädchen und  $-.54$  für Jungen; Alter des Kindes jeweils auspartialisiert): Eltern, die von anderen viel soziale Unterstützung erhalten und die eher "hegende" (nurturance) Erziehungseinstellungen haben, wählen für ihre Kinder qualitativ bessere Tagespflegestellen aus. Eltern, die gestresst sind und die mehr restriktive Einstellungen zur Kindererziehung haben, wählen eher Tagespflegestellen von niedrigerer Qualität aus. Werden diese beiden Familienindikatoren kontrolliert, so ist eine höhere Qualität der Tagespflegestelle verbunden mit einem höheren Niveau des kindlichen Spiels mit Objekten, Peers und Erwachsenen (jeweils komplexe Ratings basierend auf Beobachtungen; die Korrelationen sind mit Ausnahme der Beziehung zum Spiel mit Peers bei Jungen jeweils signifikant und reichen von .32 bis .61).

Die FDCRS - genauer die Vorform DCHERS - wurde auch im Victoria Day Care Research Project in Kanada eingesetzt (vgl. Goelman & Pence, 1987, 1988). Neben Gruppenbetreuungen (vgl. Kap. 2.1.5) wurden hier 24 lizenzierte und 25 unlizenzierte Tagespflegestellen untersucht. Lizenzierte Tagespflegestellen weisen dabei eine höhere Qualität auf als die nicht lizenzierten (Mittelwerte im Gesamtwert der DCHERS 3,4 versus 2,8). Wie bei der ECERS zeigen sich bezogen auf kindliche Aktivitäten (gemessen über die Child Observation Form) nur Beziehungen zu "information activities", d.h. zu Aktivitäten, die nicht zwingend im Umkreis von Instruktionen auftreten, in denen aber spezifische Informationen von den Erzieherinnen an die Kinder gegeben werden: Je höher die über die DCHERS erfaßte Qualität, desto umfangreicher solche Aktivitäten (Korrelation mit dem DCHERS-Gesamtwert .44, mit den Subskalen Language development .46, Learning .36 und Social development .41). Weiterhin bestehen positive Korrelationen der DCHERS mit Maßen des kindlichen Entwicklungsstandes: Der DCHERS-Gesamtwert korreliert z.B. mit .33 mit dem Peabody Picture Vocabulary Test (.51 für die Subskala Social development) und mit .32 mit dem Expressive One-Word Picture Vocabulary Test (.33 mit der Subskala Learning und .48 mit der Subskala social development).<sup>36</sup> Ähnliche Ergebnisse zeigen sich auch, wenn man anhand der DCHERS die Tagespflegestellen in solche von niedrigerer und solche von höherer Qualität aufteilt und dann diese beiden Gruppen in verschiedenen Variablen vergleicht.

In einer anderen Studie in Kanada, dem Vancouver Family Day Care Research Project, wurden 46 Tagespflegestellen untersucht (vgl. Pence & Goelman, 1991). Dabei wurde neben der DCHERS auch die HOME (Home Observation for Measurement of the Environment) eingesetzt. Aus beiden Werten wurde ein kombinierter Qualitätswert gebildet; entsprechend dieser Werte wurde die Stichprobe am Median in eine Gruppe niedrigerer Qualität und eine Gruppe höherer Qualität eingeteilt. Bei

---

<sup>36</sup> Allerdings wurde hier vermutlich der familiäre Hintergrund der Kinder nicht kontrolliert.

dem Vergleich der beiden Gruppen mit einer Reihe von anderen Variablen ergaben sich verschiedene signifikante Beziehungen: Tagespflegestellen mit einer höheren Qualität sind eher lizenziert und haben Betreuerinnen, die Fortbildungsveranstaltungen zur Tagespflege besucht haben, die ihre Tätigkeit mehr als eine professionelle auffassen, die mehr Kinder und diese länger betreuen, die mehr Kontakte zu den Eltern der betreuten Kinder haben und die eher auf eine Unterstützung durch eine Tagespflegeorganisation zurückgreifen können. Zudem berichten die Tagespflegebetreuerinnen mit höherer Qualität über eine bessere Ausstattung und über mehr die Entwicklung der Kinder stimulierende Aktivitäten. Nur im Hinblick auf kustodiale Tätigkeiten (Ruhezeiten, Mahlzeiten) zeigen sich keine signifikanten Unterschiede. Die Autoren heben diese Übereinstimmung zwischen Selbstbericht der Betreuerinnen und Fremdeinschätzung besonders hervor.

Ebenfalls in einer Studie in Kanada haben Stuart & Pepper (1988) die über die DCHERS gemessene Qualität von 161 lizenzierten Tagespflegestellen verglichen mit Persönlichkeitsvariablen (gemessen über die Personality Research Form und den Jackson Vocational Interest Survey) und weiteren Hintergrundvariablen der Betreuerinnen. Der Gesamtwert der DCHERS korreliert dabei mit einer Reihe von Persönlichkeitsvariablen: Eine höhere Qualität zeigt sich bei dominanteren bzw. sich mehr behauptenden Persönlichkeiten, bei mehr forschenden, intellektuell neugierigen Persönlichkeiten sowie bei solchen, die weniger konventionell eingestellt sind und die weniger interessiert sind in traditionellen weiblichen Aktivitäten (die Korrelationen sind niedrig bis moderat, sie liegen um .15 bis .25). Allerdings ist dabei zu berücksichtigen, daß die gesamte Stichprobe der untersuchten Tagespflegebetreuerinnen deutlich stärker als eine Vergleichsstichprobe von Collegestudentinnen durch warme, hegende Persönlichkeitsmerkmale sowie durch die Eigenschaften Regelmäßigkeit, Vorsicht und Stabilität charakterisiert sind. Eine gute Qualität der Tagespflegestelle ist zudem mit mehr betreuten Kindern verbunden ( $r = .20$ ). Die Autoren erklären dies aber dadurch, daß in dieser Stichprobe eine Tagespflegeorganisation die Kinder den einzelnen Tagespflegestellen zugewiesen hat und diese - innerhalb des Maximums von fünf Kindern - vermutlich mehr Kinder in solche Stellen schickte, von denen sie eine gute Qualität annahm. Schließlich geht eine höhere Qualität einher mit höherer formaler Schulbildung, einer gewissen Ausbildung in Betreuungsfragen (z.B. Besuch entsprechender Collegekurse) sowie einem speziellen Training in Fragen der Tagespflegebetreuung.

Schließlich haben Jones & Meisels (1987) die Effekte eines auf drei Jahre angelegten Trainingsprogramms für Tagespflegebetreuerinnen in Fragen der Betreuung behinderter Kinder mit Hilfe der DCHERS evaluiert ( $n = 13$  Tagespflegebetreuerinnen). In einem sechsmonatigen Zeitraum (zu Beginn des zweiten Trainingsjahres) zeigte sich eine signifikante Verbesserung ( $p = .01$ ) der FDCRS-Werte von 3,9 auf 4,5.



Alle berichteten Studien weisen daraufhin, daß mit der FDCRS ein objektives, reliables und valides Meßinstrument zur Erfassung der Qualität einer Tagespflegestelle vorliegt.

### **2.2.3 Family Day Home Observation Instrument**

#### **Entstehung und Anwendungskontext**

Das Family Day Home Observation Instrument wird im Rahmen eines freiwilligen Akkreditierungsprogramms in Dallas, Texas, eingesetzt (Child Care Partnership of Dallas, 1987; Mindel, 1988).

Dieses Akkreditierungsprogramm ist auf Dallas, Texas, beschränkt (vgl. Modigliani, 1991). Die Tagespflegebetreuerinnen müssen mindestens 24 Stunden eines insgesamt 36stündigen Fortbildungsprogramms von Child Care Partnership of Dallas besuchen. Sie erhalten mindestens einen Besuch von einem Berater sowie einen von einem Mitarbeiter von Child Care Partnership of Dallas. In diesem Rahmen wird das Family Day Home Observation Instrument von einem trainierten Beobachter eingesetzt, wobei für eine Akkreditierung eine bestimmte Punktzahl (vgl. weiter unten) erreicht werden muß. Eine Kommission entscheidet anschließend über die Akkreditierung. Eine Akkreditierung muß jährlich erneuert werden. Im Vergleich zu dem groß angelegten Akkreditierungsprogramm von NAFDC (vgl. Kap 2.2.1) oder dem CDA National Credentialing Program (vgl. Kap.2.1.3) handelt es sich hier um ein kleiner dimensioniertes, lokal begrenztes Programm.

Bei der Entwicklung des Family Day Home Observation Instruments wurden Materialien von anderen Organisationen gesichtet. Im engeren wurde das Instrument entworfen, um den Grad der Übereinstimmung der Merkmale einer Tagespflegestelle mit den CDA Competency Standards für Tagespflegestellen (Council for Early Childhood Professional Recognition, 1991a) zu messen. Dementsprechend orientieren sich die Items sowie ihre Gliederung (vgl. weiter unten) an den 13 CDA-Funktionsbereichen, wobei diese aber nicht direkt umgesetzt werden.

Das Instrument kann bei Tagespflegestellen für Kinder verschiedenen Alters eingesetzt werden. Eine genaue Altersangabe wird nicht gemacht, wohl aber unterscheiden einige Items nach Infants, Toddlers, Preschoolers und Schulkinder. Das Instrument wird von einem trainierten Beobachter angewandt. Dem Berichtersteller liegen aber keine Informationen über das Training vor.

#### **Aufbau des Instruments**

Neben einer Liste, mit der die Anzahl und das Alter der anwesenden Kinder erfaßt wird, besteht der Hauptteil des Family Day Home Observation Instruments aus 116 Items, die in acht Bereiche gegliedert sind. Ein gewisser Schwerpunkt liegt auf Aspekten der Sicherheit, Gesundheitsvorsorge und Ernährung. Jeder Bereich wird

durch eine kurze Beschreibung dessen, was Qualität in diesem Bereich meint, eingeleitet:

- A. Family/home child care as a business, 10 Items
- B. Working with parents, 7 Items
- C. Learning environment, 14 Items
- D. Importance of safety, health, nutrition, 47 Items
- E. Activities to enhance children's development: physical, 7 Items
- F. Activities to enhance children's development: self-social, 11 Items
- G. Activities to enhance children's development: language-thinking, 17 Items
- H. Activities to enhance children's development: creative, 3 Items

Die Items sind als kurze Statements formuliert, die mit "ja" oder "nein" beantwortet werden müssen. Darüber hinaus gibt es noch die Antwortmöglichkeit "NA" (not applicable), falls ein Statement logischerweise nicht mit "ja" oder "nein" beantwortet werden kann (z.B. wenn Fragen sich auf Kinder einer bestimmten Altersgruppe beziehen, diese aber in der Tagespflegestelle gar nicht betreut werden). Bei einigen Items stehen noch Checklisten von Indikatoren oder Subaspekten, die vorhanden sein sollten, um ein Item mit "ja" einzustufen. Bei einigen Items ist exakt vermerkt, wie viele solcher Subaspekte vorhanden sein müssen, damit ein "ja" vergeben werden darf. Die Items müssen in der Regel aufgrund von Beobachtungen beantwortet werden. Einige Items verlangen auch die Befragung der Tagespflegebetreuerin oder die Prüfung von schriftlichen Unterlagen; dies ist jeweils bei den betreffenden Items vermerkt. Die Reihenfolge, in der die Items im Instrument stehen, dürfte wohl nicht mit der Reihenfolge übereinstimmen, in der die Items bearbeitet werden können. Im Instrument wird eine Beobachtungsdauer von zwei Stunden genannt, woran sich dann die Befragung der Betreuerin sowie die Durchsicht von Dokumenten anschließt. Nelson (1989) nennt einen Zeitaufwand von insgesamt vier bis sechs Stunden.

Im folgenden vier Itembeispiele; das erste stammt aus dem Bereich B. Working with parents families und beinhaltet eine umfangreiche Checkliste, die letzten drei aus dem Bereich F. Activities to enhance children's development: self social:

B.1. Has written policies discussed and given to parents before child starts care, including; (Interview, check records) (8 / 12 required for yes)

- hours of care
- fees and payment schedule, refunds
- vacations/holidays
- philosophy, goals, objectives
- guidance methods used
- emergency procedures
- authorization for pickup of children
- policy regarding care of sick children
- permission form for trips or activities away from home
- substitute plan
- medication permission slips
- disclaimer about child's possible injury

F.3. Helps toddlers resolve conflicts by talking.

F.4. Plays and smiles with all children during routine times (feeding, bathing, dressing).

F.7. Helps children deal with anger, sadness and frustration.

Die Items des Family Day Home Observation Instrument sind von unterschiedlicher Natur. Einige Items fragen diskrete Einzelaspekte ab (wie das erste Beispiel) und lassen sich somit dem Abstraktionsniveau 1 (niedriges Abstraktionsniveau) zuordnen, andere verlangen vom Beobachter eine gewisse Generalisierung von beobachteten Einzelaspekten und sind damit auf dem Abstraktionsniveau 2 (mittleres Abstraktionsniveau) angesiedelt. Mithin werden zwar unterschiedliche Anforderungen an die Informationsverarbeitungskapazitäten der Beobachter gestellt, insgesamt aber dürften die Ratings der Beobachter im Vergleich zu anderen Instrumenten relativ wenige Abstraktionen von Einzelaspekten verlangen.

Für die Akkreditierungsentscheidung - und damit für die Kennzeichnung einer guten Qualität der Tagespflegestelle - sind die acht Bereiche nicht alle von gleicher Bedeutung. Voraussetzung für eine Akkreditierung ist, daß

- 90 % der Items aus dem Bereich D (Importance of safety, health, nutrition),
- 80 % der Items aus den Bereichen E bis H (Activities to enhance children's development, wobei diese vier Bereiche zusammen betrachtet werden) und
- 50 % der Items aus den Bereichen A (Family/home child care as a business), B (Working with parents) und C (Learning Environment)

positiv beantwortet werden.

### **Gütekriterien**

Zu den Gütekriterien liegen nur die Ergebnisse einer kleinen Studie vor (Mindel, 1988). Eine erste Version des Instruments wurde an 24 Tagespflegestellen erprobt. Nach Revisionen an dem Instrument wurde die neue Version bei neun von diesen Tagespflegestellen erneut angewandt. Als Maß für die Objektivität (Interrater-Reliabilität) wurde die durchschnittliche Übereinstimmung von zwei Beobachtern berechnet, die mit 97,5 % sehr hoch ausfällt. Hierbei ist aber zu berücksichtigen, daß nur neun Tagespflegestellen bewertet wurden.

Inhaltliche Validität wird dem Instrument aufgrund seiner Nähe zu den CDA Competency Standards zugesprochen. Bei den neun Tagespflegestellen wurde auch - im Sinne einer kriteriumsbezogenen Validität - die Korrelation zwischen den subjektiven Einschätzungen der Tagespflegestellen durch Experten (fünfstufige Skala von "5 = die Tagespflegestelle wird als exzellent beurteilt" bis "1 = die Tagespflegestelle wird als sehr negativ beurteilt") und den Gesamtwerten anhand des Family Day Home Observation Instruments berechnet.<sup>37</sup> Für die Berechnung des Gesamtwertes in dem Family Day Home Observation Instrument wurde für jeden der acht Bereiche der Anteil der mit "ja" beantworteten Items bezogen auf alle Items dieses Bereichs berechnet und daraus anschließend der einfache Durchschnitt gebildet. Die acht Bereiche gehen damit gleichgewichtet in den Gesamtwert ein (obwohl die Itemzahl pro Bereich von 3 bis 47 Items variiert). Die Korrelation zwischen den Werten des Family Day Home Observation Instrument und den subjektiven Urteilen der Experten beträgt .68 (n = 9 Tagespflegestellen). Diese Korrelation wird als befriedigend betrachtet, da zwischen der Einschätzung der Experten und der Anwendung des Family Day Home Observation Instrument einige Zeit lag und somit Veränderungen in den Tagespflegestellen sich vermindern auf die Korrelation ausgewirkt hätten.

