

EVALUATION PSYCHOLOGISCHER ENTSCHEIDUNGSHILFEN
FÜR FAMILIENGERICHTE: FAMILIENPSYCHOLOGISCHE GUTACHTEN
NACH TRENNUNG UND SCHEIDUNG

EVALUATION OF PSYCHOLOGICAL DECISION AIDS
FOR FAMILY COURTS: FAMILY PSYCHOLOGICAL ASSESSMENTS
AFTER SEPERATION AND DIVORCE

PAPIER ZUR FACHTAGUNG *PSYCHOLOGIE IM FAMILIENRECHT.*
BILANZ ODER NEUORIENTIERUNG. EIN WORKSHOP.
EVANGELISCHE AKADEMIE BAD BOLL, 9. - 11. DEZEMBER 1998

Werner G. Leitner, Dr. phil., Professur (Vertretung) für Entwicklungspsychologie am Institut für Pädagogische Psychologie und Entwicklungspsychologie der Technischen Universität Dresden, Weberplatz 5, 01062 Dresden

Die vorliegende Studie versucht 52 familienpsychologische Gutachten im Hinblick auf die grundlegende Methodik zu evaluieren. Vorab sei folgendes angemerkt:

1. Festgestellte Qualitätsdefizite sollten nicht dazu führen, daß Familiengerichte auf Entscheidungshilfen durch Sachverständige künftig vermehrt verzichten. Vielmehr wäre in einzelnen Gutachten systematische Revision zur Qualitätssicherung zu intensivieren.
2. Unabhängig von der vorliegenden Stichprobe sind auch einzelne Gutachten bekannt, die sich qualitativ sehr deutlich von der breiten Masse abheben.
3. Aus den Ergebnissen lassen sich Hypothesen generieren, die eine einstweilige Standortbestimmung umreißen. Es wird weiterführenden Studien vorbehalten sein, die im Hinblick auf Repräsentativität und Stichprobengröße weitergehende Anforderungen erfüllen, diese einstweilige Standortbestimmung zu falsifizieren oder zu verifizieren.

Nach den feststehenden Richtlinien für die Erstellung psychologischer Gutachten (FÖDERATION DEUTSCHER PSYCHOLOGENVEREINIGUNGEN, 1994, S.8) ist ein solches psychologisches Gutachten eine wissenschaftliche Leistung, die darin besteht, aufgrund wissenschaftlich anerkannter Methoden und Kriterien nach feststehenden Regeln der Gewinnung und Interpretation von Daten, zu konkreten Fragestellungen Aussagen zu machen.

Von daher muß ein solches Gutachten einigen notwendigen, wenngleich damit noch nicht hinreichenden, Bedingungen elementarer wissenschaftlicher Anforderungen genügen. In diesem Zusammenhang sei auch angemerkt, daß sich eine psychologische Begutachtung nicht nur auf einzelne, sondern stets auf mehrere wissenschaftlich haltbare Datenquellen stützen muß (vgl. FÖDERATION DEUTSCHER PSYCHOLOGENVEREINIGUNGEN, 1994, S. 11).

Die vorliegende Studie soll nach einigen qualitativen Vorbemerkungen im Hinblick auf diese Grundforderungen und deren Realisierung mit einer quantitativen Analyse vor allem darüber Aufschluß geben, inwieweit die testdiagnostische Basis, also die eingesetzten standardisierten Tests bei einer vorliegenden Stichprobe von N=52 Gutachten, den mit den Hauptgütekriterien (Reliabilität, Validität, Objektivität) beziehungsweise dem Nebengütekriterium Normierung vorliegenden elementaren Standards im Hinblick auf die Aussagegültigkeit entspricht.

Die Stichprobe basiert auf einer bundesweiten Befragung. Die Befragung erstreckte sich über den Zeitraum zwischen 1994 und 1998. Daraus ergab sich eine Stichprobe von insgesamt 52 Gutachten, von denen 50% aus Bayern und 50% aus anderen Bundesländern (einschließlich "neuer Länder") stammen. Beim Vergleich dieser beiden länderspezifischen Teilstichproben ergaben sich keine nennenswerten Unterschiede, so daß nicht anzunehmen ist, daß die erzielten Ergebnisse in besonderer

Weise durch die Zusammensetzung der Stichprobe im Hinblick auf eine bundesländerspezifische Verteilung beeinflusst waren.

