

# Zur Gültigkeit der Rosenzweig Picture-Frustration Study, Form für Kinder

Von *Theodor Ehlers*

## Zusammenfassung, Summary, Résumé

Im ersten Teil werden die vorliegenden Untersuchungen zur Kinderform der Picture-Frustration Study (P-F S) diskutiert. Die vorgelegten Befunde stützen einander wenig. Neben methodischen Mängeln der Untersuchungen lassen gewisse Besonderheiten des Tests wie der fehlende Geschlechterunterschied, das (unbekannte) Niveau der Selbstdarstellung und der (unberücksichtigte) Anregungsgehalt der Bildvorlagen auf gültigkeitsmindernde Fehlervarianzquellen schließen.

Im zweiten Teil wird eine Untersuchung beschrieben, in der die Testergebnisse von 28 neunjährigen Heimkindern beiderlei Geschlechts mit der Beurteilung der Aggressivität durch die Erzieher und mit den Aggressionen im freien Spiel verglichen wurden. Der *Rosenzweig*-Faktor E und, in gegenläufiger Richtung, die Impunitivität zeigten sehr enge Beziehungen zu den beiden Kriterien. Eine Kontrolle des Niveaus der Selbstdarstellung und des Anregungsgehaltes der Bildvorlagen erwies sich als nicht notwendig. Es ergab sich auch ein Erklärungsansatz in bezug auf die fehlende Geschlechterabhängigkeit der Aggressivitätsvariablen der P-F S.

### On the validity of the Rosenzweig Picture Frustration Study, in the form for children

In the first part the validity studies of the children's version of the P-F S are discussed. The results presented do not give each other much support. Apart from inadequacies of method in the analysis, some of the particulars of the text point to sources of error variance which reduce validity: for example the lack of sex differentiation, the (unknown) level of self-representation and the stimulation content of the pictures used (which is not taken into account).

In the second part an analysis is described in which the test results of 28 9-year-old institution children of both sexes were compared with the teacher's judgement of aggressive behaviour and the children's aggressive behaviour during play. The Rosenzweig Factor E and, conversely, the impunitivity showed very close connections with both criteria. A check on the level of self-representation and the stimulation content of the pictures used turned out to be unnecessary. As regard the absence of dependence of the aggressivity variables on sex there was an attempt at clarification.

### Valeur de la Picture-Frustration Study de Rosenzweig sous sa forme pour enfants

Dans la première partie, on discute des recherches existantes sur la version pédiatrique de la Picture-Frustration Study (P-F S). Les faits présentés manquent de cohérence. Outre les faiblesses méthodiques des recherches, certaines particularités du test, telles que l'absence de différenciation des sexes, le niveau (inconnu) de l'autoreprésentation et le contenu de stimulation (dont on ne tient pas compte) des images présentées permettent de conclure à des sources de variations dans les erreurs propres à diminuer la valeur du test.

Dans la seconde partie, on rend compte de recherches au cours desquelles les résultats de tests de 28 enfants d'un foyer, âgés de 9 ans et des deux sexes, ont été comparés aux jugements portés sur l'agressivité de l'éducateur et aux agressions dans le cadre d'un jeu libre. Le facteur E de Rosenzweig et, en sens inverse, l'absence d'impunitivité se sont révélés très étroitement liés à ces deux critères. Il s'est révélé superflu de contrôler le niveau de l'autoreprésentation et du contenu stimulant des images présentées. S'est manifesté également un début d'explication de l'absence de relations entre le sexe et les variables d'agressivité de la P-F S.

Der „Rosenzweig-Test“ gehört zu den bekanntesten psychodiagnostischen Persönlichkeitstests. Er ist in allen einschlägigen allgemeinen und speziellen Sammelwerken und Testkompendien berücksichtigt und beschrieben worden (z. B. *Rabin* u. *Haworth* 1960, *Selg* 1968, *Brickenkamp* 1975, *Schmidtchen* 1975). Seine Anwendung scheint sich dagegen in zunehmend engeren Grenzen zu halten. Die Zahl der ihn betreffenden Erfahrungsberichte ist seit etwa 1965 besonders in seinem Ursprungsland, den Vereinigten Staaten, stark zurückgegangen. Die deutschen Bearbeitungen der Erwachsenen- und der Kinderform des Tests (*Hörmann* u. *Moog* 1957, *Duhm* u. *Hansen* 1957), die auf der Basis sehr kleiner Eichstichproben und ohne Prüfung von Zuverlässigkeit und Gültigkeit erfolgten, werden immer noch in der ersten Auflage angeboten.

Zur Bekanntheit des Tests hat neben der Plausibilität seines Konzeptes sicher beigetragen, daß ein bisher schwer zugänglicher, aber aktueller Merkmalsbereich erfaßt werden soll (vgl. *Selg* 1968, S. 30).

Die geringe Verwendung des Tests ist u. a. darauf zurückzuführen, daß nur unzulängliche Anstrengungen zu seiner Validierung unternommen wurden. Die dabei vor allem unter dem Aspekt der Konstruktvalidierung zusammengetragenen Befunde werden als ergänzungsbedürftig (*Hiltmann* 1964) oder als widersprüchlich angesehen (*Schmidtchen* 1975, S. 218). Zur Abwertung des Verfahrens hat auch beigetragen, daß der Testautor selbst auf gewisse Mängel aufmerksam machte und sein Verfahren als „Versuch“ und nicht als Test verstanden wissen wollte. Sein häufig gegebener Hinweis, daß das Niveau der Selbstdarstellung des Probanden (Operation Level) nicht aus dem Testergebnis selbst erschlossen werden kann, gilt für alle Persönlichkeitstests, die verbale Reaktionen erheben. Doch ist gerade dieser Hinweis als Bestätigung eines besonderen, testspezifischen Mangels aufgefaßt worden (*Bjerstedt* 1965, S. 510, *Schmidtchen* 1975, S. 224). Auf Ablehnung kann der Test auch dadurch stoßen, daß seine Auswertung unökonomisch wird, wenn ein unvollständiges Antwortprotokoll vorliegt.

Der vorliegende Bericht befaßt sich mit der derzeit noch häufiger – z. B. im schulpyschologischen Bereich – verwendeten Kinderform. Der erste Teil versucht einen Überblick zu geben über die bisherigen Befunde hinsichtlich der Gütekriterien und über Fehlerquellen, die als mehr oder weniger testspezifisch angesehen werden können. Im zweiten Teil wird dann von einer Untersuchung berichtet, bei der erstmalig das Testergebnis mit konkretem Verhalten in Beziehung gesetzt wird.

## 1. Bisherige Erfahrungen mit der Rosenzweig Picture-Frustration Study, Kinderform

### 1.1. Kurzbeschreibung des Verfahrens

Die Kinderform der Picture-Frustration Study (P-F S) von *Rosenzweig*, *Fleming* und *Rosenzweig* (1948) soll wie die Erwachsenenform (*Rosenzweig* 1945, 1965) die Disposition zur Bevorzugung bestimmter sozialer Verhal-

tensmuster nach Frustration erfassen. Sie ist vorgesehen für den Lebensabschnitt von 4–13 Jahren<sup>1)</sup>.

Das Testmaterial besteht aus 24 cartoonartigen Bildern. Auf jedem Bild äußert sich ein Kind oder ein Erwachsener so, daß eine Situation hergestellt oder beschrieben wird, die für die zweite anwesende Person (stets ein Kind) eine Vereitelung bedeuten kann. Der Proband soll für diese zweite Person verbal reagieren. Dem Verfahren liegt die Annahme zugrunde, daß sich der Proband mit der frustrierten Person identifiziert und Merkmale des eigenen Erlebens und Verhaltens in seine Antwort einfließen läßt. Nach dieser Konzeption zählt die P-F S damit zu den sog. projektiven Verfahren.

Jede Antwort wird sowohl nach der zum Ausdruck gebrachten Aggressionsrichtung als auch nach dem Reaktionstyp<sup>2)</sup> klassifiziert. Dabei wird zwischen drei Richtungen<sup>3)</sup> (nach außen = extrapunitiv, nach innen = intro-punitiv, neutral = impunitiv) unterschieden und zwischen drei Typen (Betonung des Hindernisses = Obstacle Dominance, Ich-Verteidigung = Ego Defense, Betonung des Bedürfnisses nach Lösung des Problems = Need-Persistence).

Als interpretierbare Testvariable werden angesehen die relativen Häufigkeiten der drei Aggressionsrichtungen und der drei Reaktionstypen und die absoluten Häufigkeiten der neun Kombinationen von Richtungen und Typen zuzüglich zweier „Varianten“. Die Kombinationen werden „Faktoren“ genannt und durch Buchstaben gekennzeichnet.

Das Antwortklassifizierungssystem des Tests, das gleichzeitig die gebräuchlichsten Testvariablen definiert, ist noch einmal in Abbildung 1 dargestellt.

Abb. 1: Antwortklassifizierung und Variablenbenennung (vgl. Text)

	Reaktionstypen		
	Obstacle Dominance	Ego Defense	Need-Persistence
Aggressionsrichtungen	extrapunitiv	E'	<u>E</u> (E) e
	intropunitiv	I'	<u>I</u> (I) i
	impunitiv	M'	M m

<sup>1)</sup> Die deutsche Bearbeitung gibt einen Zeitraum vom 7. bis zum 15. Lebensjahr an.

<sup>2)</sup> In neuerer Zeit (Rosenzweig 1960) auch als Aggressionstyp bezeichnet.

<sup>3)</sup> 1970 bezeichnen Rosenzweig u. Braun die Aggressionsrichtungen als „exagression“, „introgression“ und „imagination“. Wie andere Umbenennungen einzelner Testvariablen (Rosenzweig 1960, vgl. Bjerstedt 1965) haben sich diese Bezeichnungen bisher kaum durchgesetzt.

Bei den Varianten E und I handelt es sich um solche Reaktionen, bei denen angemessenermaßen eine enge Beziehung zum Überich gegeben ist. Ihre Häufigkeit wird einmal mit unter den zugehörigen Faktoren verrechnet, zum anderen aber auch für sich allein betrachtet.

Es wird in den Manualen vorgeschlagen, aus den Faktoren durch Zusammenfassung oder Unterteilung weitere Variablen abzuleiten. Auf eine Kennzeichnung dieser Variablen, die insbesondere aus Kombinationen der Varianten mit anderen Faktoren zu sog. Superego-Patterns und aus den „Tendenzen“ bestehen, kann hier verzichtet werden. Wie später deutlich wird, haben sich für solche Variablen keine reproduzierbaren Beziehungen zu Verhaltenskriterien ergeben, wahrscheinlich infolge fehlender Zuverlässigkeit.

Als globaler Indikator der sozialen Anpassung gilt das Group Conformity Rating (GCR), die prozentuale Übereinstimmung zwischen Probanden- und Standardantwort bei 21 der Bilder.

Insgesamt ergeben sich über 25 quantifizierbare Variablen, die, bedingt durch das Klassifizierungssystem, mehr oder weniger stark voneinander abhängen.

### 1.2. Objektivität und Zuverlässigkeit

Mit Hilfe der dem Manual beigegebenen Liste mit Signierungsbeispielen ist eine hohe Auswertungsobjektivität erreichbar. Die Übereinstimmung bei der Signierung der Faktoren liegt bei geübten Auswertern deutlich über 80% (Duhm u. Hansen 1957, Rosenzweig 1960, Simons 1967a).

