

7. Kapitel

Exploration

Udo Undeutsch

Die Exploration ist wahrscheinlich die umstrittenste psychologische Erhebungsmethode. Sie wird von einigen für die ergiebigste und sicherste Erkenntnisquelle gehalten, von anderen als wertlos eingeschätzt. Ebenso unterschiedlich ist der Begriffsgebrauch. Im englischen Sprachbereich wird der Begriff für Befragungstechniken überhaupt nicht verwendet, im deutschen Sprachbereich wird der Begriff in der Psychologie in der unterschiedlichsten Weise gebraucht: manchmal synonym mit Befragung überhaupt, manchmal eingeeengt auf ganz bestimmte Befragungstechniken, wobei wieder große Unterschiede zwischen verschiedenen Autoren bestehen.

1. Begriffsbestimmung

Ein *Gespräch* ist die partnerbezogene wechselseitige Ausübung der Sprechfähigkeit im zwischenmenschlichen Kontakt in der Absicht, einen Austausch von Innerlichkeitsgehalten vorzunehmen. Dabei wechseln die Rollen der Gesprächsteilnehmer zwischen Sprecher und Hörer. Diese Rollen können unter Gesprächsteilnehmern annähernd gleichgewichtig oder mit unterschiedlichem Gewicht verteilt sein. Sind die Rollen der Gesprächspartner stärker ungleichgewichtig verteilt, so ergeben sich daraus Sonderformen der Gesprächsführung: das Lehrgespräch, das therapeutische Gespräch, das Beratungsgespräch, das Verkaufsgespräch usw. auf der einen Seite und die verschiedenen Formen der Befragung auf der anderen Seite.

Die *Befragung* kann sehr unterschiedlichen Zwecken dienen und in sehr verschiedener Form vorgenommen werden.

1. Sie kann Wissen, Meinungen und Einstellungen über außerpersönliche Sachverhalte erkunden.
2. Sie kann der Vorbereitung einer helfenden (beratenden, therapeutischen usw.) Intervention dienen.

3. Sie kann den befragten Menschen selbst zum Ziel haben: sein Erleben und Verhalten in Vergangenheit und Gegenwart und sein „Wesen“.

Befragungen mit der erstgenannten Zielrichtung werden im deutschen Sprachgebrauch als *Interview* bezeichnet. Es ist in den empirischen Sozialwissenschaften das methodische Instrument mit der weitesten Verbreitung und der größten Zahl von Anwendungsmöglichkeiten. Methodologische Übersichtsreferate sind im deutschen Sprachbereich z.B. von Scheuch 1962, Anger 1969 und Erbslöh 1972 erstattet worden.

Für Befragungen der zweiten Art wird der Begriff *Anamnese* verwendet. Thoms (1975) definiert:

„Anamnese ist eine Methode klinischer Informationssammlung und bezeichnet gleichzeitig die ermittelten Daten. Gesprächsweise oder mit Fragebogen wird dabei die Vorgeschichte eines Menschen, bezogen auf eine bestimmte Fragestellung - psychische oder körperliche Symptomatik -, erhoben.“

Die Anamnese ist eine Methode der klinischen Psychologie. Sie ist auch keineswegs auf die Befragung des Patienten beschränkt. Fremdbeobachtungen und Dokumente werden regelmäßig einbezogen.

Befragungen mit der an dritter Stelle genannten Zielrichtung können sehr verschiedene Bindungsgrade aufweisen: von der vollkommen standardisierten Befragung über die teilstandardisierte Befragung bis hin zur nicht-standardisierten (= freien, ungebundenen) Befragung. Nur die letztgenannte Befragungsart wird im deutschen Sprachgebrauch als *Exploration* bezeichnet, wobei die Wahl eines Fremdwortes, das als terminus technicus gebraucht wird, anzeigt, daß nicht jede derartige Befragung, wie sie im Alltag in unzähligen Varianten vorkommt, gemeint ist, sondern nur die fachkundig vorgenommene psychologische (oder tiefenpsychologische oder psychiatrische) Befragung. Dies unterscheidet die Exploration vom Interview des Journalisten wie auch von der Vernehmung durch Polizei, Staatsanwaltschaft oder Gericht. Der Umfang der auf das Erleben und Verhalten des untersuchten Menschen und auf diesen selbst gerichteten Befragung kann wiederum sehr unterschiedlich sein: er reicht von einer Befragung, die „das Ganze eines individuellen Lebenslaufes erfassen soll“ (Thomae 1968, 113), bis hin zu speziellen Fragestellungen, wie es die Aufklärung eines vom Befragten erlebten Ereignisses ist. Es wäre mit dem fachwissenschaftlichen Sprachgebrauch im Deutschen nicht zu vereinbaren, den Ausdruck Exploration auf die das ganze bisherige Leben eines Menschen umfassende Befragung zu beschränken. Es ist aber zweckmäßig, thematisch begrenzte Befragungen, bei denen es nur einen bestimmten Lebensbereich abzuklären gilt, als „Exploration zur Sache“ (Undeutsch 1954, 15) zu spezifizieren. Entscheidend bleibt aber auch bei diesen stark themenzentrierten Befragungen, daß es sich um offene und wenig strukturierte Befragungen handelt, in denen der befragte Mensch sein Erleben berichtend ausbreiten

kann. *Exploration ist demnach die mit psychologischer Sachkunde vorgenommene nicht-standardisierte mündliche Befragung eines einzelnen Menschen durch einen einzelnen Gesprächsführer mit dem Ziel, Aufschluß zu erhalten über „Das Individuum und seine Welt“* (Thomae 1968). Sucht man nach einem etwa bedeutungsgleichen deutschen Wort, so kommt man schon vom genauen lateinischen Wortsinn her (ex-plorare = auskundschaften, erforschen) auf den Begriff „*Erkundungsgespräch*“, der auch schon von Arnold (1957) und Pongratz (1957) vorgeschlagen worden ist.

2. Geschichte

Es läßt sich die Annahme begründen, daß das Erkundungsgespräch im Alltagsleben so alt ist wie der menschliche Sprachgebrauch selbst. Die Frage des Diomedes an seinen Gegner auf dem Schlachtfeld

„Wer doch bist Du, Edler, der sterblichen Erdenbewohner?“ (Homer: Ilias, 6. Ges., Z. 123)

(die damals dazu führte, daß die Gegner sich als Freunde aus Väterzeiten erkannten und ihre Freundschaft erneut beschlossen, statt gegeneinander zu kämpfen) ist in dieser und ähnlicher Form (freilich nicht immer mit dem damaligen erfreulichen Erfolg) eine der Grundfragen des menschlichen Alltags. Als terminus technicus ist der Begriff „Exploration“ in der klassischen Psychiatrie entstanden, wo darunter das Eruiieren psychopathologischer Phänomene beim Patienten verstanden wurde. Der Begriff wurde in weiterer Bedeutung in die Psychologie übernommen von Binet und Piaget.