Eine zufällig zustandgekommene, symmetrische Zusammensetzung der Stichprobe ergibt sich auch im Hinblick auf die Organisationsstruktur, wie eine der nachfolgenden Graphiken zeigt. Es wurde im Rahmen der Befragung darum gebeten, entsprechende Kopien des Originalgutachtens zur Verfügung zu stellen, so daß die Ergebnisse auch in dieser Weise belegbar sind.

Qualitative Analyse

Aktenanalyse, Exploration und Verhaltensbeobachtung

Damit im Hinblick auf diese Datengrundlagen von wissenschaftlichem Arbeiten gesprochen werden kann, sollte die Verarbeitung auf diesen Datenebenen mit angegebenen systematischen Methoden erfolgt sein. Eine nur beiläufige Verhaltensbeobachtung beispielsweise würde nicht als systematische Methode angesehen werden, wie nachfolgend noch näher ausgeführt wird.

Fragen

* Wurde eine systematische Aktenanalyse auf der Basis eines Kategoriensystems durchgeführt oder erfolgte die Sichtung und Darstellung der vorliegenden Akten unsystematisch und offensichtlich im freien Ermessen des Gutachters?

* Wurden die Explorationsgespräche auf der Grundlage eines systematischen Gesprächsleitfadens durchgeführt und / oder mit einem Kategoriensystem erfaßt und ausgewertet oder werden in dieser Hinsicht keine konkreten Methoden genannt?

* Wurde eine systematische (z.B. auf der Basis einer Rating Skala, eines Kategoriensystems) oder nur eine beiläufige (unsystematische, naive) Verhaltensbeobachtung durchgeführt?

Mit DORSCH et al. (1994, S. 100) ergibt sich folgendes: "Die Systematik der Verhaltensbeobachtung wird nach mehreren Gesichtspunkten vorgenommen (1) systematische (wissenschaftliche) Beobachtung - unsystematische (naive, unwissenschaftliche) Beobachtungen". Daraus geht hervor, daß eine beiläufige oder unsystematische Verhaltensbeobachtung keine wissenschaftliche Methode repräsentiert, da sie u.a. "Beobachtungs-Fehler" (vgl. DORSCH et al., 1994, S.100) nicht systematisch auszuschließen versucht. Bei der Aktenanalyse und bei der Exploration verhält sich dies ähnlich.

Bei kaum einem Gutachten der vorliegenden Stichprobe läßt sich im Hinblick auf die Aktenanalyse der nachvollziehbare Einsatz eines hierfür entwickelten wissenschaftlichen Verfahrens erkennen. Auch im Hinblick auf die Explorationsgespräche kommen entsprechende wissenschaftliche Ansätze, wie sie beispielsweise mit einem Gesprächsleitfaden von WESTHOFF und KLUCK (1991/1992) vorliegen, explizit nicht zum Einsatz. Auf welches wissenschaftlich - methodisches Paradigma sich die Gutachterin / der Gutachter jeweils stützt, wird nur sehr selten explizit thematisiert. Wissenschaftlicher Theoriebezug mit entsprechenden Quellenangaben (abgesehen von den Tests, worauf später noch ausführlich einzugehen sein wird) beziehungsweise der Verweis auf relevante Befunde wird nahezu durchgängig unzureichend transparent gemacht oder ist nicht ersichtlich.

So wird z.B. ein eventueller Einsatz einer entsprechenden Rating- Skala in keinem der N=52 Gutachten nachvollziehbar dargelegt.

Eine systematische (wissenschaftliche) Verhaltensbeobachtung ist bei keinem der Gutachten dieser Stichprobe erkennbar. In relativ unstrukturierter Form erscheinen Elemente allenfalls unwissenschaftlicher (unsystematischer) Verhaltensbeobachtung, die als wissenschaftliche Verfahrensweise nicht gewertet werden können und Beobachtungs-Fehler nicht systematisch auszuschließen versuchen.

Insgesamt gesehen sind bei der vorliegenden Stichprobe weder bei der Aktenanalyse noch bei der Exploration und den Verhaltensbeobachtungen wissenschaftliche Methoden in einer Form vorhanden, die in ihrer Gesamtheit wissenschaftlichen Grundforderungen annähernd gerecht werden.

