

Sind gültige Aussagen über die Intelligenzhöhe aufgrund des Rorschach-Versuchs möglich?

Hans Dieter Schmidt, Horst Hermann Figge
und Stephan Theodor Gehlen

Seit Hermann Rorschach fand, daß sich intelligente Vpn bei der Deutung von Klecksbildern in typischer Weise von anderen unterscheiden, machen Psychologen in aller Welt Aussagen über das Intelligenzniveau ihrer Vpn aufgrund des Rorschach-Versuchs. Unter den in Rorschachs „Psychodiagnostik“ aufgeführten Resultaten nimmt die Intelligenz den ersten Platz ein, und in dem weithin gebräuchlichen „Lehrbuch der Rorschach-Psychodiagnostik“ Bohms wird die quantitative Intelligenz als erste Aussage des „Psychogramms“ besprochen.

Zwar wird der Rorschach-Versuch nicht als Intelligenztest schlechthin postuliert (Bohm, S. 163: „Der Rorschach-Test ist kein ‚Intelligenztest‘“) doch soll er *den üblichen psychometrischen Verfahren ebenbürtig, ja, in gewissem Maße überlegen* sein (Bohm, S. 164). Es sei vorab ausdrücklich vermerkt, daß sich die vorliegende Arbeit nicht mit „qualitativen“ Intelligenzdiagnosen, sondern lediglich mit dem *Anspruch* des Verfahrens beschäftigt, zutreffende *quantitative* Intelligenzaussagen zu machen. Auf diesen Anspruch verzichtet allerdings kaum ein Psychologe, der den Rorschach verwendet. Es existieren gut umschriebene Symptomlisten für die Ermittlung des Intelligenzgrades; darüber hinaus wird häufig betont, daß der Gesamteindruck des Rorschach-Protokolls und des Psychogramms durch den kundigen Beurteiler für die Schätzung der Intelligenzhöhe von Bedeutung sei.

Bohm hält eine Einstufung der Intelligenz von Rorschach-Vpn in eine Zehn-Punkte-Skala für adäquat (S. 172) und führt — trotz einer deutlichen Abneigung gegenüber der Verwendung von Ziffern — vier Intelligenzquotienten (IQ) zur Verankerung seiner Kategorien an: Inferior (IQ 70—80), unintelligent bis gut durchschnittlich (IQ 80—105), Genial (IQ über 140). Daraus darf geschlossen werden, daß auch Bohm eine Übereinstimmung von Intelligenzschätzungen aus dem Rorschach-Versuch mit den Intelligenzquotienten aus psychometrischen Tests für wahrscheinlich hält. Selbstverständlich kann eine solche Annahme exakt geprüft werden.

Die gleiche Annahme äußert nun nicht nur der Diagnostiker, der bei der Erstellung eines psychologischen Befundes — getreu dem Prinzip der symptomatischen Mehrfachsicherung — Intelligenztest- und Rorschach-Resultat zusammen als Beleg für die ermittelte Intelligenzhöhe aufführt. Selbst Vernon, der schon 1935 bestritt, daß die Rorschach-Methode ein empirischer, objektiver Test sei und sich in diesem Zusammenhang besonders hämisch über deutsche Autoren äußert („weil sie eine so geringe Vorstellung

davon haben, was objektives Testen beinhaltet“ S. 199) hielt es für möglich, daß Schätzungen der Intelligenzhöhe aus dem Rorschach-Versuch Validität beanspruchen könnten.

Der Annahme, daß der Anspruch der Rorschach-Methode auf gültige Aussagen über das Intelligenzniveau statistisch überprüfbar sei, dürfte am allerwenigsten von denjenigen Autoren widersprochen werden, die diese Beziehung tatsächlich empirisch überprüft haben. Prüfungen der Gültigkeit des Rorschach-Versuchs als Intelligenzdiagnostikum sind sinnvollerweise sowohl auf der *Merkmalsebene* (Korrelation einzelner Rorschach-Merkmale mit Intelligenzkriterien; Faktorenanalysen der Merkmalsinterkorrelationen) als auch auf der *Interpretationsebene* (Korrelation der Diagnosen „kundiger“ Interpreten mit Intelligenzkriterien) vorgenommen worden. Wie z. B. M i e l d s (1966) erwähnt, gelingen Vergleiche auf der Ebene der Einzelmerkmale oder Merkmals-Verbände „weit weniger gut“ als solche auf der Interpretationsebene. Aber auch hier überwiegen keineswegs, wie es sich bei M i e l d s liest, eindeutige Verifikationen.

Vergleiche auf der Merkmalsebene: T u c k e r (1950) untersuchte bei 100 neurotischen Kriegsveteranen den Zusammenhang zwischen Bewegungsantworten (B, b+dyn) mit dem Wechsler-Bellevue. Er fand einen schwachen positiven Zusammenhang, den S o m m e r (1957) bestätigte. Ein ähnlicher Vergleich von B l a t t u. A l l i s o n (1963) mit dem Raven (Progressive Matrices) ergab keinen Zusammenhang. P a u k e r (1963) wies einen Zusammenhang zwischen der Anzahl der Inhaltskategorien und dem IQ nach und widerlegte damit frühere Befunde K l o p f e r s; der Vergleich fand an einer stark IQ-heterogenen Stichprobe statt. Entsprechende Prüfungen mit einer großen Zahl von Rorschach-Merkmalen liegen in großer Zahl vor. Sie zeigen recht unterschiedliche Resultate, so daß sich nicht einmal ein durchgängiges Teilergebnis darstellen läßt. A l t u s u. T h o m s o n (1949) fanden eine kurvenförmige Beziehung von B und IQ, jedoch keinen Zusammenhang zwischen F+⁰/₁₀₀ und T⁰/₁₀₀ mit dem IQ. H o l z b e r g u. B e l m o n t (1952) konnten von 45 Hypothesen über den Zusammenhang von Rorschach-Merkmalen mit dem Wechsler-Bellevue nur 4 verifizieren. L o t s o f (1953) extrahierte einen Faktor „verbale Intelligenz“, der sich mit Hd-, perspektivischen Deutungen, B und G⁰/₁₀₀ geladen fand. Die Faktorenladungen in den Rorschach-Variablen waren jedoch äußerst niedrig. W i l l i a m s u. L a w r e n c e (1953) fanden mittelhohe Ladungen von G, F, FHd, B, Hd und R (Antwortenzahl) im Faktor „intelligence“, und zwar bei einer Faktorenanalyse von Wechsler- und Rorschach-Variablen, gemessen an psychiatrischen Patienten. Eine der umfassendsten Untersuchungen stammt von A r m i t a g e et al. (1955). Auch hier waren die 503 Vpn psychiatrische Patienten, war der Wechsler-IQ das Intelligenzkriterium. Es zeigten sich für verschiedene Rorschach-Merkmale kleine, wenn auch auf dem 1⁰/₁₀₀-Niveau signifikante Korrelationen. Der Versuch der Autoren, aus den sechs am höchsten korrelierenden Einzelmerkmalen durch multiple Korrelation einen besseren Intelligenz-Schätzwert zu erhalten, führte zu einer den Zufallsbereich nicht übersteigenden Vorhersagegenauigkeit. Eine Untersuchung von C o n s a l v i u. C a n t e r (1957) führte auch unter Berücksichtigung

der Verbalisierungsfähigkeit zu keinem konkreten Ergebnis; vermutet wurde die Brauchbarkeit eines gewichteten Scores aus B+, G+, den formbestimmten Schattierungsdeutungen und der Anzahl der Inhaltskategorien. Zum Teil negative Befunde erhielten Ogdon u. Allee (1959) bei Schwachbegabten. Auf dieser Intelligenzstufe schienen das Formniveau, die V-Antworten, die Anzahl der Inhaltskategorien und die Anzahl reiner Fb (neg.) eine Intelligenzvoraussage zu gewährleisten. — Nach Durchsicht der Literatur läßt sich noch am ehesten vermuten, daß sich der Rorschach-Versuch auf der Merkmalsebene dann mit Einschränkungen zur Schätzung der Intelligenzhöhe verwenden läßt, wenn die untersuchte Stichprobe in bezug auf die Intelligenzhöhe relativ heterogen ist. So veröffentlichte Mields (1966) neuerdings 3 sehr signifikante (F+^{0/0}, G+, B+) und 3 signifikante (O+^{0/0}, R, B^{0/0}) Korrelationen von Einzelmerkmalen mit dem HAWIE-IQ, gewonnen an 24 Modell-Personen mit IQ von 79 bis 128.

