

# *Familie* und **Recht**

**Zeitschrift für die anwaltliche  
und gerichtliche Praxis**

## **Herausgeber**

Michael Klein  
Dr. Gisela Niemeyer  
Walter Schellhorn  
Pedro Schöppe-Fredenburg  
Peter Schwołow

## **Beirat**

Dr. Peter Finger  
Dr. Peter Gerhardt  
Prof. Dr. Bernd von Heintschel-  
Heinegg  
Roland Klattenhoff  
Dr. Norbert Kleffmann  
Dr. Harald Oelkers  
Christiane Winckelmann

## **Aus dem Inhalt**

- Das Internationale Privatrecht und der Versorgungsausgleich
- Zur Mängelerkennung in familienpsychologischen Gutachten
- Steuerrechtliche Auslegung des gesetzlich nicht definierten Begriffs der Berufsausbildung in der Grundsatzentscheidung des BFH vom 9. 6. 1999 – VI R 33/98

**2·2000**

Seite 49 – 96

Dr. phil. Werner G. Leitner

## Zur Mängelerkennung in familienpsychologischen Gutachten

### Vorbemerkungen

Psychologische Gutachten im Auftrag des Familiengerichts sind sehr häufig Anlaß für Klagen betroffener Elternteile. Zum einen wird über die hohen Kosten geklagt, die für ein solches Gutachten zu tragen sind. Diese Kosten bewegen sich häufig zwischen 5 000 und 10 000 DM.

Geklagt wird häufig aber vor allem über die Qualität, die den Kosten – vielen Klagen zufolge – vielfach auch nicht angemessen erscheint. Gutachter sind in den allermeisten Fällen Diplom-Psychologen. Nur in einer verschwindend geringen Anzahl werden Diplom-Pädagogen als Sachverständige bestellt. Weniger häufig sind auch psychiatrische Gutachten.

Viele solcher Klagen waren Anlaß für die nachfolgende Studie. Es wurden dabei insgesamt 52 familienpsychologische Gutachten untersucht. Nachdem die Gutachten vor Inkrafttreten des neuen Psychotherapeutengesetzes (PsychThG) vom 16. 6. 1998 erstellt worden sind, konnte nicht ermittelt werden, welche der betreffenden Diplom-Psychologen auch die Voraussetzungen für eine Approbation erfüllen oder eine Approbation haben. Nach dem neuen Psychotherapeutengesetz können Absolventen mit universitärem Abschluß der Psychologie eine psychologisch-psychotherapeutische Approbation erwerben. Nur sie führen die den approbierten Psychologen vorbehaltene und geschützte Berufsbezeichnung »Psychologischer Psychotherapeut«. Die Erteilung der Approbation setzt über das abgeschlossene universitäre Studium der Psychologie hinausgehende Qualifikationen voraus. Hierüber sind gegenüber der Approbationsbehörde entsprechende Nachweise (u. a. Fachkundenachweis) zu erbringen.

Der Verfasser dieses Beitrages ist approbierter Psychologischer Psychotherapeut. Ziel der Untersuchung war es, anhand der vorliegenden Stichprobe herauszufinden, welches Defizit solche Gutachten in der Regel am häufigsten aufweisen und so beteiligten Personen und Institutionen auch außerhalb des psychosozialen Bereiches – insbesondere im Bereich der Rechtsprechung – ein praktikables Instrumentarium an die Hand zu geben, das es ihnen ermöglicht, auch in einem für sie fachfremden und mitunter auch undurchsichtig erscheinenden Bereich, die Qualität und damit die Aussagegültigkeit solcher Gutachten zu beurteilen und

ggfs. gegen unzureichende und angesichts ihrer Qualität übertriebene Gutachten vorzugehen.

Die Brisanz dieser Thematik und die Notwendigkeit eines solchen Instrumentariums wurde dem Verfasser und Teilnehmern verschiedener Arbeitskreise nicht zuletzt durch den regen Austausch zwischen Familienrichtern, Rechtsanwälten und Vertretern psychosozialer Berufe beim 13. Deutschen Familiengerichtstag 1999 in Brühl deutlich.