Literarische Verarbeitungsprinzipien einer wissenschaftlichen Arbeit

Ein familienpsychologisches Gutachten sollte auch im Hinblick auf literarische Gestaltungsprinzipien elementare wissenschaftliche Grundforderungen erfüllen. So besteht eine unabdingbare Forderung unter anderem darin, daß im Gutachten umfangreiche Literatur- beziehungsweise Quellenangaben auch über die den Interpretationen zugrundeliegenden Theorien und Konzepte gemacht werden. Wörtliche

oder sinngemäß aus Quellen (Akten, Literatur) entnommene Passagen sind als solche im Text kenntlich zu machen. Ein Gutachten, das solche Grundforderungen wissenschaftlichen Arbeitens nicht beachtet, kann nicht den Anspruch erheben, eine wissenschaftliche Leistung (FÖDERATION DEUTSCHER PSYCHOLOGENVEREINIGUNGEN, 1994, S.8) zu sein, wie dies in den eingangs zitierten Richtlinien der *Föderation Deutscher Psychologenvereinigungen* für ein solches Gutachten (a.a.O.) ausdrücklich gefordert wird.

Insgesamt sind solche Grundforderungen nur bei einem geringen Teil der vorliegenden Gutachten angemessen berücksichtigt. Vielfach entsprechen die, ohnehin nur bei wenigen Gutachten, vorhandenen Literaturangaben allerdings nicht dem neuesten Stand und belegen eher ein Hinterherhinken um mehrere Jahre oder gar ein bis zwei Jahrzehnte hinter dem neuesten Forschungsstand.

Quantitative Analyse

Tests

Der Begriff "Test" wird im umgangssprachlichen Bereich sehr häufig für Gegebenheiten verwendet, die aus wissenschaftlicher Sicht keineswegs mit diesem Terminus belegt werden können. Als besonderes Beispiel seien hier die vielfältigen Tests in Illustrierten zu nennen, die bestenfalls Unterhaltungswert besitzen, mit einem Test im wissenschaftlichen Sinne aber wenig zu tun haben. Mit LIENERT (1969, zitiert in DORSCH, 1987, S.681 bzw. DORSCH, 1994, S.793f) läßt sich über die Testgütekriterien eine klare Entscheidung treffen, ob es sich im Einzelfall um einen Test im wissenschaftlichen Sinne handelt oder nicht: Nach LIENERT (1969) kann man zwischen Haupt- und Nebengütekriterien unterscheiden. Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) stellen die Hauptgütekriterien dar. Als Nebengütekriterien gelten Ökonomie, Nützlichkeit, Normierung (Normskalen), Vergleichbarkeit. Testverfahren, welche diese Gütekriterien nicht aufweisen, können nicht als Test im eigentlichen Sinne bezeichnet werden, da ihnen die mit wissenschaftlichen Methoden überprüften Gütekriterien wissenschaftlicher Testverfahren Grundlagen und die für eine Testkonstruktion notwendigen Kontrolluntersuchungen fehlen."

Die Gütekriterien sind mit LIENERT (1969) von elementarster Bedeutung dafür, ob ein Verfahren als solches im wissenschaftlichen Sinne überhaupt als Test bezeichnet werden kann. Letztlich sind sie also ganz elementare Standards dafür, ob oder inwieweit die mit seiner Hilfe gemachten Aussagen wissenschaftliche Aussagefähigkeit besitzen. Die Gütekriterien lassen sich untereinander im Hinblick auf ihre Wertigkeit in Haupt- und Nebengütekriterien unterteilen.

Objektivität

In der angloamerikanischen Literatur wird Objektivität im Sinne eines Testgütekriteriums nicht immer einheitlich verwendet. So hat CATTELL (vgl. DORSCH et al., 1994, S.525) beispielsweise diejenigen Testverfahren als "objektive Tests" bezeichnet, die von ihrer Absicht her nicht durchschaubar sind. Manche Autoren legen diesbezüglich ihr Augenmerk auf die Objektivität als Einheitlichkeit der Testvorlage. Von LIENERT (1969, zitiert in DORSCH et al., 1994, S.525) wurde die Objektivität als Testgütekriterium für den deutschsprachigen Raum definiert als "Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind" (a.a.O.). Bei dieser Definition steht die "interpersonelle Übereinstimmung" (a.a.O.) im Vordergrund. Diese Übereinstimmung spielt im Hinblick auf die Durchführung (Durchführungsobjektivität), die Auswertung (Auswertungsobjektivität) und die Interpretation eine Rolle.