Über die Zuverlässigkeit wird nur einmal, und zwar vom Testautor selbst, berichtet (Rosenzweig 1960). Bei zwei Altersgruppen zu je 44 Kindern erbrachte die Schätzung der inneren Konsistenz nur für die extrapunitive und die impunitiv Aggressionsrichtung Koeffizienten befriedigender Größenordnung zwischen 0,60 und 0,82. Für die Reaktionstypen und das GCR lagen die Koeffizienten unter 0,50 und waren in der Mehrzahl statistisch unbedeutsam. Von den nach drei Monaten ermittelten Retest-Koeffizienten zeigten sich die höchsten Werte, in der Größenordnung von 0,50 bis 0,69, ebenfalls für die extrapunitive und die impunitiv Aggressionsrichtung. Über die Zuverlässigkeit der einzelnen Faktoren wird nichts berichtet.

### 1.3. Untersuchungen zur Gültigkeit

In einigen Fällen sind Untersuchungen mit benachteiligten Kindern durchgeführt worden. Die Annahme war, daß ein körperliches Handicap oder eine schlechte Versorgung überdauernde Vereitelungssituationen darstellen, die die soziale Anpassung erschweren und entsprechend von der Norm abweichende Verhaltensmuster zur Folge haben. Van Roy (1961) fand bei zumeist infolge von Kinderlähmung motorisch behinderten Kindern folgende Unterschiede zu einer Kontrollgruppe: Extrapunitivität war seltener, Intro-punitivität und Impunitivität waren dagegen häufiger vertreten. Von den Reaktionstypen war die Betonung des Hindernisses häufiger. Bruant (1973) verglich Kinder mit sichtbarer Rückgratverkrümmung (Buckelbildung) mit zwei Kontrollgruppen. Die eine bestand aus vollkommen unauffälligen Kindern, die andere aus Kindern mit einer äußerlich nicht sichtbaren Mißbil-

derung des Rückgrats. Auch in diesem Falle zeigte die Versuchsgruppe gegenüber den Kontrollen weniger Extrapunitivität und mehr Impunitivität. *Szudra* (1966) verglich verschiedene Untergruppen von ambulanten und stationären Patienten einer Kinderklinik untereinander und mit den deutschen Normen. Faßt man alle seine Untergruppen zusammen, so zeigt sich für die Patienten insgesamt eine gegenüber den Normwerten erniedrigte Extrapunitivität bei häufigerem Vorkommen der beiden anderen Aggressionsrichtungen. *Johannsen* u. *Bennett* (1955) fanden bei diabetischen Kindern dagegen eine gegenüber den Altersnormen erhöhte Extrapunitivität bei erniedrigter Impunitivität. Von den Reaktionstypen war die Betonung des Hindernisses auf Kosten der Ich-Verteidigung häufiger. Außerdem zeigte sich ein erniedrigtes GCR. *Ferguson* (1954) unterteilte eine Gruppe von Kindern mit schweren Verhaltensstörungen nach der Häufigkeit ausgewählter Super-ego-Patterns. Die Untergruppe, deren Werte für eine „defektere Überich-Kontrolle“ sprachen, war in der Vergangenheit häufiger elterlicher und häuslicher Instabilität ausgesetzt gewesen.

Einige andere Untersuchungen sind mit Problemkindern, zumeist Klienten von Beratungsstellen, unternommen worden, und zwar unter der Annahme, daß eines der sie gegenüber unauffälligen Kindern auszeichnenden Merkmale eine ausgeprägtere Frustrationsanfälligkeit ist. *Rosenzweig* u. *Rosenzweig* (1952) fanden bei 162, in fünf Altersgruppen aufgeteilten Klienten einer Child-Guidance-Klinik diese Erwartung in folgenden Befunden bestätigt. Bei drei Altersgruppen zeigte sich eine gegenüber den Normwerten erhöhte Extrapunitivität. Bei ebenfalls drei Altersgruppen war Intropunitivität seltener. Von den Reaktionstypen war die Ich-Verteidigung bei zwei Altersgruppen der Klienten gegenüber den Normwerten niedriger. Die Hindernisbetonung war bei diesen Gruppen dagegen stärker ausgeprägt. Es zeigte sich auch die Tendenz zu einem niedrigeren GCR der Klienten. *Duhm* (1959) verglich ebenfalls die Ergebnisse von Problemkindern verschiedenster Symptomatik mit den Normwerten. Die 80 Problemkinder stellten sich danach als weniger intropunitiv dar. Dabei war nicht die Extrapunitivität, sondern die Impunitivität erhöht. *Szudra* unterteilte in der bereits erwähnten Untersuchung die von ihm erfaßte Patientengruppe nach der Art der Krankheits-symptome. Kinder mit Störungen im psychischen Bereich zeigten häufiger überdurchschnittliche Extrapunitivität und unterdurchschnittliche Intropunitivität als Kinder, deren Störung vor allem im somatischen Bereich lag.

Untersuchungen von Problemkindern mit homogenerer Symptomatik, aber ebenfalls noch unter der wenig spezifizierten Frustrationsempfindlichkeitshypothese, nahmen *Lord* (1952, zit. nach *Rosenzweig* 1960, S. 170) und *Weise* (1971) vor. *Lord* fand zwischen Enuretikern und unauffälligen Kindern nur einen Unterschied bei dem Faktor E: Die Enuretiker gaben seltener extrapunitive Antworten vom Typ der Ich-Verteidigung. *Weise* verglich 67 Kinder, die durch einen früherworbenen leichten Hirnschaden gekennzeichnet waren, mit den Werten der Eichstichprobe. Die nach Altersgruppen unterteilte Versuchsgruppe zeigte ebenfalls nur bei einem Faktor, einem der beiden Superegowerte, eine Abweichung von der Norm.

*Mucke* (1973) wählte aus einer Gruppe Verhaltensgestörter zwei Extremgruppen zu jeweils 30 Kindern, überwiegend Jungen, aus. Bei der einen

Gruppe handelte es sich um „offen aggressive Kinder ohne Ängste“, bei der anderen um „offen ängstliche Kinder ohne aggressive Durchbrüche“. Jede Gruppe wurde für sich mit einer Kontrollgruppe verglichen, die aus 56 verhaltensangepaßten Schulkindern bestand. Die offen ängstlichen Kinder zeigten gegenüber der Kontrollgruppe erhöhte Intropunitivität, vor allem vom Typ der Betonung des Hindernisses. Die Extrapunitivität war allgemein erniedrigt, insbesondere der Faktor E. Die offen aggressiven Kinder waren gegenüber der Kontrollgruppe dagegen häufiger extrapunitiver, seltener intro-punitiv und impunitiv. Von den Faktoren trug wiederum E am stärksten zum Unterschied bei.

Der letztgenannte Befund läßt sich bereits jenen Untersuchungen zuordnen, in denen mehr oder weniger explizit versucht wurde zu prüfen, inwieweit Art und Zahl der extrapunitiven Reaktionen im Test zur Kennzeichnung des herkömmlich als Aggressivität bezeichneten Merkmals herangezogen werden können. Zur Definition des Kriteriums wurden neben der Fremdbeurteilung, andere Testvariable, das Geschlecht, das Alter und die Delinquenz der Versuchspersonen herangezogen. Die Ergebnisse entsprachen zum Teil nicht denen von *Mucke*.

So fand *Lipman* (1959) bei 17jährigen, geistig Retardierten keine Unterschiede mit der P-F S zwischen solchen Probanden, die nach dem Urteil der Erzieher auffallend häufig offene Aggressionen zeigten, und solchen, die als weniger aggressiv eingestuft waren. Auch in der Untersuchung von *Coleman* (1967) zeigten sich keine Beziehungen zwischen Test und der durch Peer-Rating definierten Aggressivität von 10- bis 13jährigen Jungen. *Simons* (1967b) berichtet sogar eine signifikant entgegen der Erwartung ausgefallene Korrelation zwischen einer der Faktorenkombinationen und dem Lehrerurteil über Aggressivität bei 10- bis 14jährigen Schülern.

*Norman* u. *Kleinfeld* (1958) fanden zwischen 16jährigen delinquent gewordenen Jugendlichen und einer Kontrollgruppe bei Verwendung der Erwachsenenform keinerlei Unterschiede. *Lindsey* u. *Goldwyn* (1953) mußten die Feststellung machen, daß entgegen der Erwartung bei 12- bis 13jährigen Delinquenten die im Test zum Ausdruck gebrachte Extrapunitivität signifikant geringer war als bei einer Kontrollgruppe. *Vane* (1954) fand mit der Erwachsenenform bei delinquenten weiblichen Jugendlichen gegenüber einer nach Alter und sozioökonomischem Hintergrund parallelisierten Kontrollgruppe ebenfalls weniger Extrapunitivität bei stärkerer Ausprägung der beiden anderen Aggressionsrichtungen. Diese Beobachtung wurde bei Erwachsenen mit der entsprechenden Testform häufiger gemacht (vgl. *Kaswan*, *Wasman* u. *Freedman*, 1960).

Ein anderer Befund von *Lindsey* u. *Goldwyn*, allerdings mit der Erwachsenenform an Studenten gewonnen, stimmte mit der Aggressivitätshypothese besser überein. Beliebte Studenten zeigten in der P-F S eine stärkere Bevorzugung intro-punitiver Reaktionen. Dieser Befund konnte für die Kinderform von *Krieger* u. *Schwartz* (1965) bestätigt werden. *Kates* (1951) fand bei Kindern, die sich nach dem Ergebnis in einem Satzergänzungstest Gleichaltrigen unterordnen, weniger häufig extrapunitive und häufiger die impunitiv Aggressionsrichtung als bei Kindern, die Gleichaltrige eher ablehnten.

*Spielberg* u. *Rutkin* (1974) erwarteten nach der Frustrations-Aggressionshypothese einen Anstieg der Extrapunitivität nach Unterbrechung eines in Vierergruppen durchgeführten, kurz vor der Vollendung stehenden Puzzlespiels bei Achtjährigen. Die Versuchskinder unterschieden sich gegenüber Kontrollkindern deutlich im Sinne der Erwartung.

Am häufigsten ist unter dem Aspekt der Konstruktvalidierung der P-F S als Aggressivitätstest wohl der Frage nach Geschlechterunterschieden nachgegangen worden. In Analogie zu Ergebnissen konkreter Verhaltensbeobachtungen, bei denen Jungen im Vergleich zu Mädchen häufiger aggressive Handlungen zeigen (vgl. *Maccoby* u. *Jacklin* 1974), wurde erwartet, daß bei Jungen auch in der P-F S die durch den Faktor E definierte extrapunitive Ich-Verteidigung relativ überwiegt. In keinem Falle aber ließ sich diese Hypothese für die konventionelle Testauswertung bestätigen (*Habets* 1958, *Mucke* 1973, *Pichot*, *Freson* u. *Danjon* 1956, *Rosenzweig*, *Fleming* u. *Rosenzweig* 1948, *Simons* 1967a, *Spache* 1951, *Stoltz* u. *Smith* 1959).

Als weiterer Indikator im Sinne der Konstruktvalidität ist die Altersabhängigkeit der Aggressionsrichtungen von den Testautoren beschrieben worden (*Rosenzweig*, *Fleming* u. *Rosenzweig* 1948). Die Abnahme der Extrapunitivität und die entsprechende Zunahme der beiden anderen Richtungen mit steigendem Lebensalter wurden als Sozialisierungseffekt interpretiert. Bei der französischen Eichstichprobe zeigten sich gleichartige Befunde (*Pichot*, *Freson* u. *Danjon* 1956). Bei der deutschen Eichstichprobe trat die Altersabhängigkeit zahlenmäßig ebenfalls in Erscheinung, für die extrapunitive und die intropunitive Richtung aber weniger stetig. In der Untersuchung von *Simons* (1967a) ließ sie sich nur für die intropunitive Richtung statistisch absichern. *Habets* (1958) fand beim Vergleich von 9- und 13jährigen Kindern ebenfalls höhere Intropunitivitätswerte für die älteren Kinder. Gleichzeitig unterschieden sich die älteren von den jüngeren Kindern durch geringere Häufigkeiten des Faktors E und eines der Impunitivitätsfaktoren.