Ergänzend zur vorliegenden Studie sei auf die empirischen Untersuchungen psychologischer Gutachten für das FamG von Klüber (1998) und Terlinden-Arzt (1998) sowie auf einen Beitrag von Westhoff (1998) mit dem Titel »Möglichkeiten zur Verbesserung psychologischer Gutachten im Familienrecht« im Tagungsband »Psychologie im Familienrecht« der Evangelischen Akademie Bad Boll verwiesen. Hier finden sich auch weiterführende Ergebnisse über die Stichprobe und differenzierte Auswertungsaspekte in einem Beitrag des Verfassers, die den Rahmen des hier vorliegenden und pragmatisch orientierten Beitrages sprengen würden.

### Richtlinien für familienpsychologische Gutachten

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 Studie soll u. a. darüber Aufschluß geben, inwieweit die eingesetzten Tests die über die Aussagegültigkeit und wissenschaftliche Haltbarkeit entscheidenden Gütekriterien (Hauptgütekriterien: Reliabilität, Validität, Objektivität sowie das Nebengütekriterium Normierung) erfüllen und damit grundlegende Standards aufweisen, um als Tests im wissenschaftlichen Sinne gelten zu können.

Auf Grund dieser Erkenntnisse wird eine Zusammenstellung die in solchen Gutachten am häufigsten vorkommenden Tests in einer Art »Rangliste« darstellen,

die gleichzeitig zeigt, inwieweit gerade diese häufig eingesetzten Tests überhaupt die Gütekriterien erfüllen. Diese Zusammenstellung sollte eine wertvolle Hilfe sein, um gegen unzureichende Gutachten unter diesem Aspekt vorgehen zu können. Aber nicht nur die Ebene der Tests erscheint hier relevant.

#### Häufigste Mängel:

### 1. Mängel bei der Aktenanalyse, Exploration und Verhaltensbeobachtung

Damit im Hinblick auf diese Datengrundlagen von wissenschaftlichem Arbeiten gesprochen werden kann, sollte die (V)erarbeitung 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. Hierbei stellen sich folgende Fragen:

- Wurde eine systematische Aktenanalyse auf der Basis eines Kategoriensystems durchgeführt oder erfolgte die Sichtung und Darstellung der vorliegenden Akten unsystematisch und offensichtlich nach freiem 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; vgl. auch Häcker & Stapf, 1998) ergibt sich folgendes: »Die Systematik der Verhaltensbeobachtung wird nach mehreren Gesichtspunkten vorgenommen: (1) systematische (wissenschaftliche) Beobachtung – unsystematische (naive, unwissenschaftliche) Beobachtung«. 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; vgl. auch Häcker & Stapf, 1998) 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 bei-

spielsweise mit einem Gesprächsleitfaden von Westhoff und Kluck (1991/1992) vorliegen, explizit nicht zum Einsatz. Auf welches wissenschaftlich-methodische 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) bzw. der Verweis auf relevante Befunde wird nahezu durchgängig unzureichend transparent gemacht oder ist nicht ersichtlich.

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

Eine systematische (wissenschaftliche) Verhaltensbeobachtungen 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 der Verhaltensbeobachtung wissenschaftliche Methoden in einer Form vorhanden, die in ihrer Gesamtheit wissenschaftlichen Grundanforderungen gerecht werden.

### 2. Mängel bei der Verarbeitung von Fachliteratur und wissenschaftlichen Konzepten

Ein familienpsychologisches Gutachten sollte auch im Hinblick auf literarische Gestaltungsprinzipien elementare wissenschaftliche Standards erfüllen. So besteht eine unabdingbare Forderung u. a. darin, daß im Gutachten umfangreiche Literatur- bzw. 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 Anforderungen wissenschaftlichen Arbeitens mißachtet, kann nicht den Anspruch erheben »eine wissenschaftliche Leistung« (Föderation Deutscher Psychologengemeinschaften, 1994, S. 8) zu sein, wie dies in den eingangs zitierten Richtlinien der Föderation Deutscher Psychologengemeinschaften für »ein solches Gutachten« (aaO) ausdrücklich gefordert wird.

Insgesamt sind diese elementaren Anforderungen nur bei einem geringen Teil der vorliegenden Gutachten

angemessen berücksichtigt. Vielfach entsprechen die ohnehin nur bei wenigen Gutachten vorhandenen Literaturangaben auch nicht neusten Standards und belegen eher ein Hinterherhinken um mehrere Jahre oder gar ein bis zwei Jahrzehnte hinter dem neusten Forschungsstand.