Reliabilität

Bei der Reliabilität handelt es sich um die Zuverlässigkeit einer Meßmethode beziehungsweise eines standardisierten Testverfahrens, die analysiert, mit welchem Grad der Genauigkeit ein solches Verfahren Ergebnisse liefert. Die Meßgenauigkeit wird dabei unter diesem Aspekt unabhängig von der Gültigkeit seiner Resultate untersucht. Beim Konzept der Reliabilität wird davon ausgegangen, daß jede Messung gewissen Meßfehlern unterliegt, wobei die Ergebnisse aus der Sicht der klassischen Testtheorie einen wahren und einen Fehlervarianz - Anteil enthalten. Als Reliabilitätskoeffizient läßt sich der Quotient zwischen wahrer Varianz und gesamter Varianz festlegen. Im Hinblick auf die Methoden der Reliabilitätsermittlung lassen sich mit der Retest - Methode, der Paralleltest - Methode und der Testhalbierungs - Methode verschiedene Aspekte der Zuverlässigkeit unterscheiden (vgl. DORSCH et al., 1994, S.656).

Validität

Im Gegensatz zur Reliabilität wird bei der Validität (Gültigkeit) über den Grad der Genauigkeit einer Messung hinaus untersucht, inwieweit ein Verfahren tatsächlich die Verhaltensweise oder das Merkmal mißt, die/das es zu messen vorgibt (vgl. LIENERT, 1969. zitiert in DORSCH et al. 1994. S.838) . Hierbei lassen sich bestimmte Validitätsarten unterscheiden: inhaltliche Validität, kriteriumsbezogene Validität sowie Konstruktvalidität (vgl. DORSCH et al., 1994, S.838).

Normierung

Ein wichtiges Nebengütekriterium neben den drei Hauptgütekriterien Objektivität, Reliabilität und Validität ist die Normierung. Aus testtheoretischer Sicht handelt es sich bei dementsprechend vorhandenen Normenskalen um "standardisierte Testskalen zur ökonomischen Vergleichbarkeit von Testwerten" (DORSCH et al., 1994. S.519). Solche Vergleichsmaßstäbe lassen sich auf Grund der Rohwertverteilungen erstellen und bilden eine wichtige Grundlage für die Befunderstellung, deren Brauchbarkeit über Güte und Wert eines Testverfahrens nicht unwesentlich mitentscheidet.

Erfüllen die eingesetzten Tests der vorliegenden Stichprobe die Gütekriterien?

Häufigkeitsverteilungen der eingesetzten Tests bei N=52 familienpsychologischen Gutachten: Es wurden nur die Tests in die Auswertung aufgenommen, die in mindestens zwei unterschiedlichen Gutachten zum Einsatz kamen. Die nachfolgenden Übersichten geben Aufschluß über die Häufigkeit der einzelnen Verfahren im Blickfeld einer Differenzierung im Hinblick auf Herkunft und Organisationsstruktur.

N = 52 Gutachten
länderspezifische Differenzierung

Testabkürzungen in der Rangfolge der Anwendungshäufigkeit

1. FiT (16 Anwendungen)

Familie in Tieren (BREM-GRASER, 1995)

Gütekriterien nach BRICKENKAMP (1997, S.964, S.968ff)

Objektivität: nein

Reliabilität: nein

Validität: nein

Normierung: teilweise

2. FRT (16 Anwendungen)

Family-Relations-Test (BENE &, ANTHONY, 1957)

Gütekriterien nach BRICKENKAMP (1975 und 1983, 1997)

nicht verzeichnet

3. DÜSS (13 Anwendungen)

Fabelmethode (DÜSS, 1964)

Gütekriterien nach BRICKENKAMP (1975, S.520ff)

Objektivität: nein

Reliabilität: nein

Validität: nein

Normierung: ja

4. SERGT (11 Anwendungen)

Satzergänzungstest

Gütekriterien nach BRICKENKAMP (1975 und 1983, 1997)

nicht verzeichnet

5. CAT (8 Anwendungen)

Kinder-Apperzeptions-Test (BELLAK & BELLAK 1955)

Gütekriterien nach BRICKENKAMP (1997, S.942, S.949ff)

Objektivität: nein

Reliabilität: nein

Validität: nein

Normierung: nein

6. MZT (7 Anwendungen)

Mann-Zeichen-Test (ZILER, 1996)

Gütekriterien nach BRICKENKAMP (1997, S.964, S.973ff)