Vergleiche auf der Interpretationsebene: Vernon (1935) schätzte die Intelligenz von zwanzig 12—15jährigen Kindern aufgrund des Rorschach-Verrechnungsprotokolls und seiner Beobachtungen bei der Aufnahme und erzielte eine Korrelation mit dem Binet-Test von 0.78; leider fehlen genaue Angaben über die Untersuchung selbst. Taulbee (1955) teilte die nach Beck signierten Protokolle von 60 Schizophrenen mit einem IQ von 72 bis 130 in drei gleiche IQ-Gruppen und ließ deren Intelligenzhöhe durch vier Psychologen schätzen; den Beurteilern stand jede gewünschte Rorschach-Information zur Verfügung. Der Autor fand große Unterschiede in der Fähigkeit der einzelnen Beurteiler und erklärt, daß „aus Rorschach-Test-Aufzeichnungen von Schizophrenen keine Intelligenzschätzungen gemacht werden sollten“ (S. 169). Armitage et al. (1955) faßten klinische Patienten derart in Gruppen zusammen, daß ihre IQ-Werte einen definierten Abstand aufwiesen. Die Beurteilung dreier praktizierender Psychologen wurde für Original- und Verrechnungsprotokolle getrennt vorgenommen und ergab Intelligenzschätzungen, die besser waren als Zufallsschätzungen; die Ergebnisse eines Beurteilers wichen allerdings nicht signifikant vom Zufall ab. Die Urteile aufgrund der Originalprotokolle waren signifikant genauer als diejenigen aufgrund der Verrechnungen. Armitage et al. folgern, „daß — was die individuelle Voraussage betrifft — die Rorschach-Schätzungen nicht die Wechsler-Bellevue-Ergebnisse ersetzen können . . . Für die individuelle Voraussage ist die Häufigkeit genauer Schätzungen nicht groß genug, sie läßt aber auf der Gruppenbasis Voraussagen mit einiger Sicherheit zu“ (S. 328 f.). Die sich aufdrängende Vermutung, daß die Qualität des Vokabulars der Vpn eine wichtige Rolle spiele, wurde von Davis (1961) untersucht. Sie wählte 70 Vpn aus, deren Stanford-Binet-Ergebnisse sich regelmäßig von 40 bis 110 verteilten. Die Beurteiler erhielten dann unterschiedliches Informationsmaterial aus der Rorschach-Untersuchung, so z. B. Tonbandaufnahmen und alphabetische Wörterlisten. „Es wurde geschlossen, daß entweder die gesamte Rorschach-Aufnahme oder der verbalisierte Wortschatz brauchbare Zusätze zu formalen Intelligenzmaßen seien, daß aber keins von dem untersuchten Material zu einer gültigen Vorhersage im individuellen Fall gebraucht werden könnte“ (S. 157). Im Übrigen bestand

die Tendenz, niedrige Intelligenzen zu überschätzen, höhere dagegen zu unterschätzen.

Auch Trier (1958) wurde durch die Arbeit von Armitage et al. angeregt; seine 4 Beurteiler, denen man lediglich die sieben jeweils ungewöhnlichsten Wörter aus 16 Rorschach-Protokollen gab, konnten Rangreihen erstellen, die signifikant mit der Rangreihe der entsprechenden Wechsler-IQ übereinstimmten. Selbstverständlich drängt sich hier die Frage auf, ob nicht Intelligenzschätzungen aufgrund des Vokabulars von Vpn mindestens Ähnliches leisten wie der Rorschach-Versuch. Die Arbeit von Bialick u. Hamlin (1954) wird gelegentlich (so von Spitznagel, 1963) als Beweis für die Validität des Rorschach-Versuchs auf der Beurteiler-Ebene angesehen. 8 Beurteiler schätzten die Intelligenz von 25 Vpn, von denen ihnen lediglich je 5 G-Antworten und die zugehörigen „inquiries“ zur Verfügung standen. Bis auf einen Interpreten stimmte die geschätzte Rangordnung mit der aufgrund einer Messung erstellten signifikant überein. Nun stellten aber die den Beurteilern gegebenen Informationen doch wohl wiederum einen Ausschnitt aus dem verbalen Verhalten dar, der nicht einmal bestimmten Rorschach-Tafeln zugeordnet war; Spitznagel dürfte hier zu Unrecht von den G-Antworten als von einem „Faktor“ sprechen. Ob der Rorschach-Versuch als solcher Intelligenz mißt, betrachten denn auch Bialick u. Hamlin als „selbstverständlich nicht beantwortet“ (S. 242).

Die letztere Frage wurde von Michel (1961) wieder angegangen. Er untersuchte quantitative Intelligenzaussagen in bezug auf einen Teilaspekt der Objektivität (die sog. „Beurteiler-Reliabilität“) und auf Validität. Bemerkenswert ist, daß aus 28 Unterprimanern 7 Rorschach-Protokolle derart ausgewählt wurden, daß ihr Amthauer-IQ von 87 bis 142 streute. 26 Beurteiler sollten die 7 Protokolle in Rangreihen nach der Intelligenz bringen. Das außerordentlich positive Resultat wurde von Michel dahingehend interpretiert, daß der Rorschach „entgegen der Meinung mancher einseitig psychometrisch orientierter Kritiker“ eine sehr beachtliche Interpretations-Gültigkeit besitze, „welche der psychometrischer Intelligenztests nahezu gleichkommt“ (S. 59). Da es sich nun bei den erwähnten kritischen Aussagen nicht nur um „Meinungen“ sondern um Untersuchungsergebnisse handelt, ist zu überlegen, wie diese Diskrepanz zwischen den negativen amerikanischen Befunden und denjenigen Michels zustande gekommen sein könnte. Es fällt auf, daß es sich bei Michel um eine überschaubare Anzahl IQ-heterogener Materialeinheiten handelte, also um einen ausgesprochen praxisfernen Fall. In der Praxis des Rorschach-Versuchs geht es vielmehr darum, einer einzigen Materialeinheit ihren Platz in einer theoretisch unbegrenzten Zahl entsprechender Einheiten zuzuweisen. Auch Spitznagel (1963) verwendete lediglich 5 Protokolle, die jeweils um ca. 10 IQ-Punkte differierten; er schreibt dazu, „daß die Voraussetzungen durch die Heterogenität des ausgewählten Materials günstig waren“ (S. 298). Es ist auch nicht zu entscheiden, inwieweit die 7 Protokolle Michels, die dann auch Mields (1966) weiterverwendete, Rorschach-entsprechend eindeutiges, prägnantes Material enthielten. Betrachtet man z. B. die von

Lienert u. Mathaei (1958) veröffentlichten Rangreihen, so fällt auf, daß von 29 Beurteilern nur jeweils ein einziger der Vp C nicht den ersten, und Vp D nicht den letzten Rangplatz gegeben hat: C hat die weitaus meisten B (die von Bohm als „Salz in der Suppe“ bezeichnet werden), die größte Antwortenzahl, die weitaus meisten O+ und die meisten Inhaltskategorien. D dagegen hat als einziger oligophrene Details (sogar 3), als einziger keine Bewegungsantwort, die wenigsten Antworten insgesamt und die meisten G-. Es ist demnach eigentlich weniger erstaunlich, daß 27 Beurteiler sowohl den ersten als auch den letzten Rangplatz gleich verteilen, als vielmehr, daß 2 Beurteiler hier nicht konform gehen. Berücksichtigt man einmal nur die Einschätzung der 3 intelligenzmäßig mittleren Protokolle, so schmilzt die von Lienert u. Mathaei gefundene Konkordanz der 29 Beurteiler von 0.90 auf 0.44 zusammen. Validitätsuntersuchungen, von denen derart viel abhängt, sollten gerade auch an zufällig, und nicht nur an systematisch ausgewähltem Material vorgenommen werden.