Über den Zusammenhang der P-F S mit anderen Tests ist bisher kaum berichtet worden. *Coleman* (1967) konnte ebensowenig eine Beziehung zu Rorschachtestvariablen und der TAT-Karte 18 BM finden wie *Lipman* (1959) zu einer stark frustrierenden Spiegelzeichnen-Aufgabe und einem Ängstlichkeitsfragebogen. Substantielle Zusammenhänge beobachteten dagegen *Levitt* u. *Lyle* (1955) mit dem „Problem-Situations-Test“, einem Fragebogen zur Erfassung der Tendenz, punitiv zu reagieren. Die Extremgruppe von 24 elfjährigen Kindern mit hohen Punitivitätspunktwerten gaben im Vergleich zu denen mit niedrigen Punktwerten bei der P-F S signifikant häufiger Antworten im Sinne der extrapunitiven Aggression und seltener im Sinne der intropunitiven Aggression. Ebenso zeigte sich häufiger der Reaktionstyp der Ich-Verteidigung und seltener der der Betonung des Bedürfnisses nach Problemlösung. *Levitt* u. *Lyle* fanden gleichzeitig eine Korrelation zwischen IQ und Intropunitivität von  $r = 0,41$ . *Simons* (1967b) berichtet von einer geringen aber bedeutsamen Korrelation des IQ mit einem Faktor der Intropunitivität, die Korrelation mit der gesamten Intropunitivität war dagegen bedeutungslos.

#### 1.4. Diskussion der Befunde

1.4.1. *Der inhaltliche Aspekt.* Die sich mit der Gültigkeit der Kinderform befassenden Untersuchungen sind noch einmal in Tabelle 1 aufgeführt, und zwar gegliedert nach der Art der untersuchten Merkmale. In der Befundspalte werden die Einzelfaktoren nur dann erwähnt, wenn ihre Veränderung nicht bereits durch die Veränderungen der Aggressionsrichtungen und Reaktionstypen definiert wird. So bedeuten erhöhte Werte für Extrapunitivität und für Ich-Verteidigung gleichzeitig einen – meist noch stärker – erhöhten Wert für den Faktor E<sup>4</sup>).

Aus der Tabelle geht hervor, daß die Ergebnisse einerseits zwar kaum widersprüchlich ausgefallen sind, andererseits aber auch bei gleichartigen Kriteriumsgruppen nur eine begrenzte Konkordanz aufweisen. Die Übereinstimmungen wie auch die Mehrzahl der statistisch abgesicherten Befunde beschränken sich im wesentlichen auf die Aggressionsrichtungen. Diese erweisen sich damit letztlich als die einzig ergiebigen Variablen des Tests. Diese Beobachtung überrascht nicht angesichts der Tatsache, daß bisher nur für die Aggressionsrichtungen Zuverlässigkeitskoeffizienten befriedigender Größenordnung gefunden werden konnten.

Hinsichtlich der individuellen Eigenart der Ausprägung der Aggressionsrichtungen lassen sich nur zwei Konstellationen voneinander unterscheiden, die mit Verhalten zu korrespondieren scheinen. So finden sich unterdurchschnittliche Extrapunitivitätswerte bei gleichzeitig erhöhter Intropunitivität und/oder erhöhter Impunitivität vor allem bei Vpn, die als beliebt, nachgiebig, ängstlich oder in ihrer körperlichen Bewegungsfreiheit als behindert gelten. Andererseits findet sich überdurchschnittlich hohe Extrapunitivität bei gleichzeitig niedriger Intropunitivität und manchmal auch niedriger Impunitivität vor allem bei den Vpn, die als die aggressiveren gelten können. Diese Aggressivitätskonstellation tritt in der Tabelle jedoch weniger deutlich in Erscheinung als die erstgenannte. Sie scheint allerdings auch bei jenen Vpn vorzuherrschen, die durch Verhaltensstörungen unterschiedlicher Symptomatik gekennzeichnet sind, den sog. Problemkindern. Dies ist vielleicht deshalb der Fall, weil vor allem solche Störungen Anlaß zu einer Untersuchung geben, bei denen aktive oder destruktive Elemente überwiegen.

Insgesamt ergeben sich aus der tabellarischen Zusammenstellung nur Anhaltspunkte hinsichtlich der Gültigkeit der P-F S. Offenbar besitzt das Verfahren im Hinblick auf die bevorzugten sozialen Verhaltensmuster eine gewisse Diskriminationsfähigkeit in der erwarteten Richtung. Doch stützt sich diese Fähigkeit nur auf einen kleinen Teil der Testvariablen und scheint nur eine grobe Einteilung in zwei Reaktionsklassen zu ermöglichen. Auch darf der Anteil der Untersuchungen nicht übersehen werden, bei denen zwischen den Kriteriumsgruppen keine Unterschiede zu beobachten waren. Er ist wie bei der Erwachsenenform (vgl. *Seidel* 1972) relativ hoch und gibt wie bei

<sup>4</sup>) Eine Ergebnisdarstellung auf der Basis der veränderten Faktoren wäre sicher angemessener gewesen. Es mußte darauf verzichtet werden, weil in einem Teil der Untersuchungen genauere Angaben zu den Faktoren fehlen. Oft wurden nur Veränderungen der Richtungen und Typen geprüft.

Tab. 1: Untersuchungen zur Gültigkeit der P-F S. Kinderform. Kennzeichnung der Ergebnisse der Versuchsgruppen gegenüber Kontrollgruppen oder Normwerten (+ für höhere, - für niedrigere Werte)

Merkmal	Merkmalsdefinition	Untersucher	Aggr.-Richtungen Extra-punitivität	Im-	Reakt.-Typen O-D	N-P	andere Variable
Beliebtheit	Peer-Rating	<i>Krieger u. Schwartz</i> 1965	+			+	
Unterordnung	unvollständige Geschichten	<i>Kates</i> 1951	-				
Ängstlichkeit	Rating	<i>Mucke</i> 1973	-		+		
	Enuresis	<i>Lord</i> 1952	-				
Körperliche Beeinträchtigung	Fragebogen	<i>Lipman</i> 1959	keine Unterschiede				
	motorische Behinderung	<i>van Roy</i> 1961	-	+	+		
	Rückgratverkrümmung	<i>Bruant</i> 1973	-	+			
	Klinikspatienten	<i>Szudra</i> 1966	-	+			
Hirng. bedingte Verhaltensstörungen	Diabetes	<i>Johannsen u. Bennett</i> 1955	+		+		GCR -
	EEG, Anamnese, neurol. Befund	<i>Weise</i> 1971					I -
	Rating	<i>Rosenzweig u. Rosenzweig</i> 1952	+		+		GCR (-)
Psychisch bedingte Verhaltensstörungen	Rating	<i>Duham</i> 1959		+			
	Rating	<i>Szudra</i> 1966	+				E +
Aggressivität	Erzieher-Rating	<i>Mucke</i> 1973	+				
	Peer-Rating	<i>Lipman</i> 1959	keine Unterschiede				
	Lehrer-Rating	<i>Coleman</i> 1967	keine Unterschiede				
	Delinquenz	<i>Simons</i> 1967b <i>Lindsey u. Goldwyn</i> 1953	-				E minus E -

Tab. 1 (Fortsetzung)

Merkmal	Merkmaldefinition	Untersucher	Aggr.-Richtungen Extra- Intro- punitivität	Im-	Reakt.-Typen O-D E-D	N-P	andere Variable
Jungen		<i>Habets 1958, Mucke 1973, Pichot et al. 1956, Rosenzweig et al. 1948, Simons 1967a, Spache 1951, Stoltz u. Smith 1959</i>					
jüngeres Alter		<i>Rosenzweig et al. 1948</i> <i>Pichot et al. 1956</i> <i>Simons 1967a</i> <i>Habets 1958</i> <i>Spielberger u. Ruitkin 1974</i>	keine Unterschiede + + - -	- - - -			
Spielunterbrechung		<i>Coleman 1967,</i> <i>Lipman 1959</i> <i>Levitt u. Lyle 1955</i>	+ keine Unterschiede +	- -	+ -	- -	E +, M' -

der Erwachsenenform (vgl. *Davreux* 1969, *Bjerstedt* 1965) Anlaß zu einer eher skeptischen Beurteilung des Verfahrens.

*1.4.2. Der methodische Aspekt.* Ein Teil der Ergebnislosigkeit der Untersuchungen könnte auf ungeeignete Gültigkeitskriterien zurückgeführt werden. So ist der häufig vorhergesagte, aber nie nachweisbare Geschlechterunterschied wohl kaum als „theoretisch bedeutsames Kriterium“ (*Keil-Specht* u. *Denzer* 1971, S. 192) anzusehen. Denn tatsächlich fanden sich reale, durchgängige Verhaltensunterschiede zwischen den Geschlechtern nur im Bereich der körperlichen Aggressionen. Hinsichtlich der auch durch die P-F S aktualisierten verbalen Aggressionen zeigten Verhaltensbeschreibungen dagegen z. T. keine Geschlechterunterschiede (*Maccoby* 1966, *Bandura*, *Ross* u. *Ross* 1961, *McIntyre* 1972). *Feshbach* u. *Feshbach* (1972) fanden die stärkeren verbalen Aggressionen auf seiten der Mädchen. Nur *Sears*, *Rau* u. *Alpert* (1965) berichten von sowohl größerer körperlicher als auch verbaler Aggression der Jungen im Vergleich zu den Mädchen während des freien Spiels. *Feshbach* (1970) vertritt die Annahme, daß Unterschiede zwischen den Geschlechtern weniger in Bezug auf die Antriebe zu aggressiven Handlungen als vielmehr in der Art der (gesellschaftlich tolerierten) Ausdrucksmöglichkeiten bestehen. Nach diesen Befunden und Überlegungen wären Geschlechterunterschiede in der P-F S gar nicht so naheliegend.

Als ungeeignetes Kriterium für erhöhte Aggressivität ist auch das Merkmal „Delinquenz“ anzusehen, insoweit es nicht auf Delikte bezogen wird, denen aggressives Verhalten zugrunde liegt. Das war z. B. in den Untersuchungen von *Lindsey* u. *Goldwyn* und *Norman* u. *Kleinfeld* nicht der Fall.

Ein weiterer Teil der Ergebnislosigkeit der P-F S-Untersuchungen könnte mit Stichprobenproblemen zusammenhängen. So mußte bei Gegenüberstellung der Ergebnisse von Versuchsgruppen und den entsprechenden Normwerten (z. B. *Weise* 1971, *Johannsen* u. *Bennett* 1955) Vergleichbarkeit vorausgesetzt werden, weil neben dem Alter nichts über die Repräsentativität der Eichstichproben bekannt ist. Die benutzten Normwerte stützen sich darüber hinaus auf kleine Stichproben, die deutschen z. B. auf 80 Kinder pro Altersgruppe. Bei eigentlichen Gruppenvergleichen wurden in der Delinquenzuntersuchung von *Lindsey* und *Goldwyn* Kontrollgruppen verwendet, deren P-F S-Werte sich deutlich von denen der Eichstichprobe unterscheiden (vgl. *Levitt* u. *Lyle* 1955). In der Untersuchung von *Lipman* mit geistig Retardierten muß angenommen werden, daß die Probanden den Test unter anderer Voraussetzung bearbeitet haben als der Durchschnittsproband.