Objektivität: ja

Reliabilität: teilweise

Validität: ja

Normierung: ja

7. SCHLOSS (7 Anwendungen)

Schloss-Zeichen-Test

Gütekriterien nach BRICKENKAMP (1975 und 1983, 1997)

nicht verzeichnet

8. SCENO (7 Anwendungen)

Scenotest (VON STAABS, 1992)

Gütekriterien nach BRICKENKAMP (1997, S.964f, S.981)

Objektivität: teilweise

Reliabilität: teilweise

Validität: teilweise

Normierung: nein

9. FPI (5 Anwendungen)

Freiburger Persönlichkeitsinventar (FAHRENBERG, HAMPEL & SELG, 1994)

Gütekriterien nach BRICKENKAMP (1997, S.505, S.557f)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

10. FBT (5 Anwendungen)

Familien - Beziehungs - Test (HOWELLS & LICKORISH, 1994)

Gütekriterien nach BRICKENKAMP (1997, S.942, S.945f)

Objektivität: nein

Reliabilität: nein

Validität: ja

Normierung: nein

11. HAWIK (4 Anwendungen)

Hamburg - Wechsler - Intelligenztest für Kinder (TEWES, 1985)

Gütekriterien nach BRICKENKAMP (1997, S.49, S. 120ff)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

12. SF (4 Anwendungen)

Schwarzfuß - Test (CORMAN, 1992)

Gütekriterien nach BRICKENKAMP (1997, S.942, S.955ff)

Objektivität: nein

Reliabilität: nein

Validität: nein

Normierung: nein

13. VF (3 Anwendungen)

Verzauberte Familie (KOS & BIERMAN, 1995)

Gütekriterien nach BRICKENKAMP (1997, S.964, S.985f)

Objektivität: teilweise

Reliabilität: nein

Validität: nein

Normierung: teilweise

14. CFT 1 (3 Anwendungen)

Culture Fair Intelligence Test - Scale 1 (CATTELL, 1980)

Gütekriterien nach BRICKENKAMP (1997, S.49, S. 103ff)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

15. HANES (2 Anwendungen)

Hamburger Neurotizismus - und Extraversionsskala für Kinder und Jugendliche (BUGGLE & BAUMGÄRTEL, 1975)

Gütekriterien nach BRICKENKAMP (1997, S.505, S.561f)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

16. PFT (2 Anwendungen)

Rosenzweig-P-F-Test (ROSENZWEIG, 1957)

Gütekriterien nach BRICKENKAMP (1997, S.942, S.951f)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

17. CPM (2 Anwendungen)

Coloured Progressive Matrices (BECKER, SCHALLER & SCHMIDTKE, 1980)

Gütekriterien nach BRICKENKAMP (1997, S. 5 11 S. 192f)

Objektivität: ja

Reliabilität: ja

Validität: ja

Normierung: ja

18. TAT (2 Anwendungen)

Thematischer Apperzeptionstest (MURRAY, 1991)

Gütekriterien nach BRICKENKAMP (1997, S.942, S.957ff)

Objektivität: ja

Reliabilität: ja

Validität: nein

Normierung: nein

Anmerkungen zum Family - Relations - Test (FRT)

Das zusammen mit dem im Hinblick auf die Gütekriterien völlig unzureichendem Test "Familie in Tieren" (BREM-GRÄER, 1995) insgesamt am häufigsten eingesetzte Verfahren, der Family - Relations - Test von BENE und ANTHONY (1957), ist im Testhandbuch von BRICKENKAMP (1997) explizit nicht verzeichnet. Seine Spitzenposition in der Rangfolge verdankt das Verfahren insbesondere der Tatsache, daß er in Gutachten der Gesellschaft für wissenschaftliche Gerichts- und Rechtspsychologie (GWG) ausgesprochen häufig zum Einsatz kommt. Zwölf der insgesamt 16 Anwendungen dieses Verfahrens betreffen solche von der GWG erstellten Gutachten. Insbesondere bei diesem Testverfahren läßt sich erkennen, daß ganz offenbar ausgeprägte organisationspezifische Besonderheiten beim