Nelson (1989) fand eine Korrelation von .82 zwischen den Gesamtwerten von 32 Tagespflegestellen in der FDCRS (Family Day Care Rating Scale, vgl. Kap. 2.2.2) und dem Family Day Home Observation Instrument. Beide Instrumente decken somit einen vergleichbaren Satz von Qualitätsaspekten von Tagespflegestellen ab. Insgesamt aber bedarf es weiterer Studien, die sich der Reliabilität und Validität des Family Day Home Observation Instruments zuwenden.

---

<sup>37</sup> Unklar ist, ob die gleichen Experten sowohl das subjektive Urteil abgaben als auch das Family Day Home Observation Instrument anwendeten. Für die Überprüfung der Validität wäre es wünschenswert, wenn die beiden Qualitätseinschätzungen von verschiedenen Personen durchgeführt würden.

### **2.2.4 Evaluating Home Based Day Care** **Entstehung und Anwendungskontext**

Das Instrument "Evaluating Home Based Day Care" wurde in Pittsburg, Pennsylvania, vom Louise Child Care Center entwickelt (Louise Child Care Center, 1988; Rodriguez 1978, 1982); abgekürzt wird auch von dem "Louise Instrument" gesprochen.<sup>38</sup> Das Instrument soll denen helfen, die eine systematische Beobachtung und Evaluation von Tagespflegestellen durchführen. Ziel ist es, zu einer möglichst objektiven Einschätzung der Qualität einer Tagespflegestelle zu kommen, auch um die Beobachter von "the burden of purely subjective evaluation" (Louise Child Care Center, 1988, S. 1) zu befreien. Das Louise Instrument wurde für eine Fremdbeobachtung einer Tagespflegestelle durch einen trainierten Beobachter (z.B. ein Supervisor) entwickelt, allerdings dürfte es auch aufgrund seiner einfachen Struktur für eine Selbstevaluation durch die Tagespflegebetreuerin oder für eine Evaluation durch Eltern, die eine qualitativ gute Tagespflegestelle für ihr Kind suchen, geeignet sein. Angaben zu einem Training liegen dem Berichtersteller nicht vor.

Als theoretischer Hintergrund des Louise Instruments wird auf die CDA Competency Standards (Council for Early Childhood Professional Recognition, 1991a) verwiesen. Die Items wurden mit Bezug auf diese CDA Standards ausgewählt und entsprechend der 13 CDA Funktionsbereiche gruppiert. Dabei wurden allerdings einige der CDA Funktionsbereiche zusammengefaßt, so daß die Items in nur neun Bereiche gruppiert werden (vgl. weiter unten). Aufgrund dieses Bezugs auf die CDA Competency Standards wird dem Instrument inhaltliche Validität zugesprochen (vgl. z.B. Clifford et al., in Druck). Aus diesem Kontext kann auch erschlossen werden, daß das Louise Instrument für Tagespflegestellen für Kinder im Alter von der Geburt bis zum Schuleintritt angemessen ist.

#### **Aufbau des Instruments**

Das Louise Instrument besteht aus 206 Items, die im Sinne von Checklisten aufgebaut sind. Die Items sind in neun Bereiche gruppiert:

- Safe environment, 11 Items
- Health and nutrition, 37 Items
- Learning environment, 24 Items

---

<sup>38</sup> Das Louise Instrument wird im Kontext der strikten Gesundheits- und Sicherheitsregularien in Pennsylvania eingesetzt. Aus diesem Grund enthält es keine Basisstandards in diesen Bereichen (vgl. Clifford et al., in Druck). Gegenwärtig wird das Instrument einer Revision unterzogen, um auch einen Einsatz in anderen US-Staaten zu ermöglichen (vgl. Modigliani, 1991, S. 4).

- Curriculum and activities, 23 Items
- Social and emotional, 51 Items
- Guidance and discipline, 29 Items
- Families, 10 Items
- Program Management, 11 Items
- Professionalism, 10 Items

In der Regel sind die Items unter allgemeineren Qualitätskriterien zusammengefaßt, als deren Operationalisierungen sie betrachtet werden können. Pro Kriterium gibt es unterschiedlich viele Items. Beurteilt werden aber nicht die Kriterien selbst, sondern nur die Einzelaspekte. Eine Beurteilung auf der Ebene der allgemeineren Kriterien ist dann auf analytischem Wege möglich. Für die Items gibt es die Antwortmöglichkeiten "ja" (Aspekt ist vorhanden), "nein" (Aspekt ist nicht vorhanden), "n.o." (nicht beobachtet) und "n.a." (trifft nicht zu), wobei die letzte Kategorie in der Regel dann benutzt werden soll, wenn ein Item aufgrund des Alters der betreuten Kinder nicht zutrifft. Vermutlich dürfte auch bei dem Louise Instrument die Reihenfolge der Items im Instrument nicht mit der Reihenfolge übereinstimmen, mit der die Items bei der Beobachtung abgearbeitet werden.

Desweiteren sind die Items innerhalb von jedem der neun Bereiche drei Qualitätsniveaus zugeordnet:

- **Basic care:** Darunter wird eine Betreuung verstanden, die hauptsächlich nur die körperlichen Bedürfnisse der betreuten Kinder befriedigt. Dieses Niveau wird als Minimalstandard betrachtet.
- **Intuitively good care:** Dieses Qualitätsniveau wird als durchschnittlicher Betreuungsstandard betrachtet. Die Betreuung erfolgt dabei in etwa analog der Betreuung, wie sie ein Kind auch zu Hause durch die Eltern erfahren würde.
- **Informed care:** Dieses Niveau entspricht einer "sophisticated" Betreuungsqualität, die sich dann ergibt, wenn eine Tagespflegebetreuerin mit einer "intuitively good care" ein spezielles Training in Betreuungsfragen erhält.

Die drei Ebenen sind hierarchisch angeordnet, d.h., eine höhere Ebene setzt immer voraus, daß alle Items der darunter liegenden Ebenen erfüllt sind. Diese drei Qualitätsniveaus sind vor allem dann von Bedeutung, wenn Qualitätsbeurteilungen auf der Ebene der Einzelitems durchgeführt werden. In der Regel sollen die Items aufgrund von (objektiven) Beobachtungen eingeschätzt werden, so daß nur ein Minimum von subjektiver Beurteilung durch den Beobachter erforderlich ist. Im Instrument finden sich keine Angaben, inwieweit auch auf Befragungen der Betreuerinnen zurückge-

griffen werden kann. Dem Berichterstatter liegen keine Informationen darüber vor, wieviel Zeit für die Anwendung des Louise Instrument benötigt wird.

Im folgenden werden als Beispiele für den Aufbau des Louise Instruments jeweils alle Items genannt, die unter zwei der allgemeineren Qualitätskriterien stehen. Das erste Beispiel stammt aus dem Bereich "Safe environment" und ist hier dem Niveau "Basic care" zugeordnet; das zweite Beispiel stammt aus dem Bereich "Social and emotional" und ist der Ebene "Informed care" zugeordnet:

*Can provide a safe environment without being overly restrictive*

- 1) always knows where the children are
- 2) is always within hearing distance of the children
- 3) prevents children from hurting one another
- 4) prevents situations which are physically dangerous

*Can cope with children's aggression*

- 16) allows time for children to solve their own conflicts and only intercedes when they cannot
- 17) limits the aggression but avoids causing the child to lose face
- 18) tries to understand the cause of the child's aggression
- 19) tries to interpret to the child the causes of his anger
- 20) expresses her understanding of his emotion
- 21) makes a distinction between angry feelings and aggressive acting out
- 22) tries to get the children to substitute words for blows
- 23) interpretes one child's anger to another

Wie die Beispiele zeigen, beziehen sich die einzuschätzenden Items auf diskrete Einzelaspekte. Der Charakter des Louise Instruments als Sammlung von Checklisten von Einzelaspekten stellt - relativ zu den anderen Instrumenten - nur geringe Anforderungen an die Informationsverarbeitungskapazitäten der Beobachter. Insofern lassen sich die Items dieses Instruments eher einem niedrigeren Abstraktionsniveau zuordnen. Dieses Instrument dürfte relativ einfach anzuwenden sein.

Die Berechnung eines Gesamtwertes des Louise Instruments für eine Tagespflegestelle - das gleiche Vorgehen kann prinzipiell auch für die einzelnen Bereiche durchgeführt werden - geschieht, indem die Anzahl der "nein" Antworten ausgezählt und diese Summe von 206 (Gesamtzahl aller Items) abgezogen wird.<sup>39</sup> Entsprechend des

---

<sup>39</sup> Dieses "indirekte" Vorgehen wurde gewählt, um Tagespflegestellen nicht dafür zu "bestrafen", daß einzelne Items nicht beantwortet wurden. Clifford et al. (in Druck) weisen darauf hin, daß in diesem Falle eine Tagespflegestelle bevorzugt würde, bei der mehrere Items nicht be-

Gesamtwertes werden drei Qualitätsniveaus unterschieden:

- 0 - 43 Punkte: basic care
- 44 - 105 Punkte: intuitive care
- 106 - 206 Punkte: informed care.

Diese drei Qualitätsniveaus sind aber nicht zu verwechseln mit den drei weiter oben beschrieben, gleich benannten Niveaus, in die die Einzelitems im Instrument eingeteilt sind. Denn für die Gesamtbeurteilung spielt es keine Rolle, ob z.B. ein Wert von 39 durch Items zustande kommt, die dem Niveau "basic care" oder die dem Niveau "intuitive care" zugeordnet sind.

### *Gütekriterien*

Zu der beschriebenen Version des Louise Instruments liegen keine Daten zur Reliabilität und Validität vor (vgl. Clifford et al., in Druck). Aufgrund der Orientierung an den CDA Competency Standards wird dem Instrument inhaltliche Validität zugesprochen. Reliabilität und Validität wurden an einer früheren Version von 1976 überprüft (Rodriguez, 1978, 1982). Dem Berichtersteller liegen aber keine Informationen darüber vor, inwieweit die Gütekriterien der früheren Version auf die vorliegende Version übertragen werden können. Nach Rodriguez (1982) bestand die ältere Version aus 68 evaluativen Kriterien; Angaben zum Antwortmodus fehlen aber. Insofern scheint die Übertragbarkeit der Daten zu den Gütekriterien zumindest problematisch.

Zur Überprüfung der Interrater-Reliabilität führten zwei trainierte Beobachter (fünf Stunden Training) unabhängig voneinander Beobachtungen in sechs Tagespflegestellen durch. Dabei ergaben sich Koeffizienten von .74 bis 1.00, im Durchschnitt .88. Berechnet wurden hier offensichtlich die Anteile übereinstimmender Einschätzungen. Zur Überprüfung der kriteriumsbezogenen Validität haben zunächst neun Experten 69 ihnen sehr gut bekannte Tagespflegestellen nach subjektiven Kriterien als "below average", "average" oder "above average" eingeschätzt. Anschließend haben sie dann die frühere Version des Louise Instruments in den gleichen Tagespflegestellen angewandt. Die beiden Qualitätsbewertungen korrelieren auf der Ebene des Gesamtwertes des Louise Instruments mit .72. Ebenfalls wurden die Ergebnisse auf der Ebene der Einzelitems verglichen. Allerdings ist eine solch hohe Korrelation nicht überraschend, da die gleichen Personen sowohl das subjektive Rating abgaben als

---

antwortet wurden. Bezieht sich allerdings eine Nicht-Beantwortung darauf, daß Kinder einer bestimmten Altersgruppe nicht in der Tagespflegestelle betreut werden, dann dürfte dies für eine Beurteilung der Qualität nicht von Bedeutung sein.



auch das Louise Instrument anwendeten. Insofern muß dieses Validitätsmaß mit Vorsicht betrachtet werden.

## **2.3 Instrumente für häusliche Umwelten**

### **2.3.1 Home Observation for Measurement of the Environment**

#### **Entstehung und Anwendungskontext**

Die HOME (Home Observation for Measurement of the Environment) ist ein Verfahren für eine allgemeine Einschätzung des Stimulationsgehalts der häuslichen Umwelt (Caldwell & Bradley, 1984). Das Instrument wurde von B.M. Caldwell und R.H. Bradley entwickelt, mehrfach revidiert und liegt zur Zeit in drei Versionen vor: für Kinder unter drei Jahren (HOME Inventory for Families of Infants and Toddlers), für Kinder im Alter von drei bis sechs Jahren (HOME Inventory for Families of Preschoolers) und für Schulkinder (HOME Inventory for Families of Elementary Children). Als Begründung für die Auswahl der Items weisen die Autoren auf einige allgemeine Charakteristika einer die kindliche Entwicklung stimulierenden Umwelt hin, eine systematische und theoretisch orientierte Legitimierung der Items sowie der verschiedenen Subskalen wird jedoch nicht gegeben. Es wird nur daraufhin gewiesen, daß die Items aufgrund von Faktorenanalysen aus einem größeren Itempool ausgewählt wurden.<sup>40</sup>

HOME wird durch einen trainierten Beobachter während eines Besuchs in der Wohnung des Kindes angewandt. Voraussetzung ist, daß das Kind während des Besuches wach ist und mit der Mutter interagiert bzw. interagieren kann. Dabei ist mit "Mutter" nicht zwingend die leibliche (oder soziale) Mutter gemeint, sondern die Person, die zu Hause in der Regel das Kind betreut. Übergeordnetes Entscheidungskriterium ist dabei die Sicht des Kindes: "The intent of the assessment procedure is to get a picture of what the child's world is like from his or her perspective -- i.e., from where he or she lies or sits or stands or moves about and sees, hears, smells, feels, and tastes the world." (Caldwell & Bradley, 1984, S. 95) Daß dennoch durchgängig im Instrument von "Mutter" gesprochen wird, hängt damit zusammen, daß in der Regel im nordamerikanischen Kulturkreis die Mutter die häusliche Betreuung übernimmt. Für die Anwendung von HOME ist zwar ein Training erforderlich, aufgrund dessen, daß jeweils nur "ja - nein" Antworten gegeben werden müssen (vgl.

---

<sup>40</sup> An einigen Stellen erscheint die Auswahl und Zuordnung der Items zu Subskalen eher als willkürlich. Zudem enthält das Instrument Items, deren Qualitätsbezug zumindest im bundesrepublikanischen Raum fraglich ist. So z.B., wenn in der Version für Kinder im Kindergartenalter - in der Subskala "Modeling and encouragement of social maturity" - das Nicht-Vorhandensein eines Fernsehers (Item 39) zu einer negativen Beurteilung führt. Das Gleiche gilt für die häusliche Umwelt von alleinerziehenden und alleinlebenden Müttern oder Vätern, die automatisch im Hinblick auf den sozialen Kontext bei Mahlzeiten eine negative Bewertung erhalten (Item 50).

weiter unten), sei aber das Training nicht allzu umfangreich. Schon mit einem geringen Trainingsaufwand seien Beobachterübereinstimmungen von 90 % und mehr zu erreichen (vgl. Caldwell & Bradley, 1984, S. 11f.).

Im folgenden wird nur noch auf die Versionen für Kinder im vorschulischen Alter eingegangen.