Auch die Bedingungen der Testdurchführung könnten die Vergleichbarkeit der Untersuchungsergebnisse in Frage stellen. So wurde die Antwortregistrierung manchmal von den Probanden, manchmal vom Testleiter selbst vorgenommen. Nach *Mirmow* (1952) wird durch die letztgenannte Bedingung der Anteil der extrapunitiven Antworten erheblich vermindert. *Simons* (1967a) stellte allerdings keine Medianunterschiede fest zwischen den durch die Testleiter bei seiner Stichprobe von 714 Kindern registrierten Antworten und der deutschen Eichstichprobe. Eine Untersuchung von *Zuk* (1956) erbrachte Anhaltspunkte dafür, daß Kinder im Rahmen einer wenig eingeschränkten sozialen Atmosphäre (Ferienlager) stärker extrapunitiv reagieren

als in Schul- und im Rahmen von Leistungssituationen. Die meisten P-F S-Untersuchungen wurden unter Leistungsbedingungen durchgeführt.

Neben solchen Fehlern, die sich auf die Anlage der Untersuchungen beziehen und sich zu Ungunsten des Gültigkeitsnachweises ausgewirkt haben mögen, sind Mängel des Tests selbst zu berücksichtigen.

Als konzeptspezifischer Mangel der P-F S wird, wie eingangs bereits erwähnt, die Unklarheit darüber angesehen, welche Art der Selbstdarstellung im Einzelfall vorliegt. Die Testautoren vermuteten, daß gerade Kinder naiv-realistisch auf die Aufgabe reagieren. Sie schloßen aber nicht aus, daß die Bilder unterdrückte Bedürfnisse und selbstkritische Darstellungen provozieren können (*Rosenzweig, Fleming u. Rosenzweig 1948, Rosenzweig 1960, 1963*).

Für den durch das Manual nahegelegten Interpretationsansatz, daß bei Kindern die Selbstdarstellung realistisch ist, spricht nach *Rosenzweig (1960, S. 168)* die Beobachtung mit einer Vorform der P-F S. Bei 40 untersuchten Kindern im Alter von 9 bis 11 Jahren wurde das Testergebnis von 3 Lehrern signifikant häufiger als charakteristisch für das jeweilige konkrete Verhalten eines Kindes angesehen.

Die durch den starken Effekt von Verfälschungsinstruktionen nachgewiesene Durchschaubarkeit des Verfahrens (z. B. *Mucke 1972, Simons 1967b*) läßt auch eine andere Art der Selbstdarstellung als relativ wahrscheinlich erscheinen. Die Instruktion, wie ein „gutes Kind“ zu antworten, führt zu einer starken Abnahme der Extrapunitivität, insbesondere des Faktors E, und zu einem Anstieg der Impunitivität, z. T. auch der Intropunitivität. *Simons (1967b)* berichtet auch, daß 11- bis 14jährige im Gespräch erkennen ließen, die P-F S prüfe ihr Verhalten. Er sieht in dieser Beobachtung ein Argument für eine beschönigende Selbstdarstellung. Die bei Delinquenten beobachtete unterdurchschnittliche Extrapunitivität ist ebenfalls als Einstellungseffekt interpretiert worden (*Vane 1954, vgl. auch Rosenzweig 1963*).

Auch die Iteminhalte sind in Bezug auf die Gültigkeit diskutiert worden. *Wilson und Frumkin (1968)* kamen aufgrund eines Skalierungsexperimentes, das allerdings mit der Erwachsenenform durchgeführt wurde, zu dem Schluß, daß die dargestellten Frustrationssituationen sowohl hinsichtlich ihrer Thematik als auch hinsichtlich ihrer Intensität nicht als repräsentativer Querschnitt des Alltagslebens angesehen werden können. Nur bei zweien der Bilder gaben die 100 College-Studenten an, sie beschrieben Situationen, in denen sie sich schon einmal befunden hätten. Nur bei 12 Bildern wurde die Situation mehrheitlich als frustrierend aufgefaßt.

*Simons (1967a)* wies auf den die Antwortvarianz dominierenden Einfluß des Reizmaterials hin: Bei den über 700 sechs- bis elfjährigen Jungen und Mädchen deckten bei 10 Bildern die beiden häufigsten Antworten jeweils mehr als 66 % der Gesamtvarianz ab. *Simons* zog daraus den Schluß, daß dieser starke Reizeinfluß die interindividuelle Variation im Hinblick auf eine „gruppenspezifische Diskriminationsfähigkeit“ des Verfahrens zu stark vermindert. Diese Schlußfolgerung hat keine allgemeine Gültigkeit. Sie gilt vor allem, wenn der Reizeinfluß bei allen Items gleichartig ist, d. h. Antwortreaktionen derselben Diagnoseklassen zur Folge hat. Tatsächlich sind bestimmte Antworttypen recht häufig, andere dagegen recht selten (vgl. *Rosen-*

zweig, Fleming u. Rosenzweig 1948, Habets 1958). So betrug bei der deutschen Eichstichprobe von Duhm u. Hansen der Anteil der extrapunitiven Aggressionsrichtung mit ihren drei Faktoren über 57 % aller Signierungen. Der Faktor E wurde allein in über 33 % aller Fälle signiert. Der Faktor E war auch die einzige Signierungsmöglichkeit, von der bei allen Bildern Gebrauch gemacht wurde. Einer der übrigen Faktoren trat als Signierung nur bei 12 Bildern auf. Der Vorwurf fehlender Gruppenspezifität von Simons scheint danach nicht ungerechtfertigt zu sein. Er findet eine Stützung durch die im vorangehenden Abschnitt herausgestellte Beobachtung, daß sich bisher nur zwei (unterschiedliche) Reaktionsmuster abzeichnen, die mit Verhaltenskriterien korrelieren. Die unterschiedliche Häufigkeit der Antworttypen ist sicher auch mitverantwortlich für die teilweise fehlende Meßgenauigkeit der Testvariablen.

Unbeachtet bei der Gültigkeitsdiskussion blieb bisher die kinderformspezifische Besonderheit, daß bei den Bildvorlagen manchmal ein Kind, manchmal ein Erwachsener als frustrierende Person (Frustrator) auftritt. Zwischen dieser Art der Reizbedingung und der Art der Probandenäußerungen besteht aber eine charakteristische Wechselwirkung. Bei der deutschen Eichstichprobe lag der durchschnittliche Anteil der extrapunitiven Aggressionsrichtung für die neun Bilder mit einem Kind als Frustrator bei 68 %. Für die restlichen Bilder mit einem erwachsenen Frustrator lag der Anteil dieser Aggressionsrichtung bei 51 %. Für den Faktor E fällt der Unterschied bei %-Anteilen von 52 bzw. 22 noch deutlicher auf. Bei amerikanischen Kindern fand Spache (1951) ebenfalls erhöhte Extrapunitivität bei Bildern mit kindlichem Frustrator. Bei Bildern mit erwachsenem Frustrator zeigte sich erhöhte Intropunitivität und stärkere Betonung des Bedürfnisses nach Problemlösung<sup>5</sup>). Sandidge und Friedland (1973, zit. nach Maccoby und Jacklin 1974, S. 238 und 574) beobachteten mit vergleichbaren Bildvorlagen u. a. auch eine Interaktion zwischen dem Geschlecht der dargestellten Person und den verbalen Reaktionen der Probanden. Jungen und Mädchen gaben aggressivere Antworten, wenn die dargestellte Person ein Junge war.

Die Untersuchung von Spielberg und Rutkin (1974) legt nahe, daß die unterschiedlichen Reaktionen in Bezug auf die Art des Frustrators im Test mit realem Verhalten korrespondieren. Die achtjährigen Jungen wurden entweder von Gleichaltrigen oder Erwachsenen bei der Fertigstellung des Puzzles gestört. In der P-F S war abgesehen vom bereits erwähnten Haupteffekt Extrapunitivität häufiger, wenn der Störer ein Gleichaltriger war.

Wenn Kind und Erwachsener unterschiedliche Verhaltensmuster provozieren, so sind auch intraindividuelle Unterschiede hinsichtlich der Neigung, diese Verhaltensmuster zu zeigen, nicht auszuschließen. Die Nichtberücksichtigung dieser situationsspezifischen Unterschiede bei der konventionellen Auswertung der P-F S könnte eine weitere Fehlerquelle bei Gültigkeitsuntersuchungen darstellen.

<sup>5</sup>) In höherem Lebensalter scheint sich diese Beziehung umzukehren. Rosenzweig und Braun (1970) fanden mit einer speziell für Jugendliche entwickelten Testform bei 16- bis 18jährigen mehr extrapunitive und weniger intropunitive Reaktionen bei Bildern mit erwachsenem gegenüber Bildern mit gleichaltrigem Frustrator.

Insgesamt lassen sich bei einem größeren Teil der bisher durchgeführten Gültigkeitsuntersuchungen Mängel aufzeigen. Andererseits lassen sich die unbefriedigenden Befunde auch durch testspezifische Besonderheiten erklären. Auf jeden Fall machen die derzeit vorliegenden Erfahrungen deutlich, daß der Gültigkeitsanspruch der Kinderform nur teilweise gerechtfertigt ist. Dem Verfahren ist eine gewisse Diskriminationsfähigkeit in Bezug überdauernder Frustrationsempfindlichkeit nicht abzuspüren. Doch ist die Zahl der angenommenen Klassifizierungsmöglichkeiten der resultierenden Verhaltensmuster utopisch.

Eine differenziertere Einschätzung der Leistungsfähigkeit des Verfahrens, gerade auch im Hinblick auf die Situationspezifität, wird unter Umständen dann möglich, wenn man als eindeutiges Gültigkeitskriterium konkretes Verhalten heranzieht. Dies ist überraschenderweise bisher nicht geschehen.

## 2. Eigene Untersuchung

### 2.1. Fragestellung

Der Gültigkeitsanspruch der P-F S setzt letztlich voraus, daß es überdauernde und relativ situationsunabhängige Dispositionen für soziales Verhalten gibt. An diesem Traitkonzept des sozialen Verhaltens sind seit der Untersuchung von *Hartshorne* u. *May* (1928) immer wieder Zweifel geäußert und begründet worden, in neuerer Zeit insbesondere von *Mischel* (1968). Befunde, die für die Anwendbarkeit des Traitkonzeptes auf den fraglichen Verhaltensbereich sprechen, haben *Gormly* u. *Edelberg* (1974) vorgelegt. Sie berichten von einer relativ engen, wenig situationsabhängigen Konkordanz der Ergebnisse verschiedener Ratingmethoden hinsichtlich des Merkmals Aggressivität.

Unter diesem Aspekt scheinen Verhaltensbeurteilungen geeignete Kriteriumsvariable für die P-F S darzustellen. Sie sind auch verschiedentlich zu diesem Zwecke erhoben worden (vgl. Tabelle 1), haben aber zu keinen eindeutigen Ergebnissen geführt. Nur in zwei von insgesamt fünf Fällen fanden sich bedeutsame Kovariationen mit einzelnen Testvariablen in der erwarteten Richtung. Allerdings sind Ratingverfahren, wie auch *Gormly* und *Edelberg* zugeben, nicht frei von methodenspezifischen Fehlerquellen, so daß Urteilerkonkordanzen nicht unbedingt zutreffende Verhaltensbeschreibungen definieren. Eine fehlende Beziehung zum Testergebnis muß daher nicht mangelnde Verhaltensrelevanz des Tests bedeuten. Es scheint daher ratsam zu sein, neben Beurteilungen auch das Verhalten selbst als Kriteriumsvariable zu berücksichtigen.