Einsatz bestimmter Tests nicht von der Hand zu weisen sind. Auf Grund seines häufigen Einsatzes beziehungsweise solchen organisationsspezifischen Besonderheiten, insbesondere im Hinblick auf dieses Testverfahren, sollen hierzu noch einige ergänzende Anmerkungen gemacht werden: Beim FRT handelt es sich um ein Verfahren, daß in einer Übersetzung von FLÄMIG und WÖRNER (1977) in Fassungen für vier- bis fünfjährige sowie für sechs- bis elfjährige Kinder vorliegt (vgl. BEELMANN, 1995, S.38). BEELMANN referierte und diskutierte bei der Tagung der Fachgruppe Entwicklungspsychologie der Deutschen Gesellschaft für Psychologie e.V. in Leipzig im Jahre 1995 über "neuere Untersuchungen mit dem Family - Relations - Test". Hierbei wurde deutlich, daß die Validität dieses Verfahrens zum gegenwärtigen Zeitpunkt keineswegs als gesichert gelten kann. Im Rahmen seines Vortrages und der anschließenden Diskussion bezeichnete BEELMANN den Umgang mit diesem Verfahren in der diagnostischen Praxis als "haarsträubend" und verwies in diesem Zusammenhang unter anderem darauf, daß aus ökonomischen Gründen bei der praktischen Durchführung häufig instruktionsinadäquate Modifikationen vorgenommen werden.

Fazit

Testdiagnostische Basis

Selbst wenn nur eines der Hauptgütekriterien nicht hinreichend gesichert ist, kann dies die Aussagegültigkeit der auf dieser testdiagnostischen Grundlage gemachten Aussagen ganz erheblich in Zweifel ziehen. Bei einer beeinträchtigten Reliabilität (Zuverlässigkeit) fällt dies vor allem dann ins Gewicht, wenn das Verfahren - wie nahezu ausschließlich vorliegend - lediglich im Rahmen einer Statusdiagnose in einem eng begrenzten Zeitraum oder nur zu einem bestimmten Untersuchungszeitpunkt durchgeführt wurde.

Auf dieser Basis gewonnene Ergebnisse können damit nicht den Anspruch hinreichender Aussagegültigkeit im Rahmen einer entscheidungsorientierten Diagnostik erheben, da sie nicht nachweislich gültig, zuverlässig und objektiv sind.

Mit BRICKENKAMP (1997, S.XXIV) "(kann man) vereinfachend... sagen, daß mit abnehmender Zuverlässigkeit und Gültigkeit der hypothetische Charakter der Testergebnisse zunimmt und nach einer Absicherung durch andersartige Ergebnisse verlangt". Andersartige wissenschaftliche begründete und nachvollziehbare aussagegültige Befunde sind den Gutachten der vorliegenden Stichprobe aber ebenfalls kaum zu entnehmen. Der spekulative Charakter entsprechender Aussagen wird vor allem dann deutlich, wenn selbst die Hauptgütekriterien nicht hinreichend gesichert sind. Dies ist selbst oder gar insbesondere bei sehr häufig eingesetzten Testverfahren (z.B. Familie in Tieren: Rangplatz Nr. 1) mit zudem zentraler Position für die Beantwortung vorgegebener Fragestellungen leider sehr häufig der Fall.

Gütekriterienindex

Ein quantitatives Fazit über die Qualität der testdiagnostischen Basis läßt sich auch über die Ermittlung eines Wertes (Gütekriterienindex) für die Erfüllung der Gütekriterien bei den insgesamt 117 Testanwendungen der vorliegenden Stichprobe ziehen. So wurde jeweils ein Punkt für die Erfüllung beziehungsweise ein halber Punkt für die teilweise Erfüllung oder kein Punkt für die Nichterfüllung eines der Gütekriterien Objektivität, Reliabilität, Validität und Normierung (nach BRICKENKAMP, 1997) vergeben. Bei nur einem von 25 der insgesamt 117 Testanwendungen - mit erreichtem Gütekriterienindex von "4" (Maximalwert) beziehungsweise von 3,5 "- kann von gesicherter Aussagegültigkeit ausgegangen werden. In 92 Fällen hingegen lag der entsprechende Index beim Wert "2" oder darunter oder war nach BRICKENKAMP (1975/1983/1997) nicht zu ermitteln.

Gütekriterienindex
N=117 **Testanwendungen**

Nicht nur in Fällen, bei denen unkonventionelle Verfahren zur Anwendung kamen, die in einschlägigen Testhandbüchern nicht verzeichnet sind, sollte es aber Aufgabe der Sachverständigen sein, über die Erfüllung der Gütekriterien der eingesetzten Tests im Gutachten Rechenschaft abzulegen und damit die Aussagegültigkeit der testdiagnostischen Basis auch für das Gericht nachvollziehbar zu erörtern. Dies wäre gleichsam ein ganz wesentlicher Beitrag zur Transparenz der Aussagegültigkeit von

Entscheidungshilfen für das Gericht und zur Qualitätssicherung bzw. Qualitätsverbesserung, die es nachdrücklich anzustreben gilt.