Die ersten thematischen Ansätze zur Erforschung von Individuen und ihrer Welten durch Befragung und andere biographischen Methoden finden sich bei W. Stern (1900, 3. Aufl. 1921) und seinen Mitarbeitern (Baade und Lipmann 1909, Margis 1911). In der wissenschaftlichen Forschung ist die Erhebungsmethode der Exploration intensiv angewandt worden etwa innerhalb der Child-Guidance-Untersuchung von MacFarlane (1938), in den Jahren 1938 bis 1947 von Kinsey und seinen Mitarbeitern (1948 und 1953) zur Erforschung des sexuellen Verhaltens des Menschen. Bezüglich des Einsatzes der Exploration in der Persönlichkeitsforschung verdient Pfahler (1939) der Vergessenheit entrissen zu werden. Sodann ist aber vor allem auf Thomae zu verweisen, der durch die Art des Einsatzes der Exploration und der Auswertung von Explorationsbefunden gerade jene Gebiete der Persönlichkeitserforschung erschlossen hat, die bisher von der Wissenschaft beiseite gelassen worden waren, nämlich das „alltägliche“ wie auch das „krisenhafte“ Verhalten des Menschen in „natürlichen“ Situationen (1968). Nach Thomae bildet die Exploration

„einen der wenigen Zugänge zu einer durch den methodischen Zugriff noch nicht veränderten seelischen Wirklichkeit“ (1968, 113).

Sein Hauptwerk „Das Individuum und seine Welt“ (1968) kann man nach seinen eigenen Worten

„als einen Beitrag zur Technik der Auswertung von systematisch gewonnenen Explorationsprotokollen und von Protokollen über Verhaltensbeobachtungen aus unterschiedlich langen biographischen Einheiten ansehen“ (117).

Für charakterdiagnostische Zwecke hat die Exploration eine zentrale Rolle gespielt in der deutschen Wehrmachtspanychologie (1927-1945; s. hierzu Kreipe 1936, Walther 1941, Beck 1942, Kröber 1942, Mierke 1944, 66-70). In abgewandelter Form wurde die Exploration von der US-amerikanischen Militärpsychologie übernommen (Assessment 1948). Seit Herbst 1944 wurde die Exploration zusätzlich zu einer Serie von Fähigkeitstests und einem Persönlichkeitsfragebogen bei der Auswahl von Militärpiloten der schwedischen Luftwaffe (Trankell 1956) und seit 1951 bei der Auswahl der Piloten der SAS (Trankell 1959) verwendet.

Der Wert der wissenschaftlich ausgestalteten Exploration für die Beurteilung der Eignung von Bewerbern wurde natürlich auch für Berufe der freien Wirtschaft entdeckt. In England hatte schon vor dem Zweiten Weltkrieg Oldfield im Auftrag des National Institute for Industrial Psychology die Erfahrungen aus Einstellungsgesprächen gesammelt und ausgewertet und daraus eine Methodik des Einstellungsgesprächs entwickelt (1951). Parkinson (1957) schrieb eine geistreiche Parodie über das in England bei Behörden wie Industrieunternehmen übliche Bewerbergespräch. Eine noch größere Rolle spielte das Bewerberinterview (employment interview, selection interview) im US-amerikanischen Wirtschaftsleben, weil dort - wegen des geringen Ausleseeffekts des dortigen Schul- und Bildungssystems - die Bewerber viel weniger durch Schulbildung und Berufsausbildung vorsortiert sind, weil eine viel größere Fluktuation der Arbeitskräfte besteht und weil, vor allem in den unteren Bevölkerungsschichten, Ehrlichkeit und Redlichkeit gegenüber dem Arbeitgeber sehr zu wünschen übrig lassen. Nach einer Erhebung von Spriegel und James (1958) gaben im Jahre 1930 93% von 236 befragten Firmen an, daß sie ihre Bewerber vor der Einstellung einem Interview unterziehen. Bellows und Estep schätzten 1954 die Zahl der jährlich in den USA zur Bewerberauslese durchgeführten Interviews auf 150 Millionen. Dort ist deshalb schon seit Jahrzehnten eine reiche Literatur über Explorationstechnik in Monographien (Kephart 1952, Bellows & Estep 1954, Fear 1953, 2. Aufl. 1978, The McGraw Hill course in effective interviewing, 1973) sowie in Handbüchern der Industrie-psychologie (Roethlisberger und Dickson 1939, Bellows 1949, 3. Aufl. 1961) vorhanden.

Eine andere Entwicklungslinie leitet sich her aus der sehr viel älteren Kriminaltaktik, die mit Vernehmungspanychologie (H. Gross 1893, 1898) und Aussagepsychologie (W. Stern 1902) wichtige Beiträge zur Explorationstechnik ver-

heimlicher Sachverhalte und zur Bewertung von Explorationsdaten geliefert hat. Auf die Bedeutung der „experimentell gestalteten“ „Exploration zur Sache“ hat Undeutsch für die Aussagepsychologie (1954, 15 und 1967, 117) und für die Beurteilung der Schuldfähigkeit (1965, ebenso Thomae und Schmidt 1967, 354-356) hingewiesen. Ebenso hat Undeutsch zu wiederholten Malen auf die Bedeutung der Exploration in der Fahrereignungsdiagnostik (für die Abschätzung der Rückfallwahrscheinlichkeit bei Verkehrsdelinquenten unter Alkoholeinfluß sowie bei mehrfach durch Verstöße in nüchternem Zustand auffällig gewordenen Verkehrsdelinquenten) hingewiesen.

Welche Bedeutung die Exploration im methodischen Arsenal der Psychologie heute erlangt hat, kann man am besten daran erkennen, daß im „Handbuch der Psychologie“ in den bisher erschienenen Bänden keine andere Methode der Datengewinnung so häufig behandelt worden ist wie sie. Es sind ihr vier umfassende Artikel gewidmet worden:

In ihrer allgemeinsten Form, aber ausschließlich im Hinblick auf ihre Verwendung in den Sozialwissenschaften, behandelt die Befragung der Artikel von

Anger: Befragung und Erhebung. 1969. 7/1, 567-617.

Die Bedeutung der Exploration in der Persönlichkeitsdiagnostik behandelt der Artikel von

Schraml: Das psychodiagnostische Gespräch. (Exploration und Anamnese). 1964, 6, 868-897.

Ihre Bedeutung in der Form der Anamnese für die Klinische Psychologie der Artikel von

Kemmler und Echelmeyer: Anamnese-Erhebung. 1978. 8/2, 1628-1648.

Ihre Bedeutung für die Forensische Psychologie der Artikel von

Friedrichs: Die aussagepsychologische Exploration. 1967, 11, 3-25.

Die Zahl der empirischen Untersuchungen zur Methode der Exploration ist rund um die Welt in den letzten Jahren gewaltig angewachsen.