Michel hatte also keineswegs eine „hinsichtlich der Intelligenz relativ homogene Gruppe“ (1963, S. 136) untersucht, denn zwischen SW 87 und SW 142 (Amthauer) liegen 90.3% aller Fälle. Ein Teil der hier gegen die Studie von Michel erhobenen Bedenken wurde nun von Mields (1964, 1966) widerlegt. 8 Beurteiler sollten 24 Rorschach-Psychogramme von Vpn mit IQ zwischen 79 und 128 in eine Rangreihe bringen. Es ergaben sich sämtlich sehr signifikante Korrelationen mit dem HAWIE-IQ (0.75 bis 0.80). Wenn man einmal den Gehalt der von Mields (1966, S. 61 f.) referierten Einwände Hörmanns und Spitznagels, bei Validierungsversuchen auf der Interpretationsebene handele es sich stets um die Validierung einer Kombination Diagnostiker+Test, als beim Rorschach-Versuch notwendiges Übel hinnimmt, so bleibt doch zu bemerken, daß die Schwierigkeit einer Relativierung der Rorschach-Resultate intelligenzheterogener Vpn wohl kaum dem Schwierigkeitsgrad vergleichbar ist, den die Einstufung eines einzelnen Versuchsergebnisses in ein theoretisches Gesamt (und das ist die Situation des Praktikers) besitzt. Ein weiterer Einwand könnte sein, daß die Beurteiler bei Mields über das Vergleichskriterium (HAWIE) informiert waren. Gerade bei erfahrenen Psychologen ist es nämlich denkbar, aus der intimen Kenntnis beider Verfahren (HAWIE und Rorschach) eine bessere Schätzung vom Rorschach auf den HAWIE zu leisten, als wenn eine unabhängige Intelligenzschätzung verlangt würde. Immerhin weisen die Mieldschen Koeffizienten Höhen auf, die auch für Rangkorrelationen beachtlich zu nennen sind.

Planung der Untersuchung

Eine Überprüfung des Anspruchs auf gültige Schätzung der Intelligenzhöhe mit Hilfe des Rorschach-Versuchs muß nach dem Ausgeführten sowohl auf der Merkmals- als auch auf der Interpretationsebene geschehen. Wir vermuten dabei, daß sich bei beiden Arten des Vorgehens die der Nullhypothese entsprechende Aussage zurückweisen läßt. Zumindest wollen wir erwarten, daß die Zusammenhänge zwischen dem Intelligenzkriterium

und relevanten Rorschach-Merkmalen einerseits, und zwischen dem Kriterium und der Intelligenzschätzung durch Beurteiler andererseits *mehr als zufällig* sind.

Es werden 12 *Hauptthesen* gebildet, die sich auf das *Rorschach-Syndrom* für die quantitative Intelligenzschätzung direkt beziehen und die die erwarteten Korrelationen der einzelnen Symptome mit dem HAWIE-IQ bzw. den beiden Gruppen-IQ formulieren:

Mit dem Gesamt-IQ werden positiv korrelieren: F +⁰/₀, Anzahl der G (ggfs. der G+), Anzahl der B, O⁰/₀, Anzahl der Inhaltskategorien. Mit dem Verbal-IQ sollen positiv korrelieren: F +⁰/₀, G (G+), B, O⁰/₀, Anzahl der Inhaltskategorien. Mit dem Handlungs-IQ sollen positiv korrelieren: F +⁰/₀, V⁰/₀.

Außerdem werden insgesamt 68 weitere Hypothesen aufgestellt, die einen Bezug einzelner Rorschach-Variablen zu bestimmten HAWIE-Untertests aufgrund eines vergleichbaren Symptomwerts erwarten lassen. Die Basis für diese Hypothesen ergibt sich einmal aus der Beschreibung der Variablen in der jeweiligen Literatur, zum anderen aus einer bereits zum Gegenstand früherer Untersuchungen gemachten Vermutung, daß bestimmten Klassen von Rorschach-Deutungen durch die Signierungsbedingungen (in der Regel müssen gewisse *verbale* Qualifikationen erfüllt sein) eine engere Beziehung zum Wortschatz und damit auch zum Intelligenzniveau zukommt. Zu diesen Variablen kann man z. B. die B, b, dyn, die Schattierungsdeutungen und u. U. auch bestimmte besondere Phänomene wie Impressionen und Deskriptionen zählen.

Insgesamt werden so 38 Rorschach-Scores, deren Überprüfung uns sinnvoll erscheint, der Untersuchung auf der Merkmalsebene zugrundegelegt; dabei ist bezüglich der Variablen, die nicht zum eigentlichen „Intelligenz-Syndrom“ gehören, einschränkend anzumerken, daß aufgrund des nur ähnlichen, nicht aber identischen Symptomwerts numerisch hohe Korrelationskoeffizienten nicht wahrscheinlich sind und daher auch nicht erwartet werden.

Die Prüfung auf der *Merkmalsebene* ist sinnvoll, weil weder die Methode der direkten „klinischen“ Validierung (vgl. H ö r m a n n, 1964, S. 100) noch ein Blinddiagnoseverfahren zu einer hohen Vorhersagegenauigkeit führen können, wenn die wichtigsten Aspekte (d. h. die Einzelmerkmale) auf die sich ein solches Vorgehen stützen muß, nicht oder ungenügend valid sind. Außerdem haben R o r s c h a c h, B o h m und K l o p f e r (S. 267 ff.) eindeutige Beziehungen von Einzelmerkmalen zum Intelligenzniveau postuliert.

Eine Prüfung auf der *Interpretationsebene* versteht sich nach (theoretischen) ganzheitspsychologischen Erwägungen von selbst; außerdem soll ja unsere Untersuchung von vornherein ein großes Maß an *Praxisnähe* aufweisen.

Da eine Reihe von Prüfungen an Schizophrenen (T a u l b e e), Neurotikern (A r m i t a g e et al.) und unterdurchschnittlich Normalen (D a v i s) vor-

genommen wurde, soll unsere *Stichprobe* aus *überdurchschnittlich intelligenten Normalen* bestehen. Die Stichprobe soll auf jeden Fall erheblich größer sein als die bei früheren (insbesondere deutschen) Arbeiten mit Normalen verwendete. Sie soll in mehrfacher Hinsicht möglichst *homogen* sein, damit sich die erwarteten positiven Resultate nicht mit Hinblick auf vermischte Intelligenzfaktoren (vgl. *Wewetzer*, 1961), unterschiedliches Alter (*Eisdorfer*, 1963) und soziale Variable leicht relativieren lassen. Es soll insbesondere darauf geachtet werden, daß es sich nicht um Vpn handelt, deren IQ alle einen definierten (*Lienert* u. *Mathaei*, *Michel*) oder ähnlichen Abstand (*Mields*) voneinander haben oder sonstwie ausgelesen sind; innerhalb der gewählten Population soll die Auswahl der Stichprobenmitglieder in bezug auf die Intelligenzhöhe zufällig sein.

Die *Signierung* soll — ebenso wie in der Praxis — vom jeweiligen V1 vorgenommen, anschließend jedoch einer Kontrolle unterworfen werden.

Bei der Beurteilung des Intelligenzniveaus durch Experten soll folgendes beachtet werden:

1. Den Beurteilern soll sowohl das vollständige, signierte *Originalprotokoll* als auch das zugehörige *Verrechnungsprotokoll* zur Verfügung stehen — die wahrscheinliche Mitverwertung von Wortschatz u. ä. wird ebenso wie in der Praxis gewährleistet.

2. Auf keinen Fall sollen die Beurteiler die Protokolle nach der Intelligenz in Rangreihen legen. Dies ist vermutlich zu „leicht“ und entspricht nicht der Praxis. Vielmehr soll die Intelligenzhöhe jeder *einzelnen* Vp geschätzt werden.

3. Die Beurteiler sollen auf keinen Fall den IQ direkt schätzen, sondern — entsprechend *Bohm* — eine verbal verankerte Skala benutzen. In der Rorschach-Praxis werden keine IQ-Aussagen gemacht; die Angabe des IQ würde für manchen Rorschach-Experten ungewohnt, für manchen womöglich ärgerlich sein. Zudem gewährleistet die Vorstellung der IQ-Skala keine vom Testkriterium unabhängige Schätzung, sondern begünstigt die Überlegung, wie wohl der betreffende Proband auf die Items des betreffenden Intelligenztests reagieren würde.