### 3. Tests, die die Gütekriterien nicht hinreichend erfüllen

Gravierende und fachlich unausweichlich belegbare Mängel bestehen – wie bereits angedeutet – vor allem auch auf der Ebene der eingesetzten – oder auch nicht eingesetzten – Tests. Hierbei wird es auch erforderlich sein, sich in einem Exkurs zumindest kurz über die Bedeutung der einschlägigen Gütekriterien zu informieren. Ein entsprechender Hinweis oder eine Diskussion in den Gutachten unterbleibt meist nahezu vollkommen – ein weiterer gravierender Mangel, der entsprechende Unzulänglichkeiten auf der Testebene in den Gutachten aber auch sehr deutlich aufdecken würde.

Im umgangssprachlichen Bereich wird der Begriff »Test« 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, 1998 zitiert in *Dorsch*, 1987, S. 681 bzw. *Dorsch*, 1994, S. 793 f; vgl. auch *Häcker & Stapf*, 1998) 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 im eigentlichen Sinne nicht als Test bezeichnet werden, da ihnen die mit wissenschaftlichen Methoden überprüften Grundlagen und die für eine Testkonstruktion notwendigen Kontrolluntersuchungen fehlen.«

#### Gütekriterien wissenschaftlicher Testverfahren – ein notwendiger Exkurs

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. Untereinander lassen sich die Gütekriterien im Hinblick auf ihre Wertigkeit in Haupt- und Nebengütekriterien unterteilen.

#### Objektivität

In der angloamerikanischen Literatur wird Objektivität als eines Testgütekriteriums nicht immer einheitlich verwendet. So hat *Cattell* (vgl. *Dorsch et al.*, 1994, S. 525; vgl. auch *Häcker & Stapf*, 1998) 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; vgl. auch *Häcker & Stapf*, 1998) wurde die Objektivität als Testgütekriterium für den deutschsprachigen Raum definiert als »Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind« (aaO). Bei dieser Definition steht die »interpersonelle Übereinstimmung« (aaO) 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 bzw. eines standardisierten Testverfahrens, die angibt, 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 Testhalbierungsmethode verschiedene Aspekte der Zuverlässigkeit unterscheiden (vgl. *Dorsch et al.*, 1994, S. 656 f).

Nach den Standards für pädagogisches und psychologisches Testen, ausgearbeitet vom »Committee to Develop Standards for Educational and Psychological Testing« der »American Educational Research Association (AERA)«, der »American Psychological Association

on, (APA)« und dem »National Council on Measurement in Education (NCME)« in der deutschen Fassung von Häcker, Leutner und Amelang (1998, S. 23) »(wird) unter Reliabilität ... das Ausmaß verstanden, in dem Testwerte frei von Meßfehlern sind«. Aus diesen Ausführungen wird deutlich wie zuverlässig (oder besser unzuverlässig) Werte von Tests in psychologischen Gutachten für das Familiengericht sind, bei denen die Tests keine Reliabilität aufweisen.

### 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; vgl. auch Häcker & Stapf, 1998). Hierbei lassen sich bestimmte Validitätsarten unterscheiden: inhaltliche Validität, kriteriumsbezogene Validität sowie Konstruktvalidität (vgl. Dorsch et al., 1994, S. 838 f; vgl. auch Häcker & Stapf, 1998).

Nach den Standards für pädagogisches und psychologisches Testen, ausgearbeitet vom »Committee to Develop Standards for Educational and Psychological Testing« der »American Educational Research Association (AERA)«, der »American Psychological Association (APA)« und dem »National Council on Measurement in Education (NCME)« in der deutschen Fassung von Häcker, Leutner und Amelang (1998, S. 10) »(gilt) die Ermittlung der Validität eines Tests ... als wichtigster Aspekt der Testevaluation. Es wird dabei beurteilt, wie angemessen, wie bedeutsam und wie nützlich die spezifischen Schlußfolgerungen sind, die aus solchen Testwerten gezogen werden können. Der Prozeß, in dem empirische Belege für die Richtigkeit solcher Schlußfolgerungen akkumuliert werden, wird als Validierung bezeichnet«. (Häcker, Leutner und Amelang, 1998, S. 10). Aus diesen Ausführungen läßt sich ebenfalls sehr leicht erschließen, »wie angemessen, wie bedeutsam und wie nützlich« folglich Schlußfolgerungen aus Tests in psychologischen Gutachten für das Familiengericht sind, die keine Validität aufweisen. Im dazugehörigen Glossar dieser Standards wird Validität demzufolge definiert als »das Ausmaß, in dem eine bestimmte Schlußfolgerung aus einem Test angemessen oder bedeutsam ist« (aaO, S. 110).