Im vorliegenden Falle sollte der Zusammenhang zwischen P-F S, Beurteilungen der Aggressivität und dem aggressiven Verhalten selbst überprüft werden. Das Traitkonstrukt Aggressivität wurde ausgewählt, weil Anhaltspunkte für überdauernde individuelle Unterschiede bestehen, die P-F S besonders häufig zur Diagnostik dieses Konstruktes verwendet wird und hinsichtlich dieses Konstruktes noch am ehesten eine spezifische Merkmalskonstellation erwarten läßt. Außerdem stand ein Beobachtungsbogen zur Registrierung aggressiven Verhaltens zur Verfügung, der bei dieser Gelegenheit erprobt werden sollte.

Es wurde davon ausgegangen, daß ein vorhandener Zusammenhang zwischen Test- bzw. Ratingergebnis einerseits und dem Verhalten andererseits um so deutlicher in Erscheinung tritt, je freier die Versuchsperson in der Wahl ihrer Verhaltensweisen während der Phase der Verhaltensregistrierung ist. Ein Optimum hinsichtlich der Realisierbarkeit von sozialen Verhaltensweisen, die ein Kind bei realistischer Selbstdarstellung sich zuschreiben würde, schien uns das freie (ungelenkte) Spiel mit einem gewissen Angebot an (bekannten) Spielpartnern zu gewährleisten.

Ein weiteres Problem war bei der Auswahl der Versuchspersonen und der Rater zu berücksichtigen. Um für die Ratingvariable den Einfluß unsystematischer Beurteilungsfehler gering zu halten, sollten möglichst viele Kinder von möglichst wenig Ratern beurteilt werden. Diese Forderung war im vorliegenden Falle nur erfüllbar bei Heimkindern mit ihren Erziehern als Urteiler oder bei Schulkindern, wo die Lehrer entsprechende Urteile abgeben konnten. Wir entschieden uns gegen die zuletzt genannte Konstellation, unter anderem wegen der Begrenzung des fraglichen Verhaltens in der Schulsituation. Bei einer Heimkinderstichprobe stellt sich unter Umständen allerdings das Problem der Verallgemeinerbarkeit der Ergebnisse.

Im einzelnen sollte also überprüft werden, ob als „aggressiv“ beurteilte Heimkinder im Vergleich zu Kindern, die als „wenig aggressiv“ beurteilt werden, während des freien Spieles häufiger aggressives Verhalten zeigen und P-F S-Ergebnisse aufweisen, die nach dem Testkonzept im Sinne erhöhter Aggressivität interpretiert werden. Als P-F S-Merkmale für Aggressivität gelten zum Beispiel überdurchschnittlich hohe Extrapunitivitätswerte, insbesondere auch eine große Häufigkeit des Faktors E bei gleichzeitig unterdurchschnittlichen Werten bei den beiden anderen Aggressionsrichtungen (vgl. auch Tabelle 1).

In Bezug auf die Eigenart des Stimulusmaterials der P-F S sollte dabei die Frage beachtet werden, ob etwa vorhandene Kriteriumsbeziehungen enger ausfallen für Bilder, bei denen der Frustrator ein Kind ist, gegenüber Bildern, bei denen der Frustrator als Erwachsener auftritt.

Auch der Geschlechterabhängigkeit des aggressiven Verhaltens sollte Beachtung geschenkt werden. Da für die P-F S bisher keine Geschlechterunterschiede aufgezeigt werden konnten, stellt sich hier die Frage, ob es Anteile aggressiven Verhaltens beim freien Spiel gibt, für die keine Geschlechterunterschiede bestehen, und ob diese Anteile einen engen Zusammenhang mit den Testvariablen aufweisen. Im ersten Teil des Berichts war die Möglichkeit diskutiert worden, daß hinsichtlich verbaler Aggressionen keine Geschlechterunterschiede bestehen und daß entsprechend keine Unterschiede bei der P-F S auftreten, weil dort Merkmale verbaler Aggressionen erhoben werden.

## 2.2. Methode

2.2.1. *Versuchspersonen.* Bei den Versuchspersonen (Vpn) handelte es sich um 12 Bewohner eines Kinderheims der Wohlfahrtverbände in einer westdeutschen Großstadt und um 16 Kinder aus einem in der Nähe des Heimes gelegenen Kinderdorf. Die Hälfte der Stichprobe bestand aus

Mädchen. Die Aufnahme in die Institutionen war erfolgt wegen Verwaisung, wegen chronischer Erkrankung der Mutter oder wegen gestörter Familienverhältnisse. Das Durchschnittsalter der Kinder betrug 9,0 Jahre, bei einer Standardabweichung von 0,6 Jahren. Der bisherige Heimaufenthalt lag zwischen 3,0 und 5,0 Jahren. Jedes Kind hatte mindestens das letzte Drittel, in der Regel fast die Hälfte seines Lebens im Heim verbracht. Der durchschnittliche IQ betrug 98,6 Punkte bei einer Standardabweichung von 6,4 Punkten, lag also im Normbereich. Zwei der Kinder zeigten hin und wieder Verhaltensstörungen (Bettnässen, hysterische Symptome). Bei zwei anderen Kindern lagen Anzeichen für eine verlangsamte geistige Entwicklung vor.

*2.2.2. Einteilung nach der beurteilten Aggressivität.* Im vorliegenden Versuch wurde als unabhängige Variable das Ratingergebnis benutzt. Definiert wurde sie durch Extremgruppen. Die 28 Vpn waren aus einer größeren Gruppe von ca. 40 Kindern ausgewählt worden. Jedes Kind dieser Ausgangsgruppe war durch zwei Beurteiler eingestuft worden, und zwar alle aus dem Kinderheim stammenden Vpn durch den Heimleiter und eine Erzieherin, alle Vpn aus dem Kinderdorf durch dieselben zwei Erzieherinnen. Die Einstufung war auf einer 6-Punkte-Skala erfolgt, deren Pole mit „kaum aggressiv“ (= 1 Punkt) und mit „sehr aggressiv“ (= 6 Punkte) gekennzeichnet waren.

Der Begriff Aggressivität war vorher sinngemäß als Bereitschaft zu Verhaltensweisen definiert worden, mit denen die direkte oder indirekte Schädigung eines Mitmenschen herbeigeführt wird. Es wurde weiter darauf hingewiesen, daß der Ausprägungsgrad der Aggressivität durch die Gesamtheit der aggressiven Handlungen bestimmt sei.

Von den beurteilten Kindern wurden die sieben Jungen und die sieben Mädchen mit den geringsten Punktwerten auf der Aggressivitätsskala als Extremgruppe „nicht aggressive Vpn“ ausgewählt. Die pro Kind in den beiden Beurteilungen erzielte Punktsumme lag im Höchstfalle bei 4, im Mittel bei 2,9 Punkten. Entsprechend wurden die sieben Jungen und die sieben Mädchen mit den höchsten Punktsummen (im Mindestfalle 9, im Durchschnitt 10,6 Punkte) zur Extremgruppe „aggressive Vpn“ zusammengefaßt. Es wurden also nur konkordant beurteilte Kinder berücksichtigt. Bei der Auswahl zeigte es sich, daß die Jungen gegenüber den Mädchen allgemein als aggressiver eingestuft worden waren. Diese Geschlechterabhängigkeit der Beurteilung wurde bei der Zusammenstellung der Extremgruppen nicht ganz eliminiert.

*2.2.3. Die Spielgruppen.* Unter Berücksichtigung der institutionellen Herkunft, der Extremgruppenzugehörigkeit und des Geschlechtes wurden Spielgruppen zu je vier Kindern gebildet. Jede Spielgruppe bestand aus zwei Jungen und zwei Mädchen, die aus der gleichen Institution kamen, sich also kannten, und die sich auch hinsichtlich des Alters, des Heimaufenthaltes und des Intelligenzgrades möglichst wenig unterschieden. Um die noch wenig untersuchte Wechselwirkung von individuellen Unterschieden der Aggressivität auf das Verhalten auszuschalten, spielten bis auf einen Fall immer nur Vpn derselben Extremgruppe miteinander. Aus den 28 Vpn wurden also sieben Spielgruppen gebildet: drei mit aggressiven Kindern, drei mit nicht

aggressiven Kindern und eine mit zwei aggressiven und zwei nicht aggressiven Kindern jeweils unterschiedlichen Geschlechts.

**2.2.4. Spielsituation.** Zum Spielen wurden die größten Spielräume im Heim und im Kinderdorf hergerichtet, so daß eine möglichst große freie Fläche zur Verfügung stand. Es wurden altersgemäße Spielsachen (Autos, Puppen, Bausteine usw.) und die im Raum befindlichen Turngeräte zur Verfügung gestellt. Das Spiel fand damit in einer vertrauten Umgebung statt. Die reine Spielzeit betrug eine Stunde. Jede Spielgruppe hatte in dieser Zeit den Raum für sich allein.

**2.2.5. Der Verhaltensbeobachtungsbogen.** Der zur Verhaltensregistrierung benutzte Beobachtungsbogen wurde freundlicher Weise von *B. Ehlers* zur Verfügung gestellt. Er war zu Ausbildungszwecken im Rahmen des Kinderbeobachtungspraktikums entwickelt worden. Als Anregung diente dabei ein von *Koch* (1967) beschriebenes Verfahren.

In der hier verwendeten Form enthielt der Bogen eine Liste von 39 Verhaltensweisen. Diese waren durch Ratings nach fünf Verhaltenskategorien unterteilt worden: Aggression (I), Ignorieren (II), Beobachten (III), Assoziativer Kontakt (IV) und Kooperativer Kontakt (V). Die Liste und ihre Unterteilung sind in Tabelle 2 dargestellt.

Tab. 2: Items des Beobachtungsbogens

---

<b>I Aggression</b>	20. Um Hilfe bitten
1. Schadenfreude zeigen	21. Mitleid zeigen
2. Gegenstand wahrnehmen	22. Gegenstand erbitten
3. Körperl. Auseinandersetzung	23. Vormachen
4. Prahlern	24. Kritik üben
5. Auseinandersetzung mit Worten	25. Anweisung geben
6. Etwas verbieten	26. Befolgen von Weisungen
	27. Anlächeln
<b>II Ignorieren</b>	28. Etwas zeigen
7. Aus dem Wege gehen	29. Vorschlag machen
8. Auf Ansprechen und Berührung nicht reagieren	
9. Hilfe verweigern	<b>V Kooperativer Kontakt</b>
10. Etwas ablehnen	30. Gegenstand austauschen
	31. Zusammen spielen
<b>III Beobachten</b>	32. Leistungsvergleich
11. Zuschauen	33. Sich gegenseitig verbessern
12. Bei der Hand fassen	34. Teilen
13. Zuhören	35. Freude über fremde Leistung
14. Sich daneben stellen/setzen	36. Gegenstand abgeben
	37. Gemeinsames Ziel setzen
<b>IV Assoziativer Kontakt</b>	38. Helfen
15. Sich unterhalten	39. Sich helfen lassen
16. Ansprechen	
17. Nachgeben	
18. Nachahmen	
19. Zärtlichkeit erweisen	

---

Während der Beobachtungszeit ist die Häufigkeit des Auftretens der Items in Form von Strichlisten vom Beobachter festzuhalten. Mit einer verbalen Erläuterung der Verhaltensweisen wird eine ausreichende Konkordanz zwischen Beobachtern eines Kindes bereits nach einer Beobachtungsstunde erreicht. Es hat sich darüber hinaus gezeigt, daß eine Beobachtungszeit von 40 Minuten (davon 20 Minuten eigentliche Registrierzeit) ausreicht, um eine durchweg zufriedenstellende innere Konsistenz der Registrierungen pro Kategorie zu erreichen. Gürge (1970) korrelierte die von einem Beobachter bei 40 Kindergartenkindern während des freien Spiels erhobenen Häufigkeitssummen aus dem ersten und dritten Beobachtungsabschnitt mit den entsprechenden Summen aus dem zweiten und vierten Abschnitt (1 Abschnitt = 10 Minuten). Es ergaben sich für die fünf Kategorien die folgenden, nach Spearman-Brown korrigierten Koeffizienten: 0,61 (I), 0,51 (II), 0,80 (III), 0,72 (IV) und 0,84 (V).