3. Qualitative Charakterisierung

Die Exploration gehört zu den qualitativen Methoden par excellence. Deren methodische Eigentümlichkeiten sind herausgearbeitet z.B. in den Artikeln von W. Salber (1960, 1969). Die dort herausgestellten methodischen Grundzüge der qualitativen Methoden haben samt und sonders auch für alle explorativen Verfahren Geltung.

Im Sinne der Methodensystematik von Cattell (1957) gehört die Exploration zu den Verfahren, welche der Gewinnung von „Life-data“ dienen. Sie ist die Hauptmethode einer „psychologischen Biographik“, die nach Thomae als eine Synthese zwischen „ideographischer“ und „nomothetischer“ Forschung aufgefaßt werden kann und deren Ziel es ist, „eine psychologische Analyse des menschlichen Verhaltens im natürlichen Ablauf des Lebens zu erarbeiten“ (1968, 105).

Die Exploration *unterscheidet* sich von anderen Methoden der Datenerhebung.

Sie unterscheidet sich vom *Experiment* dadurch, daß das von ihr erfaßte Erleben und Verhalten der Untersuchungssituation vorausliegt und daher Bestandteil des natürlichen Ablaufs des Verhaltens des Individuums bleibt, während das Experiment in die Lebensumstände des Untersuchten eingreift, bestimmte Bedingungen herstellt und diese planmäßig verändert. Sie kann aber andererseits von der experimentellen Methodik profitieren, indem sie die Bedingungen der Befragung planmäßig verändert, was alsdann zu der Beobachtung Gelegenheit gibt, welche Veränderungen dadurch in den Mitteilungen der Befragten hervorgerufen werden (Undeutsch 1954, 15). Diese Veränderungen können zur Beurteilung des Wahrheitsgehalts der gemachten Mitteilungen herangezogen werden.

Sie unterscheidet sich vom *Test* dadurch, daß sie sich nicht darauf beschränkt, nur einen durch standardisierte Testbedingungen definierten ganz bestimmten Ausschnitt aus dem Verhaltensrepertoire in Anspruch zu nehmen. In Tests werden in der Regel Auslösereize dargeboten, die dem Untersuchten entweder kaum oder jedenfalls weit weniger vertraut sind als dem Untersucher, während in der Exploration typischerweise eine „Gemeinsamkeit der Erlebnissphäre“ (Thomae 1968, 112) besteht.

Zudem ist die Testsituation im allgemeinen durch eine gewisse „Neutralität“ und Konstanz der Situation gekennzeichnet. Die in den hohen Testwiederholungskoeffizienten zutage tretende hohe Stabilität und Konstanz des Reagierens der Individuen sind eher Ausdruck einer Versuchstechnik, welcher es gelingt, aus allen „existentiell“ begründeten Varianten des Verhaltens, wie sie sich schon im Laufe eines Tages, erst recht aber im Laufe eines Lebens ergeben, auf eine relativ neutrale Sphäre auszuweichen. Das Ziel der persönlichkeitsdiagnostischen Verfahren, nämlich bestimmte Persönlichkeitszüge mit konstanter Zuverlässigkeit zu erfassen, wird in den Tests geradezu dadurch erreicht, daß diese gegen den Aufforderungscharakter existentiell bedeutsamer Aspekte der jeweiligen Situation immun gemacht werden. Als Folge davon haben diese diagnostischen Verfahren, welche von dieser Konzeption der „existentiellen Neutralität“ ausgehen, für die Erfassung des realen individuellen Verhaltens und seiner „dispositionellen“ Hintergründe nur bedingt Wert. In

der Exploration ist es demgegenüber möglich, das Verhalten des Individuums in der Vielfalt der banalen wie der existentiell bedeutsamen Situation seines realen Lebensvollzugs zu erfassen.

Auch der standardisierte *Fragebogen* läßt der persönlichkeitspezifischen Aktivitätentfaltung keinen Spielraum. Die erwähnte „Gemeinsamkeit der Erlebnissphäre“ ist auch bei den Fragebögen stärker beeinträchtigt, da diese so konstruiert sind, daß vom Befragten ein hoher Prozentsatz der gestellten Fragen verneinend beantwortet werden muß.

„Die Vermutung, eine ganze Serie von ‚neurotischen Symptomen‘ zu haben, wirkt auf ‚Normale‘ nicht gerade kontaktstiftend“ (Thomae, ebd.).

Thomae führt zusammenfassend aus:

„Die Gewinnung dieser Verhaltens-Daten hat dem ideographischen Prinzip zu folgen, das möglichst die ‚unverzerrte‘ psychische Wirklichkeit zu erfahren und zu erfassen strebt. Deshalb können Tests und Fragebögen nicht die *primäre* Quelle dieser Forschung sein, da sie in jedem Falle ja eine Veränderung des realen Verhaltens (Umschaltung von einer sinnbezogenen ‚erfüllten‘ Situation auf eine relativ sinnarme, nur durch Zusatzmotivationen zu stabilisierende) darstellen“ (1968, 106).

Am schwierigsten ist naturgemäß die Abgrenzung der Exploration gegen das *Interview*. Thomae (1968; 112) stellt zunächst die Gemeinsamkeiten heraus: Beide stellen eine Form der verbalen Kommunikation dar. Gemeinsam ist beiden weiterhin, daß Auslösereize (z.B. Fragen) verwendet werden, die einer beiden Partnern in ähnlicher Weise vertrauten Erlebnissphäre angehören. Dennoch besteht zwischen Interview und Exploration ein Unterschied, den es mit Wellek (1958) und Thomae (1968, 112) festzuhalten gilt. Nicht nur ist das Interview in seinen verschiedenen Formen mehr oder minder standardisiert, sondern es ist auch ausschließlich oder doch vorwiegend auf Wissen, Meinungen, Vorstellungen von außerhalb des befragten Subjekts gelegenen „Dingen“ ausgerichtet. Demgegenüber ist es für die Exploration typisch, daß sie auf den befragten Menschen selbst gerichtet ist, der nicht nur als Reflektor, sondern in seiner Eigenschaft als Subjekt, als Mitmensch, als Partner Gegenstand des ganzheitlichen explorativen Interesses ist.

Von allen anderen Datenerhebungsverfahren unterscheidet sich die Exploration dadurch, daß sie nicht wie diese die Antwortmöglichkeiten des Untersuchten auf ein Konzept einengt, das den Erwartungen einer bestimmten Theorie oder den Erfordernissen einer bestimmten Methodologie entspricht. Dadurch bleibt den anderen Verfahren der Zugang zur vollen Breite menschlichen Verhaltens verschlossen.

„Da eine Fremdbeobachtung dieses Verhaltens aus äußeren Gründen meist nicht möglich ist, stellt die Exploration einen der wenigen Zugänge zu einer durch den methodologischen Zugriff noch nicht veränderten seelischen Wirklichkeit dar“ (Thomae 1968, 113).