### **Aufbau des Instruments**

Mit dem HOME werden zwar auch einige allgemeine Rahmendaten erhoben (z.B. Alter des Kindes, Haushaltszusammensetzung, Erwerbstätigkeit der Mutter), der Hauptteil besteht aber aus 45 Items in sechs Dimensionen bei der Version für Kinder unter drei Jahren (HOME Inventory for Families of Infants and Toddlers) bzw. aus 55 Items in acht Dimensionen bei der Version für Kinder im Kindergartenalter (HOME Inventory for Families of Preschoolers). Die Items verteilen sich auf die Dimensionen wie folgt:

- HOME Inventory for Families of Infants and Toddlers:<sup>41</sup>
  - I. Emotional and verbal responsivity of the mother, 11 Items
  - II. Avoidance of restriction and punishment, 8 Items
  - III. Organization of the physical and temporal environment, 6 Items<sup>42</sup>
  - IV. Provision of appropriate play materials, 9 Items
  - V. Maternal involvement with the child, 6 Items
  - VI. Opportunities for variety in daily stimulation, 5 Items
  
- HOME Inventory for Families of Preschoolers:
  - I. Stimulation through toys, games and reading materials, 11 Items
  - II. Language stimulation, 7 Items
  - III. Physical environment: safe, clean and conducive to development, 7 Items
  - IV. Pride, affection, and warmth, 7 Items
  - V. Stimulation of academic behavior, 5 Items
  - VI. Modeling and encouragement of social maturity, 5 Items

---

<sup>41</sup> In einem Antwortblatt wird durchgängig von "parents" gesprochen, bei der Beschreibung der Items aber von "mother". Auf dem Antwortblatt wird die zweite Dimension als "Acceptance of child's behavior" bezeichnet, in der Itembeschreibung als "Avoidance of restriction and punishment".

<sup>42</sup> Einige Items dieser Dimension beschreiben nicht die häusliche Umwelt selber, sondern eher "Verbindungen" mit anderen Bereichen: Kommt das Kind aus der Wohnung heraus und kann andere Erfahrungen machen; erhält das Kind regelmäßige medizinische Untersuchungen?

VII. Variety of stimulation, 9 Items

VIII. Physical punishment, 4 Items

Im Manual zu HOME werden die Items ausführlich beschrieben; in dem Antwortbogen sind sie jedoch nur als sehr kurze Statements formuliert. Die Statements müssen mit "+" (ja, der jeweils angesprochene Aspekt ist vorhanden) oder "-" (nein, der angesprochene Aspekt ist nicht vorhanden) beantwortet werden, eine andere Möglichkeit hat der Beobachter nicht. Die Autoren haben Wert darauf gelegt, daß so viele Items wie möglich aufgrund von Beobachtungen während des Besuchs bewertet werden können. Da aber nicht alles während der Besuchszeit beobachtet werden kann, müssen zusätzlich die Mütter befragt werden. Etwa ein Drittel der Items wird aufgrund von Befragungen der Mütter beurteilt. Es gibt keine festgelegte Reihenfolge der Beobachtungen oder der Fragen an die Mütter. Vielmehr soll der Beobachter ein Gespräch mit der Mütter initiieren, in dem dann die verschiedenen Aspekte angesprochen werden, so daß der Beobachter die für die Beantwortung der Items benötigten Informationen erhält. Die Anwendung des Instruments dauert insgesamt etwa eine Stunde. Alle Items müssen beantwortet werden, bevor die Wohnung verlassen wird.

Im folgenden sind zwei Itembeispiele aus der HOME Version für Kinder im Kindergartenalter wiedergegeben; in Klammern steht jeweils die ausführliche Itembeschreibung aus dem Manual. Das erste Beispiel stammt aus der Dimension III. Physical environment, das zweite aus dem Bereich V. Stimulation of academic behavior:

20. Outside play environment appears safe.

(20. *Child's outside play environment appears safe and free of hazards.* (No outside play area requires an automatic "no") Once again the interviewer should use good judgement on the scoring. Examples of typical hazards are: broken glass lying around, junk cars abandoned in the yard or along the side of the street, open ditches of a house so close to a street that a child could not safely play in the yard, and boards with nails sticking out of them.)

37. Child is encouraged to learn to read a few words.

(37. *Child is encouraged to learn to read a few words.* This is another area where ability increases with age. At the lowest level (age three) credit is given for making an attempt to teach the child to recognize her printed name. Another acceptable procedure is to use books to teach the child the association between "D" for dog and the association of a picture with a dog).

Die beiden Beispiele verdeutlichen, daß die Items zwar eher Einzelaspekte der häuslichen Umwelt ansprechen, andererseits aber auch hier der Beobachter ein seine

Beobachtungen zusammenfassendes Urteil abgegeben muß. Im Vergleich zu anderen Instrumenten beziehen sich seine Einschätzungen auf ein mittleres Abstraktionsniveau.

Die Ergebnisse des Gesamtinstruments wie auch in den einzelnen Dimensionen werden zusammengefaßt, indem jeweils die Summe der mit einem "+" (Aspekt ist vorhanden) eingeschätzten Items gebildet wird. Diese Summen können auf das Antwortblatt übertragen werden. Dort sind auch drei Qualitätsniveaus eingetragen ("lowest middle", "middle half", "upper fourth" bei der HOME-Version für Kinder unter drei Jahren; "lowest fourth", "middle half", "upper fourth" bei der Version für Kinder im Kindergartenalter), die Auskunft darüber geben, wie die erreichten Punktzahlen einzuschätzen sind.

### **Gütekriterien**

Caldwell & Bradley (1984, S. 13-86) berichten ausführlich über psychometrische Charakteristika der beiden HOME-Versionen. An dieser Stelle soll nur auf die Hauptmerkmale eingegangen werden. Die Ergebnisse basieren auf 174 Familien bei der HOME-Version für Kinder unter drei Jahren und 117 Familien bei der HOME-Version für Kinder im Kindergartenalter. Zum Teil sind die Fallzahlen bei einzelnen Analysen aber geringer. Zur Beobachterübereinstimmung stehen hier keine weiteren Ergebnisse. Es wird nur von den Autoren angemerkt (vgl. weiter oben), daß bereits nach einem kurzen Training Beobachterübereinstimmungen von 90 % und mehr erreicht werden könnten.

Bei einer Itemanalyse für die *HOME-Version für die Kinder unter drei Jahren* ergaben sich moderate Korrelationen zwischen den Items und der Gesamtskala sowie den Subskalen und den jeweils zu einer Subskala gehörigen Items. Die höchste Korrelation beträgt .80 (Item 38 mit der Subskala V. Maternal involvement with the child), die niedrigste .05 (Item 20 mit der Gesamtskala); negative Korrelationen gibt es keine. 35 der 45 Item-Subskala-Korrelationen sind größer/gleich .40. Einigen Items wird überwiegend zugestimmt. Die Autoren behalten angesichts des Charakters von HOME als Screeninginstrument diese Items aber dennoch bei, da ein "-" bei solchen Items u.U. sehr deutlich auf die Notwendigkeit von Interventionen hinweist. Die internen Konsistenzen (Kuder-Richardson formula 20) betragen .89 für die Gesamtskala und für die Subskalen .72 (Subskala I), .67 (II), .89 (III), .77 (IV), .69 (V) und .44 (VI). Für 91 Familien wurde HOME zu drei Zeitpunkten angewandt: als das Kind 6, 12 und 24 Monate alt war. Es ergeben sich moderate bis höhere Test-Retest-Korrelationen von .27 bis .77 bei den Subskalen und von .62 (6 zu 12 Monate) bis .77 (12 zu 24 Monate) bei dem Gesamtwert. Angesichts der zum Teil geringen Anzahl von Items pro Subskala und der Zeitabstände kann somit von relativ

stabilen Werten über die Zeit ausgegangen. Insgesamt weisen diese Ergebnisse auf eine gute Reliabilität der HOME hin.

Die sechs Subskalen sind nicht unabhängig voneinander, allerdings ergeben sich zum Teil nur moderate Skaleninterkorrelationen (.07 bis .68), so daß die einzelnen Subskalen doch unterschiedliche Aspekte der häuslichen Umwelt messen. Die Autoren weisen noch daraufhin, daß eine Faktorenanalyse fünf Faktoren mit Eigenwerten größer 1 ergeben hätte. Weitere Informationen zu dieser Faktorenanalyse fehlen allerdings.

Die HOME-Werte stehen in interpretierbaren Beziehungen zu einer Vielzahl von anderen Meßwerten: Es gibt in der Regel positive, aber moderate Korrelationen zwischen der Gesamtskala und den Subskalen einerseits und Indikatoren des sozialen Status andererseits (u.a. Bildungsstand Mutter und Vater, Beschäftigung von Mutter und Vater). Korrelationen über .50 sind aber selten, so daß die HOME-Werte zwar nicht unabhängig vom sozialen Status, gleichwohl aber nicht mit diesem identisch sind (n = 91 Familien). Durchgängig zeigen sich niedrige bis moderate positive Korrelationen zwischen HOME-Werten und Intelligenzmaßen (Bayley Infant Scales, Stanford-Binet Intelligenztest). Die HOME-Werte von Kinder im Alter von 6 Monaten korrelieren z.B. mit .50 mit den Intelligenzwerten (Stanford-Binet) der Kinder im Alter von 36 Monaten. HOME-Werte im Alter von 24 Monaten korrelieren mit .57 mit Intelligenzwerten (Stanford-Binet) im Alter von 54 Monaten. Ähnliche Korrelationen zeigen sich zwischen den HOME-Werten und der Sprachleistung (Illinois Test of Psycholinguistic Abilities; gemessen im Alter von 37 Monaten). Die HOME-Werte von Kindern im Alter von 6 Monaten korrelieren mit .39 mit den späteren Sprachleistungen, im Alter von 24 Monaten sogar mit .61 (jeweils HOME-Gesamtwert). HOME-Werte von Kinder im Alter von 6 Monaten sind auch dazu geeignet, zwischen Kindern zu unterscheiden, deren Intelligenz sich zwischen dem 6. und 36. Lebensmonat steigert (increaser), verringert (decreaser) oder deren Intelligenzniveau gleich bleibt (non-changer). Mit den sechs HOME Skalen konnten diskriminanzanalytisch 88 % der increaser, 68 % der non-changer und 69 % der decreaser richtig vorhergesagt werden. Caldwell & Bradley (1991, S. 56ff.) weisen noch auf anderen Untersuchungen hin, in denen HOME angewandt wurde und in denen sich eine gute Qualität von HOME erwiesen hätte. U.a. wurde überprüft, ob mit HOME Werten im Alter von 12 Monaten spätere Probleme in der Grundschule (u.a. Sitzenbleiben, Zuweisung zu einem Programm für Lernbehinderte) vorhergesagt werden können. Nur in 22 % der Fälle erwiesen sich die Voraussagen aufgrund der HOME Werte als falsch. Insgesamt weisen die verschiedenen Untersuchungen daraufhin, daß die HOME valide die Qualitätsaspekte der häuslichen Umwelt erfaßt.

Bei einer Itemanalyse für die *HOME-Version für Kinder im Kindergartenalter* (n =

117 Familien) zeigten sich zwar einige Probleme bei den Item-Total- und den Item-Subskala-Korrelationen,<sup>43</sup> insgesamt aber ergeben sich moderate positive Korrelationen. Die höchste Item-Subskala-Korrelation beträgt .76 (Item 19), die niedrigste .24 (Item 48), negative Korrelationen gibt es keine. 39 der 55 Item-Subskala-Korrelationen sind größer/gleich .40. Die internen Konsistenzen (Kuder-Richardson formula 20) betragen .93 für die Gesamtskala und für die Subskalen .88 (Subskala I), .65 (II), .83 (III), .75 (IV), .60 (V), .53 (VI), .69 (VII) und .59 (VIII). Für 33 Familien wurde HOME zu zwei Zeitpunkten angewandt: im Alter von 3 Jahren und im Alter von 4 1/2 Jahren des Kindes. Es ergeben sich niedrigere bis höhere Test-Retest-Korrelationen von .70 bei der Gesamtskala und von .05 (Skala VIII) bis .70 (Skala I) bei den Subskalen. Insgesamt kann damit von einer zufriedenstellenden Reliabilität dieser HOME-Version ausgegangen werden. Die acht Subskalen erfassen unterschiedliche Aspekte der häuslichen Umwelt. Die Skaleninterkorrelationen fallen eher niedrig aus (in der Mehrzahl unter .50). Es wird zwar auf eine Faktorenanalyse der Items bei einer früheren Version verwiesen, weitere Hinweise auf die Faktorenstruktur werden aber nicht gegeben.

Wie bei der HOME-Version für die jüngeren Kinder korrelieren auch die HOME-Werte für die älteren Kinder nur niedrig mit Indikatoren des sozialen Status (in der Regel liegen die Korrelationen unter .50). Beide HOME-Versionen messen somit zu einem erheblich Teil Aspekte der häuslichen Umwelt, die unabhängig vom sozialen Status sind. Die HOME-Werte der Kinder im Alter von 4 1/2 Jahren ( $n = 51$ ) korrelieren mit .58 mit ihrer gleichzeitig erfaßten Intelligenz (Stanford-Binet). Die Korrelation der HOME-Werte von 34 Kinder im Alter von fünf bis sechs Jahren mit ihrer Intelligenz gemessen im Alter von sechs bis zehn Jahren beträgt ebenfalls .58 (jeweils HOME-Gesamtwerte). Für 66 Kinder wurden ihre HOME-Werte im Alter von drei bis fünf Jahren mit Schulleistungstestwerten (SRA Achievement Test) im Alter von sechs bis zehn Jahren korreliert. Auch hier ergibt sich eine Korrelation von .58 zwischen den jeweiligen Gesamtwerten. Die verschiedenen Ergebnisse weisen auf eine valide Erfassung der Qualitätsaspekte der häuslichen Umwelten hin. Wie bereits weiter oben ausgeführt (vgl. 2.2.2) haben Pence & Goelman (1991) im Vancouver Family Day Care Research Project sowohl die HOME als auch die DCHERS (Vorform der FDCRS) eingesetzt (Korrelation der beiden Gesamtwerte .34;  $n = 46$  Tagespflegestellen). Aus beiden Qualitätsmaßen wurde dann ein kombiniertes Maß gebildet, das - im Sinne einer Konstruktvalidität - mit verschiedenen anderen Variablen in Beziehung steht.

---

<sup>43</sup> Bei verschiedenen Items sind die Item-Total-Korrelationen größer als die Korrelationen zwischen den Items und ihren Subskalen, was zumindest Fragen an die Faktorenstruktur dieser HOME-Version eröffnet.

Als Fazit läßt sich festhalten, daß beide HOME-Versionen als reliable und valide Instrumente zur Erfassung der Qualität häuslicher Umwelten angesehen werden können.

### *3. Zusammenfassung und Vergleich der Instrumente*

Im Hauptteil dieses Arbeitsberichts (Kapitel 2) wurde ein Überblick über in den USA entwickelte Screeninginstrumente zur Einschätzung der Qualität frühkindlicher Betreuungs- und Erziehungsformen gegeben. Die Intention für die Erstellung dieses Überblicks basiert auf der gegenwärtigen Diskussion in der Bundesrepublik, in der das Hauptaugenmerk auf Fragen der quantitativen Erweiterung frühkindlicher Betreuungs- und Erziehungsformen liegt, während Fragen der Sicherung der pädagogischen Qualität weniger Aufmerksamkeit gewidmet wird. Als eine Konsequenz dieser Situation kann der ausgeprägte Mangel an empirisch fundierten Instrumenten angesehen werden, mit denen in einer zeitökonomischen, aber gleichfalls forschungstechnischen Gütekriterien genügenden Weise die Qualität der pädagogischen Arbeit in frühkindlichen Betreuungs- und Erziehungsformen analysiert und bewertet werden kann. Dies gilt sowohl für Instrumente, die sich auf ausgewählte Aspekte der frühkindlichen Umwelten (z.B. räumlich-materiale Merkmale oder Merkmale der Interaktionsprozesse im engeren) beziehen, als auch für solche, die mehr darauf abzielen, einen (differenzierten) Gesamteindruck der Qualität solcher Umwelten zu liefern. Die letzte Gruppe von Instrumenten - die hier als Screeninginstrumente bezeichnet werden - bildet den Gegenstand des vorliegenden Arbeitsberichts.