Im vorliegenden Falle wurde als eigentliche Versuchsvariable für jede Vp der auf die einzelne Kategorie entfallende Anteil (Prozent) am gesamten registrierten Verhalten berechnet. Es interessierten vor allem die auf die Kategorie Aggression entfallenden Verhaltensanteile.

Um die Geschlechterunterschiede des aggressiven Verhaltens zu kontrollieren, wurden zusätzlich die absoluten Häufigkeiten für die körperlichen und die verbalen Aggressionen getrennt erfaßt. Als körperliche Aggressionen wurden die Items Nr. 2 (Gegenstand wegnehmen) und Nr. 3 (körperliche Auseinandersetzung) angesehen, als verbale Aggressionen die übrigen vier Verhaltensweisen der Kategorie I.

**2.2.6. Die Beobachter.** Als Beobachter stellten sich zwei männliche und zwei weibliche Psychologiestudenten mittlerer Semesterzahl im Rahmen eines Praktikums zur Verfügung. Ihnen war die Klassifikation der Vpn nach der Aggressivität unbekannt. Die Reihenfolge der zu beobachtenden Spielgruppen wurde zudem hinsichtlich des fraglichen Merkmals unsystematisch gehalten.

Jeder der vier Beobachter hatte jeweils ein Mitglied der Spielgruppe zu beobachten. Um geschlechterspezifische Wechselwirkungen möglichst gering zu halten, hatte der eine männliche Beobachter stets einen Jungen, der andere ein Mädchen zu beobachten. Entsprechend wurde bei den weiblichen Beobachtern verfahren. Von Spielgruppe zu Spielgruppe wurden zwischen den gleichgeschlechtlichen Beobachtern diese Zuordnungen getauscht.

Um den Kindern die Gewöhnung an die Beobachtungssituation zu erleichtern, wurden sie so gut es ging von den Beobachtern getrennt. In der einen Institution konnten die Beobachter auf einer Empore plaziert werden. Im anderen Falle saßen sie auf einem großen Mattenhaufen.

**2.2.7. P-F S.** Die Kinderform des Tests wurde im Einzelversuch von einem fünften Psychologiestudenten durchgeführt, der an den Beobachtungen nicht teilnahm. Die Antworten wurden vom Testleiter registriert. Die Testauswertung erfolgte doppelt, und zwar durch den Verfasser und eine in der diagnostischen Praxis erfahrene Diplompsychologin unabhängig voneinander und in Unkenntnis der Klassifikation der Vpn nach der Aggressivität. Es war vorher festgelegt worden, daß lediglich die Ergebnisse der Psycho-

login in die eigentliche Datenanalyse einbezogen werden sollten. Die Auswertung erstreckte sich nur auf die Faktoren (vgl. Abbildung 1). GCR, Superego-Patterns und „Tendenzen“ wurden nicht berücksichtigt.

2.2.8. *Versuchsdurchführung.* Die Datenerhebung erfolgte im Rahmen einer Semesterarbeit von H. P. Steden (1970). Den in den Spielraum geführten Kindern wurde folgende Instruktion gegeben: „Liebe Kinder, hier sind vier Leute, die mal sehen wollen, wie Ihr spielen könnt, wenn Ihr alles das tun könnt, was Euch Spaß macht!“ Nach einer Viertelstunde Anwärmzeit begann die eigentliche Beobachtung, und zwar für 40 Minuten (Registrierzeit). Im Anschluß an die Spielstunde wurde im Sekretariat des Heimes bzw. des Kinderdorfes das P-F S-Protokoll aufgenommen. Während dieser Testuntersuchung machten die restlichen Kinder der Spielgruppe in einem Nebenraum unter Aufsicht Schulaufgaben.

### 2.3. Ergebnisse

2.3.1. *Beziehung zwischen Verhaltensbeurteilung und -beobachtung.* In Tabelle 3 sind die Ergebnisse der Verhaltensbeobachtung dargestellt. Die Unterschiede zwischen den als „aggressiv“ und den als „nicht aggressiv“ beurteilten Kindern wurden hinsichtlich der statistischen Bedeutsamkeit ihrer zentralen Tendenzen mit dem *Mann-Whitney-U-Test* überprüft.

Tab. 3: Ergebnisse der Verhaltensbeobachtung bei 14 als aggressiv und 14 als nicht aggressiv beurteilten Kindern. Darstellung der durchschnittlichen Anteile ( $\bar{X}$ ) am Gesamtverhalten, der Standardabweichung (SD) und der Zufallswahrscheinlichkeit (p) des Unterschieds

	„aggressive“		„nicht aggressive“		p
	$\bar{X}$	SD	$\bar{X}$	SD	
I Aggression	35,8	9,5	12,6	7,7	< 0,002
II Ignorieren	7,4	4,1	3,7	5,5	< 0,05
III Beobachten	8,4	7,4	10,6	9,2	
IV Assoziativer Kontakt	37,6	12,8	39,8	11,9	
V Kooperativer Kontakt	10,8	6,7	32,7	8,8	< 0,002

Unterschiede zwischen den beiden Gruppen traten vor allem bei den zwei hier als relevant anzusehenden Verhaltenskategorien in Erscheinung. So entfiel bei den aggressiven Kindern im Durchschnitt mehr als ein Drittel des registrierten Gesamtverhaltens auf die Kategorie Aggression. Bei den nicht aggressiven waren es dagegen nur knapp 13 Prozent. Umgekehrt kam es bei den nicht aggressiven Vpn zu häufigen kooperativen Kontakten. Sie machten hier fast ein Drittel des Gesamtverhaltens aus, bei den aggressiven Kindern dagegen nur etwas über 10 Prozent.

Die Unterteilung der Aggressionen nach körperlichen und verbalen ergab, daß bei beiden Aggressionsarten die Gruppenunterschiede in gleicher Weise hervortraten. Die durchschnittlichen Häufigkeiten an registrierten körperlichen und verbalen Aggressionen betragen bei den aggressiven Vpn 7,5 bzw.

11,6, bei den nicht aggressiven 1,4 bzw. 4,6. Für beide Verhaltensweisen lag die Zufallswahrscheinlichkeit des Unterschieds zwischen den Gruppen unter 0,2 Prozent.

Die den statistisch hochbedeutsamen Unterschieden entsprechende enge inhaltliche Beziehung zwischen Verhaltensbeurteilung und Verhaltensbeobachtung kann zahlenmäßig deutlich gemacht werden, wenn man die Vpn in Vierfelderschemata einordnet, wie das in Abbildung 2 geschehen ist. Dabei ist das Beurteilungskriterium wieder durch die Extremgruppenzugehörigkeit definiert. Zur Einordnung nach dem Verhalten wurden die entsprechenden Medianwerte (Mdn) der Gesamtstichprobe herangezogen.

Abb. 2: Aufteilung der Vpn nach Beurteilung und a) dem Anteil an Aggression, b) dem Anteil an kooperativem Kontakt, c) der Differenz zwischen beiden Anteilen

	Aggression		Koop. Kontakt		Aggression	Aggression
	> Mdn	< Mdn	> Mdn	< Mdn	> Koop. Kontakt	< Koop. Kontakt
„aggressiv“	13	1	1	13	14	0
„nicht aggressiv“	1	13	13	1	0	14
	a		b		c	

Die Abbildung zeigt, daß Kinder, deren Anteil an Aggression oberhalb des Mdn oder deren Anteil an Kooperation unterhalb des Mdn liegen, jeweils bis auf einen Fall der als aggressiv beurteilten Extremgruppe angehören (a, b). Werden die Anteile an Aggression und Kooperation pro Kind gegeneinander abgewogen, ergibt sich sogar eine fehlerlose Klassifikation (c).

2.3.2. *Beziehung zwischen Beurteilung bzw. Verhalten und P-F S.* Die Übereinstimmung der beiden Auswerter hinsichtlich der Antworteinstufungen eines Protokolls variierte zwischen 81 und 100 Prozent und betrug im Mittel 94 Prozent bei der P-F S.

Die weitestgehende Entsprechung von beurteilter und beobachteter Aggressivität bedeutete, daß die bisher benutzte Unterteilung der Vpn in „aggressive“ und „nicht aggressive“ auch als eine angemessene Kriteriumsdefinition für konkretes Verhalten anzusehen war und daher beibehalten werden konnte. Die Unterschiede zwischen den beiden Gruppen bei der P-F S gehen aus Tabelle 4 hervor.

Aggressive Vpn gaben gegenüber nicht aggressiven Vpn vor allem mehr extrapunitive Antworten vom Typ der Ich-Verteidigung (Faktor E) ab. Auch die beiden anderen extrapunitiven Reaktionsmöglichkeiten kamen statistisch bedeutsam (E') bzw. zahlenmäßig häufiger vor, so daß hinsichtlich der gesamten Aggressionsrichtung ein deutlicher Unterschied zwischen den beiden Gruppen bestand. Weiterhin zeigten die aggressiven Vpn gegenüber den nicht aggressiven durchgängig und statistisch hoch bedeutsam seltener impunitiv Reaktionen (M', M, m). Bei der intropunitiven Aggressionsrichtung bestand dagegen nur ein Unterschied, und zwar für den Faktor i. Er fand sich bei den aggressiven Vpn weniger häufig. Diese Ergebniskonstellationen bedeuten

Tab. 4: Ergebnisse der P-F S. Darstellung der durchschnittlichen Punktwerte ( $\bar{X}$ ), ihrer Standardabweichung (SD) und der Zufallswahrscheinlichkeit (p) des Unterschieds zwischen den Gruppen

	Faktoren	„aggressive“		„nicht aggressive“		p
		$\bar{X}$	SD	$\bar{X}$	SD	
1.	E'	1,82	1,20	0,75	1,02	< 0,05
2.	E	10,61	2,42	4,57	3,47	< 0,002
3.	c	3,36	1,52	2,29	1,33	
4.	I'	1,04	1,35	0,67	0,71	
5.	I	3,00	1,65	3,14	1,42	
6.	i	1,64	0,69	2,89	1,25	< 0,02
7.	M'	1,43	1,00	3,68	1,72	< 0,002
8.	M	0,36	0,63	2,36	1,22	< 0,002
9.	m	0,75	0,58	3,71	1,16	< 0,002

gleichzeitig, daß hinsichtlich der Reaktionstypen keine durchgängig gleichsinnigen Veränderungen der Faktoren vorlagen.

Die Zusammenhänge zwischen der vorgegebenen Klassifikation der Vpn und ihrer Einstufung an Hand der von *Duhm* u. *Hansen* (1957) mitgeteilten deutschen Normen werden zahlenmäßig durch die Kontingenztafeln a und b in der Abbildung 3 veranschaulicht. Ausgewählt wurden für diese Darstellung die beiden Testvariablen, für die sich die bedeutsamsten Gruppenunterschiede ergeben hatten, der Faktor E und die Impunitivität. Für beide Variablen wurde eine Aufgliederung der Vpn nach der Position ihrer Ergebnisse zu den als Normen mitgeteilten Quartilgrenzen ( $Q_1$ , Mdn,  $Q_3$ ) vorgenommen.