Für die Erforschung der Existenzweise und der Funktion bestimmter „inhaltlicher“ Strukturen des „subjektiven Lebensraumes“, des Selbst und der am meisten relevanten „thematischen“ und „instrumentellen“ Einheiten innerhalb der menschlichen Existenz ist die Exploration unentbehrlich und unersetzbar, weil die untersuchten Phänomene dort untersucht werden müssen,

„wo sie aller Voraussicht nach anzutreffen sind, nämlich im ‚alltäglichen‘ wie u.U. auch im ‚krisenhaften‘ Verhalten des Menschen. Dieses in der erforderlichen Extensität und Intensität von fremden Beobachtern aus zu erfassen, wird jedoch durch hoffentlich kaum überwindbare ethische Vorentscheidungen verwehrt. In dieser Situation muß die Wissenschaft versuchen, dem einzig verfügbaren Zeugen, nämlich dem Subjekt des Geschehens, mehr Glauben zu schenken. Seine Aussagen, die wir durch die Exploration zu erfassen und zu vertiefen versuchen, stellen von hier aus gesehen die einzige sichere Quelle für die Erschließung des Verhaltens in ‚natürlichen‘ Situationen dar“ (Thomae 1968, 222f.).

Das schließt nicht aus, daß die Exploration in bestimmten ihrer Formen (Anamnese) und in bestimmten Phasen (gegen Ende) eine gewisse Standardisierung erfahren kann. Auch die Exploration kann systematisch auf die relevanten Gesichtspunkte der jeweils vorliegenden Fragestellung eingehen, wobei auch eine weitgehende Standardisierung der einzelnen Fragen, unter Umständen auch der Reihenfolge, vorgenommen werden kann. Prinzipiell müssen alle Fragen aber „offene Fragen“ bleiben.

Auch eine Quantifizierung von Explorationsergebnissen ist keineswegs ausgeschlossen. So kann z.B. im Verlauf einer Exploration der bisherigen Verkehrsgeschichte allmählich die Zahl der Verkehrsunfälle erfragt werden, an denen der Befragte beteiligt war (v. Klebelsberg 1970, S. 45), oder es kann zur Prognose der Rückfallwahrscheinlichkeit in eine erneute Fahrt in alkoholbedingt fahruntüchtigem Zustand die größte im letzten Jahr vor der Untersuchung aufgenommene Menge an alkoholischen Getränken exploriert werden (Welzel 1976). Außer solchen trivialen Quantifizierungsmöglichkeiten bestehen zahlreiche Skalierungsmöglichkeiten von formalen Grundkategorien des in der Exploration erfassbaren Verhaltens eines Menschen in den natürlichen zeitlichen Einheiten seines Lebens. Bei einem von Thomae unternommenen Versuch, das von verschiedenen Menschen geschilderte Geschehen unter Abstraktion von seinem jeweiligen Gehalt deskriptiv zu erfassen, ergaben sich mehrere formale Aspekte des biographischen Geschehens, die auf die natürlichen Einheiten des individuellen Bios - der Handlung, des Tageslaufes und der mehr oder minder großen faßbaren Ausschnitte des Lebenslaufes (z.B. Formen der Auseinandersetzung mit einer beruflichen oder familiären Situation) - anwendbar sind. Die Skalen können, je nach dem Ausmaß der vorhandenen Informationen, wenige (z.B. 5) oder viele (z.B. 9) Stufen haben (Thomae 1968, 124-216).

4. *Methodische Prinzipien des explorativen Gesprächs*

Ziel der in der Form des Gesprächs angestellten Erkundungsbemühungen ist das konkrete mitmenschliche „Individuum und seine Welt“. Dieses menschliche Individuum wird veranlaßt, Auskunft über sich und „seine“ persönlich-individuelle „Welt“ zu geben, über das, was es erlebt hat und der Erinnerung und der Erwähnung für wert hält, über seinen „psychologischen Lebensraum“ (K. Lewin 1936, S. 14ff.; im deutschen Originaltext 1969, 40-50). Das ist „alles, was vom Standpunkt des Psychologen für diese Person existiert“ (40), der

„Inbegriff dessen, was jeweils für das betreffende Individuum wirksam ist. Für die Aufgabe der begrifflichen Ableitung des Geschehens kann man Wirksamkeit als das Kriterium für psychologische Existenz verwenden: ‚*Wirklich ist, was wirkt*‘“ (41).

Dies erfordert die uneingeschränkte Kooperationsbereitschaft des Befragten. Seine Bereitschaft muß vom Explorator gewonnen werden. Alle explorations-technischen Anleitungen beschäftigen sich daher damit, wie diese grundlegende Voraussetzung am besten herzustellen sei.

Bedingungen für das Zustandekommen dieser Bereitschaft werden zunächst in der *Persönlichkeit des Explorators* gesucht. Verlangt werden: persönliche Reife, Reichtum an Vorstellungs- und Miterlebensmöglichkeiten (Rowe 1963, Blakeney & McNaughton 1971) und an Lebenserfahrung, Kontaktfähigkeit und charakterliche Werte wie Vertrauenswürdigkeit (für alles Vorstehende: Taft 1955, Wiley & Jenkins 1964, Steinkamp 1966) und eine Haltung dem Mitmenschen gegenüber, die C. G. Jung vor vielen Jahren (1932) einmal wie folgt gekennzeichnet hat:

„Will der Arzt die Seele eines Anderen führen, oder sie auch nur begleiten, so muß er mit ihr Föhlung haben. Diese Föhlung kommt nie zustande, wenn der Arzt verurteilt, ob er das nun mit so viel Worten laut tut, oder unausgesprochen im Stillen, ändert nichts an der Wirkung. Auch das Umgekehrte, nämlich dem Patienten unbesehen Recht geben, hilft nichts, es wirkt ebenso entfremdend wie das Verurteilen. Föhlung entsteht nur durch vorurteilslose Objektivität. Das klingt beinahe wie etwas Wissenschaftliches. Man könnte es mit einer rein intellektuellen, abstrakten Einstellung verwechseln. Was ich aber meine, ist etwas ganz anderes: Es ist etwas Menschliches, etwas wie eine Hochachtung vor der Tatsache, vor dem Menschen, der an dieser Tatsache leidet, vor dem Rätsel eines solchen Menschenlebens“ (Ges. W. 1963, 11, 366f.).

Darüber hinaus stellt ein *sozialpsychologischer Bedingungs-zusammenhang* eine grundsätzlich wichtige Voraussetzung für das Zustandekommen dieser Bereitschaft dar: Jeder Mensch teilt sich lieber und leichter gegenüber einem anderen Menschen mit, wenn er überzeugt sein darf, daß der andere ihn versteht, denn nicht verstanden zu werden, birgt in sich fast unvermeidlich die Gefahr, falsch und daher ungerecht beurteilt zu werden.