Auch wenn im Vergleich zu Spezialinstrumenten für Einzelaspekte der Betreuungs- und Erziehungsumwelten der Einsatz eines solchen Screeninginstruments mit einem "Informationsverlust" verbunden ist, so dürften doch die Vorteile von Screeninginstrumenten (routinemäßiger und zeitökonomischer Einsatz) ihnen einen breiten Anwendungsbereich sichern. Dazu zählen im Bereich der Forschung sowohl der Einsatz in Surveystudien als auch die Verwendung in solchen Studien, die sich Spezialaspekten der Betreuung und Erziehung im frühen Kindesalter zuwenden und in denen Screeninginstrumente die Funktion übernehmen, allgemeinere Umwelt- und Qualitätsaspekte mit einzubeziehen bzw. zu kontrollieren. Daneben können diese Screeninginstrumente auch in Aus- und Fortbildungsprozessen des pädagogischen Personals sowie im Rahmen von Selbstevaluationen der Praktiker eingesetzt werden.

In der wissenschaftlichen Diskussion - wie sie sich seit den 50er Jahren entwickelt hat - wird ein breiter Qualitätsbegriff benutzt, der die Funktion einer frühkindlichen Betreuungs- und Erziehungsform in der Unterstützung der physischen, sozialen, emotionalen und intellektuellen Entwicklung eines Kindes, seines Wohlbefindens sowie der Eltern in ihrer Elternrolle sieht. Dieser breite Qualitätsbegriff schließt nicht nur im engeren bildungs- und erziehungsbezogene Aspekte mit ein, sondern umfaßt zugleich Aspekte der Pflege, Betreuung und physischen Gesundheit der Kinder. Die praktische Relevanz der Screeninginstrumente beruht nicht zuletzt darauf, daß sie Operationalisierungen einer bestimmten, inhaltlichen Vorstellung der



Qualität bzw. einer *guten* Qualität von frühkindlichen Betreuungs- und Erziehungsformen darstellen und somit Handlungshinweise für Veränderungsprozesse in der Praxis geben können. Allerdings setzt diese Funktion voraus, daß die hinter den Instrumenten stehenden Konzepte einer guten Qualität offengelegt und somit diskutierbar werden. In dieser Hinsicht zeigen einige der diskutierten Instrumente Schwächen (vgl. weiter unten).

Während im deutschsprachigen Raum ein ausgeprägtes Defizit an solchen Screeninginstrumenten besteht, wird in den USA seit längerem vermehrte Aufmerksamkeit der Qualitätsfrage und der Entwicklung entsprechender Instrumente zur Qualitätseinschätzung gewidmet. Dies muß auch gesehen werden auf dem Hintergrund des anders gelagerten gesellschaftlichen Kontextes der Betreuung und Erziehung in der frühen Kindheit und eines "marktorientierten" Systems, das auch in diesem gesellschaftlichen Bereich eine quantitative Versorgung mit entsprechenden Betreuungsplätzen über die Mechanismen von Angebot und Nachfrage steuert. Nicht zuletzt aufgrund dieses anders gelagerten Kontextes dürfte eine einfache bzw. unreflektierte Übernahme der Instrumente für die Situation in der Bundesrepublik nicht möglich sein.

Der in Kapitel 2 enthaltene Überblick über Screeninginstrumente zur Erfassung von Qualitätsmerkmalen frühkindlicher Betreuungs- und Erziehungsformen hat zum Ziel, diese Instrumente im deutschsprachigen Raum bekannt zu machen und damit potentiell für verschiedene Anwendungsbereiche Auswahlhilfen zu geben. Betrachtet wurden dabei solche Instrumente, die sich auf Betreuungs- und Erziehungsformen für Kinder im vorschulischen Alter beziehen, die im engeren Sinne als "Meß"-Instrumente charakterisiert werden können, die für eine breitere, nicht auf ein bestimmtes Projekt beschränkte Anwendung geeignet sind, die eine Gesamteinschätzung der Qualität geben und die relativ neue Instrumente darstellen bzw. derzeit in Gebrauch sind. Der Berichterstatter erhebt keinen Anspruch auf eine vollständige Erfassung aller diesbezüglichen Instrumente. Gleichwohl aber wird davon ausgegangen, daß das Kapitel 2 einen guten Überblick über die gegenwärtig in den USA benutzten Instrumente gibt.

Berichtet wurde über acht Instrumente zur Beurteilung von institutionellen Gruppenbetreuungen (vgl. die folgende Tabelle 1), vier bzw. fünf Instrumente für Tagespflegestellen<sup>44</sup> sowie ein Instrument zur Beurteilung der häuslichen Umwelten (vgl. Tabelle 2). Die Hauptaspekte der verschiedenen Instrumente werden im folgenden

---

<sup>44</sup> In Kapitel 2.2 wurden nur vier Instrumente vorgestellt. Allerdings kann das CDA Assessment Observation Instrument (vgl. Kapitel 2.1.3) sowohl in institutionellen Gruppenbetreuungen als auch für Tagespflegebetreuungen eingesetzt werden.

in Tabellenform zusammengefaßt. Anschließend werden die verschiedenen Instrumente kurz untereinander verglichen.

Im Kopf der beiden Tabellen stehen jeweils die Instrumente mit den Nummern der Kapitel, in denen sie ausführlich besprochen werden. Die Tabellen sind in drei inhaltliche Bereiche eingeteilt, die auch der ausführlicheren Darstellung in Kapitel 2 zugrunde liegen:

- **Konzept - Einsatzbereich:** In diesem Teil der Tabellen wird zunächst die Altersgruppe der Kinder genannt, für deren Betreuungsform das jeweilige Instrument zur Qualitätsbeurteilung geeignet ist. Es folgen Angaben zum theoretischen Hintergrund sowie zum Anwendungsziel und zum Einsatzbereich (z.B. Fremdbeobachtung, Selbstevaluation). Den Abschluß dieses Teils bilden Angaben zu den Anwendern der Instrumente (z.B. trainierte Beobachter, Laien) sowie zum erforderlichen Training.
- **Aufbau:** Unter dieser Überschrift sind die Informationen zum formalen Aufbau der Instrumente zusammengefaßt. Den Beginn bilden Angaben zur Anzahl der Items in den Instrumenten sowie zu ihrer Gliederung in Subbereiche, Komponenten oder Dimensionen. Allerdings werden die Instrumente nur unzureichend durch die Anzahl der Items charakterisiert, wenn nicht zugleich der Itemtyp, die Komplexität der Items und das Niveau der für die Beurteilung der Items durch den Anwender erforderlichen Informationsverarbeitungsprozesse berücksichtigt werden. Hierbei werden drei "Abstraktionsniveaus" unterschieden, wobei ein geringes Abstraktionsniveau (1) Items indiziert, die eher als Checklisten aufgebaut sind, während ein höheres Abstraktionsniveau (3) solche Items charakterisiert, bei denen der Beobachter eine Vielzahl von Einzelinformationen zu einem zusammenfassenden Urteil verarbeiten muß (vgl. Kapitel 2.1.1). Anschließend wird auf die zur Bearbeitung der Items benötigten Informationsquellen und den für die Bearbeitung des Instruments benötigten Zeitaufwand verwiesen. Zum Abschluß werden die für die jeweiligen Instrumente vorgesehenen Zusammenfassungen der Informationen im Sinne von Skalenbildungen angesprochen. Damit zusammen hängt auch die Frage, ob explizit quantitative Kriterien für eine gute Qualität - gemessen über die Instrumente - angegeben werden.
- **Gütekriterien:** Der dritte Teil der Tabellen wendet sich den forschungstechnischen Gütekriterien der Instrumente zu. Allerdings liegen hierzu nur eingeschränkte Informationen vor; in mehreren Fällen wurden keine Überprüfungen der technischen Qualität der Instrumente durchgeführt. Dies ist durch ein "-" indiziert. Oftmals wurden die Gütekriterien an Vorformen überprüft, die anschließend - z.B. vor der Veröffentlichung - modifiziert wurden, um eine Verbesserung der technischen Qualität zu erreichen. Dies ist in den Tabellen aber nur dann vermerkt, wenn zwischen der Instrumentenversion, an der die Gütekriterien überprüft wurden, und der Version, die der ausführlichen Beschreibung zugrunde liegt, vermutlich größere Unterschiede bestehen. Zuerst werden Daten zur Objektivität/Interrater-Reliabilität vorgestellt; danach folgen Angaben zur Reliabilität (Test-Retest-Reliabilität, interne Konsistenz) und zur Validität (inhaltliche Validität, kriteriumsorientierte Validität, Konstruktvalidität).<sup>45</sup> Den Abschluß bilden Analysen zur Aufdeckung der Faktorenstruktur

---

<sup>45</sup> "Kriteriumsorientierte Validität" bezieht sich auf alle Vergleiche der Instrumente mit anderen Maßen, die aber ebenfalls die Qualität der jeweiligen Betreuungsform indizieren. "Konstruktvalidi-

der Instrumente.

---

tät" umfaßt demgegenüber Vergleiche mit anderen Konstrukten wie z.B. dem Entwicklungsstand der betreuten Kinder oder Hintergrundmerkmalen der Betreuungspersonen.

Tabelle 1:

## Überblick über Instrumente für institutionelle Gruppenbetreuungen

	2.1.1 Early Childhood Classroom Observation	2.1.2 Classroom Practi- ces Inventory - CPI	2.1.3 CDA Assessment Observation In- strument	2.1.4 Assessment Profi- le for Early Childhood Pro- grams	2.1.5 Early Childhood Environment Ra- ting Scale - ECERS	2.1.6 Infant/Toddler Environment Ra- ting Scale - ITERS	2.1.7 Preschool Class- room Implementa- tion Rating In- strument - PCI- N/S	2.1.8 Child Care Faci- lity Schedule - CCFS
<u>Konzept - Ein- satzbereich</u>								
Altersgruppe der betreuten Kinder	0 bis 8 Jahre	4 bis 5 Jahre	0 bis 5 Jahre	0 bis 10 Jahre (verschiedene Instrumente)	2 ½ bis 6 Jahre	0 bis 2 ½ Jahre	3 bis 5 Jahre	0 bis 6 Jahre
Theoretischer Hintergrund	NAEYC - Akkredi- tierungskriterien für High Quality Early Childhood Programs	NAEYC - Leitli- nien für Develop- mentally Approp- riate Practices	CDA - Competency Standards	kein explizites Konzept; Verweis auf NAEYC-Akkre- ditierungskrite- rien, CDA-Compe- tency Standards sowie Items der ECERS	breites Verständ- nis von Umwelt; Verweis auf einige allgemeine Leitfragen und päd. Orientierun- gen; kein expli- zites, ausformu- liertes Konzept	Adaptation der ECERS für Grup- penbetreuungen für Kinder unter 2 ½ Jahren	Entwicklungsthe- orie von Piaget; Übereinstimmung mit den NAEYC - Leitlinien für Developmentally Appropriate Prac- tices	Verweis auf ver- schiedene Samm- lungen von Quali- tätskriterien, speziell: NAEYC- Akkreditierungs- kriterien
Anwendungsziel/ Einsatzbereich	Selbst- und Fremdevaluation im Rahmen des Akkreditierungs- programms; Quali- tätseinschätzung für Gruppen	Fremdevaluation, entwickelt in einem Projekt zu akademischen Ori- entierungen in institutionellen Gruppenbetreuun- gen; Qualitäts- einschätzung ein- iger Aspekte der päd. Arbeit in Gruppen	Fremdevaluation im Rahmen des CDA National Creden- tialing Programs; Qualitätsein- schätzung von Gruppen, hier: Fokus auf die Qualifikationen der Erzieherin	umfassendes Sys- tem für Selbst- evaluation der Arbeit in Ein- richtungen, 4 Komponenten; Tei- le für Gruppen (Infant-, Pre- school-, School- age-Gruppen) so- wie für Einrich- tung insgesamt	Selbst- und Fremdevaluation; Qualitätsein- schätzung für Gruppen	Selbst- und Fremdevaluation; Qualitätsein- schätzung für Gruppen	Fremdevaluation der Implementa- tion eines auf der Entwicklungs- theorie von Piag- et basierenden Curriculums; breiterer Einsatz bei Auslassung einiger Items	Fremdevaluation von Einrichtun- gen; für Selbst- evaluation werden Modifikationen nötig; Qualitäts- einschätzung von Einrichtungen insgesamt
Anwender/Training	Laien (z.B. Er- zieherinnen), trainierte Beob- achter; keine Informationen über das Training	trainierte Beob- achter; keine Angaben zum Um- fang des Train- ings	Laien (Berater) und trainierte Beobachter (CDA- Repräsentant), für letzteren existiert ein Trainingsfilm; weitere Informa- tionen zum Train- ing liegen nicht vor.	Laien; ca. drei- tägiges (Selbst-) Training; bei Einsatz in For- schung ca. zwei- tägiges Training	Laien und traini- erte Beobach- ter; verschiedene Trainingsmaterie- lien	Laien und traini- erte Beobach- ter; verschiedene Trainingsmaterie- lien	trainierte Beob- achter; keine Angaben zum Um- fang des Train- ings	trainierte Be- obachter; keine Angaben zum Um- fang des Train- ings

Tab. 1, Fortsetzung

	Early Childhood Classroom Observation	Classroom Practices Inventory - CPI	CDA Assessment Observation Instrument	Assessment Profile for Early Childhood Programs	Early Childhood Environment Rating Scale - ECERS	Infant/Toddler Environment Rating Scale - ITERS	Preschool Classroom Implementation Rating Instrument - PCI-H/S	Child Care Facility Schedule - CCFS
<b>Aufbau</b>								
Anzahl Items und Subbereiche/Subskalen	68 Items in 5 Dimensionen/Komponenten	26 Items in 2 bzw. 3 Bereichen	43 Items in 11 Bereichen	insgesamt für alle Teile 523 Items in 21 Dimensionen bei 4 Teilinstrumenten; für Einrichtung: 145 Items in 5 Dim.; für Infant-Gruppen: 128 Items in 5 Dim.; für Preschool-Gruppen: 147 Items in 6 Dim.; für School-Age Gruppen: 103 Items in 5 Dim.	37 Items in 7 Bereichen bzw. Subskalen	35 Items in 7 Bereichen bzw. Subskalen	54 Items in 12 Bereichen	74 Items in 7 Bereichen
Itemtyp/Abstraktionsniveau	komplexe Items, die auf dreistufiger Ratingskala beantwortet werden müssen, zusätzlich Indikatoren; bei Skalenstufen 1 und 2 (Aspekt nicht klar erfüllt) Kommentare erforderlich; höheres Abstraktionsniveau (3)	komplexe Items, die auf fünfstufiger Ratingskala beantwortet werden müssen; höheres Abstraktionsniveau (3)	komplexe Items, die auf dreistufiger Ratingskala beantwortet werden müssen, zusätzlich Indikatoren; bei Skalenstufen 1 und 2 (Aspekt nicht oder nur manchmal erfüllt) Begründungen erforderlich; höheres Abstraktionsniveau (3) erforderlich	kurze Statements, die mit "Ja" oder "nein" beantwortet werden müssen; mittleres Abstraktionsniveau (2)	komplexe Items, die auf siebenstufigen Ratingskalen beantwortet werden müssen; höheres Abstraktionsniveau (3)	komplexe Items, die auf siebenstufigen Ratingskalen beantwortet werden müssen; höheres Abstraktionsniveau (3)	komplexere Items, die auf dreistufiger Ratingskala beantwortet werden müssen; in der Regel Kommentare erforderlich; mittleres Abstraktionsniveau (2)	kurze Statements, die auf dreistufiger Ratingskala beantwortet werden müssen; mittleres Abstraktionsniveau (2)
Informationsquellen	Überwiegend Beobachtungen; u.U. Befragung von Erzieherinnen	Beobachtungen	Beobachtungen	Beobachtungen (Überwiegend); Durchsicht von Dokumenten; Befragung von Gruppenerzieherinnen und Einrichtungsleiterinnen	Beobachtungen, z.T. Befragung von Erzieherinnen	Beobachtungen, z.T. Befragung von Erzieherinnen	Beobachtungen, z.T. Befragung von Erzieherinnen	Beobachtungen und Befragung von Erzieherinnen/ Einrichtungsleiterinnen