Um alle geforderten Bedingungen sauber zu gewährleisten, soll nicht auf vorhandene Unterlagen eines Psychologischen Instituts oder einer anderen Stelle zurückgegriffen werden; es sollen auch nicht bereits publizierte Modellfälle weiter verarbeitet werden (wie bei *Mields*). Vielmehr müssen sämtliche Daten eigens erhoben werden.

Methoden und Ergebnisse

Versuchspersonen. Die Versuchsgruppe bestand aus 80 Vpn im Alter von 18—27 Jahren (Altersmittelwert 21.4). 40 Vpn (33 männl., 7 weibl.) waren kaufmännische Lehrlinge, die von einem Luftverkehrsunternehmen zu Luftverkehrskaufleuten ausgebildet wurden; 31 von ihnen waren Abiturienten oder Absolventen einer Höheren Handelsschule. 40 Vpn (16 männl., 24 weibl.) waren Studenten der Anfangssemester verschiedener Fakultäten der

Universität Bonn. Sämtlichen Vpn waren beide Verfahren zuvor unbekannt gewesen. Die HAWIE-IQ streuten insgesamt zwischen 100 und 129 ($M=116.6$; $SD=6.1$). Ein Vergleich mit dem t-Test ergab keinerlei Unterschiede zwischen der Studenten- und der Lehrlingsgruppe hinsichtlich Alter und Intelligenzniveau; auch zwischen beiden Geschlechtergruppen bestanden keine Unterschiede.

Versuchsablauf. An allen 80 Vpn wurde im Einzelversuch zunächst der Rorschach und dann der HAWIE angewendet. Die Untersuchungen fanden im Büroraum des Betriebspsychologen einerseits, und in einem Testraum des Psychologischen Instituts der Universität Bonn andererseits statt. Die Untersuchungsbedingungen wurden für alle Einzelversuche möglichst konstant gehalten: Einleitendes Kontaktgespräch, Verwendung neuen Testmaterials, gleiche Abfolge, gleiche Sitzordnung, Zeitnahme durch Armbanduhr, lang-schriftliche Aufzeichnung durch den VI, gleiche Instruktionen und Zwischenanweisungen beim unstandardisierten Rorschach-Versuch (insbesondere wurde der Eindruck vermieden, es handele sich um einen Schnelligkeitstest oder ähnliches), identischer VI für je 40 Vpn.

Signierung und Verrechnung. Alle Protokolle wurden — mit Ausnahme der zusätzlichen Erfassung von Tier- und Objektbewegungen und der verarbeiteten Objektbewegungen — unter der Rubrik „besondere Phänomene“ streng nach B o h m signiert. Beide VI verglichen dann die getrennt erstellten Signierungen und räumten Differenzen aus. Eine weitere Kontrolle bestand darin, daß die Signierungen getrennt für jede einzelne Rorschach-Tafel geprüft und miteinander verglichen wurden. Die Verrechnung erfolgte auf einem Verrechnungsbogen, der alle von B o h m geforderten und einige weitere Daten enthielt. Auch die Verrechnung wurde von beiden VI wechselseitig kontrolliert. Ein weiterer Rorschach-Experte billigte die Art der Signierung und Verrechnung.

Prüfung des Zusammenhangs auf der Merkmalsebene. Die für die korrelationsstatistische Verarbeitung erforderliche lineare Regression ist bei den meisten Rorschach-Variablen problematisch (M u r s t e i n, 1960; C r o n b a c h, 1949). Nur folgende Merkmale waren annähernd normal verteilt: $R+^0/0$, $G+^0/0$, $G^0/0$, $F+^0/0$, V , $V^0/0$, $E^0/0$, Summe der Inhaltskategorien; P o i s s o n -ähnlich verteilten sich: R , R/Min , $R+$, G , Summe der 'Zw., $O^0/0$, B , $B+$, $B+b+dyn+dyn-verarb.$, $b+dyn+dyn-verarb.$, $B+BFb+BHD$, F , $F+$, T , $T^0/0$, fb . Es schien vertretbar, diese Variablen zu normalisieren und eine T-Transformation vorzunehmen. Lediglich D , Dd , O und $O+$ verteilten sich weitgehend nicht-normal. Aus den Produkt-Moment-Korrelationen sämtlicher Merkmale wurde (trotz der Bedenken von M u r s t e i n gegen die Verwendung abhängiger und relativierter Werte) eine Interkorrelationsmatrix erstellt, die auch die Interkorrelationen sämtlicher HAWIE-IQ- und Untertest-Standardwerte enthielt*).

*) Sämtliche Korrelationsmatrizen wurden von der Großrechenanlage (IBM 7090) des Instituts für Instrumentelle Mathematik (IIM-IAM) der Universität Bonn erstellt.

Da die Produkt-Moment-Matrix eine große Zahl normalisierter und eine kleine Zahl nicht-normalverteilter Merkmale bzw. nicht interpretierbarer Korrelationen enthält, wurde eine weitere Korrelationsmatrix erstellt, deren Elemente entsprechende *Spearman*-sche Rangkorrelationskoeffizienten waren. Sämtliche Korrelationen wurden über *t* (*Student*) auf ihre Signifikanz (5%, 1%) geprüft; bei $df=78$ sind r_{xy} und ρ_{xy} ab 0.220 auf dem 5%-Niveau, und ab 0.287 auf dem 1%-Niveau gesichert.

Die Rangreihenkoeffizienten weichen, wie zu erwarten, weder allgemein noch bezüglich irgendeiner der Hypothesen erheblich von den *Pearson*-schen Koeffizienten ab. Zum besseren Vergleich mit *Mields* und anderen Autoren soll daher nur die ρ -Matrix angegeben werden (Tab. 1). Getrennt für die beiden Teilstichproben (kaufm. Lehrlinge und Studenten) wurden ebenfalls r_{xy} - und ρ_{xy} -Matrizen erstellt. Die entsprechenden Differenzen zwischen beiden Teilstichproben erwiesen sich dabei als so gering, daß sich eine statistische Prüfung oder eine gesonderte Darstellung erübrigte.

Von den insgesamt 68 *Nebenypothesen*, die Zusammenhänge zwischen Rorschach-Merkmalen und dem IQ bzw. einzelnen HAWIE-Untertests vorhersagten, wurden nur 5 verifiziert. Bei einer weiteren bleibt die Interpretation wegen der nicht normalen Verteilung der Rorschachvariablen problematisch. Zu einer einzigen Hypothese ergab sich das statistisch gesicherte Gegenteil. Insgesamt 61 Hypothesen konnten nicht bestätigt werden.

Von den 12 *Hauptypothesen*, die sich auf das von *Rorschach* aufgestellte Intelligenzsyndrom beziehen, konnte nur eine verifiziert werden: Zwischen dem $F+$ und dem Gesamt-IQ besteht ein numerisch geringer, aber statistisch gesicherter Zusammenhang. Völlig entgegen der Erwartung korrelierten die Ganzantworten (G) mit dem IQ signifikant negativ.

Die Korrelationskoeffizienten streuen um einen mittleren Wert von -0.005 ($SD=0.14$). Sie sind weitaus zu niedrig, als daß sich eine statistische Korrelationsanalyse (Faktorenanalyse) lohnen würde.

Prüfung des Zusammenhangs auf der Interpretationsebene. Es erwies sich als unmöglich, Rorschach-Experten bzw. Beurteiler zu finden, die sich bereit fanden, alle 80 Materialarbeiten durchzuarbeiten. Daher sollte jeder Beurteiler jeweils 20 Protokolle zur Intelligenzbeurteilung erhalten. Als *Beurteiler* stellten sich zur Verfügung: Fünf Mitarbeiter des Psychologischen Instituts der Universität Bonn, ein Mitarbeiter der Erziehungsberatungsstelle des Instituts und zwei Hauptexamenskandidaten, von denen einer bereits seit Jahren mit dem Rorschach beruflich arbeitet, und der andere in den Vereinigten Staaten ein Jahr lang einem bekannten Rorschach-Experten assistiert hatte. Da alle Beurteiler „kundig“ sein sollten, wurde (ebenso wie bei *Mields*) mit nicht mehr als 8 Experten gearbeitet. Keiner der Beurteiler hielt sich durch die gestellte Aufgabe für überfordert. In einigen Fällen wurde zuvor geklärt, daß nicht die Intelligenzkapazität schlechthin, sondern die zur Zeit der Testaufnahme aktualisierte Intelligenz einzuschätzen sei.