„A subject is inevitably hesitant to discuss things which seem to be both outside of the experience of the interviewer, and beyond his knowledge“ (Kinsey u.a., 1948, 60).

So ist es denn auch gewiß kein Zufall, daß z.B. die ergiebigsten Explorationen mit Homosexuellen von homosexuellen Wissenschaftlern geführt worden sind. Das gleiche ist natürlich auch bei anderen Personengruppen zutage getreten: z.B. bei Alkoholabhängigen, Drogensüchtigen oder auch bei Ordensleuten. Cospers (1969) untersuchte die Möglichkeit des Interviewer-Bias systematisch, indem er das Trinkverhalten seiner 28 Interviewer sehr detailliert erhoben hat. Alle Interviewer waren Alkoholkonsumenten und wurden ausführlich geschult, um das Interview möglichst zu standardisieren. Trotzdem konnten eindeutige Effekte nachgewiesen werden: stärker trinkende Interviewer erhalten höhere Quantitätsangaben zum Trinken; die Angaben zur Häufigkeit des Konsums variieren nicht mit dem Trinken der Interviewer. Dies ist ein Hinweis darauf, daß Quantitätsangaben zum Trinken schwieriger zu erhalten sind als Häufigkeitsangaben. Es läßt sich recht genau angeben, worauf sich das Vertrauen auf die Verständnisfähigkeit und -bereitschaft beim Explorierten zu gründen pflegt:

1. Eine (vermutete, angenommene, erlebte) - mindestens partielle - Wesensverwandtschaft zwischen Explorator und Exploriertem. So hat sich z.B. in Untersuchungen von Ledvinka (1971, 1972) die Auswirkung von Rassen-gleichheit bzw. -Verschiedenheit auf die vom Explorierten gemachten Mitteilungen gezeigt.
2. Eine (vermutete, angenommene, erlebte) aus - mindestens partieller - gleichartiger Lebenserfahrung oder wenigstens
3. aus Sachkunde und Berufserfahrung erwachsene Vertrautheit mit dem explorierten Lebensgebiet.

Die Vertrautheit mit dem explorierten Lebensgebiet muß der Explorator gegenüber dem Explorierten während der Exploration unter Beweis stellen:

1. durch sachkundige, intime Kenntnis des betreffenden Lebensgebietes verrätende Fragerichtung,
2. durch Nachweis seiner Vertrautheit mit typischen Zusammenhängen zwischen einzelnen Gegebenheiten des betreffenden Lebensgebietes,
3. durch insider-typischen Wortgebrauch,
4. durch Vermeidung aller Äußerungen des Erstaunens, der Überraschung, der positiven oder negativen Bewertung des vom Explorierten berichteten Verhaltens oder Erlebens.

Der Explorator muß mit der „Welt“, in der der Explorierte lebt, vertraut sein: mit seinem äußeren Lebensraum, mit seiner sozialen Schicht und seinem Berufsstand, mit seinen Lebensbedingungen, mit seiner Sprache und seinen Wertvorstellungen. Er muß aber auch vertraut sein mit dem Lebensgebiet oder

dem Sachgebiet, dem die Exploration gilt. Es genügt auch nicht, daß der Explorator solche Vertrautheit behauptet, sondern diese muß während der Exploration, in deren Gestaltung und ihrem Verlauf, in Erscheinung treten, muß für den Explorierten auf Schritt und Tritt spürbar werden, damit die „Gemeinsamkeit der angesprochenen Erlebnissphäre“ (Thomae 1968, 112) zum Tragen kommen kann. Kinsey u.a. (1948) gibt dafür ein eindrucksvolles Beispiel:

„A single phrase from an understanding interviewer is often sufficient to make the subject understand this, and such an interviewer wins a record where none would have been disclosed to the uneducated investigator. A specific illustration will make this more apparent.

This is the case of the older Negro male whose first answers were wary and evasive. When questioned concerning his occupation, he listed a variety of minor jobs which, taken in connection with his manner of response, seemed to spell underworld activities. We followed up our clue by immediately asking the subject whether he had ever been married. We were not satisfied with his denial of marriage, and followed with a question as to whether he had ever lived common law. The easy use of a vernacular term made him feel freer to talk, and when he admitted that he had so lived, we asked how old he was when he first lived common law. When he said that he was then fourteen, our first suspicion concerning his underworld activity was confirmed, and we immediately followed up by asking how old the woman was. At this, he smiled and admitted that she was thirty-five. Then we remarked, easily and without surprise: „She was a hustler, wasn't she?“ This was the final step necessary for winning complete confidence. The subject stopped short in his reply, opened his eyes wide, smiled in a friendly fashion, and said, ‚Well, sir, since you appear to know something about these things, I'll tell you straight.‘ The extraordinary record that we then got of his history as a pimp could not have been obtained if the subject had not comprehended that we understood the world in which he lived“ (60f.).

Das erfordert vom Explorator viel Einarbeitung und Vorbereitung. Als Beispiel sei erwähnt, daß Kinsey und seine Mitarbeiter vor Beginn ihrer allein aus Explorationen gewonnenen Erhebung mehrere Jahre darauf verwendet haben, sich die erforderlichen Sachkenntnisse auf dem von ihnen erforschten Lebensgebiet anzueignen, und ein weiteres volles Jahr, um sich in die Explorations-technik einzüben (1948, 61).

„Very often the interviewer's capacity to secure an accurate history depends upon his knowledge of the correlations that usually exist between certain items, and his readiness to demand an explanation of any inconsistency that appears in a particular history. To illustrate again: one starts by asking the girl how old she was when she turned her first trick (but one does not ask how old she was when she was first paid as a prostitute). She is then asked how many of the tricks return after their first contacts with her. Considerably later in the interview there is a question concerning the frequency with which she rolls her tricks (robs her customers). The girl who reports that few of the men ever return, and who subsequently says that she never robs any of the men, needs to be caught up abruptly and assured that you know that it doesn't work that way. If she

doesn't roll any of the men, why don't they return to her? This question is likely to bring a smile from the girl and an admission that since you appear to know how these things work, she will tell you the whole story, which means that she robs every time there is any possibility of successfully doing so" (Kinsey u.a., 1948, S. 61).