Tab. 1, Fortsetzung

	Early Childhood Classroom Observation	Classroom Practices Inventory - CPI	CDA Assessment Observation Instrument	Assessment Profile for Early Childhood Programs	Early Childhood Environment Rating Scale - ECERS	Infant/Toddler Environment Rating Scale - ITERS	Preschool Classroom Implementation Rating Instrument - PCI-H/S	Child Care Facility Schedule - CCFS
<u>Aufbau, Fortsetzung</u>								
Bearbeitungszeit (bei Fremdbeobachtung)	1 bis 2 Stunden, plus u.U. Zeit für Befragungen	2 ½ Stunden	2 ½ bis 3 Stunden	für gesamte Selbstevaluation etwa eine Woche; während eines Vormittags 1 bis 4 Gruppen (rotierendes Vorgehen)	2 Stunden	2 Stunden	keine Zeitangaben	keine Zeitangaben
Skalenbildung/Qualitätsniveaus	keine Skalenbildungen vorgesehen; kein quantitatives Kriterium für gute Qualität	additive Skalenbildungen, Gesamtskala und vier Subskalen; keine direkten Angaben zu Qualitätsniveaus; durch Mittelwertbildung kann ursprüngliche fünfstufige Ratingskala zur Beurteilung herangezogen werden	keine Skalenbildungen vorgesehen; kein quantitatives Kriterium für gute Qualität	additive Skalenbildungen, Gesamtskala und verschiedene Subbereiche; verschiedene Auswertungshilfen (Dimension Score Sheet, Profile); quantitatives Kriterium für exzellente und durchschnittliche Qualität	additive Skalenbildungen, Gesamtwert und 7 Subskalen; Auswertungsblatt (classroom profile); keine direkten Angaben zu Qualitätsniveaus; durch Mittelwertbildung kann ursprüngliche siebenstufige Ratingskala zur Bewertung herangezogen werden	additive Skalenbildungen, Gesamtwert und 7 Subskalen; keine direkten Angaben zu Qualitätsniveaus; durch Mittelwertbildung kann ursprüngliche siebenstufige Ratingskala zur Bewertung herangezogen werden	additive Skalenbildungen; keine Angaben zu Qualitätsniveaus	keine Angaben zu Skalenbildungen und Qualitätsniveaus
Gütekriterien								
Objektivität/Interrater-Reliabilität	.89 (n=5 Gruppen; vermutlich Korrelation der Werte von 2 Beobachtern)	Korrelation von zwei Beobachtern .86 bei Gruppen aus 10 Einrichtungen; durchschnittliche Übereinstimmung exakt: 64 %, innerhalb von einer Abweichung von einer Skalenstufe: 98 %	-	nach zweitägigem Training für Forschungsstudien Beobachterübereinstimmung von über 85 %	Korrelationen der Werte von 2 Beobachtern in verschiedenen Studie um .80 und höher	Korrelationen der Werte von 2 Beobachtern .84 für Gesamtwert; .58 bis .89 bei Subskalen (n=30 Gruppen); Beobachterübereinstimmung bei 29 Items über 70 % (bei einer Skalenstufe Abweichung)	-	bereichsspezifische Korrelationen der Werte von zwei Beobachtern über .80 (in Griechenland), für Gesamtwert .94 (in Nigeria); gute Beobachterübereinstimmung auf Itemebene
Test-Retest-Reliabilität					.96 (n=31 Gruppen, 2 Wochen Abstand)	.79 Gesamtwert, .58 bis .82 bei Subskalen (n=18 Gruppen, 3 bis 4 Wochen Abstand)		

Tab. 1. Fortsetzung

	Early Childhood Classroom Observation	Classroom Practices Inventory - CPI	CDA Assessment Observation Instrument	Assessment Profile for Early Childhood Programs	Early Childhood Environment Rating Scale - ECERS	Infant/Toddler Environment Rating Scale - ITERS	Preschool Classroom Implementation Rating Instrument - PCI-H/S	Child Care Facility Schedule - CCFS
<b>Gütekriterien. Fortsetzung</b>								
Interne Konsistenz (Cronbach's Alpha)	Verläuferversion: .70 (stand. .86) für Gesamtwert (n=387 Gruppen)	Gesamtskala .96, Subskalen .88 bis .96 (n=207 Gruppen aus 58 Einrichtungen)	-	Gesamtwerte für die 4 Komponenten jeweils über .90; Halbierungsreliabilität von .73 bis .94	USA: Gesamtskala .83, Subskalen .32 bis .79 (n=25 Gruppen); BRD: Gesamtskala .93, Subskalen .42 bis .76 (n=43 Gruppen)	Gesamtskala .83, Subskalen .26 bis .79 (n=30 Gruppen)	Gesamtskala .96 (Anzahl Gruppen nicht bekannt)	-
Inhaltliche Validität	Bezug zu NAEYC-Kriterien, die über Expertenbefragungen abgesichert wurden	Bezug zu NAEYC-Guidelines for Developmentally Appropriate Practices	Bezug zu CDA-Competency Standards	Bezug zu NAEYC-Kriterien, CDA-Competency Standards und ECERS Itemsatz	Expertenratings (n=7 Experten)	hohe Konsistenz mit NAEYC-Kriterien und CDA-Competency Standards sowie weiteren Instrumenten; Expertenratings (n=5 Experten)	-	-
Kriteriumsorientierte Validität	-	-	-	Korrelationen zu ECERS: .64 (n=46 Gruppen), ITERS: .49 bei Infant-Gruppen (n=39) und .66 bei Toddler-Gruppen (n=28)	Vergleich mit globalen Qualitätsratings (n=18 Gruppen), Korrelation um .74; moderate Korrelationen mit Aspekten des Day Care Environment Inventory (n=8 Einrichtungen)	Vergleich mit globalen Qualitätsratings (n=12 Gruppen), in 10 von 12 Gruppen gleiche Einteilung in "niedrigere" und "höhere" Qualität	-	-

Tab. 1, Fortsetzung

	Early Childhood Classroom Observation	Classroom Practices Inventory - CPI	CDA Assessment Observation Instrument	Assessment Profile for Early Childhood Programs	Early Childhood Environment Rating Scale - ECERS	Infant/Toddler Environment Rating Scale - ITERS	Preschool Classroom Implementation Rating Instrument - PCI-H/S	Child Care Facility Schedule - CCFS
<u>Gütekriterien, Fortsetzung</u> Konstruktvalidität	-	Plausible Beziehungen (Gruppen aus 10 Einrichtungen) z.B. zu akademischen Orientierungen der Mütter und kindlichem Entwicklungsstand (Kreativität, Angst)	-	-	bewährtes Instrument in verschiedenen Studien; plausible Korrelationen mit kindlichem Entwicklungsstand, Maßen der verbalen Interaktion, Gruppengröße, Erzieher-Kind-Relation, Merkmalen der Erzieherinnen	plausible Beziehungen von zweifaktorenanalytisch gewonnenen Skalen zu Maßen des kindlichen Entwicklungsstandes, Gruppengröße, Erzieher-Kind-Relation, Merkmalen der Erzieherinnen	Korrelation zu Schulreifezeit zu Schulbeginn (weitere Informationen dazu liegen nicht vor)	-
Faktorenstruktur	Vorläuferversion (n=387 Gruppen): 3 Faktoren, die 48 % der gesamten Varianz aufklären; teilweise Reproduktion der theoretisch unterstellten Struktur	4 Faktoren (n=207 Gruppen aus 58 Einrichtungen), die 69 % der gesamten Varianz aufklären; zu größeren Teilen Reproduktion der theoretisch unterstellten Struktur	-	-	verschiedene Faktorenanalysen (USA: n=185, 313 Gruppen; BRD: n=43, 68; Hong Kong: n=53); verschiedene Faktorenomodelle; 2 bis 5 Faktoren, die je nach Studie zusammen 40 bis 67 % der gesamten Varianz aufklären; keine direkte Reproduktion der theoretischen Struktur, aber sinnvoll zu interpretierende Faktoren; gewisse Stabilität der Faktoren über verschiedene Datensätze, Faktorenomodelle und Länder	2 Faktoren (n=330 Gruppen), keine Information zum Anteil aufgeklärter Varianz an der gesamten Varianz	-	-



**Tabelle 2:** Überblick über Instrumente für Betreuungen in Tagespflege und in häuslichen Umwelten

	2.1.3 CDA Assessment Observation Instrument	2.2.1 Assessment Profile for Family Day Care	2.2.2 Family Day Care Rating Scale - FDCRS	2.2.3 Family Day Home Observation Instrument	2.2.4 Evaluating Home Based Day Care (Louise Instrument)	2.3.1 Home Observation for Measurement of the Environment - HOME
<b>Konzept - Einsatzbereich</b>  Altersgruppe der betreuten Kinder	0 bis 5 Jahre	keine exakte Altersangabe; nach den Items von 0 bis 12 Jahre	0 bis 6 Jahre	keine exakte Altersangabe; nach den Items von 0 Jahre bis Schulalter	keine exakte Altersangabe; aufgrund der Nähe zu den CDA-Competency Standards 0 - 5 Jahre	verschiedene Versionen: 0 bis 3 Jahre, 3 bis 6 Jahre, Schulalter
<b>Theoretischer Hintergrund</b>	CDA-Competency Standards	kein explizites Konzept; narrative Beschreibung einer guten Tagespflegestelle in einem Study Guide	Adaptation der ECERS für Betreuungen in Tagespflegestellen; Übereinstimmung mit CDA-Competency Standards	CDA-Competency Standards	CDA-Competency Standards	kein explizites Konzept, Bezug auf allgemeine Merkmale einer die Entwicklung stimulierenden Umwelt
<b>Anwendungsziel/Einsatzbereich</b>	Fremdevaluation im Rahmen des CDA National Credentialing Programs; Fokus auf die Qualifikationen der Tagespflegebetreuerin	Selbst- und Fremdevaluation im Rahmen des NAFCO-Akkreditierungsprogramms	Selbst- und Fremdevaluation; u.a. auch Einsatz in regionalem Akkreditierungsprogramm	Fremdbeobachtung im Rahmen eines regionalen Akkreditierungsprogramms	Fremdbeobachtung (speziell im Rahmen von Supervisionen); auch für Selbstevaluation oder für Eltern geeignet	Fremdbeobachtung in der häuslichen Umwelt des Kindes (Stimulationsgehalt der häuslichen Umwelt)
<b>Anwender/Training</b>	Laien (Berater) und trainierte Beobachter (CDA-Repräsentant); für letzteren existiert ein Trainingsfilm, weitere Informationen zum Training liegen nicht vor	Laien (Tagespflegebetreuerin) und trainierter Beobachter (NAFCO-Vertreter); keine Angaben zum Umfang des Trainings für den NAFCO-Vertreter	Laien und trainierte Beobachter; verschiedene Trainingsmaterialien	trainierte Beobachter; keine Angaben zum Umfang des Trainings	trainierte Beobachter; keine Angaben zum Training	trainierte Beobachter; geringer Trainingsaufwand erforderlich; keine weiteren Angaben zum Training
<b>Aufbau</b>  Anzahl Items und Subbereiche/Subskalen	43 Items in 11 Bereichen	277 Items in 7 Dimensionen	32 Items in 6 Bereichen bzw. Subskalen (8 zusätzliche Items für behinderte Kinder)	116 Items in 8 Bereichen	206 Items in 9 Bereichen	45 Items in 6 Dimensionen bei Version für Kinder unter 3 Jahren; 55 Items in 8 Dimensionen bei Version für Kinder von 3 bis 6 Jahren

Tab. 2, Fortsetzung

	CDA Assessment Observation Instrument	Assessment Profile for Family Day Care	Family Day Care Rating Scale - FDCRS	Family Day Home Observation Instrument	Evaluating Home Based Day Care	Home Observation for Measurement of the Environment - HOME
<b>Aufbau, Fortsetzung</b>						
Itemtyp/Abstraktionsniveau	komplexe Items, der auf dreistufiger Ratingskala beantwortet werden müssen, zusätzlich Indikatoren; bei Skalenstufen 1 und 2 (Aspekt nicht oder nur manchmal erfüllt) Begründungen erforderlich; höheres Abstraktionsniveau (3) erforderlich	kurze Statements, die mit "ja" oder "nein" beantwortet werden müssen; mittleres Abstraktionsniveau (2)	komplexe Items, die auf siebenstufigen Ratingskalen beantwortet werden müssen; höheres Abstraktionsniveau (3)	kurze Statements, die mit "ja" oder "nein" beantwortet werden müssen, zusätzlich bei einigen Items Checklisten; niedriges (1) bis mittleres (2) Abstraktionsniveau	eher Checklisten/kurze Statements, die mit "ja" oder "nein" beantwortet werden müssen; Items sind 3 Niveaus zugeordnet (basic care, intuitively good care, informed care); niedriges Abstraktionsniveau (1)	kurze Statements, die mit "+" oder "-" beantwortet werden müssen; mittleres Abstraktionsniveau (2)
Informationsquellen	Beobachtungen	Beobachtungen (überwiegend); Durchsicht von Dokumenten; Befragung der Tagespflegebetreuerin	Beobachtungen, z.T. Befragung der Tagespflegebetreuerin	Beobachtungen, z.T. Befragung der Tagespflegebetreuerin und Durchsicht schriftlicher Unterlagen	Beobachtungen	Beobachtungen; bei etwa einem Drittel der Items Befragung der Mütter
Bearbeitungszeit (bei Fremdbeobachtung)	2 ½ bis 3 Stunden	(mind.) 6 Stunden	2 Stunden	2 Stunden für Beobachtung, insgesamt 4 - 6 Stunden	keine Zeitangabe	1 Stunde
Skalenbildung/Qualitätsniveaus	keine Skalenbildungen vorgesehen; kein quantitatives Kriterium für gute Qualität	additive Skalenbildungen, Gesamtskala und verschiedene Subbereiche; verschiedene Auswertungshilfen (Dimension Score Sheet; Profile); quantitatives Kriterium für gute Qualität (pro Dimension mindestens 80 % der Items "ja")	additive Skalenbildungen, Gesamtwert und 6 Subskalen; keine direkten Angaben zu Qualitätsniveaus; durch Mittelwertbildung kann ursprüngliche siebenstufige Ratingskala zur Bewertung herangezogen werden	additive Skalenbildungen; quantitatives Kriterium für gute Qualität (je nach Bereich müssen für Akkreditierung 50 bis 90 % der Items mit "ja" beantwortet werden)	additive Skalenbildungen, Gesamtwert und Bereichswerte; für Gesamtwert drei quantitative Qualitätsniveaus	additive Skalenbildungen, Gesamtwert und Bereichswerte; jeweils drei quantitative Qualitätsniveaus
<b>Gütekriterien</b>						
Objektivität/Intra-ter-Reliabilität		im Akkreditierungsprozeß durchschnittliche Beobachterübereinstimmung von 92 %	Korrelationen der Werte von 2 Beobachtern in verschiedenen Studien um .80 und höher; Beobachterübereinstimmungen in verschiedenen Studien um 80 % und höher	Beobachterübereinstimmung 97,5 % (n=9 Tagespflegestellen)	Vorläuferversion: Durchschnittliche Beobachterübereinstimmung 88 % (n=6 Tagespflegestellen)	keine genauen Angaben; nach kurzem Beobachtertraining Beobachterübereinstimmungen um 90 % und mehr