Tabelle I

	IQ _{tot}	IQ _V	IQ _H	AW	AV	ZN	RD	GF	ZS	BO	BE	MT	FL
R	-17	-07	-17	07	06	-20	01	-02	10	07	-26	-06	-31
R/Min.	-15	-02	-22	17	07	-11	-03	17	-01	-09	-15	-10	-16
Min.	-01	-03	04	-06	02	-12	10	-18	04	17	-12	04	-13
R+	-05	03	-06	09	10	-10	10	-05	14	09	-18	-02	-23
R+ ^{0/0}	36	31	30	09	18	28	25	-10	08	06	28	12	26
G	-29	-23	-26	05	-04	-19	-14	-08	-07	-03	-23	-04	-28
G+	-16	-13	-15	07	01	-08	-05	-14	-04	-03	-13	04	-21
G+ ^{0/0}	25	22	17	06	15	29	16	-14	04	-01	19	11	20
G ^{0/0}	-10	-18	-04	-03	-09	06	-20	-06	-23	-07	02	05	11
D	-03	06	-06	05	10	-17	12	-01	20	09	-16	-05	-25
Dd	-18	00	-23	21	03	-09	-01	14	-01	-12	-19	-14	-10
Σ Zw	-08	-03	-09	-04	-16	-02	06	-06	13	10	-16	-01	-28
DZw	03	07	-03	-03	-12	07	11	04	20	12	-23	05	-17
B	-03	01	-05	07	08	-06	04	-03	07	03	-07	03	-30
B+	-01	01	-02	07	08	-07	04	00	06	03	-06	04	-26
B ^{0/0}	10	06	10	04	19	08	-01	-02	06	04	11	13	-17
B, b, dyn	15	19	06	20	19	06	14	03	12	11	00	13	-21
b, dyn	26	27	13	26	13	14	20	07	15	13	04	16	-11
B, BFb, BHd	00	03	-02	09	14	-05	04	-01	09	07	-06	07	-30
F+	-09	01	-10	06	02	-14	15	-06	18	10	-27	-09	-20
F+ ^{0/0}	24	27	17	07	05	30	27	-11	02	03	20	00	24
T ^{0/0}	21	21	14	09	08	15	05	11	00	05	-03	13	21
V	05	08	03	15	16	-05	10	01	25	17	-13	01	-22
V ^{0/0}	19	10	19	-04	08	18	04	-06	05	03	15	14	15
O	-17	-03	-22	11	-05	-12	-03	16	11	-04	-20	-08	-39
O+	-04	12	-14	15	12	04	00	13	16	-06	-10	-06	-32
O ^{0/0}	-09	02	-14	10	-08	00	-06	20	10	-08	-04	-03	-26
E ^{0/0}	-20	-17	-16	-08	-05	00	-20	04	-20	-09	05	-10	-14
F(Fb), Σ Hd	-19	03	-28	24	-01	-16	11	10	01	-18	-23	-16	-22
F	-11	-04	-11	04	04	-20	07	-07	17	10	-28	-03	-27
T	-05	05	-10	12	09	-13	05	05	14	11	-28	-02	-23
Σ Inh. kat.	-12	-08	-09	02	-09	-08	05	-02	12	07	-14	-05	-22
Anat ^{0/0}	-16	-10	-15	03	15	-20	-11	12	-02	-10	-09	-13	-03
fb	-25	-23	-20	-07	-08	-16	-18	-02	-02	01	-17	-09	-28
FFb/Σ fb	36	23	32	01	16	05	24	-14	24	17	16	26	03
Σ Schocks	-25	-16	-19	-17	-16	-14	-02	-01	-22	-06	-08	-15	06
Σ bes. Phän.	-08	-04	-04	12	-14	-05	-07	13	01	10	-06	-03	-06
Σ Komb.	04	-04	04	-01	-02	02	-19	-09	07	11	02	14	-20

(E^{0/0} ist das Verhältnis der nicht formbestimmten zur Gesamtzahl der Deutungen — Formel nach F u c h s)

B o h m folgend, wurde eine verbal verankerte numerische Schätzskala, in diesem Fall mit 16 Punkten, vorgegeben. Die Beschriftung der Skala be-

rücksichtigte den Sprachgebrauch W e c h s l e r s wie auch B o h m s : Extrem hohe Intelligenz (1, 2), Sehr hohe Intelligenz (3, 4), Hohe Intelligenz (5, 6), Gut durchschnittliche Intelligenz (7, 8), Knapp durchschnittliche Intelligenz (9, 10), Niedrige Intelligenz (11, 12), Sehr niedrige Intelligenz (13, 14), Extrem niedrige Intelligenz (15, 16). Die Instruktion der Beurteiler enthielt die Mitteilung, daß die Mitte zwischen „Gut“- und „Knapp durchschnittlich“ als Skalenmitte anzusehen sei, die einem IQ von 100 entspräche. Eine IQ-Beurteilung wurde aus den oben erwähnten Gründen nicht verlangt, in Fällen, in denen IQ angegeben wurden, aber auch nicht verweigert. Von den Beurteilern wurde die vorgelegte Skala als verständlich und Rorschach-adäquat angesehen.

Nach Zufall wurden 4 Gruppen von je 20 Rorschach-Protokollen ausgewählt. Auf diese Weise wurde jede Vp von jeweils zwei Experten beurteilt und es wurde vermieden, daß allen Beurteilern dieselben „Modellpersonen“ vorgelegt wurden. Durch Prüfung mit F-Test und t-Test wurde sichergestellt, daß alle 4 Teilstichproben die gesamte IQ-Stichprobe in völlig gleicher Weise repräsentierten. Einige Beurteiler erleichterten sich die Intelligenzschätzung, indem sie gewisse Protokolle gar nicht beurteilten. Dies wurde toleriert, obgleich darin eine erhebliche Vereinfachung der Beurteiler-Aufgabe zu sehen ist.

Die Schätzwerte jedes einzelnen Beurteilers auf der 16stufigen Skala wurden mit der Rangreihenkorrelation nach S p e a r m a n auf ihren Zusammenhang mit den IQ-Werten geprüft. Es zeigte sich, daß die Hälfte der Korrelationen negativ, die Hälfte positiv ist (vgl. Tab. 2). Die Intelligenzbeurteilungen keines der 8 Experten korrelierten mehr als zufällig mit dem Ergebnis des HAWIE.

T a b e l l e 2

Korrelationen zwischen Intelligenzschätzungen von Rorschach-Experten und dem HAWIE-IQ (Schätzungen auf einer 16stufigen Skala)

Beurteiler	rho	N	Beurteiler	rho	N
A 1	.31	16	A 2	.26	20
B 1	-.10	20	B 2	.44	19
C 1	-.18	20	C 2	.42	18
D 1	-.14	20	D 2	-.14	20

Da sich gezeigt hatte, daß ein Teil der Beurteiler die vorgegebene Skala praktisch auf 8 Klassen reduziert hatte, wurden weitere Korrelationen berechnet, bei denen zunächst nur die Rorschach-Beurteilungsskala (Tab. 3) und dann auch die IQ-Skala (Tab. 4) auf je 8 Klassen reduziert wurden. Die Resultate zeigen, daß sich auch bei dieser Berechnungsart kein „besseres“ Bild ergibt. Wir versuchten noch eine „synthetische Gruppe“ aus den Beurteilern A 1 und A 2 zu bilden, da wenigstens diesen beiden schwach positive Koeffizienten zugeordnet werden konnten; auch dieser „gemeinsame“ Koeffizient erreicht mit 0.33 aber nicht die Signifikanzgrenze. Die

Untersuchung ist im übrigen nicht so geplant, daß durch größere synthetische Gruppen höhere Korrelationskoeffizienten erhältlich wären.

Tabelle 3

Korrelationen zwischen Intelligenzschätzungen von Rorschach-Experten und dem HAWIE-IQ (Schätzungen auf einer 8-stufigen Skala).

Beurteiler	rho	N	Beurteiler	rho	N
A 1	.37	16	A 2	.13	20
B 1	-.07	20	B 2	.44	19
C 1	-.18	20	C 2	.42	18
D 1	-.23	20	D 2	-.25	20

Tabelle 4

Korrelationen zwischen Intelligenzschätzungen von Rorschach-Experten und dem HAWIE-IQ nach Reduzierung beider Skalen auf 8 Klassen.