### 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; vgl. auch Häcker & Stapf, 1998). 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.

### Inwieweit erfüllen die eingesetzten Tests der Stichprobe die Gütekriterien?

Für den Pragmatiker dürfte es nun vor allem von Interesse sein, welche Tests in solchen Gutachten denn üblicherweise besonders häufig vorkommen und inwieweit gerade diese die Gütekriterien erfüllen. Aus diesem Grund wurde anhand der vorliegenden Stichprobe eine Art »Rangliste« für die Häufigkeit des Einsatzes erstellt und gleichzeitig anhand einschlägiger Fachliteratur untersucht, inwieweit diese Verfahren die Gütekriterien hinreichend erfüllen, um als Test im wissenschaftlichen Sinne gelten zu können. Eigentlich wäre es ja anzunehmen gewesen, daß vor allem die besonders aussagegültigen Testverfahren auch besonders häufig angewendet werden, also die Verfahren, die die Gütekriterien möglichst gut erfüllen. Überraschenderweise war bei den vorliegenden psychologischen Gutachten aber eher das Gegenteil der Fall: Besonders häufig wurden Tests eingesetzt, die die Gütekriterien nur unzureichend erfüllen oder in einschlägiger Fachliteratur kaum verzeichnet sind.

Es wurden nur die Tests in die Gesamtauswertung (18 Tests) aufgenommen, die in mindestens zwei unterschiedlichen Gutachten zum Einsatz kamen. Die nachfolgende Übersicht gibt Aufschluß über Tests mit sieben und mehr Anwendungen und wird für den Pragmatiker von besonderem Interesse sein, um besonders häufig verwendete Tests, die die Gütekriterien nur unzureichend erfüllen, in solchen Gutachten auffindig machen zu können.

### Häufigkeitsrangskala - Tests N=52 familienpsychologische Gutachten

1. **Familie in Tieren (BREM-GRÄSER, 1995), 16 Anwendungen**  
Gütekriterien nach BRICKENKAMP (1997, 964, 968 ff):  
Objektivität: nein      Reliabilität: nein  
Validität:    nein      Normierung: teilweise
2. **Family-Relat.-Test (BENE & ANTHONY, 1957), 16 Anwendungen**  
Gütekriterien nach BRICKENKAMP: nicht verzeichnet
3. **Fabeln (DÜSS, 1964), 13 Anwendungen**  
Gütekriterien nach BRICKENKAMP (1975, 520ff):  
Objektivität: nein      Reliabilität: nein  
Validität:    nein      Normierung: ja

**4. Satzergänzungstest, 11 Anwendungen**

Gütekriterien nach BRICKENKAMP: nicht verzeichnet

**5. Kinder-Apperzeptions-Test****(BELLAK & BELLAK, 1955), 8 Anwendungen**

Gütekriterien nach BRICKENKAMP (1997, 942, 949 f):

Objektivität: nein      Reliabilität: nein

Validität: nein      Normierung: nein

**6. Mann-Zeichen-Test (ZILER, 1996),****7 Anwendungen**

Gütekriterien nach BRICKENKAMP (1997, 964, 973 ff):

Objektivität: ja      Reliabilität: teilweise

Validität: ja      Normierung: ja

**7. Schloss-Zeichen-Test, 7 Anwendungen**

Gütekriterien nach BRICKENKAMP: nicht verzeichnet

**8. Scenotest (VON STAABS, 1992),****7 Anwendungen**

Gütekriterien nach BRICKENKAMP (1997, 964, 981):

Objektivität: teilweise      Reliabilität: teilweise

Validity: teilweise      Normierung: nein

**Anmerkungen zum Family-Relations-Test (FRT)**

Das zusammen mit dem im Hinblick auf die Gütekriterien völlig unzureichendem Test »Familie in Tieren« (Brem-Gräser, 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 Gutachten. Insbesondere bei diesem Testverfahren läßt sich erkennen, daß ausgeprägte organisationsspezifische Besonderheiten beim Einsatz bestimmter Tests offenbar kaum von der Hand zu weisen sind.