Tab. 2. Fortsetzung

	CDA Assessment Observation Instrument	Assessment Profile for Family Day Care	Family Day Care Rating Scale - FDCRS	Family Day Home Observation Instrument	Evaluating Home Based Day Care	Home Observation for Measurement of the Environment - HOME
<u>Gültigkeitskriterien, Fortsetzung</u> Test-Retest-Reliabilität	-	-	-	-	-	.62 bis .77 bei Gesamtskala, .05 bis .77 bei Subskalen (Abstand bis zu 1 1/2 Jahre; n=91 bei Kindern unter 3 Jahren, n=33 bei Kindern von 3 bis 6 Jahren)
Interne Konsistenz (Cronbach's Alpha)	-	-	Subskalen von .70 bis .93 (n=55 Tagespflegestellen); ähnlich hohe Werte in verschiedenen Studien, Gesamtwert über .90	-	-	Gesamtskala um .90; Subskalen von .44 bis .89 (n=174 Kinder unter 3 Jahren, n=117 Kinder von 3 bis 6 Jahren)
Inhaltliche Validität	Bezug zu CDA-Competency Standards	-	hohe Übereinstimmung mit CDA-Competency Standards; Bezug zur ECERS, deren inhaltliche Validität über Expertenrating gesichert wurde	Bezug zu CDA-Competency Standards	Bezug zu CDA Competency Standards	-
Kriteriumsorientierte Validität	-	Korrelationen von je .62 mit FDCRS und Family Day Home Observation Instrument (n=32 Tagespflegestellen)	Korrelationen von .82 mit Family Day Home Observation Instrument und .62 mit Assessment Profile for Family Day Care (n=32 Tagespflegestellen); Korrelation von .34 mit HOME (n=46); Korrelation von .64 mit globaler Qualitätseinschätzung (n=161)	Korrelation mit globaler Qualitätseinschätzung durch Experten .68 (n=9 Tagespflegestellen); Korrelation mit FDCRS .82 (n=32)	Vorläuferversion: Korrelation mit globaler Qualitätseinschätzung durch Experten .72 (n=69 Tagespflegestellen)	Korrelation von .34 mit FDCRS (n=66 Tagespflegestellen)

Tab. 2, Fortsetzung

	CDA Assessment Observation Instrument	Assesment Profile for Family Day Care	Family Day Care Rating Scale - FDCRS	Family Day Home Observation Instrument	Evaluating Home Based Day Care	Home Observation for Measurement of the Environment - HOME
<u>Gütekriterien, Fortsetzung</u> Konstruktvalidität	-	-	bewährtes Instrument in verschiedenen Studien; plausible Korrelationen mit Merkmalen der Eltern, Rahmenbedingungen der Tagespflegestelle und Merkmalen der Tagespflegebetreuerin, Aspekten des kindlichen Spiels, der verbalen Interaktion sowie des kindlichen Entwicklungsstandes	-	-	bewährtes Instrument in verschiedenen Studien; plausible Korrelationen z.B. zum sozialen Status, zum kindlichen Entwicklungsstand (Intelligenz, Sprachleistungen, Schulreife testwerte, auch Schulprobleme)
Faktorenstruktur	-	-	-	-	-	-

## ***Instrumente zur Qualitätsbeurteilung institutioneller Gruppenbetreuungen (Tabelle 1)***

### **- Konzept - Einsatzbereich**

In der Regel decken alle besprochenen Instrumente institutionelle Gruppenbetreuungen für die gesamte frühkindliche Altersspanne von der Geburt bis zum Schuleintritt ab, wobei es zum Teil innerhalb eines Instrumentenkonzepts verschiedene Versionen für Gruppenbetreuungen von Kindern unter ca. drei Jahren und von Kindern über drei Jahren gibt (Assessment Profile for Early Childhood Programs; ECERS bzw. ITERS). Falls ein und dasselbe Instrument die gesame Altersspanne abdeckt, sind einzelne Items speziell auf die unterschiedlichen Altersgruppen zugeschnitten. Allerdings bleibt in diesen Fällen zu fragen, inwieweit die Spezifität der Gruppenbetreuungen für jüngere und ältere Kinder in *einem* Instrument erfaßt ist bzw. erfaßt werden kann. Zwei Instrumente beziehen sich ausschließlich auf Gruppenbetreuungen von Kindern im Kindergartenalter. In beiden Fällen läßt sich dies auch damit erklären, daß hier im engeren eine Beurteilung spezifischer Aspekte der pädagogischen Arbeit - "entwicklungsangemessene" Spielorientierung vs. akademische Orientierung der pädagogischen Arbeit; entsprechend den NAEYC Guidelines for Developmentally Appropriate Practices - im Mittelpunkt steht und somit eine Beschränkung auf eine eingeschränktere Altersgruppe erforderlich wird.

Sieht man von den theoretischen Orientierungen der beiden zuletzt genannten Instrumente einmal ab, so wird in der Regel als theoretischer Hintergrund der Instrumente auf die NAEYC - Accreditation Criteria for High Quality Early Childhood Programs und/oder auf die CDA - Competency Standards verwiesen. Zwar sind beide Kriteriensätze in den USA als Operationalisierungen einer guten Qualität frühkindlicher Betreuungs- und Erziehungsumwelten breit anerkannt, dennoch aber können sie nicht als konsistente und theoretisch begründete Entwürfe betrachtet werden. Vielmehr stellen sie ad hoc für das Akkreditierungsprogramm bzw. für das CDA National Credentialing Program entwickelte Sammlungen von Qualitätskriterien dar, für deren Entwicklung die vorhandene Forschungsliteratur gesichtet sowie Experten des frühkindlichen Bereichs befragt wurden. Dies gilt auch für die Qualitätskriterien, die in der ECERS und der ITERS enthalten sind und die in den USA von der Fachöffentlichkeit ebenfalls als Operationalisierungen einer guten Qualität betrachtet werden. Wünschenswert wäre somit für alle angesprochenen Instrumente eine stärkere theoretische Konzeptualisierung einer guten Qualität frühkindlicher Betreuungs- und Erziehungsumwelten, um somit auch die unterschiedlichen Ansätze und Reichweiten der verschiedenen Instrumente transparenter zu machen.

Die Hälfte der Instrumente kann sowohl in Fremd- als auch in Selbstevaluationsprozessen eingesetzt werden. Die andere Hälfte ist von der Anlage her zunächst nur

für eine Fremdevaluation gedacht; entsprechende Modifikationen scheinen aber auch einen Einsatz in einer Selbstevaluation zu ermöglichen. Die Instrumente werden von Laien (z.B. den Erzieherinnen) oder von trainierten Beobachtern angewandt. Im Hinblick auf ein u.U. notwendiges Training liegen dem Berichtersteller nur wenige Informationen vor. Im allgemeinen kann aber davon ausgegangen werden, daß aufgrund des vorgesehenen Einsatzes in Selbstevaluationsprozessen bzw. des "relativ" einfachen Aufbaus der Instrumente der Trainingsaufwand sich in Grenzen halten dürfte. Inklusive praktischer Übungen dürfte ein Gesamtumfang von zwei Tagen die Obergrenze bilden.

#### - *Aufbau*

Die Instrumente sind unterschiedlich lang; das kürzeste enthält nur 26 Items (CPI), das längste 523 Items (Assessment Profile for Early Childhood Programs). Im Falle des Assessment Profiles for Early Childhood Programs muß aber berücksichtigt werden, daß für die Beurteilung einer einzelnen Gruppe nur 128 (Infant-Gruppe) bzw. 147 (Preschool-Gruppe) Items benötigt werden. Auch im Hinblick auf die interne Gliederung der Instrumente zeigen sich erhebliche Unterschiede (2 bis zu 12 Subbereiche). Allerdings sagt die reine Anzahl der Items nur relativ wenig über die "Länge" eines Instruments aus. Vielmehr muß dazu auch die Art und Komplexität der Items berücksichtigt werden. Hier zeigt sich eine negative Beziehung: Je mehr Items - d.h. einzuschätzende Aspekte - in einem Instrument vorhanden sind, desto weniger komplex sind die Items. Oder umgekehrt: Je komplexer die Items und je mehr Anforderungen sie an den Beobachter stellen, desto weniger Items sind in dem Instrument enthalten. So sind z.B. die 147 Items für Preschool-Gruppen im Assessment Profile kurze Statements, die nur mit "ja" oder "nein" beantwortet werden müssen, während die 37 Items der ECERS komplexe siebenstufige Ratingskalen sind, für deren Beantwortung der Beobachter jeweils eine Vielzahl von Einzelinformationen zu einem einzigen Urteil zusammenfassen muß. Diese negative Beziehung zwischen Länge eines Instruments und Komplexität der Items hat auch zur Folge, daß die für die Ausfüllung eines Instruments benötigte Zeit nur geringfügig zwischen den Instrumenten variiert. Soweit Zeitangaben vorhanden sind, wird bei einer Fremdbeobachtung für die Qualitätsbeurteilung einer Gruppe um zwei bis drei Stunden benötigt. Aus diesem Grunde können alle Instrumente als zeitökonomisch anzuwendende Verfahren zur Qualitätsbeurteilung charakterisiert werden.

Für die Bearbeitung der Items werden bei allen Instrumenten hauptsächlich Beobachtungen herangezogen. In keinem Fall sind damit aber minuziöse Beobachtungen gemeint, bei denen z.B. im Sinne eines time-samplings oder event-samplings feinste Beobachtungseinheiten in einer hochstrukturierten Weise protokolliert werden müssen. Vielmehr bezeichnet Beobachtung hier eine Tätigkeit, bei der der Beobach-

ter sich über einen gewissen Zeitraum einen Eindruck von der Gruppe bzw. vom Gruppengeschehen verschafft und anschließend, durch das jeweilige Instrument angeleitet, Urteile - in unterschiedlichem Komplexitätsgrad (vgl. weiter oben) - zu den beobachteten Aspekten abgibt. Insofern besteht bei allen Instrumenten in mehr oder weniger starkem Maße die Möglichkeit einer durch die Subjektivität des Beobachters verzerrten Qualitätsbeurteilung. Dieses Problem dürfte da besonders nahe liegen, wo z.B. Erzieherinnen im Rahmen einer Selbstevaluation ihre eigene Praxis einschätzen. Allerdings zeigt eine kleine, im Rahmen einer Staatsexamensarbeit durchgeführte Studie, daß - am Beispiel der ECERS - nach einem kurzen Training die Selbstbeurteilungen der Erzieherinnen nur unwesentlich von den Beurteilungen eines außenstehenden Beobachters abweichen (Sander, 1986). Insgesamt dürfte hier aber auf ein Problem dieser Instrumente verwiesen sein, auf das bei einem Training im Umgang mit den Instrumenten besonders geachtet werden muß. Bei den meisten Instrumenten werden neben der Beobachtung noch Befragungen der Erzieherinnen und - beim Assessment Profile for Early Childhood Programs - eine Durchsicht von schriftlichen Materialien erforderlich. Auch hierbei handelt es sich nicht um strukturierte, hoch standardisierte Befragungen der Erzieherinnen. Vielmehr sollen in unstrukturierten Gesprächen mit den Erzieherinnen die für die Beurteilung der verschiedenen Aspekte benötigten Informationen eingeholt werden.

Die Instrumente unterscheiden sich danach, ob nur die Informationen auf der Itemebene interpretiert werden können oder ob auch Zusammenfassungen der Einzelinformationen zu Skalen-/Dimensionswerten oder zu einem Gesamtwert vorgesehen sind. Damit hängt auch zusammen, ob - bezogen auf Skalenwerte oder auf Gesamtwerte - quantitative Kriterien/Schwellenwerte angegeben werden, ab wann von einer qualitativ guten Gruppenbetreuung gesprochen werden kann. Für das CCFS liegen dem Berichtersteller hierzu keinerlei Informationen vor. Die Early Childhood Classroom Observation und das CDA Assessment Observation Instrument verzichten auf Zusammenfassungen der Informationen zu Skalen. Hier wird davon ausgegangen, daß die einzelnen Items nicht gleichgewichtet sind und deshalb additive Zusammenfassungen wenig aussagekräftig seien. Dementsprechend wird ein quantitativer Schwellenwert für gute Qualität - z.B. eine bestimmte Punktzahl bei Addition der Antworten auf die dreistufigen Ratingskalen - abgelehnt. Alle anderen Instrumente sehen additive Zusammenfassungen der Informationen vor. Im Assessment Profile for Early Childhood Programs werden auf der Ebene der Dimensionen explizit Qualitätsniveaus angegeben, z.B. indizieren 70 bis 89 % "ja" - Antworten auf die Items einer Dimension ein durchschnittliches Qualitätsniveau. Die anderen Instrumente verzichten zwar auf explizite Angaben zu Qualitätsniveaus, allerdings können bei additiven Zusammenfassungen auf Subskalen- oder Gesamtwertebene und anschließender Bildung von Durchschnittswerten die ursprünglichen fünf- (CPI) bzw. siebenstufigen (ECERS, ITERS) Ratingskalen zur Qualitätsbeurteilung herangezogen

werden. Für das PCI-H/S liegen dem Berichtersteller hierzu keine Informationen vor. Auch wenn zum Teil Bedenken gegenüber additiven Zusammenfassungen der Items - wie sie für die Early Childhood Classroom Observation und für das CDA Assessment Observation Instrument geäußert werden - dem Berichtersteller als nicht unberechtigt erscheinen, so bieten doch solche zusammenfassenden Skalenbildungen eine Möglichkeit, eine institutionelle Gruppenbetreuung insgesamt oder ihre Subbereiche zusammenfassend zu beurteilen und durch explizite oder implizite Angaben von Qualitätsniveaus eine stärker verobjektivierte Qualitätsbeurteilung zu erreichen. Hervorgehoben werden soll noch, daß für die Interpretation der Ergebnisse beim Assessment Profile for Early Childhood Programs und bei der ECERS durch graphische Veranschaulichungen der Ergebnisse Hilfen gegeben werden.

### ***Gütekriterien***

Im Hinblick auf die forschungstechnischen Gütekriterien fällt bei Betrachtung von Tabelle 1 auf, daß bei mehreren Instrumenten hierzu so gut wie keine bzw. nur unzureichende Informationen vorliegen. Dies gilt für das CDA Assessment Observation Instrument, das PCI-H/S und das CCFS. Eingeschränkte Informationen zur technischen Qualität sind für die Early Childhood Classroom Observation (für eine Vorläuferversion) und das Assessment Profile for Early Childhood Programs vorhanden. Umfangreichere Informationen liegen nur für das CPI, die ECERS und die ITERS vor. Im allgemeinen läßt sich aber festhalten, daß in der Regel die forschungstechnischen Gütekriterien weiterer Überprüfung bedürfen.