Beurteiler	rho	N	Beurteiler	rho	N
A 1	.45	16	A 2	.16	20
B 1	-.08	20	B 2	.37	19
C 1	-.21	20	C 2	.33	18
D 1	-.21	20	D 2	-.16	20

Obwohl für die Frage der Validität ohne besonderen Belang, wurden auch die Übereinstimmungen der (jeweils zwei) Experten bestimmt. Wie zu erwarten, ist diese sogenannte „Beurteiler-Reliabilität“ wesentlich höher (Tab. 5). Erstaunlich ist jedoch, daß bei zwei der vier Gruppen eine mangelhafte Objektivität zu verzeichnen ist. Insgesamt ergibt sich, daß die Urteile, in denen Beurteiler miteinander übereinstimmen, zum großen Teil nicht mit den (objektiven) Testwerten übereinstimmen. Betrachtet man den Zusammenhang zwischen Rorschach-Urteil und HAWIE-IQ nur in jenen Fällen, in denen die beiden Beurteiler übereinstimmen, so wird auch hier im günstigsten Fall nicht einmal die 5%-Signifikanzgrenze erreicht.

Tabelle 5

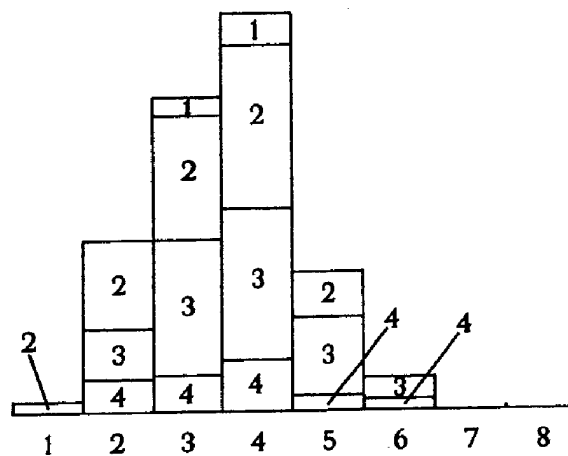
Übereinstimmung je zweier Rorschach-Experten bzgl. Intelligenzhöhe

Beurteiler	rho	N
A 1 und A 2	.68 **)	16
B 1 und B 2	.28	19
C 1 und C 2	.29	18
D 1 und D 2	.58 **)	20

**) $p < 1\%$.

Die Leistungen der Rorschach-Beurteiler können aufgrund graphischer Darstellungen noch ein wenig näher charakterisiert werden. So wurde die Herkunft der Stichprobe aus der in bezug auf die Intelligenz oberen Hälfte der Gesamtbevölkerung überwiegend richtig erkannt; die Abweichungen einzelner Schätzungen von den zugehörigen Meßwerten nach oben und unten sind jedoch beträchtlich, und es läßt sich keine den IQ-Maßen entsprechende Ordnung erkennen. Es ergibt sich ein allgemeiner Trend bei den Beurteilern, Schätzungen zum Skalenmittelpunkt hin zu verlegen; in 93 Fällen liegen Schätzungen näher am Mittelpunkt, und in nur 18 Fällen weiter von ihm entfernt als die entsprechenden Testwerte. Dieser Fehler zur Mitte ist nicht identisch mit dem „Fehler der zentralen Tendenz“, da ein Kennzeichen des letzteren die stark reduzierte Streuung ist. In unserem Falle neigen aber die Rorschach-Beurteiler dazu, die Probanden intelligenzheterogener einzuschätzen als sie in Wirklichkeit sind d. h. die Streuung um die zum Mittelpunkt hin verschobenen Mittelwerte der Urteilsverteilungen ist erhöht; die Häufigkeitsverteilungen der Intelligenzschätzungen sind flacher und breiter als die der Messungen. Der beschriebene Trend zeigt sich bei jedem einzelnen Beurteiler. Durchgängig werden hohe Intelligenzen unterschätzt und niedrige Intelligenzen überschätzt.

Die folgende Abbildung zeigt eine Gesamtdarstellung aller 153 Intelligenzschätzungen aufgrund des Rorschach-Versuchs. Die Skala von 1 bis 8 ist die reduzierte Schätzskala, „1“ bezeichnet dabei hohe, „8“ niedrige Intelligenz aufgrund der Expertenschätzung. Die Zahlen in den Blöcken geben an, aus welcher der entsprechenden IQ-Klassen die Vpn stammen; die Blockhöhe entspricht der Häufigkeit von Fällen pro Klasse.



Die beiden untersuchten Teilstichproben (kaufm. Lehrlinge und Studenten) stammen zwar in bezug auf die Intelligenzhöhe aus der gleichen Population, doch zeigte sich bei einem routinemäßig vorgenommenen Vergleich ein einziger Unterschied: Die Studenten gaben signifikant mehr Antworten bzw. Deutungen. Betrachtet man nun die Leistung der Rorschach-Experten getrennt nach beiden Teilstichproben, so ergibt sich, daß die Stu-

dentengruppe durchschnittlich höher und auch „richtiger“ eingestuft wurde als die der (gleich intelligenten) Lehrlinge. Auch zeigte sich, daß Rorschach-Protokolle mit weniger als 20 Antworten im Durchschnitt weniger zutreffend eingestuft werden als Protokolle mit mehr als 20 Antworten.

Wie schon erwähnt, hatten sich einige Beurteiler die Aufgabe durch Weglassen gewisser Protokolle erleichtert. Verschiedene Protokolle scheinen also verschieden leicht zu beurteilen zu sein. Stellt man nun einmal jene Intelligenzurteile zusammen, die mit gewissen Einschränkungen abgegeben worden waren, so zeigt sich, daß sie nicht signifikant „schlechter“ ausfielen als die uneingeschränkt abgegebenen. Daraus darf gefolgert werden, daß das subjektive Sicherheitsgefühl bei der Intelligenzschätzung aus dem Rorschach-Versuch in keinem merklichen Zusammenhang mit der Güte des Urteils selbst steht.

D i s k u s s i o n

Alle Ergebnisse widerlegen die Annahme, daß gültige Aussagen über die Intelligenzhöhe aufgrund des Rorschach-Versuchs möglich seien; für die durch unsere Stichprobe repräsentierte Population sind entsprechende Hypothesen sowohl auf der Merkmals- als auch auf der Beurteiler-Ebene zurückgewiesen worden, wenn man den Hamburg-Wechsler-Intelligenztest für Erwachsene als verbindliches Kriterium akzeptiert.

Fast alle Resultate widersprechen den von M i e l d s an einer kleineren Stichprobe von Modellpersonen erhobenen Befunden (Merkmalskorrelationen bis zu 0.71 und Beurteiler-Korrelationen bis zu 0.80 mit dem HAWIE).

Eine Möglichkeit, diese Diskrepanz für die Merkmalsebene zu interpretieren, sehen wir nicht; auf diesem Gebiet stimmen unsere Resultate im wesentlichen mit amerikanischen Arbeiten und auch den Erwartungen der meisten Autoren überein.

Selbst für den Skeptiker unerwartet niedrig sind aber die Zusammenhänge zwischen den Intelligenzschätzungen der Rorschach-Experten und dem Testkriterium ausgefallen; hier herrscht doch ein weit größerer Optimismus vor. M i e l d s' Modellpersonen (zuletzt 24 weibliche Probanden) wurden auf eine Art beurteilt, die wir bereits als wenig praxisnah charakterisiert hatten: Sie sollten nach der Intelligenzhöhe in eine *Rangordnung* gebracht werden. Dies dürfte leichter sein als eine direkte Schätzung der „absoluten“ Intelligenzhöhe aufgrund einer Schätzskala oder gar des geschätzten IQ selbst. Um diese leichteren Bedingungen nachzuahmen (hinzu kam, daß die IQ der M i e l d s schen Modell-Personen bis auf wenige Ausnahmen gleichmäßig den Bereich von $IQ = 79$ bis $IQ = 128$ belegten) wurde ein neuer Beurteilungsversuch geplant und ausgeführt.