Auf Grund seiner Häufigkeit in den vorliegenden familienpsychologischen Gutachten sollen zu diesem Testverfahren noch einige ergänzende Anmerkungen gemacht werden:

Beim FRT handelt es sich um ein Verfahren, das 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 »neuere Untersuchungen mit dem Family Relations Test«. Hierbei wurde deutlich, daß die Validität dieses Verfahrens zum gegenwärtigen Zeitpunkt kei-

neswegs als gesichert gelten kann. Im Rahmen seines Vortrages und der anschließenden Diskussion bezeichnete Beelmann den Umgang mit diesem Verfahren in der diagnostischen Praxis zudem als »haarsträubend« und verwies in diesem Zusammenhang u. a. darauf, daß aus ökonomischen Gründen bei der praktischen Durchführung häufig instruktionsinadäquate Modifikationen vorgenommen werden.

**Weitere Schlußfolgerungen**

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 wissenschaftlich 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 sogar insbesondere bei sehr häufig eingesetzten Testverfahren (z. B. Familie in Tieren: Rangplatz Nr.1) der Fall.

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 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.

Nachdem sehr viele psychologische Gutachten der vorliegenden Stichprobe auf relevanten Ebenen ganz elementaren wissenschaftlichen Anforderungen unzurei-

chend genügen, erscheint die Mehrzahl als Entscheidungsgrundlage für Umgangs- und Sorgerechtsregelung, die sich auf die künftige Lebenswelt und das Erziehungsumfeld von Kindern sehr nachhaltig auswirken, in der vorliegenden Form ungeeignet (vgl. Seibert, 1999).

Die in diesem Rahmen gemachten Aussagen und Empfehlungen entbehren bisweilen der erforderlichen wissenschaftlichen Grundlage sogar bis hin auf elementarste Anforderungen, die an wissenschaftliches Arbeiten gestellt werden müssen.

Neben der unzureichenden wissenschaftlichen Fundierung der psychologischen Untersuchung sind aber insgesamt auch elementare erziehungswissenschaftliche Grundlagen, die für solche Fragestellungen überaus bedeutsam wären, den Gutachten der vorliegenden Stichprobe kaum zu entnehmen. Auch vorliegende Befunde zum Parental-Alienation-Syndrom (PAS) (vgl. u. a. Gardner, 1992; Ward & Harvey, 1998; Koeppl & Kodjoe, 1998; Leitner & Schoeler, 1998; Leitner, 1999) sind in keinem einzigen der vorliegenden Gutachten explizit thematisiert.

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). Anzudenken wäre vielleicht auch die Bestellung von Sachverständigen mit pädagogisch-erziehungswissenschaftlich und therapeutischer Qualifikation bzw. mit entsprechender Kompetenz auch für eine interventionsdiagnostische bzw. interventionsgeleitete Vorgehensweise, die sich nicht allein in der Diagnose erschöpft.

Die in der GWG-Broschüre »Ein Institut stellt sich vor« (GWG, ohne Jahresangabe, S. 16) gemachten Ausführungen mögen angesichts dieser Ergebnisse eher verwundern. Die von dieser Organisation 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 vor allem auch für Gutachten dieser Organisation kaum hinreichend bestätigen.

Es wäre wünschenswert, wenn sich Weinerts (1995, 11) nachfolgend zitierte Ausführungen zur pädagogischen Diagnostik auch in der psychologischen Diagnostik umsetzen ließen: »Bereits Cronbach & Gleser (1965)

stellten den in der Literatur allgemein anerkannten Grundsatz auf, daß Diagnostik nie Selbstzweck sein sollte, sondern Mittel zur Vorbereitung einer Entscheidung. Darüber hinaus müsse gelten, daß pädagogische Diagnostik nur verantwortungsbewußter Umgang mit Daten zu einem verantwortbaren Zweck sein darf«. Den Begriff »pädagogische Diagnostik« brachte Ingenkamp (1968) in die Diskussion ein (vgl. aaO). »Die Teildisziplin »pädagogische Diagnostik« unterscheidet sich mit Weinert (1995, 11) von anderen Disziplinen, wie etwa der psychologischen Diagnostik, durch die Verwendung anderer Theorieansätze« und spezifischer Anwendungsbezüge. Wenngleich es die pädagogische Diagnostik gibt und Erziehungsfragen für Entscheidungen »zum Wohl des Kindes« einen zentralen Stellenwert haben, werden bislang nahezu ausschließlich psychologische Gutachten in Auftrag gegeben.