Wenn Angaben zur Objektivität der Instrumente vorliegen, so zeigt sich durchgängig eine hohe Objektivität: Korrelationen der Gesamtwerte der Einschätzungen von zwei unabhängigen Beobachtern oder die Übereinstimmung von zwei Beobachtern in der Bewertung der einzelnen Items indizieren ein zufriedenstellendes Ausmaß an Objektivität. Im Hinblick auf die Reliabilität muß zwischen Test-Retest-Reliabilität und interner Konsistenz (Cronbach's Alpha) unterschieden werden. Die Test-Retest-Reliabilität wurde dabei nur selten untersucht. Nur für die ECERS und die ITERS liegen hierfür Informationen vor, wobei die ECERS in dieser Hinsicht eine sehr hohe und die ITERS eine hohe Reliabilität aufweisen. Für sechs der acht Instrumente liegen Angaben zur internen Konsistenz vor. Zum Teil werden dabei - zumindest auf der Ebene der Gesamtwerte - hohe bis sehr hohe Konsistenzen erreicht, wobei dann allerdings die hohen Itemzahlen berücksichtigt werden müssen. Für CPI, ECERS und ITERS liegen Informationen über die internen Konsistenzen auf Subskalenebene vor. Bei der ECERS und ITERS zeigen sich dabei bei einigen Subskalen eher niedrige und problematische Werte, die auf Inhomogenitäten in diesen Subskalen hinweisen. Beim CPI liegen die internen Konsistenzen höher, allerdings sind die hier jeweils in einer Dimension einzuschätzenden Aspekte auch deutlich homogener



konzipiert.

Bei der Validität wird in Tabelle 1 zwischen inhaltlicher und kriteriumsbezogener Validität sowie Konstruktvalidität unterschieden. Im Hinblick auf inhaltliche Validität stellen zunächst die Early Childhood Classroom Observation, das CPI und das CDA Assessment Observation Instrument direkte Operationalisierungen spezifischer Kriteriensätze dar, deren inhaltliche Wichtigkeit für eine Qualitätseinschätzung frühkindlicher Betreuungs- und Erziehungsumwelten zum Teil durch umfangreiche Expertenbefragungen (bei den NAEYC - Akkreditierungskriterien) sichergestellt wurde. Während beim Assessment Profile for Early Childhood Programs nur auf eine Nähe zu diesen Kriteriensätzen verwiesen wird, sichern Expertenratings bei der ECERS und der ITERS eine inhaltliche Validität - auch wenn die Anzahl der einbezogenen Experten als niedrig anzusehen ist. Informationen zur kriteriumsorientierten Validität, d.h. darüber, inwieweit die durch ein Instrument gemessene Qualität mit der über ein anderes vergleichbares Instrument oder mit einer globalen Einschätzung erhobenen Qualität übereinstimmen, liegen nur für Assessment Profile for Early Childhood Programs, ECERS und ITERS vor. Hier zeigen sich zufriedenstellende Übereinstimmungen, wobei zu berücksichtigen ist, daß unterschiedliche Zugriffe zum Teil auch unterschiedliche Schwerpunkte beinhalten. Für vier Instrumente liegen darüber hinaus Hinweise auf Zusammenhänge mit anderen Konstrukten vor (CPI, ECERS, ITERS und PCI-H/S), wobei hier nicht zuletzt aufgrund des breiten Einsatzes und der damit verbundenen größeren Breite von Informationen die ECERS und die ITERS als im Sinne der Konstruktvalidität bewährte Instrumente betrachtet werden können.

Nur bei vier Instrumenten wurde auf empirischem Wege die Faktorenstruktur untersucht. Bei der Early Childhood Classroom Observation, dem CPI und der ITERS liegen Faktorenanalysen nur aus jeweils einer Studie vor, während bei der ECERS die Faktorenstruktur in verschiedenen Studien und Ländern sowie mit verschiedenen statistischen Verfahren untersucht wurde. Dabei zeigen sich zwar Unterschiede zur unterstellten theoretischen Struktur, gleichwohl aber ergeben sich interpretierbare Faktoren, die über verschiedene Studien, Länder und Analysemethoden eine gewisse Stabilität aufweisen.

Bei einer Gesamtbeurteilung der technischen Qualität der verschiedenen Instrumente - die als eine Voraussetzung für alle Anwendungsbereiche anzusehen ist - muß berücksichtigt werden, daß dann, wenn Autoren sich der Mühe ausführlicherer Analysen unterziehen, auch die "Gefahr" besteht, Schwächen und Problemzonen aufzudecken. Werden keine Studien zu den Gütekriterien durchgeführt, dann bleiben solche Schwächen unentdeckt. Speziell für eine Anwendung in Forschungsstudien scheint es sinnvoller, mit einem Instrument zu arbeiten, über dessen technische Qua-

lität man Informationen hat - auch wenn diese u.U. Problemzonen indizieren -, als daß man mit einem Instrument arbeitet, über dessen technische Qualität letztlich keine empirischen Informationen vorliegen. In dieser Perspektive läßt sich festhalten, daß gegenwärtig zum CPI sowie zur ECERS und ITERS die meisten Informationen über die technische Qualität vorliegen. Diese drei Instrumente weisen eine zufriedenstellende technische Qualität auf und können deshalb für eine Anwendung empfohlen werden, wobei allerdings das CPI in seiner Reichweite und aufgrund seiner theoretischen Orientierung als eingeschränkter zu betrachten ist.

### ***Instrumente zur Qualitätsbeurteilung von Betreuungen in Tagespflegestellen und in häuslichen Umwelten (Tabelle 2)***

Zu einem größeren Teil werden für diese Instrumente vergleichbare Anmerkungen erforderlich wie für die Instrumente zur Beurteilung institutioneller Gruppenbetreuungen, so daß die folgenden Ausführungen kürzer ausfallen können.

#### ***- Konzept - Einsatzbereich***

Für die HOME gibt es verschiedene Versionen zur Einschätzung des Anregungsgehalts der häuslichen Umwelt in Abhängigkeit vom Alter des Kinder. Alle Instrumente zur Qualitätsbeurteilung von Tagespflegebetreuungen beinhalten in *einem* Instrument die gesamte Altersspanne von der Geburt bis zum Schuleintritt, zum Teil auch bis zum Alter von 12 Jahren. In der Regel enthalten die Instrumente aber spezielle Items für Betreuungen in den verschiedenen Alterssubgruppen innerhalb dieser Altersspanne. Im allgemeinen wird dabei ein besonderer Schwerpunkt auf die Betreuungssituation der jüngsten Kinder gelegt. Darin spiegelt sich, daß Tagespflege eher eine Betreuungsform für Kinder unter drei Jahren als für ältere Kinder darstellt.

Die meisten Instrumente zur Beurteilung von Tagespflegestellen nennen als ihren theoretischen Hintergrund die CDA - Competency Standards bzw. verweisen auf eine Übereinstimmung ihres Konzepts mit diesen Standards. Für das Assessment Profile for Family Day Care fehlen explizite Angaben zur theoretischen Orientierung. Wie bei den Instrumenten zur Beurteilung institutioneller Gruppenbetreuungen wäre somit auch hier eine explizitere theoretische Fundierung der Instrumente wünschenswert. Dasselbe gilt auch für HOME.

Alle Instrumente sind für den Einsatz in Fremdevaluationsprozessen geeignet. Das Assessment Profile for Family Day Care, die FDCRS sowie das Louise Instrument (Evaluating Home Based Day Care) können ebenfalls in Selbstevaluationsprozessen eingesetzt werden. Die Instrumente werden von Laien (z.B. Tagespflegebetreuerin)

und von trainierten Beobachtern angewandt. Dem Berichtersteller liegen nur wenige Informationen über ein Training im Umgang mit diesen Instrumenten vor. Allerdings dürfte sich insgesamt der notwendige Trainingsumfang in Grenzen halten.

#### - *Aufbau*

Die Instrumente sind wiederum unterschiedlich lang (32 Items bei der FDCRS bis 277 Items beim Assessment Profile for Family Day Care) und in unterschiedlich viele Subbereiche differenziert (sechs Subbereiche bei FDCRS sowie HOME bis zu 11 Subbereiche beim CDA Assessment Observation Instrument). Auch hier zeigt sich die bereits schon erwähnte negative Beziehung zwischen Anzahl Items und Komplexität der Items: Instrumente mit einer geringen Anzahl von Items (z.B. die FDCRS mit 32 Items oder das CDA Assessment Observation Instrument mit 43 Items) haben zum Teil sehr komplexe Items, die hohe Anforderungen an die Informationsverarbeitungskapazitäten der Beobachter stellen. Demgegenüber haben Instrumente mit einer hohen Anzahl von Items - z.B. das Assessment Profile for Family Day Care mit 277 Items oder das Louise Instrument (Evaluating Home Based Day Care) mit 206 Items - eher Items, die im Sinne von Checklisten aufgebaut sind bzw. die weniger Anforderungen an zusammenfassende Urteile durch den Beobachter stellen. Allerdings ergeben sich hier im Vergleich zu den Instrumenten für institutionelle Gruppenbetreuungen größere Unterschiede im Zeitaufwand für die Anwendung der Instrumente. Für HOME wird bei 45 bzw. 55 Items, die vom Beobachter ein mittleres Abstraktionsniveau verlangen, nur etwa eine Stunde benötigt, für das CDA Assessment Observation Instrument und die FDCRS etwa zwei bis drei Stunden. Deutlich mehr Zeit - um vier bis sechs Stunden - müssen für das Assessment Profile for Family Day Care und das Family Day Home Observation Instrument veranschlagt werden. Über die für die 206 Items des Louise Instruments (Evaluating Home Based Day Care) benötigte Zeit liegen dem Berichtersteller keine Informationen vor; vermutlich dürfte aber auch für dieses Instrument ein höherer Zeitaufwand erforderlich sein. Mit vier bis sechs Stunden Bearbeitungszeit dürfte auch die Obergrenze für ein Instrument erreicht sein, das als zeitökonomisch anwendbar zu betrachten ist.

Hauptinformationsquelle für die Bearbeitung der Instrumente sind wiederum Beobachtungen. Außer bei dem CDA Assessment Observation Instrument und dem Louise Instrument (Evaluating Home Based Day Care) werden noch Befragungen der Tagespflegebetreuerinnen bzw., im Fall von HOME, der Mütter erforderlich. Das Assessment Profile for Family Day Care und das Family Day Home Observation Instrument sehen noch die Durchsicht von schriftlichen Unterlagen vor. Auf die sich für ein Training ergebenden Konsequenzen wurde schon weiter oben verwiesen. Mit Ausnahme des CDA Assessment Observation Instruments (vgl. die Ausführungen bei

den Instrumenten für Gruppenbetreuungen) sehen alle Instrumente Informationszusammenfassungen zu Subskalen bzw. zu Gesamtwerten vor. Zusätzlich werden für alle diese Instrumente quantitative Qualitätsniveaus angegeben, d.h. Bereiche für Subskalenwerte oder Gesamtwerte, die unterschiedliche Qualitätsniveaus indizieren. Eine Ausnahme bildet die FDCRS; hier kann aber bei Durchschnittsbildungen auf Subskalen- oder Gesamtwertebene als Bewertungsmaßstab die ursprüngliche siebenstufige Ratingskala herangezogen werden. Solche Zusammenfassungen der Informationen zu Skalen sowie die explizite Angabe von Qualitätsniveaus ermöglichen eine zusammenfassende und verobjektivierte Qualitätsbeurteilung. Graphische Hilfen für die Interpretation der Ergebnisse gibt es nur beim Assessment Profile for Family Day Care.

### **Gütekriterien**

Die Angaben zu den technischen Gütekriterien der in Tabelle 2 enthaltenen Instrumente fallen noch dürftiger aus als bei den Instrumenten zur Beurteilung institutioneller Gruppenbetreuungen. Für das CDA Assessment Observation Instrument liegen keinerlei Angaben zur technischen Qualität vor. Einige Informationen - insgesamt aber nur sehr eingeschränkt - sind für das Assessment Profile for Family Day Care, das Family Day Home Observation Instrument und das Louise Instrument (Evaluating Home Based Day Care) vorhanden. Ausführlichere Informationen liegen nur für die FDCRS and für HOME vor. Für alle Instrumente sind aber weitere Analysen zur technischen Qualität erforderlich.

Die Angaben zur Objektivität - soweit sie vorliegen - weisen auf hohe Beobachterübereinstimmungen hin (über 80 %), wobei allerdings für das Family Day Home Observation Instrument und das Louise Instrument (Evaluating Home Based Day Care) die Beobachterübereinstimmungen nur bei sehr wenigen Tagespflegestellen und zudem bei dem zweiten Instrument nur bei einer Vorläuferversion überprüft wurden. Für die FDCRS liegen Angaben zu den Korrelationen der Meßwerte von zwei Beobachtern aus verschiedenen Studien vor. Bei Korrelationen von .80 und mehr ist von einer guten Objektivität auszugehen.

Die Reliabilität wurde nur bei der FDCRS und bei HOME analysiert. Die internen Konsistenzen für die jeweiligen Gesamtwerte liegen um .90, wobei die beiden Instrumente mit 32 bis 55 Items eher als mittellange Instrumente zu charakterisieren sind. Bei den Subskalen liegen die internen Konsistenzen zum Teil niedriger, was nicht nur auf geringere Itemzahlen, sondern vermutlich auch auf Inkonsistenzen auf der Skalenebene zurückzuführen ist. Insgesamt aber dürften die berichteten Werte auf eine zufriedenstellende interne Konsistenz verweisen. Darüber hinaus werden für HOME Test-Retest-Reliabilitäten berichtet. Diese liegen für den Gesamtwert bei .62

bis .77, auf Subskalenebene bei einigen Skalen jedoch deutlich niedriger. Dabei sollte aber auch der Zeitabstand der beiden Messungen (bis zu 1 1/2 Jahren Abstand) berücksichtigt werden.

Im Hinblick auf die inhaltliche Validität stellt das CDA Assessment Instrument eine direkte Operationalisierung der CDA - Competency Standards dar; drei der anderen Instrumente zur Beurteilung von Tagespflegestellen nennen einen engen Bezug zu diesem Satz von Qualitätskriterien. Eine Absicherung der inhaltlichen Validität über Expertenbefragungen liegt jedoch bei keinem Instrument vor. Untereinander korrelieren das Assessment Profile for Family Day Care, die FDCRS und das Family Day Home Observation Instrument moderat bis hoch (.62 bis .82). Diese Korrelationen verweisen auf eine gute kriteriumsorientierte Validität, wobei zugleich die verschiedenen Instrumente jeweils unterschiedliche Qualitätsaspekte einer Tagespflegestelle thematisieren. Die FDCRS, das Family Day Home Observation Instrument und das Louise Instrument (Evaluating Home Based Day Care) korrelieren weiterhin in einer befriedigenden Höhe mit globalen Qualitätseinschätzungen, wobei allerdings zum größten Teil die gleichen Personen sowohl das globale Rating abgaben als auch das jeweilige Instrument anwandten. Eine eher niedrige Korrelation zeigt sich noch zwischen FDCRS und HOME. Hierbei ist aber zu berücksichtigen, daß HOME hier nicht für häusliche Umwelten, sondern für Tagespflegestellen eingesetzt wurde und daß beide Instrumente konzeptuell unterschiedlich sind. HOME und FDCRS sind auch die einzigen Instrumente, für die Informationen zum Zusammenhang mit anderen Konstrukten vorliegen. Beide Instrumente erweisen sich hier im Sinne einer Konstruktvalidität als bewährte Instrumente.

Analysen zur Aufdeckung der Faktorenstruktur wurden bei keinem der Instrumente durchgeführt, wodurch sich im Hinblick auf empirische Qualitätsdimensionen von Tagespflegebetreuungen ein ausgesprochenes Defizit ergibt.