Aus den 80 Protokollen wählten wir solche aus, die den IQ des Versuchs von M i e l d s entsprachen; dies gelang uns für 20 Fälle, da der niedrigste IQ unserer Stichprobe 100 betrug; einen weiteren Fall mit $IQ = 90$ fügten wir aus einer laufenden Untersuchung hinzu; in Fällen, in denen mehrere Protokolle pro IQ zur Auswahl standen, wurde eine Zufallsauswahl getroffen, doch wählten wir im Zweifelsfall Protokolle mit mehr als 20 Deu-

tungen. Alle 21 Rorschach-Fälle legten wir 7 als qualifiziert angesehenen Beurteilern vor; es handelte sich dabei um Absolventen eines 3-semesterigen Rorschach-Kurses. Diese Beurteiler verfügten zwar nicht über die gleiche Rorschach-Erfahrung wie diejenigen unserer Hauptuntersuchung, doch nahmen sie die Beurteilungen unter den von M i e l d s (1966, S. 58) angegebenen Bedingungen vor; dazu gehörte auch die Information, daß der HAWIE-IQ als Vergleichskriterium fungieren solle. Wir vermuteten höhere Zusammenhänge zwischen Intelligenzschätzung und tatsächlichem IQ.

Zwischen den 7 Beurteilern ergab sich zunächst eine auf dem 1%-Niveau gesicherte Übereinstimmung (Konkordanz) bezüglich der Einstufung der 21 Modell-Personen ($W = .47$; bei M i e l d s : $W = .86$). Die Korrelationen der Beurteiler mit dem Testkriterium erwiesen sich als gegenüber dem Hauptversuch kaum erhöht; lediglich ein einziger rho-Koeffizient ist auf dem 5%-Niveau signifikant:

0.454 / 0.368 / 0.262 / 0.094 / 0.042 / 0.026 / -0.203 (vgl. Tab. 2, 3, 4)
Bei Bildung einer „synthetischen Gruppe“ beträgt rho 0.212 (nicht sign.).

Nach diesen Resultaten scheint es nicht mehr viel Hoffnung zu geben, die Diskrepanz zu den M i e l d s'schen Befunden durch unterschiedliche Methodik erklären zu können; es drängt sich die Vermutung auf, daß entweder die von M i e l d s verwendeten Protokolle oder die herangezogenen Beurteiler von außergewöhnlicher Art gewesen sein könnten.

Selbstverständlich könnte man sowohl M i e l d s (der sich seit einiger Zeit um eine Interpretations-Validität des Rorschach bemüht) als auch uns (die wir bestimmten Ansprüchen des Rorschach-Versuchs gegenüber skeptisch eingestellt sind) Voreingenommenheit vorwerfen. Daß das „Wissen“ um einen Zusammenhang zwischen Wechsler-IQ und Rorschach-Intelligenzschätzung eine gewichtige Rolle spielen kann, mag folgende Untersuchung zeigen:

Wir stellten aus den Akten des Psychologischen Instituts Bonn sämtliche Examensgutachten (Fall-Klausuren) eines bestimmten Zeitabschnitts zusammen, in denen die Versuchsleiter sowohl den Wechsler-IQ bestimmt als auch den Rorschach angewendet hatten. Im Psychologischen Befund des Examensgutachtens wird die jeweilige Aussage über die Intelligenzhöhe in der überwiegenden Mehrzahl der Fälle sowohl durch den standardisierten Test als auch durch das Rorschach-Ergebnis belegt. Die 58 herangezogenen Gutachten betrafen Probanden mit überdurchschnittlicher Intelligenz (IQ-Mittelwert: 113.2) bei normaler Streuung ($SD = 13.8$). Die Rangreihenkorrelationen zwischen einigen relevanten Rorschach-Merkmalen und dem IQ zeigten sich gegenüber unserer Hauptuntersuchung erhöht (sie erreichten für F + % mit .31 die 5%- und für B mit .42 die 1%-Grenze) doch erreichten sie nicht annähernd die von M i e l d s (1966, S. 62) berichtete Höhe. Im Falle von Examensgutachten mögen sich zwar die geringere Versiertheit der Beurteiler korrelationsmindernd, die Erwartung positiver Zusammenhänge, die vermehrte Beurteilungszeit und die persönliche Kenntnis des Probanden sowie die Identität des Versuchsleiters jedoch korrelationsfördernd ausgewirkt haben. Es ist geradezu verwunderlich, daß sich bei Versuchs-

leitem, denen das Intelligenz-Testergebnis bereits bekannt ist, nicht auch auf der Merkmalebene höhere Korrelationen mit dem Rorschach einstellen.

Hilfreich für eine Interpretation der Tatsache, daß sich auf keiner methodischen Ebene Zusammenhänge zwischen Intelligenzmessung und Rorschach-Intelligenzschätzung aufweisen lassen, könnte eine Betrachtung der *Motivationslage* der Versuchsperson beim Intelligenztest einerseits, und beim Rorschach-Versuch andererseits sein.

Während sich ein Intelligenztest wie der HAWIE gewissermaßen aus seinen Aufgaben heraus für die Vp als Intelligenzprüfung definiert, wird beim Rorschach (wie C o h e n , 1962, ausführt) durch die mangelnde Determiniertheit von Instruktion und Testmaterial eine Verunsicherung der Vp hervorgerufen, die eine hohe Leistungsmotiviertheit zumindest nicht begünstigt. Zwischen Test und projektivem Versuch scheint ein starkes Gefälle an aktualisierter Leistungsmotivation zu bestehen.

Gerade das Leistungsmotiv, wie es der Intelligenztest durch standardisierte Aufgaben und Instruktionen (die häufig auch Hinweise auf Zeitbegrenzungen enthalten) für jede Vpn fast optimal anspricht, wird im Verlaufe des Klecksdeuteversuchs keineswegs verstärkt: Auf die Frage „Muß ich mehr sagen?“ antwortet der V1 z. B. „Wie sie wollen!“, und selbstverständlich sagt der V1 im Sinne der Intelligenzhöhenschätzung die Unwahrheit, wenn er behauptet, „Es gibt keine richtigen oder falschen Antworten“ oder ähnliches. Außerdem gibt B o h m (S. 24) für Zwischenfragen den Rat „... wenn die Vp fragt, ob dies ein Intelligenztest sei ... ist es zweckmäßig, diese Frage sofort wahrheitsgemäß zu verneinen“. Hier offenbart sich das Dilemma eines Verfahrens, das kein Intelligenztest sein und doch die Intelligenzhöhe bestimmen will.

Sieht man die Anzahl der Deutungen als groben Indikator für Leistungsmotivation an, — die Befunde von C a l d e n u. C o h e n (1953, zit. n. C o h e n , 1962) scheinen dies zu bestätigen — so ergibt sich für unsere Untersuchung eine erhöhte Leistungsbereitschaft bei der Studentengruppe. Dies entspricht den Beobachtungen der Untersucher; die Lehrlinge waren darüber informiert, daß der Versuch wissenschaftlichen Zwecken diene und für ihre Personalakte ohne Bedeutung war.

Bemerkenswert ist, daß gerade jener Untertest (FL), der am geringsten zum Gesamt-IQ beiträgt und von W e c h s l e r (Fußnote S. 111) als sehr empfindlich „gegen momentane Schwankungen der Aufmerksamkeit und des Interesses“ bezeichnet wird, am häufigsten mit Rorschach-Merkmalen signifikant korreliert — allerdings fast durchweg negativ. Der signifikant negative Zusammenhang von FL und R (Antwortenzahl im Rorschach) ließe sich möglicherweise aus dem für Erwachsene äußerst geringen Aufforderungscharakter des FL erklären: Vpn mit hohem Quantitätsehrgeiz (einer beim FL nicht sehr erfolversprechenden Haltung) würden beim Klecksdeuten zu einer hohen Deutungszahl, dementsprechend aber auch zu einer absolut hohen Zahl von R +, G, D, Summe d. Zw., F, B, B +, V, Summe d. Schattierungsdeutungen, fb und T kommen, als deren Bedingung die Gesamtzahl der Deutungen nach der begründeten Auffassung G l i c k s t e i n s angesehen werden muß (M u r s t e i n , S. 268).

Unterstellt man, daß die Antwortenzahl den Grad der Leistungsmotivation anzeigt, so läßt sich auch für unsere Untersuchung ein Zusammenhang zwischen Intelligenzhöhe und Leistungsmotiviertheit nicht feststellen (vgl. McClelland et al., 1953; Heckhausen, 1965).