Abb. 3: Klassifikation der Vpn auf der Basis der deutschen Normen für a) Faktor E, b) Impunitivität, c) Kombination von E und Impunitivität

	Faktor E			Impunitivität			Kombination		
	> $Q_3$	Mdn	< $Q_1$	> $Q_3$	Mdn	< $Q_1$	+	0	-
„aggressive“	8	6	0	0	5	9	11	3	0
„nicht aggressive“	1	1	12	12	2	0	1	1	12
	a			b			c		

Wie die Abbildung zeigt, ergab sich sowohl für den Faktor als auch für die Aggressionsrichtung eine deutliche zahlenmäßige Beziehung zu den Kriteriumsgruppen. Wurden die Quartilbewertungen von E und Impunitivität konfigural miteinander kombiniert, so fiel die Übereinstimmung noch etwas überzeugender aus (c). Der Kombination lag dabei folgende Punktbewertung zugrunde:

Pluspunkt (+) bei  $E > Q_3$  oder Impunitivität  $< Q_1$ ; Minuspunkt (-) bei  $E < Q_1$  oder Impunitivität  $> Q_3$ .

2.3.3. *Geschlechterunterschiede.* Die im vorliegenden Falle interessierenden Geschlechterunterschiede sind in Tabelle 5 dargestellt. Ihre statistische Überprüfung erfolgte gleichsam unter konservativen Bedingungen, da die Art der Vpn-Auswahl eine erhebliche Varianz innerhalb der Geschlechtergruppen bewirkte.

Tab. 5: Geschlechterabhängigkeit der Beurteilung und ausgewählter Variablen der Beobachtung und der P-F S. Darstellung der Punktwerte bzw. Häufigkeiten

	Jungen (N = 14)		Mädchen (N = 14)		Unterschied (n. s. = nicht signifikant)
	$\bar{X}$	SD	$\bar{X}$	SD	
Beurteilung	7,3	4,2	6,1	3,9	n. s.
körperliche Aggression	6,4	5,4	2,7	2,9	p < 0,05
verbale Aggression	8,3	4,2	8,0	5,6	n. s.
E	8,5	4,4	6,6	4,0	n. s.
Impunitivität	6,0	4,5	6,3	3,9	n. s.

Bei der Beurteilungsvariable wurde die bereits früher erwähnte Tatsache deutlich, daß die Jungen gegenüber den Mädchen als aggressiver eingestuft wurden. Der Unterschied war hier unwesentlich. Beim Verhalten ließ sich ein deutlicher Geschlechterunterschied nur für die körperlichen Aggressionen nachweisen, die bei den Jungen häufiger registriert wurden. Verbale Aggressionen zeigten Jungen und Mädchen dagegen gleichhäufig. Von den beiden P-F S-Variablen, die die engste Kriteriumsbeziehung aufgewiesen hatten, zeigte keine eine statistisch bedeutsame Geschlechterabhängigkeit. Rein zahlenmäßig fand sich der Faktor E bei den Jungen etwas häufiger als bei den Mädchen.

Über den Zusammenhang der durch Kombination von E und Impunitivität gebildeten P-F S-Variable mit der körperlichen und der verbalen Aggression informiert die Abbildung 4. Der Zusammenhang ist in beiden Fällen eng und für die verbale Aggression nur unwesentlich ausgeprägter als für die körperliche.

Abb. 4: Aufteilung der 28 Vpn nach dem Ergebnis der Kombination von E und Impunitivität und nach dem Ausprägungsgrad der körperlichen bzw. der verbalen Aggression

		Kombination				Kombination			
		+	0	-		+	0	-	
körperliche Aggression	> Mdn	9	3	1	verbale Aggression	> Mdn	11	2	0
	< Mdn	3	1	11		< Mdn	2	2	11

**2.3.4. Der Frustratoreffekt.** Um die Auswirkungen dieser testspezifischen Bedingung zu untersuchen, wurde für jede Vp der Punktwert des Faktors E einmal für die neun Bilder mit einem Kind als Frustrator und zum anderen für die restlichen fünfzehn Bilder mit einem erwachsenen Frustrator festgestellt. Für die Bilder mit dem Kind als Frustrator ergab sich ein Gruppenschnitt von 4,7 Punkten bei einer Standardabweichung von 2,1 Punkten, für die größere Zahl der übrigen Bilder dagegen nur ein Durchschnitt von 2,9 Punkten mit einer Standardabweichung von 2,5 Punkten. Bereits für diese absoluten Zahlenwerte ist der Unterschied statistisch hochbedeutsam ( $p < 0,01$ ). Trotz dieses Mittelwertunterschiedes korrelieren die Variablen zu  $r = 0,72$ . Dabei ist zu berücksichtigen, daß dieser Koeffizient, bedingt durch Varianzerhöhung infolge der Art der Vpn-Auswahl, eine Überschätzung des wahren Zusammenhangs darstellt.

Die Beziehung der beiden E-Variablen zu den Kriteriumswerten wird zahlenmäßig durch entsprechende Aufteilung der Vpn in Abbildung 5 veranschaulicht. Beide Variablen zeigten jeweils gleichartige enge Zusammenhänge mit Beurteilung, körperlicher und verbaler Aggression.

*Abb. 5:* Klassifikation der Vpn auf der Basis des Faktors E für die Bilder mit einem Kind als Frustrator und für Bilder mit einem Erwachsenen als Frustrator

	Kind		Erwachsener		
	E > Mdn	E < Mdn	E > Mdn	E < Mdn	
Beurteilung	„aggressive“	13	1	12	2
	„nicht aggressive“	1	13	2	12
körperliche Aggression	> Mdn	12	2	11	3
	< Mdn	2	12	3	11
verbale Aggression	> Mdn	12	2	12	2
	< Mdn	2	12	2	12

## 2.4. Diskussion

Die engen Beziehungen zwischen den Ergebnissen der Beurteilung, der Beobachtung und der P-F S sprechen dafür, daß es überdauernde Merkmale des individuellen Verhaltens gibt, die unter das Traitkonstrukt „Aggressivität“ gefaßt werden können. Sie unterstützen damit gleichzeitig die Annahme, daß sowohl das Ratingverfahren als auch der *Rosenzweig-Test* geeignete Methoden darstellen, eine Person diesem Konstrukt zuzuordnen. In Bezug auf den Test deuten die Ergebnisse außerdem daraufhin, daß testspezifische Varianzquellen die Gültigkeit weniger beeinträchtigen als bisher vermutet

werden konnte. Die Reaktionen zu den Bildvorlagen entsprachen einer realistischen Selbstdarstellung.

Die Anlage der Untersuchung war am ehesten der von *Mucke* (1973) vergleichbar. Auch die Ergebnisse zeigen große Übereinstimmung. Von den Faktoren trennte bei *Mucke* ebenfalls E am schärfsten die „aggressiven“ von den „nicht aggressiven“ Vpn. Ebenso zeigten von den Aggressionsrichtungen Extrapunitivität und Impunitivität die deutlichsten Unterschiede. Diese Befundbestätigung spricht zusammen mit den Hinweisen, daß E und Impunitivität zu den zuverlässigsten P-F S-Variablen gehören, dafür, daß mit Hilfe der konfiguralen Kombination von Faktor und Aggressionsrichtung eine zutreffende Klassifikation kindlicher Probanden nach ihrer Aggressivität gelingt.

Das gegen den Gültigkeitsanspruch der P-F S angeführte Argument der fehlenden Geschlechterunterschiede konnte zumindest in Frage gestellt werden. Im Unterschied zu den konkreten körperlichen Aggressionen zeigten die konkreten verbalen Aggressionen auch keinen Geschlechterunterschied. Außer dem Aufzeigen einer leichten zahlenmäßigen Tendenz konnte allerdings kein Nachweis dafür erbracht werden, daß das geschlechterunabhängige Testergebnis enger mit den geschlechterunabhängigen verbalen als mit den geschlechterabhängigen körperlichen Aggressionen verknüpft war: Es bleibt offen, ob die gleichartige Beziehung der P-F S zu beiden Aggressionsarten infolge der Vorauswahl der Vpn als methodisches Artefakt zustande gekommen ist oder im Sinne der Kanalisationshypothese von *Feshbach* (1970) interpretiert werden muß.

Ebenso ist der Befund hinsichtlich der Auswirkung der Frustratorperson nicht eindeutig zu bewerten. Zwar ließ sich der an Hand der Daten des deutschen Testmanuals beschriebene Effekt bestätigen, daß Bilder mit einem Kind als Frustrator häufiger E-Antworten provozieren als Bilder mit einem Erwachsenen als Frustrator. Die Erwartung, daß die Bildserie mit dem kindlichen Frustrator auch die engeren Beziehungen zum konkreten Verhalten gegenüber Kindern aufweist, wurde dagegen nicht bestätigt. Dafür könnte ebenfalls die Vorselektion der Vpn als Ursache angesehen werden. Andererseits ist es auch vorstellbar, daß die beiden Bildserien deshalb unterschiedliche E-Häufigkeiten produzieren, weil die Situation mit dem Erwachsenen nicht etwa Hemmungsmechanismen wirksam werden läßt, sondern real als weniger frustrierend erlebt wird. Beide Bildserien würden sich dann nur hinsichtlich der Intensität der Vereitelung unterscheiden, ihre Beziehung zu den Außenkriterien könnte gleichartig sein. Für diese Möglichkeit spricht die hohe Korrelation der beiden Bildserien. Auf mögliche Intensitätsunterschiede weist zudem die Tatsache hin, daß der kindliche Frustrator häufiger als eine Art Angreifer auftritt (bei etwa 6 der 9 Bilder) als der erwachsene Frustrator, der nur in etwa 6 von 15 Fällen direkt beleidigt, straft oder vorenthält.

Geht man davon aus, daß beide Bildserien einen unterschiedlich hohen, sonst aber gleichartigen Anregungsgehalt für verbale Reaktionen besitzen, so kann die gleichhohe Gültigkeitsbeziehung beider Bildserien auch als Hinweis dafür gedeutet werden, daß für die untersuchte Stichprobe keine Wechselwirkung zwischen Merkmalsausprägung und Anregungsgehalt der Bildvorlagen in Bezug auf das Merkmal besteht. Der fragliche Tatbestand, daß

aggressive Reaktionen zu einer Bildvorlage mit einem Kriterium für Aggressivität erst dann korrelieren, wenn die Bildvorlage selbst in höherem Maße Feindseligkeit darstellt, ist für den TAT beschrieben und gedeutet worden (Kaplan 1967, James u. Mosher 1967).

Ein Problem stellt auch die Verallgemeinerbarkeit der Befunde dar, und zwar sowohl von der Art der Stichprobe als auch von der Art der Versuchsdurchführung her. Einmal ist daran zu denken, daß die beobachteten Kinder sich in einem Lebensalter befanden, wo Aggressivität noch relativ ungehemmt in spontanes Verhalten umgesetzt wird. Es ist nicht zu erwarten, daß sich für ältere Gruppen ähnlich enge Kriteriumsbeziehungen ergeben. Weiter ist zu bedenken, daß die beiden Extremgruppen von zusammen 28 Kindern aus einer Ausgangsstichprobe von nur etwas über 40 Kindern zusammengestellt werden konnten. Jedes zweite Kind, insbesondere jeder zweite Junge, war von seinen Erziehern auf der Aggressivitätsskala einem der beiden Extreme zugeordnet worden. Obwohl sich in der betreffenden Altersstufe das Verhalten beim freien Spiel im fraglichen Merkmalsbereich stark in Bezug auf Aggression und Kooperation polarisiert, dürfte der Prozentsatz an extrem aggressiven und extrem nicht aggressiven Kindern in der Regel geringer sein. Es ist daher nicht auszuschließen, daß auch die Art der erfaßten Verhaltensdisposition „Aggressivität“ etwas von dem Regelfall abweicht.