Im Hinblick auf eine Gesamtbeurteilung der technischen Qualität der in Tabelle 2 angesprochenen Instrumente läßt sich festhalten, daß umfangreichere Informationen nur für FDCRS und HOME vorliegen. Bei beiden Instrumenten kann insgesamt von einer zufriedenstellenden technischen Qualität ausgegangen werden, wobei aber auch hier noch weitere Analysen erforderlich sind (z.B. im Hinblick auf Faktorenstrukturen).

## Literatur

- Abbott-Shim, M. & Sibley, A. (1987a). *Assessment Profile for Early Childhood Programs: Manual, Administration*. Atlanta, GA: Quality Assists.
- Abbott-Shim, M. & Sibley, A. (1987b). *Assessment Profile for Early Childhood Programs: Preschool, Infant, School-age*. Atlanta, GA: Quality Assists.
- Abt. (1980). *National Day Care Study Documentation and Data*. Cambridge, MA: Abt Associates.
- Bowlby, J. (1951). *Maternal Care and Mental Health*. Geneva: World Health Organization.
- Bredenkamp, S. (1986). The Reliability and Validity of the Early Childhood Classroom Observation Scale for Accrediting Early Childhood Programs. *Early Childhood Research Quarterly*, 1, 103 - 118.
- Bredenkamp, S. (1989). *Measuring Quality Through a National Accreditation System for Early Childhood Programs*. Paper presented at the AERA Annual Conference, San Francisco, CA.
- Bredenkamp, S. (Ed.). (1987). *Developmentally Appropriate Practice in Early Childhood Serving Children From Birth Through Age 8*. Washington, D.C.: National Association for the Education of Young Children.
- Bronfenbrenner, U. & Crouter, A.C. (1983). The Evolution of Environmental Models in Developmental Research. In P.H. Mussen (Ed.), *Handbook of Child Psychology (4<sup>th</sup> Ed.)*. Vol. 1 (ed. by W. Kessen) (pp. 357 - 414). New York: Wiley.
- Bronfenbrenner, U. (1981). *Die Ökologie der menschlichen Entwicklung. Natürliche und geplante Experimente*. Stuttgart: Klett-Cotta.
- Bundesminister für Bildung und Wissenschaft (Hrsg.). (1990). *Grund- und Strukturdaten 1990/91*. Bad Honnef: Bock.
- Bundesministerium für Jugend, Familie und Gesundheit (Hrsg.). (1980). *Das Modellprojekt "Tagesmütter" - Abschlußbericht der wissenschaftlichen Begleitung*. Stuttgart: Kohlhammer.
- Caldwell, B.M. & Bradley, R.H. (1984). *Home Observation for Measurement of the Environment*. Little Rock, AK: University of Arkansas at Little Rock, College of Education, Center for Child Development and Education.
- Child Care Partnership of Dallas. (1987). *Family Day Home Observation Instrument*. Dallas, TX: Autor.
- Clifford, R.M., Harms, T., Pepper, S. & Stuart, B. (in Druck). Assessing Quality in Family Day Care. In D.L. Peters & A.R. Pence (Ed.), *Family Day Care: Current Research for Informed*

*Public Policy*. New York: Teachers College Press.

Clifford, R.M., Russel, S.D., Fleming, J., Peisner, E.S., Harms, T. & Cryer, D. (1989). *Infant/Toddler Environment Rating Scale. Reliability and Validity Study. Final Report*. Chapel Hill, NC: Frank Porter Graham Child Development Center, University of North Carolina.

*Competence. News for the CDA Community*, 8 (2). (1991).

*Competence. News for the CDA Community*, 9 (1). (1992).

Council for Early Childhood Professional Recognition. (1990). *Preschool Caregivers in Center-Based Programs. Child Development Associate Assessment System and Competency Standards*. Washington, D.C.: Autor.

Council for Early Childhood Professional Recognition. (1991a). *Family Day Care Providers. Child Development Associate Assessment System and Competency Standards*. Washington, D.C.: Autor.

Council for Early Childhood Professional Recognition. (1991b). *Advisor Report Form*. Washington, D.C.: Autor.

Council for Early Childhood Professional Recognition. (1991c). *Infant/Toddler Caregiver in Center-Based Programs. Child Development Associate Assessment System and Competency Standards*. Washington, D.C.: Autor.

Council for Early Childhood Professional Recognition. (1992). *CDA Assessment Observation Instrument for CDA Advisors*. Washington, D.C.: Autor.

Doherty, G. (1991). *Factors Related to Quality in Child Care: A Review of Literature*. Ontario: Queen's Printer for Ontario.

Fiene, R. & Melnick, S.A. (1991). *Quality Assessment in Early Childhood Programs. A Multi-Dimensional Approach*. Paper presented at the AERA Annual Conference, Chicago, IL.

Frede, E.C. & Miller, A.K. (1989). *The Preschool Classroom Implementation Rating Instrument: High Scope*. The Virginia Department of Education.

Frede, E., Miller, A.K., Eads, G., Lee, B., MacNeill, C. & Barton, A. (1991). *A Comparison of Early Classroom Practices*. Paper presented at the AERA Annual Conference, Chicago, IL.

Fried, L., Roßbach, H.G., Tietze, W. & Wolf, B. (1992). Elementarbereich. In K. Ingenkamp, R.S. Jäger, H. Petillon & B. Wolf (Hrsg.), *Empirische Pädagogik von 1970 bis 1990. Eine Bestandsaufnahme der Forschung in der Bundesrepublik Deutschland. Bd. I* (S. 197 - 263). Weinheim: Deutscher Studien Verlag.

Goelman, H. & Pence, A.R. (1987). Effects of Child Care, Family, and Individual Characteristics on Children's Language Development: The Victoria Day Care Research Project. In D.A. Phillips (Ed.), *Quality in Child Care: What Does Research Tell Us?* (S. 89 - 104). Washington, D.C.:

NAEYC.

Goelmann, H. & Pence, A.R. (1988). Children in Three Types of Day Care: Daily Experiences, Quality of Care and Developmental Outcomes. *Early Child Development and Care*, 33, 67 - 76.

Hagen, A. & Roßbach, H.G. (1987). Zur Erfassung und Beurteilung der pädagogischen Umwelt im Kindergarten. In N. Kluge & L. Fried (Hrsg.), *Spielen und Lernen mit jungen Kindern* (S. 203 - 222). Frankfurt: Peter Lang.

Harms, T. & Clifford, R.M. (1980). *Early Childhood Environment Rating Scale*. New York: Teachers College Press.

Harms, T., Clifford, R.M. & Padan-Belkin, E. (1982). *Day Care Home Environment Rating Scale*. Unpublished manuscript, Frank Porter Graham Child Development Center, University of North Carolina, Chapel Hill, NC.

Harms, T. & Clifford, R.M. (1983). Assessing Preschool Environments with the Early Childhood Environment Rating Scale. *Studies in Educational Evaluation*, 8, 261 - 269.

Harms, T. & Clifford, R.M. (1986). *Rationale Underlying the ECERS*. Unpublished paper, Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill, NC.

Harms, T. & Clifford, R.M. (1989). *Family Day Care Rating Scale*. New York: Teachers College Press.

Harms, T., Cryer, D. & Clifford, R.M. (1990). *Infant/Toddler Environment Rating Scale*. New York: Teachers College Press.

Harms, T. & Clifford, R.M. (in Druck). Studying Educational Settings. In B. Spodek (Ed.), *Handbook of Research in Early Childhood Education*. 2<sup>nd</sup> Edition. New York: Teachers College Press.

Hohmann, M., Baret, B. & Weikart, D. (1978). *Young Children in Action*. Ypsilanti, MI: High/Scope Press.

Howes, C. (1990). Can the Age of Entry into Child Care and the Quality of Child Care Predict Adjustment in Kindergarten? *Developmental Psychology*, 26, 292 - 303.

Howes, C. & Olenick, M. (1986). Family and Child Care Influences on Toddler's Compliance. *Child Development*, 57, 202 - 216.

Howes, C. & Stewart, P. (1987). Child's Play With Adults, Toys, and Peers: An Examination of Family and Child-Care Influences. *Developmental Psychology*, 23, 423 - 430.

Howes, C., Phillips, D.A. & Whitebook, M. (in Druck). Thresholds of Quality: Implications for the Social Development of Children in Center Based Child Care. *Child Development*, 63.



- Hyson, M.C., Hirsh-Pasek, K. & Rescorla, L. (1990). The Classroom Practices Inventory: An Observation Instrument Based on NAEYC's Guidelines for Developmentally Appropriate Practices for 4- and 5-Year-Old Children. *Early Childhood Research Quarterly*, 5, 475 - 494.
- Jones, S.N. & Meisel, S.J. (1987). Training Family Day Care Providers to Work With Special Needs Children. *Topics in Early Childhood Special Education*, 7, 1 - 12.
- Louise Child Care Center. (1988). *Evaluating Home Based Day Care*. Pittsburg, PA: Autor.
- McCartney, K. (1984). Effect of Quality of Day Care Environment on Children's Language Development. *Developmental Psychology*, 20, 244 - 260.
- Mindel, C.H. (1988). *Child Care Partnership Family Day Home Observation Instrument: Reliability and Validity Analysis*. Dallas, TX: Child Care Partnership of Dallas.
- Modigliani, K. (1991). *Assessing the Quality of Family Child Care. A Comparison of Five Instruments*. Hayward, CA: Mervyn's Public Affairs Office.
- Moore, G.T. (1986). Effects of the Spatial Definition of Behavior Settings on Children's Behavior: A Quasi-Experimental Field Study. *Journal of Environmental Psychology*, 6, 205 - 231.
- Moss, P. (1988). *Child Care and Equality of Opportunity. Consolidated Report to the European Commission*. Brussels: Commission of the European Communities.
- Moss, P. (1990). *Childcare in the European Community 1985 - 1990*. European Commission Childcare Network.
- National Academy of Early Childhood Programs. (1985). *Early Childhood Classroom Observation*. Washington, D.C.: National Association for the Education of Young Children.
- National Association for Family Day Care. (ohne Jahr). *Accreditation*. Washington, D.C.: Autor.
- National Association for the Education of Young Children (1991a). *Accreditation Criteria & Procedures of the National Academy of Early Childhood Programs. Revised Edition*. Washington, D.C.: Autor.
- National Association for the Education of Young Children. (1991b). *Guide to Accreditation by the National Academy of Early Childhood Programs. Revised Edition*. Washington, D.C.: Autor.
- Nelson, M.J. (1989). *Concurrent Validity of Three Family Day Care Assessment Instruments*. San Antonio, TX: City of San Antonio Children's. (ERIC ED 312 036).
- Pence, A.R. & Goelmann, H. (1991): The Relationship of Regulation, Training, and Motivation to Quality of Care in Family Day Care. *Child and Youth Care Quarterly*, 20, 83 - 101.
- Phillips, C.B. (1991a). *Field Advisor's Guide for the CDA Professional Preparation Program - a Companion to ESSENTIALS for Child Development Associates Working With Young Children*.

- Washington, D.C.: Council for Early Childhood Professional Recognition.
- Phillips, C.B. (Ed.) (1991b). *Essentials for Child Development Associates Working with Young Children*. Washington, D.C.: Council for Early Childhood Professional Recognition.
- Phillips, D., McCartney, K. & Scarr, S. (1987). Child-Care Quality and Children's Social Development. *Developmental Psychology*, 23, 537 - 543.
- Prescott, E., Kritechsky, S. & Jones, K. (1972). *The Day Care Environment Inventory*. Washington, D.C.: U.S. Department of Health, Education and Welfare.
- Quality Assist. (ohne Jahr). *Assessment Profile for Early Childhood Programs. Validity & Reliability*. Atlanta, GA: Autor.
- Rodriguez, D. (1978). A Tool for Evaluating Family Day Care Mothers. *Child Welfare*, LVII, 55 - 58.
- Rodriguez, D.T. (1982). Assessment of Home Day Care Services. *Child Care Quarterly*, 11, 291 - 297.
- Roßbach, H.G. (1981). *Evaluation von Fortbildungsveranstaltungen für Erzieher. Zur Entwicklung und Erprobung eines Fragebogens*. Münster: Lit.
- Roßbach, H.G. (1990). Assessing the Quality of Kindergarten Environments with the Early Childhood Environment Rating Scale. In J.M. Pieters, K. Breuer & P.R.J. Simons (Eds.), *Learning Environments. Contributions from Dutch and German Research* (S. 77 - 90). Berlin: Springer.
- Roßbach, H.G., Clifford, R.M. & Harms, T. (1991). *Dimensions of Learning Environments: Cross-national Validation of the Early Childhood Environment Rating Scale*. Paper presented at the AERA Annual Conference, Chicago, IL.
- Ruopp, R., Travers, J., Glantz, R. & Coelen, C. (1979). *Children at the Center. Final Report of the National Day Care Study*. Cambridge, M.A.: Abt Associates.
- Sander, C. (1986). *Erprobung einer deutschen Version der Early Childhood Environment Rating Scale*. Unveröffentlichte Staatsexamensarbeit, Fachbereich Erziehungswissenschaft, Universität Münster.
- SAVI. (1979). *Head Start Program Performance Standards - Self Assessment/Validation Instrument*. Washington, D.C.: U.S. Department of Health and Human Services.
- Sibley, A. & Abbott-Shim, M. (1989). *Assessment Profile for Family Day Care. Study Guide*. Atlanta, GA: Quality Assist.
- Sibley, A. & Abbott-Shim, M. (1991). *Assessment Profile for Family Day Care*. Atlanta, GA: Quality Assist.

Stallings, J.A. (1977). *Learning to Look. A Handbook on Classroom Observation and Teaching Models*. Belmont, CA: Wadsworth Publishing Company.

Stallings, J.A., Wilcox, M. & Travers, J. (1976). *Phase II Instruments for the National Day Care Cost-Effects Study: Instrument Selection and Field Testing*. Menlo Park, CA: Stanford Research Institute.

Stuart, B. & Pepper, S. (1988). The Contribution of the Caregivers' Personality and Vocational Interest to Quality in Licenced Family Day Care. *Canadian Journal of Research in Early Childhood Education*, 2, 99 - 109.

Tietze, W. (1985). *Sozialisationsrelevante Strukturbedingungen in frühkindlichen Betreuungsformen - Konzeptionelle Überlegungen zur Erfassung von Settingstrukturen in der IEA-Preprimary Study*. Unveröffentlichtes Manuskript, Fachbereich Erziehungswissenschaft, Universität Münster.

Tietze, W. & Roßbach, H.G. (1986). *Bewertungsanleitung zur Einschätzung der Ausstattung und pädagogischen Arbeit in Kindergärten*. Deutsche Erprobungsfassung der Early Childhood Environment Rating Scale von Th. Harms und R.M. Clifford. Unveröffentlichtes Manuskript, Fachbereich Erziehungswissenschaft, Universität Münster.

Tietze, W. & Roßbach, H.G. (1991). Die Betreuung von Kindern im vorschulischen Alter. *Zeitschrift für Pädagogik*, 37, 555 - 579.

Tietze, W., Roßbach, H.G. & Jakobs, H. (in Vorbereitung). *Die Betreuungssituation junger Kinder in der Bundesrepublik Deutschland*. (Arbeitstitel).

Tsiantis, J., Caldwell, B., Dragonas, T., Jegedo, R.O., Lambidi, A., Bonaag, C. & Orley, J. (1991). Development of a WHO Child Care Facility Schedule (CCFS): a Pilot Collaborative Study. *Bulletin of the World Health Organization*, 69, 51 - 57.

Whitebook, M., Howes, C. & Phillips, D.A. (1990). *Who cares? Child Care Teachers and the Quality of Care in America. The National Child Care Staffing Study*. Oakland, CA: Child Care Employee Project.