Gutachten insgesamt

Nachdem viele Gutachten der vorliegenden Stichprobe auf relevanten Ebenen ganz elementaren wissenschaftlichen Anforderungen unzureichend genügen, erscheinen sie als Entscheidungsgrundlage für Umgangs- und Sorgerechtsregelungen, die sich auf die künftige Lebenswelt und das Erziehungsumfeld von Kindern sehr nachhaltig auswirken, in der vorliegenden Form als eher inadäquat. Die in diesem Rahmen gemachten Aussagen und Empfehlungen entbehren bisweilen der erforderlichen wissenschaftlichen Grundlage sogar bis hin auf elementarste Forderungen, die an wissenschaftliches Arbeiten gestellt werden müssen.

Neben der unzureichenden wissenschaftlichen Fundierung der psychologischen Untersuchung sind aber auch elementare erziehungswissenschaftliche Grundlagen (erziehungs-wissenschaftliche Theorien und Konzepte, die für solche Fragestellungen sehr bedeutsam wären) den Gutachten der vorliegenden Stichprobe kaum zu entnehmen. Auch neuere Befunde zum Parental-Alienation-Syndrome (PAS) oder gar überhaupt das Konzept an sich (vgl. GARDNER, 1992, vgl. auch neuere Publikationen zu PAS von BAKALAR, 1998, LEITNER 1998, WARD & HARVEY, 1998; KOEPEL & KODJOE, 1998; LEITNER & SCHOELER, 1998) sind in keinem der vorliegenden Gutachten explizit berücksichtigt. Wie der ehemalige Bundesminister der Justiz dem Verfasser in einem persönlichen Schreiben vom 28. Februar 1997 mitteilte, hält er die "Anregung, verstärkt erziehungswissenschaftliche Gutachten einzuholen ... für bedenkenswert". Diese richte sich "allerdings in erster Linie an die Gerichte" (SCHMIDT-JORTZIG, 1997, persönliches Schreiben an den Unterzeichner). "Die Prüfung, ob die Zuziehung eines Sachverständigen mit erziehungswissenschaftlicher Qualifikation oder eines psychologischen Sachverständigen erforderlich ist", obliegt also dem jeweiligen Gericht (SCHMIDT-JORTZIG, 1997). Ratsam wäre vielleicht die Hinzuziehung eines Sachverständigen mit erziehungswissenschaftlicher und psychologischer Qualifikation.

Die in der GWG-Broschüre "Ein Institut stellt sich vor" (GWG, ohne Jahresangabe, S. 16) gemachten Ausführungen mögen angesichts dieser Ergebnisse eher verwundern, zumal 50% der vorliegenden Stichprobe Gutachten der GWG waren. Die von der GWG in der vorliegenden Broschüre vorgegebenen Standards, u.a. "Nachvollziehbarkeit des diagnostischen Prozesses für Betroffene und Auftraggeber", "Transparenz des diagnostischen Vorgehens für die Betroffenen" sowie insbesondere "strikte Beachtung der Zulässigkeit der diagnostischen Verfahren" lassen sich durch die Ergebnisse der vorliegenden Analyse kaum hinreichend bestätigen.

Den Familiengerichten wird ans Herz zu legen sein, Einwendungen gegen solche Gutachten sehr ernst zu nehmen, insbesondere dann, wenn diese durch Expertisen ausgewiesener Fachleute gestützt werden, die die Qualität eines bestimmten Gutachtens sehr deutlich in Zweifel ziehen.

In keinem Fall jedoch sollten diese Ergebnisse dazu verleiten, künftig auf erziehungswissenschaftlichen und psychologischen Sachverstand in diesem sensiblen und für die Zukunft von Kindern aus entwicklungspsychologischer und erziehungswissenschaftlicher Perspektive so einschneidenden Bereich zu verzichten. Es wird insgesamt also weniger auf die Frage hinauslaufen dürfen, ob Sachverständige beauftragt werden sollten als vielmehr darum, wer beauftragt wird. Möglichkeiten einer Evaluation und Qualitätssicherung werden insgesamt nachhaltig zu intensivieren sein.