Auch der Einfluß der Versuchsdurchführung auf die Kriteriumsbeziehungen ist zu beachten. So könnte der hohe Bekanntheitsgrad der Vpn untereinander als eine der Realität Rechnung tragende Bedingung die Zusammenhänge zwischen den Variablen optimiert haben. Andererseits könnte durch die Zusammensetzung der Spielgruppen nach gleicher Merkmalsausprägung eine gegenseitige Verstärkung bewirkt worden sein mit der Folge einer Verhaltenspolarisierung beim Spiel und der anschließend durchgeführten P-F S. Auf diese Weise würde sich eine gewisse Überschätzung der wahren Verhältnisse ergeben.

### 3. Literatur

- Bandura, A., Ross, D. und Ross, S. A.*: Transmission of aggression through imitation of aggression models. *Journal of Abnormal and Social Psychology* 1961, 63, 575–582
- Bjerstedt, A.*: Rosenzweig Picture-Frustration Study. Review. In: *O. K. Buros* (Hrsg.), *The sixth mental measurement yearbook*. Highland Park: Gryphon Press 1965, 511–516
- Brickenkamp, R.*: *Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe 1975
- Bruant, G.*: Les composantes psychologiques de l'attitude cyphotique. *Bulletin de Psychologie* 1973/74, 27, 331–345
- Coleman, J. C.*: Stimulus factors in the relation between fantasy and behavior. *Journal of Projective Technique and Personality Assessment* 1967, 31, 67–73
- Davreux, L.*: Quelques réflexions sur le test de Rosenzweig. *Revue de Psychologie et des Sciences de l'Éducation* 1969, 4, 448–458
- Duhm, E.*: Die Reaktionen von Problemkindern im Rosenzweig P-F-Test. *Psychologische Rundschau* 1959, 10, 283–291
- Duhm, E. und Hansen, J.*: *Der Rosenzweig P-F Test. Form für Kinder. Handanweisung der deutschen Bearbeitung*. Göttingen: Hogrefe 1957

- Ferguson, R. G.:** Some developmental factors in childhood aggression. *Journal of Educational Research* 1954, 48, 15–27
- Feshbach, S.:** Aggression. In: *P. H. Mussen* (Hrsg.), *Carmichael's manual of child psychology*. New York: Wiley 1970, Vol. 2, 159–260
- Feshbach, N. und Feshbach, S.:** Children's aggression. In: *W. W. Hartup* (Hrsg.), *The young child: Reviews of research*. Vol. 2. Washington, D.C.: National Association for the Education of Young Children 1972, 284–302
- Gormly, J. und Edelberg, W.:** Validity in personality trait attribution. *American Psychologist* 1974, 29, 189–193
- Gürge, W.:** Soziale Verhaltensweisen von Kindergartenkindern. Marburg: Unveröffentlichte Semesterarbeit 1970
- Habets, J. J.:** Enige bevindingen over de Rosenzweig "Picture-Frustration Study" voor kinderen. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden* 1958, 13, 205–228
- Hartshorne, H. und May, M. A.:** Studies in the nature of character. Vol. 1. Studies in deceit. New York: Macmillan 1928
- Hörmann, H. und Moog, W.:** Der Rosenzweig P-F Test. Form für Erwachsene. Handanweisung der deutschen Bearbeitung. Göttingen: Hogrefe 1957
- Hiltmann, H.:** Wortassoziationen und verbale Ergänzungsverfahren. III. Der Rosenzweig Picture-Frustrations-Test. In: *R. Heiss* (Hrsg.), *Handbuch der Psychologie*, 6. Band, *Psychologische Diagnostik*. Göttingen: Hogrefe 1964, 545–549
- James, P. P. und Mosher, D. L.:** Thematic aggression, hostility guilt and aggressive behavior. *Journal of Projective Techniques and Personality Assessment* 1967, 31, 61–67
- Johannsen, D. E. und Bennett, E. M.:** The personality of diabetic children. *The Journal of Genetic Psychology* 1955, 87, 175–185
- Kaplan, M. F.:** The effect of cue relevance, ambiguity, and self-reported hostility on TAT responses. *Journal of Projective Techniques and Personality Assessment* 1967, 31, 45–50
- Kaswan, J., Wasman, M. und Freedman, L. Z.:** Aggression and the Picture-Frustration Study. *Journal of Consulting Psychology* 1960, 24, 446–452
- Kates, S. L.:** Suggestibility, submission to parents and peers, and extrapunitive, intropunitive, and impunitiveness in children. *Journal of Psychology* 1951, 31, 233–241
- Keil-Specht, H. und Denzer, S.:** Möglichkeiten der Diagnostik aggressiven Verhaltens. In: *A. Schmidt-Mummendey und H. D. Schmidt* (Hrsg.), *Aggressives Verhalten*. München 1972, 189–208
- Koch, I.:** Untersuchungen zum Training sozialer Verhaltensweisen im Vorschulalter. *Probleme und Ergebnisse der Psychologie* 1967, 22, 7–36
- Krieger, L. und Schwartz, M. M.:** The relationship between sociometric measures of popularity among children and reaction to frustration. *Journal of Social Psychology* 1965, 66, 291–296
- Levitt, E. E. und Lyle, W. H.:** Evidence for the validity of the children's form of the P-F S. *Journal of Consulting Psychology* 1955, 19, 381–386
- Lindzey, G. und Goldwyn, R. M.:** Validity of the Rosenzweig P-F Study. *Journal of Personality* 1953, 22, 519–547
- Lipman, R. S.:** Some test correlates of behavioral aggression in institutionalized retardates with particular reference to the Rosenzweig P-F Study. *American Journal of Mental Deficiency* 1959, 63, 1038–1045
- Maccoby, E. E. (Hrsg.):** The development of sex differences. Stanford, Calif.: Stanford University Press 1966
- Maccoby, E. E. und Jacklin, C. N.:** The psychology of sex differences. Stanford: Stanford University Press 1974
- McIntyre, A.:** Sex differences in children's aggression. *Proceedings of the 80th Annual Convention of the APA* 1972, 7, 93–94
- Mirmow, E. L.:** The Rosenzweig Picture-Frustration Study. In: *D. Brower und L. E. Abt* (Hrsg.), *Progress in clinical psychology*, Vol. I, 1, New York 1952, 209–221
- Mischel, W.:** Personality and assessment. New York: Wiley 1968

- Mucke, H.*: Zur Problematik der sozialen Verhaltensdiagnostik in der ambulanten Einzelsituation. In: *H. Rennert, K. Liebner und H. D. Rösler* (Hrsg.), *Zu aktuellen Problemen der medizinischen Psychologie*. Halle/Saale 1972, 148–155
- Mucke, H.*: Validierungsuntersuchung zum Picture-Frustration Test (Kinderform) von Rosenzweig. *Wissenschaftliche Zeitschrift der Martin-Luther-Universität Halle-Wittenberg, mathematisch-naturwissenschaftliche Reihe* 1973, 22, 113–119
- Norman, R. D. und Kleinfeld, G. J.*: Rosenzweig P-F Study results with minority group juvenile delinquents. *The Journal of Genetic Psychology* 1958, 92, 61–67
- Pichot, P., Freson, V. und Danjon, S.*: Le test de frustration de Rosenzweig (forme pour enfants). Standardisation et étalonnage de la version française. *Revue de Psychologie Appliquée* 1956, 6, 111–138
- Rabin, A. und Haworth, M.* (Hrsg.): *Projective techniques with children*. New York: Grune and Stratton 1960
- Rosenzweig, S.*: The picture-association method and its application in a study of reactions to frustration. *Journal of Personality* 1945, 14, 3–23
- Rosenzweig, S.*: The Rosenzweig Picture-Frustration Study, Children's Form. In: *A. Rabin und M. Haworth* (Hrsg.), *Projective techniques with children*. New York: Grune and Stratton 1960, 149–176
- Rosenzweig, S.*: Validity of the Rosenzweig Picture-Frustration Study with felons and delinquents. *Journal of Consulting Psychology* 1963, 27, 535–536
- Rosenzweig, S.*: *Revised scoring manual for the Rosenzweig Picture-Frustration Study (form for adults)*. St. Louis, Mo.: published by the author 1965
- Rosenzweig, S. und Braun, S. H.*: Sex differences in reaction to frustration among adolescents as explored by the Rosenzweig P-F Study. *The Journal of Genetic Psychology* 1970, 116, 53–61
- Rosenzweig, S., Fleming, E. E. und Rosenzweig, L.*: The children's form of the Rosenzweig Picture-Frustration Study. *Journal of Psychology* 1948, 26, 141–191
- Rosenzweig, S. und Rosenzweig, L.*: Aggression in problem children and normals as evaluated by the Rosenzweig Picture-Frustration Study. *Journal of Abnormal and Social Psychology* 1952, 47, 683–687
- Schmidtchen, St.*: *Psychologische Tests für Kinder und Jugendliche*. Göttingen: Hogrefe 1975
- Sears, R. R., Rau, L. und Alpert, R.*: *Identification and child rearing*. Stanford, Calif.: Stanford University Press 1965
- Seidel, P. A.*: *Untersuchungen zur Validität des Rosenzweig Picture-Frustrations-Tests (R-P-F-T). Korrelation der RP-FT-Variablen mit zweiunddreißig Merkmalen als Außenkriterien*. Dissertation, Hamburg 1972
- Selg, H.*: *Diagnostik der Aggressivität*. Göttingen: Hogrefe 1968
- Simons, H.*: Zur gruppenspezifischen Diskriminationsfähigkeit der Kinderform des Rosenzweig Picture-Frustration Test. *Diagnostica* 1967, 13, 15–29 (a)
- Simons, H.*: Über die Auswirkungen unterschiedlicher Instruktionsbedingungen im Rosenzweig Picture-Frustration Test auf die Antworten von Schülern. *Archiv für die gesamte Psychologie* 1967, 119, 16–25 (b)
- Spache, G.*: Sex differences in the Rosenzweig P-F Study, children's form. *Journal of Clinical Psychology* 1951, 7, 235–238
- Spielberg, L. und Rutkin, R.*: The effects of peer vs. adult frustration on boys of middle childhood. *Journal of Psychology* 1974, 87, 231–235
- Steden, H. P.*: *Zuverlässigkeits- und Gültigkeitsuntersuchung eines Verhaltensbeobachtungsbogens bei Gruppen von aggressiven und nicht aggressiven Kindern*. Marburg: unveröffentlichte Semesterarbeit 1970
- Stoltz, R. E. und Smith, M. D.*: Some effects of socio-economic, age and sex factors on children's responses to the Rosenzweig Picture-Frustration Study. *Journal of Clinical Psychology* 1959, 15, 200–203
- Szudra, E.*: Der Rosenzweig Picture-Frustration Test und seine Eignung für die klinische Psychodiagnostik des Kindesalters. *Monatsschrift für Kinderheilkunde* 1966, 114, 507–511
- Vane, J. R.*: Implications of the performance of delinquent girls on the Rosenzweig Picture-Frustration Study. *Journal of Consulting Psychology* 1954, 18, 414

- van Roy, F.*: L'enfant infirme, son handicap, son drame, sa guérison. Paris. Neuchâtel: Delachoux et Niestlé 1961
- Weise, M.*: Der Rosenzweig Picture-Frustration Test bei Kindern mit leichtem hirnanorganischen Residualsyndrom. Praxis der Kinderpsychologie und Kinderpsychiatrie 1971, 20, 170–172
- Wilson, M. E.* und *Frumkin, R. M.*: Underlying assumptions of the Rosenzweig Picture-Frustration Study: A critical appraisal. Educational and Psychological Measurement 1968, 28, 586–594
- Zuk, G. H.*: The influence of social context on impulse and control tendencies in pre-adolescents. Genetic Psychology Monographs 1956, 54, 117–166

*Anschrift des Verfassers:*

Fachbereich Psychologie  
d. Philipps-Universität  
Gutenbergstr. 18  
3550 Marburg/Lahn