Zu den genannten persönlichen und sozialpsychologischen Voraussetzungen muß auf seiten des Explorierten hinzukommen die *Bereitschaft, sich mitzuteilen*. Diese Bereitschaft muß zu Beginn der Exploration geweckt und während des gesamten Verlaufes unterhalten werden durch die gewissermaßen werbe-psychologischen Maßnahmen (B. Spiegel 1965) des Informierens und des Motivierens: des Informierens über den Zweck der Untersuchung und des Motivierens zu aufrichtigen, ehrlichen, wahrheitsgemäßen und vollständigen Angaben. Die konkrete Ausgestaltung des Informierens und des Motivierens im Zuschnitt auf den konkreten Einzelfall wird je nach dem Untersuchungszusammenhang und je nach den Besonderheiten des Einzelfalles sehr verschieden aussehen können und müssen. Hier sind die Unterschiede sehr groß zwischen der Gewinnung der Mitarbeit für eine wissenschaftliche Untersuchung auf der einen Seite und der sehr viel schwerer zu gewinnenden Mitarbeit für eine Untersuchung, deren Ergebnis nicht von vornherein und zwangsläufig dem Untersuchten zugute kommt. Handelt es sich um die Mitarbeit an einer wissenschaftlichen Untersuchung, so sind oftmals schon die Befriedigung darüber, vor einem verständigen Zuhörer über sich selbst sprechen zu können, und die Genugtuung darüber, an einem anerkannt wichtigen Unternehmen mitzuwirken, eine ausreichende und im allgemeinen leicht zu erreichende Motivationsbasis (Kinsey 1948, 36; Whyte 1955; Bain 1960; Thomae 1968, 114). Sehr viel schwieriger ist die Motivation zu wahrheitsgemäßer und vollständiger Auskunft zu erzielen, wenn der Befragte - zu Recht oder zu Unrecht - gerade davon unerwünschte oder sogar direkt nachteilige Konsequenzen für sich zu befürchten hat. Das ist nicht nur der Fall bei allen Bewerberexplorationen, sondern in noch viel höherem Maße bei Fahrereignungsuntersuchungen, bei denen der Untersuchte selbstverständlich die Entdeckung seiner Ungeeignetheit mit allen Mitteln zu verhindern trachtet, und erst recht bei allen Untersuchungen im gerichtlichen Auftrag - sei es z.B. bei der Sorgerechtszuteilung oder bei der Rekonstruktion des Herganges eines Schadensereignisses zum Zwecke der Schadenersatzregelung im zivilrechtlichen Bereich oder bei der Rekonstruktion der Entstehung und des Ablaufes einer tatbestandsmäßig rechtswidrigen Handlung, der Motive und der geistig-seelischen Verfassung des Täters dabei im strafrechtlichen Bereich. Dennoch lehrt die Erfahrung, daß es auch in solchen Situationen - sogar in der Regel - möglich ist, eine aufrichtige Kooperationsbereitschaft des Befragten zu erzielen, wenn der Fragende psychologisch geschult ist und die vernehmungpsychologischen Richtlinien (Inbau & Reid 1974) richtig anwendet:

„Since he is trained in his task, he will typically sympathize with the suspect, provide face-saving rationalizations for any crimes that might have been committed, and indi-

cate subtly that he can understand how someone might murder but that lying is the lowest form of degeneracy: Such a procedure may sound farfetched, yet in the hand's of an expert it is remarkably convincing" (Orne, Thackray, and Paskewitz 1972, 746).

Die konkrete Ausgestaltung des Informierens und Motivierens ist für die einzelnen Untersuchungsanlässe und für die einzelnen Personen zu verschieden, um im Rahmen dieses Artikels dargestellt werden zu können. Vortreffliche Anleitungen dazu sind in der Literatur leicht zugänglich. Weithin bekannte Anleitungen zur Explorationstechnik sind: Richardson, Dohrenwend, and Klein, 1965; Bingham, and Moore, 1. ed. 1931, 4. ed. with the collaboration of J. W. Gustad, 1959; Kahn & Cannell 1957 (10. printing 1966); Fear 1953, sec. ed. 1978. Weniger bekannt ist, daß es gerade für die Exploration schwieriger, heikler Sachverhalte, bezüglich deren beim Befragten starke Tendenzen zur Verheimlichung bestehen, sehr gute und ganz aufs Praktische ausgerichtete Anleitungen gibt: Kinsey, Pomeroy & Martin 1948, 35-62, deutsche Ausgabe 1964, 22-52; Inbau & Reid 1974.

Als Voraussetzung für die Objektivität und die Differenziertheit der Auswertung der Explorationsbefunde ist es erforderlich, Aufnahmen der Exploration auf Tonträgern und maschinenschriftliche Übertragungen anzufertigen. Froehlich (1958) verglich für 97 anamnestiche Explorationen, die zur Vorbereitung von psychologischen Beratungen durchgeführt worden waren, die Aufzeichnungen der therapeutischen Berater mit den Tonbandaufnahmen dieser Explorationen. Seine Untersuchung kam zu dem Ergebnis, daß weniger als ein Drittel der Inhalte im schriftlichen Bericht auftauchte, diese aber mit 75% bis 94% Korrektheit. Es ließ sich jedoch keine klare Beziehung zwischen der Wichtigkeit der Daten und ihrer Wiedergabe durch den Therapeuten im schriftlichen Bericht erkennen. Froehlich fand in seiner Erhebung keinen signifikanten Unterschied zwischen erfahrenen und weniger erfahrenen Beratern. Zu einem gleichartigen Ergebnis kommt Thomae (1968):

„Als Ergebnis langer Erfahrungen in der Handhabung der explorativen biographischen Anamnese muß aber leider darauf verwiesen werden, daß Gedächtnisprotokolle für eine systematische Auswertung völlig unzureichend sind und das Mitschreiben bzw. -stenographieren erhöhte Anforderungen an den Untersucher stellt und das Gespräch sehr stört. Auch hierbei kann unter Auswertung aller früher genannten Motivationen die zunächst unmöglich erscheinende Zustimmung erreicht werden. In der Regel wird bei einem gut geführten Gespräch das mitlaufende Tonband vergessen“ (115).

Durch die vorstehenden Ausführungen ist auch

„die in Deutschland und überhaupt im kontinentalen Europa vorherrschende Ansicht, Exploration sei eine Kunst, welche man nicht lehren und lernen und damit auch nicht empirisch untersuchen könne“ (Schraml 1964, 869),

widerlegt. Alle praktischen, künstlerischen und wissenschaftlichen Fertigkeiten sind lehrbar und lernbar, wenn der Erfolg natürlich auch je nach den

mitgebrachten Persönlichkeits- und Begabungseigenschaften unterschiedlich sein wird. Das ist für die Explorationstechnik nicht anders als für die Ausbildung in ausübender Musik oder in Mathematik, ohne daß den Wissenschaftscharakter mathematischer Methoden jemals einer aus diesem Grund zu bestreiten sich veranlaßt gesehen hätte (W. Metzger 1942). Hinweise zur Didaktik der Ausbildung finden sich u.a. bei Schraml (1964, 889), Rechetnick & Barkus (1966), Soudijn u.a. (1970), Thorne (1970), Wolfe (1970), Meier (1972), Schuller & Rosemeier (1973), Froehlich & Bishop (1973). Von erfolgreicher Ausbildung für die Durchführung von Explorationen berichtet Kinsey (1948, 61 f.). Ausbildungserfolge in der Auswertung von biographischen Explorationen beweist ein Versuch von U. Lehr (1964), der S. 26 näher geschildert wird. Es waren Explorationsausschnitte, die Monate oder Jahre später gewonnen worden waren, Ausschnitten aus einer Exploration der gleichen Personen zu einem früheren Zeitpunkt und zu anderen Themen zuzuordnen. Die Trefferquote betrug bei 30 Studenten des ersten Semesters der Psychologie 67,6%, bei den in psychologischer Diagnostik ausgebildeten Hauptdiplomkandidaten dagegen 835%.