Bei alledem sollten diese Ergebnisse nicht dazu verleiten, künftig auf relevanten Sachverstand in diesem sensiblen und für die Zukunft so einschneidenden Bereich zu verzichten. Es wird insgesamt also weniger auf die Frage hinauslaufen dürfen, ob Sachverständige beauftragt werden sollten. Vielmehr wird es darum gehen, wer beauftragt wird. Möglichkeiten einer Evaluation und Qualitätssicherung werden insgesamt nachhaltig zu intensivieren sein. Hierzu hofft der Verfasser einen auch für die anwaltliche und richterliche Praxis relevanten Beitrag geleistet zu haben, um in begründeten Fällen Einwände gegen ein solches Gutachten adäquat akzentuieren und thematisieren zu können.

## Literatur

- Beelmann, W., Neuere Untersuchungen mit dem Family Relations Test – Normierungen und Anwendung in Scheidungsfamilien. In: Etrich, K.U., Friedrich, G., Fries, M. und Witruk, E. (Hrsg.), Abstractband zu den Tagungen der Fachgruppen Entwicklungspsychologie und Pädagogische Psychologie. Leipzig 1995, 38.
- Boerner, K., Das psychologische Gutachten. Weinheim und Basel 1991.
- Brickenkamp, R., Handbuch psychologischer und pädagogischer Tests. Göttingen 1975, 1997.
- Brickenkamp, R., Erster Ergbd. zum Handbuch psychologischer und pädagogischer Tests. Göttingen 1983.
- Dorsch, F., Häcker, H. und Stapf, K.-H., Psychologisches Wörterbuch. Bern 1987, 1994.
- Etrich, K. U., Friedrich, G., Fries, M. und Witruk, E. (Hrsg.), Abstractband zu den Tagungen der Fachgruppen Entwicklungspsychologie und Pädagogische Psychologie. Leipzig 1995.
- Föderation Deutscher Psychologinnenvereinigungen, Richtlinien für die Erstellung Psychologischer Gutachten. Bonn 1994.
- Gardner, R. A., The Parental Alienation Syndrome. A Guide for Mental Health and Legal Professionals. Cresskill, New Jersey 1992.