Aufgrund unserer Ergebnisse kann der Feststellung Eysencks (1965, S. 4) wonach es keinen Beweis gibt „für irgendeinen eindeutigen Zusammenhang zwischen beliebigen projektiven Testmerkmalen und Intelligenzeigenarten und Fähigkeiten, die unabhängig davon gemessen, geschätzt oder klassifiziert wurden“, nicht widersprochen werden.

Zusammenfassung

Der trotz einer Fülle negativer Ergebnisse von Validitätsuntersuchungen bestehende Anspruch des Rorschach-Verfahrens auf gültige quantitative Intelligenzaussagen wurde an 80 Fällen empirisch überprüft. Auf der Merkmalsebene wurden von 12 Haupthypothesen, die sich auf das Rorschach-Intelligenzsyndrom beziehen, nur eine, und von 68 Nebenhypothesen nur 5 verifiziert. Die Interkorrelationen von Rorschach-Merkmalen und Intelligenztestvariablen (HAWIE) ergaben einen mittleren Wert von 0.0. Auf der Interpretationsebene ergab sich, daß zutreffende Intelligenzschätzungen von kundigen Beurteilern aufgrund des Rorschach-Versuchs nicht möglich sind. Es zeigte sich eine allgemeine Tendenz, niedrige Intelligenzen zu überschätzen, und höhere Intelligenzen zu unterschätzen, und zwar um so stärker, je extremer der betreffende IQ ist; insgesamt wurden die Vpn als zu intelligenz-heterogen eingestuft. Die Ergebnisse wurden im Zusammenhang mit den völlig entgegengesetzten Befunden von Mields (1966) interpretiert; dabei wurden weitere Resultate vorgelegt. In der Diskussion wurde ferner auf die erheblichen Unterschiede zwischen dem Rorschach-Versuch und standardisierten Intelligenztests in bezug auf die Motivationslage der Vpn hingewiesen.

Literatur

- Altus, W. D. and Thompson, G. M. (1949) The Rorschach as a Measure of Intelligence. *J. consult. Psychol.* 13, 341—347
- Armitage, S. G., Greenberg, P. D., Pearl, D., Berger, D. G. and Daston, P. C. (1955) Predicting Intelligence from the Rorschach. *J. consult. Psychol.* 19, 321—329
- Bialick, I. and Hamlin, R. M. (1954) The Clinician as Judge: Details of Procedure in Judging Projective Material. *J. consult. Psychol.* 18, 239—242
- Blatt, S. J. and Allison, J. (1963) Methodological Considerations in Rorschach Research. The W Response as an Expression of Abstractive and Integrative Strivings. *J. proj. Tech. pers. Assess.* 27 (3), 269—278
- Bohm, E. (1957) *Lehrbuch der Rorschach-Psychodiagnostik*. 2. Aufl., Bern und Stuttgart
- Cohen, R. (1962) Die Psychodynamik der Testsituation. *Diagnostica* 8, 3—11
- Cohen, R., Bortz, J. u. Schümer, R. (1965) Die kommunikative Funktion von Rorschach-Deutungen. *Diagnostica* 11, 56—73
- Consalvi, C. and Canter, A. (1957) Rorschach Scores as a Function of Four Factors. *J. consult. Psychol.* 21, 47—51

- Cronbach, L. J. (1949) Statistical Methods Applied to Rorschach Scores: A Review. *Psychol. Bull.* 46, 393—429
- Davis, H. S. (1961) Judgments of Intellectual Level from Various Features of the Rorschach Including Vocabulary. *J. proj. Tech.* 25, 155—157
- Eisdorfer, C. (1963) Rorschach Performance and Intellectual Functioning in the Aged. *J. Gerontol.* 18, 358—363
- Eysenck, H. J. (1965) Persönlichkeitstheorie und psychodiagnostische Tests. *Diagnostica* 11, 3—27
- Fröhlich, W. D. (1965) *Forschungsstatistik*. 4. Aufl., Bonn
- Fuchs, C. (1942) Hohe Intelligenz, Versuch ihrer experimentellen Erfassung mit dem Rorschach-Test. *Z. Psychol.* 152, 30—112
- Heckhausen, H. (1965) Leistungsmotivation. In: Thoma, H. (Hg.) *Handbuch der Psychologie*, Bd. 2, Göttingen, 602—704
- Hörmann, H. (1964) Theoretische Grundlagen der projektiven Tests. In: Heiß, R. (Hg.) *Handbuch der Psychologie*, Bd. 6, Göttingen, 71—105
- Holzberg, J. D. and Belmont, L. (1952) The Relationship between Factors in the Wechsler-Bellevue and Rorschach Having Common Psychological Rationale. *J. consult. Psychol.* 16, 23—29
- Klopfer, B. and Kelley, M. D. (1946) *The Rorschach Technique*. New York
- Lienert, G. A. und Matthaei, F. K. (1958) Die Konkordanz von Rorschach-Ratings. *Z. diagn. Psychol.* 6, 228—241
- Lots of, E. J. (1953) Intelligence, Verbal Fluency, and the Rorschach Test. *J. consult. Psychol.* 17, 21—24
- McClelland, D. C., Atkinson, J. W., Clark, R. A. and Lowell, E. L. (1953) *The Achievement Motive*. New York
- Michel, L. (1961) Zuverlässigkeit und Gültigkeit von quantitativen Intelligenz-Diagnosen aus dem Rorschach-Test. *Diagnostica* 7, 44—60
- Michel, L. (1963) Der Rorschach-Test als Intelligenz-Diagnostikum. *Schweiz. Z. Psychol., Beih.* 45, 129—139
- Mields, J. (1964) Möglichkeiten der Intelligenz-Diagnose bei Team-Work. *Diagnostica* 10, 15—31
- Mields, J. (1966) Zur Frage der Interpretations-Reliabilität und Interpretations-Validität des Rorschach-Tests. *Diagnostica* 12, 52—66
- Murstein, B. I. (1960) Factor Analyses of the Rorschach. *J. consult. Psychol.* 24, 262—275
- Ogdon, D. P. and Allee, R. (1959) Rorschach Relationship with Intelligence among Familial Mental Defectives. *Am. J. ment. defic.* 63, 889—896
- Pauker, J. D. (1963) Relationship of Rorschach Content Categories to Intelligence. *J. proj. Tech. pers. Assess.* 27, 220—221
- Rorschach, H. (1954) *Psychodiagnostik*. 7. Aufl., Bern
- Sommer, R. (1957) Rorschach Animal Responses and Intelligence. *J. consult. Psychol.* 21, 358
- Spitznagel, A. (1963) Objektivität der Rorschach-Interpretation. *Schweiz. Z. Psychol., Beih.* 45, 287—300
- Spitznagel, A. und Vogel, H. (1964) Formdeutungsverfahren. In: Heiß, R. (Hg.) *Handbuch der Psychologie*, Bd. 6, Göttingen, 556—632
- Tauble, E. S. (1955) The Use of the Rorschach Test in Evaluating the Intellectual Levels of Functioning in Schizophrenics. *J. proj. Tech.* 19, 163—169

- Trier, T. R. (1958) Vocabulary as a Basis for Estimating Intelligence from the Rorschach. *J. consult. Psychol.* 22, 289—291
- Tucker, J. E. (1950) Rorschach Human and other Movement Responses in Relation to Intelligence. *J. consult. Psychol.* 14, 283—286
- Vernon, P. E. (1935) The Significance of the Rorschach Test. *Brit. J. med. Psychol.* 15, 199—217
- Wechsler, D. (1964) Die Messung der Intelligenz Erwachsener. Deutsche Bearbeitung von A. Hardesty und H. Lauber. 3. Aufl., Bern und Stuttgart
- Wewetzer, K. H. (1961) Zum Problem der faktoriellen Gültigkeit projektiver Testverfahren. *Schweiz. Z. Psychol.* 20, 238—252
- Williams, H. L. and Lawrence, J. F. (1953) Further Investigation of Rorschach Determinants Subjected to Factor Analysis. *J. consult. Psychol.* 17, 261—264

Anschrift der Verfasser:

Dr. H. D. Schmidt, H. H. Figge und S. T. Gehlen
Psychol. Institut der Universität
5300 Bonn
Am Hof 4