5. Auswertung

Die methodischen Einwände gegen eine verbreitete Anwendung der Exploration ergeben sich im wesentlichen aus Bedenken gegen den Wahrheitsgehalt (die Ehrlichkeit und die objektive Richtigkeit) der Mitteilungen. Es handelt sich dabei vor allem um zwei prinzipielle Fehlerquellen: Verfälschung und Verheimlichung. Der Tatbestand der Verfälschung liegt vor, wenn der Befragte zwar richtige Angaben machen könnte, sich aber entschließt, stattdessen eine bewußt entstellte Darstellung zu geben, oder etwas tatsächlich Nichtvorhandenes zu behaupten, während der Tatbestand der Verheimlichung vorliegt, wenn er sich entschließt, den erfragten Sachverhalt ganz oder teilweise wahrheitswidrig in Abrede zu stellen. Die Tendenzen zur Verfälschung und zur Verheimlichung spielen naturgemäß eine außerordentlich unterschiedliche Rolle, je nachdem ob die Exploration zu einem neutralen wissenschaftlichen Zweck, bei zugesicherter Anonymität bei der Verwertung der Explorationsergebnisse, geschieht oder in einer Situation des realen Lebens, in der für den Untersuchten elementare vitale Interessen auf dem Spiele stehen. Möglichen Verfälschungs- und Verheimlichungstendenzen muß schon bei der Erhebung der Explorationsbefunde mit explorationstechnischen Maßnahmen entgegengewirkt werden. Dennoch ist es in Situationen, die dem Befragten zu einer absichtlichen Verfälschung oder Verheimlichung Anlaß geben können, erforderlich, den gegebenen Bericht an Hand der in der Aussagepsychologie erarbeiteten Kriterien für den Wahrheitsgehalt von Aussagen (Undeutsch 1967) zu überprüfen (Thomae und Schmidt 1967, 354; Böcher 1968; Kunkel 1978, 96-110). Eine weitere Überprüfungsmöglichkeit besteht im Vergleich der

Mitteilungen mit den Ergebnissen empirischer Forschungen auf dem betreffenden Lebensgebiet. Ein Beispiel dafür war oben schon mit Kinseys Hinweisen für die Exploration von weiblichen Prostituierten („Beischlafdiebstahl“) gegeben worden. Die Überprüfung des Wahrheitsgehaltes der Mitteilungen durch den Vergleich mit den Ergebnissen empirischer Forschungen ist nicht nur während der Exploration erfolgreich einzusetzen, sondern dient auch nachträglich zur Überprüfung des Wahrheitsgehaltes der Explorationsbefunde. Darauf hat ebenfalls schon Kinsey (1948) hingewiesen:

„When 90 to 95 per cent of the persons *in any social level report histories which agree with the patterns shown in Chapter 10, they not only establish the nature of the group Patterns, but establish the validity of their own reports as well“ (129).*

Andere Beispiele gibt Kunkel (1978, 65-89, 99-102, 104f.): Die Behauptung eines Kraftfahrers z.B., die abgeurteilte Trunkenheitsfahrt sei die erste seines Lebens gewesen, ist an sich schon unwahrscheinlich wegen der hohen Dunkelziffer von Trunkenheitsfahrten (1:400). Sie wird noch unwahrscheinlicher, wenn man in Rechnung zieht, daß die BAK sehr hoch war und daß er nichtsdestoweniger eine längere Strecke unauffällig gefahren war und auch selbst bekundet, er habe sich voll fahrtüchtig gefühlt. Da auch das Fahren unter Alkoholeinfluß gelerntes Verhalten ist, muß ein Fahrer bereits häufiger unter Alkoholeinfluß gefahren sein, bis es ihm gelingt, eine längere Strecke mit einer so hohen BAK unauffällig zurückzulegen. Beispiele für diese Art der Überprüfung von Explorationsangaben lassen sich in großer Zahl aus fast allen Lebensbereichen finden.

Für die Auswertung von Explorationsdaten für die persönlichkeitspsychologische Forschung hat Thomae (1968) eine umfassende und größtenteils bahnbrechende Anleitung gegeben. über die oben bereits erwähnte Einstufung des explorativ erhobenen biographischen Geschehens unter mehreren formalen Gesichtspunkten (1968, 124-216) hinaus hat Thomae Listen von Dimensionen entwickelt zur Kennzeichnung der inhaltlichen Aspekte (unter Gesichtspunkten der Expansion, des strukturellen Aufbaus und der qualitativen Dimensionen) des subjektiven Lebensraumes (S. 223-256. Tab. 17-20) und des Selbst (S. 256-282, Tab. 22-28). Er hat weiter gezeigt, daß explorativ gewonnenes biographisches Material sieben fundamentalen thematischen Einheiten zugeordnet werden kann (292-328) und daß aus diesem Material die instrumentellen Einheiten oder Aspekte des personalen Geschehens, die er als „Daseinstechniken“ bezeichnet, herausgearbeitet werden können. „Daseinstechniken“ sind die persönlichkeitspezifischen Arten und Weisen, wie sich das Individuum sein Leben innerlich wie äußerlich „möglich“ bzw. „erträglich“ zu machen sucht. Thomae hat über 20 Typen oder Klassen solcher instrumentellen Einheiten herausgearbeitet, die den verschiedenen Ebenen einer biographischen Analyse (Tagesläufen und den größeren biographischen Abschnitten) gemeinsam sind, und diese zu 5 Typen oder Grundklassen funda-

mentaler instrumenteller Einheiten zusammengefaßt (349-366). Bei Anwendung dieser Auswertungskategorien kann das in Explorationen gewonnene Material wichtige und einzigartige Beiträge zur Entwicklungsforschung, zur Persönlichkeitsforschung und zur Individualdiagnose der untersuchten Menschen liefern.

Die Verarbeitung der Explorationsbefunde zu einer Diagnose oder Prognose setzt weiter voraus, daß der diagnostische oder prognostische Wert des Befundes - für sich allein oder in Verbindung mit anderen Befunden - aus empirischen Untersuchungen bekannt ist. Ein mustergültiges Beispiel für die Verarbeitung der prognostischen Bedeutsamkeit von biographischen Daten und ihrer Gewichtung für die Erstellung einer Prognose hat Kunkel für die „Rückfallprognose bei Trunkenheitstätern im Straßenverkehr“ (1977) gegeben.