- Gesellschaft für wissenschaftliche Gerichts- und Rechtspsychologie, GWG* – Ein Institut stellt sich vor. Ohne Angabe von Erscheinungsjahr und Ort.
- Häcker, H., Leitner, D. und Amelang, M., Standards für pädagogisches und psychologisches Testen. Ausgearbeitet vom Committee to Develop Standards for Educational and Psychological Testing of, The American Educational Research Association (AERA), The American Psychological Association (APA), The National Council on Measurement in Education (NCME). Supplements 1 (1998) der Diagnostica und der Zeitschrift für Differentielle und Diagnostische Psychologie. Bern 1998.
- Häcker, H. und Stapf, K.H., Dorsch Psychologisches Wörterbuch. Bern 1998.
- Ingenkamp, K.-H., Pädagogische Diagnostik. Weinheim, 1975.
- Klüber, A., Psychologische Gutachten für das FamG. Eine empirische Untersuchung über Nachvollziehbarkeit und Verständlichkeit des diagnostischen Prozesses sowie ausgewählte Aspekte des Kindeswohls. Lengerich 1998.
- Kodjoe, U. & Koepfel, P., The Parental Alienation Syndrome (PAS), Der Amtsvormund 1 (1998) 10 – 140.
- Leitner, W.G., Pädagogisch-psychologische Diagnostik. Interventionsgeleitete Befunderstellung im exemplarischen Einzelfall. Bamberg 1995.
- Leitner, W.G., Evaluation psychologischer Entscheidungshilfen für Familiengerichte: Familienpsychologische Gutachten nach Trennung und Scheidung. Evangelische Akademie Bad Boll, Psychologie im Familienrecht, Bad Boll 1998, S. 152 – 165.
- Leitner, W.G., Interventionsgeleitete Einzelfallhilfe. Intervention-guided single case-help in counselling. 2., vollständig überarbeitete und erweiterte Aufl. Theorie und Forschung Psychologie. Regensburg 1999.
- Leitner, W.G., Kindliche Lebenswelten unter besonderer Berücksichtigung der Situation von Scheidungswaisen im familiären und schulischen Kontext. In: Seibert, N. Kindliche Lebenswelten. Eine mehrperspektivische Annäherung. Bad Heilbrunn 1999, S. 127 – 156.
- Leitner, W.G., Intervention-Guided Single Case-Help and Parental Alienation Syndrome (PAS). Differential Diagnosis and Treatment Approaches. In: Sebre, S., Rascevska, M. and Miezite, S. (1999). Identity and Self-Esteem. Interactions of Students, Teachers, Family and Society. Riga 1999, S. 253 – 260.
- Leitner, W.G., Psychological Expert Assessments for Family Court Decisions in Germany. In: Sebre, S., Rascevska, M. and Miezite, S. (1999). Identity and Self-Esteem. Interactions of Students, Teachers, Family and Society. Riga 1999, S. 261 – 273.
- Leitner, W.G. und Goethe, S., Fragebogen: Sachverständigengutachten für gerichtliche Entscheidungen zum Sorge-/Umgangsrecht. In: ISUV-Report 73 (1997) 9, S. 24.
- Leitner, W.G. und Goethe, S., Fragebogen: Sachverständigengutachten für gerichtliche Entscheidungen zum Sorge-/Umgangsrecht. In: ISUV-Report 75 (1998) 3, S. 36.
- Leitner, W.G. und Linsler, J., Das Recht des Kindes in Trennungs- und Scheidungsfamilien. Tätigkeit einer neuen Bundesarbeitsgemeinschaft. Abstract zur Europäischen Fachtagung zur Familienforschung. Universität Bamberg 1994.
- Leitner, W.G. und Linsler, J., Elternverlust – unabdingbares oder vermeidbares Kinderunglück nach Trennung oder Scheidung? In: Zeitschrift für Familienforschung, Familie und Familienprobleme im Wandel. Hrsg. im Auftrag des Staatsinstituts für Familienforschung an der Universität Bamberg. Sonderheft 1/1994, 6.Jg., S. 315 – 322.
- Leitner, W.G. und Schoeler, R., Maßnahmen und Empfehlungen für das Umgangsverfahren im Blickfeld einer Differentialdiagnose bei Parental Alienation Syndrom (PAS) unterschiedlicher Ausprägung in Anlehnung an Gardner (1992/1997). In: Der Amtsvormund (DAVorm). Monatsschrift des Deutschen Instituts für Vormundschaftswesen 11/12, 1998, 71. Jg., S. 849 – 868.
- Lienert, G. A., Testaufbau und Testanalyse. Weinheim/Berlin 1969, München 1998.
- Schmidt-Jortzig, E., Persönliche Korrespondenz 1997.
- Seibert, N., Perspektive Schulpädagogik. Kindliche Lebenswelten. Eine mehrperspektivische Annäherung. Bad Heilbrunn 1999.
- Terlinden-Arzt, P., Psychologische Gutachten für das FamG. Eine empirische Untersuchung über diagnostische Strategien sowie ausgewählte Aspekte des Kindeswohls. Lengerich 1998.
- Ward, P. & Harvey, J. D., Familienkriege – die Entfremdung von Kindern, ZfJ 85 (1998) 6, S. 237 – 245
- Weinert, B., Diagnostik – Was ist das eigentlich. In: Nauck, B., Schuleingangsdiagnostik. Braunschweig 1995, S. 9 – 24
- Westhoff, K., Möglichkeiten zur Verbesserung psychologischer Gutachten im Familienrecht. Evangelische Akademie Bad Boll, Psychologie im Familienrecht, Bad Boll 1998, S. 68 – 79
- Westhoff, K. und Kluck, M.-L., Psychologische Gutachten schreiben und beurteilen. Heidelberg 1991.
- Westhoff, K. und Kluck, M.-L., Ein Beispiel für entscheidungsorientierte Hypothesenbildung bei gerichtlichen Fragen zur elterlichen Sorge. In: WESTHOFF, K. (Hrsg.), Entscheidungsorientierte Diagnostik. Bonn 1992, S. 28 – 34.
- Westhoff, K. und Kluck, M.-L., Entwicklung eines Leitfadens für ein entscheidungsorientiertes Gespräch bei gerichtlichen Fragen zur Umgangsregelung. In: Westhoff, K. (Hrsg.), Entscheidungsorientierte Diagnostik. Bonn 1992, S. 35 – 44.
- Westhoff, K., Terlinden-Arzt, P. und Klüber, A., Entscheidungsorientierte psychologische Gutachten für das FamG. Vorabdruck 1999.

Dr. phil. Werner G. Leitner, Approbierter Psychologischer Psychotherapeut, Markusplatz 14, 96047 Bamberg, e-mail: werner.leitner@ppp.uni-bamberg.de