6. Leistungsfähigkeit der explorativen Methoden

Für die Bemessung der Leistungsfähigkeit messender diagnostischer Verfahren wird in der klassischen Testtheorie unterschieden zwischen der Leistungsfähigkeit des betreffenden Verfahrens als Meßinstrument (Meßgenauigkeit in bezug auf eine gegebene Population = Reliabilität) und seiner Leistungsfähigkeit für den vorgesehenen Zweck (Vorhersagegenauigkeit, Tauglichkeit = Validität). Die übliche Behandlung der Objektivität als eines weiteren Gütekriteriums (z.B. Lienert 1969, 13) ist gewiß verfehlt; Objektivität ist nur eine Komponente der Reliabilität, genauer: eine von mehreren Arten reliabilitätssenkender Fehler (Ekman 1947, 158f.).

Es gibt eine unübersehbar große Zahl von Arbeiten, in denen diese Testparameter für einzelne explorative Techniken oder für einzelne Explorationsdaten berechnet worden sind. Es bedarf zuvor aber einer kritischen Besinnung darauf, ob die Leistungsfähigkeit eines so andersartigen methodischen Instruments, wie es die explorativen Techniken darstellen, durch diese Maßwerte überhaupt sinnvoll gekennzeichnet werden kann. Es darf nicht aus dem Auge verloren werden, daß die klassische Testtheorie, der die Begriffe Reliabilität und Validität entstammen, zur Bestimmung der Leistungsfähigkeit von messenden Verfahren entwickelt worden ist, d.h. von Verfahren, die in ihrer Durchführung und Auswertung vollständig oder weitgehend standardisiert sind, die in einer vom „natürlichen“ Leben abgehobenen, künstlichen, gleichartig und konstant gehaltenen Situation durchgeführt werden und die sich darauf beschränken, einen mehr oder minder klar umgrenzten Persönlichkeitsbereich oder sogar nur eine spezielle Begabungskomponente zu erfassen. Grundlegend für die klassische Testtheorie ist die Voraussetzung der Verfügbarkeit „äquivalenter“, d.h. vergleichbarer Messungen derselben Eigenschaft. In allen genannten Hinsichten sind die explorativen Techniken nicht nur an-

ders, sondern geradezu entgegengesetzt beschaffen. Schon die projektiven diagnostischen Verfahren haben sich für eine Bemessung ihrer diagnostischen Leistungsfähigkeit nach den Kriterien der klassischen Testtheorie als weitgehend unzugänglich erwiesen. Die unter Anwendung dieser testtheoretischen Verfahren bei den projektiven Tests erzielten Ergebnisse sind bekanntlich allesamt sehr unbefriedigend geblieben. Der Grund dafür ist ein doppelter: Zum einen ist eine Berechnung der Kennzahlen der klassischen Testtheorie nur sinnvoll, wenn die Befähigung des Psychologen zur Durchführung und zur Auswertung dieser Verfahren von völlig untergeordneter Bedeutung ist, so daß das Verfahren von jedem Psychologen „gleich gut“ angewandt werden kann - oder es für die Durchführung und die Auswertung überhaupt nicht einmal eines Psychologen bedarf. Wenn nämlich zwischen den Psychologen erhebliche persönliche Unterschiede hinsichtlich ihrer Qualifikation (auf Grund von Begabung, Ausbildung, Übung und Erfahrung) für die Durchführung und die Auswertung eines Verfahrens bestehen, so kann jede Maßzahl für die Leistungsfähigkeit eines Verfahrens nur einen Durchschnittswert aus den Leistungen von Testanwendern der allerverschiedensten Qualitätsstufen darstellen. Es besagt aber selbstverständlich nichts gegen den Wert eines Verfahrens, wenn sich bei Überprüfungen seiner Leistungsfähigkeit herausstellt, daß einige damit hervorragend zu arbeiten verstehen, während viele andere nur sehr mäßige Erfolge damit erzielen und einige andere sogar zu vorwiegend falschen Ergebnissen damit gelangen. Der Validitätskoeffizient ist in solchen Fällen nur ein nichtssagender Mittelwert - nichtssagend, weil er die viel belangreichere Tatsache überdeckt, daß es Testanwender gibt, die mit dem Test hervorragende Ergebnisse zu erzielen verstehen, während es gleichzeitig andere gibt, in deren Händen die Testanwendung lediglich Unfug ist. Ein Beispiel dafür ist die Untersuchung von Magnusson (1959) zur Validität des TAT:

An 63 männlichen Studenten aus einem Hochschulinternat wurde der TAT durchgeführt. Die Testergebnisse wurden von 4 Psychologen ausgewertet, die ihre Kompetenz für TAT-Auswertungen erklärt hatten. Die gleichen Studenten wurden von je 10 bis 18 Kommilitonen, die seit 2 Jahren im Internat mit ihnen zusammenlebten, in 19 Persönlichkeitsvariablen eingestuft. Sowohl die Psychologen als auch die Kommilitonen hatten für die Einstufung bei allen 19 Variablen eine 7stufige Skala zu verwenden. Für jedes Individuum wurde für jede Persönlichkeitsvariable der Mittelwert der Einstufungen durch seine Kommilitonen berechnet. Diese Mittelwerte bildeten die Kriterienvariable. Die Einstufungen der 4 Psychologen wurden mit den Kriterienvariablen korreliert. Diese Korrelationskoeffizienten sind Validitätskoeffizienten.

Die Validitätskoeffizienten fielen für die 4 Psychologen sehr unterschiedlich aus. Bei dem Psychologen B trat kein einziger negativer Korrelationskoeffizient auf, von seinen positiven Korrelationen waren 8 auf dem 10%-Niveau oder einem noch höheren Niveau signifikant. Der Psychologe A hatte hingegen ebenso viele positive wie negative Koeffizienten. Die Verteilung der von ihm erreichten Validitätskoeffizienten entspricht genau einer Zufallsverteilung (89-91, 105-107).

Für diagnostische Verfahren dieser Art gilt der Satz von Vernon (1957):

„... it is hardly possible to dissociate the test from the tester. One clinical psychologist does well with interviewing or with Rorschach, another with Thematic Apperception or drawings, another with deterioration tests or expressive movements and so on“ (205).

Ein anderes sehr eindrucksvolles Beispiel ist die Beurteilung des intellektuellen Niveaus mit Hilfe der graphologischen Diagnostik.

In einer Untersuchung von Michel (1969), an der 7 Diplom-Psychologen, die zusätzlich eine gründliche graphologische Ausbildung genossen hatten und über ausreichende praktische Erfahrung auf diesem Gebiet verfügten, teilgenommen hatten, betrug der Korrelationskoeffizient zwischen den graphologischen Schätzungen des Intelligenzniveaus und den Gesamtstandardwerten des IST, die als Kriterienvariable genommen worden sind, beim besten Graphologen + .57, beim schlechtesten - .29. Dieser Befund ist interessant, denn er besagt, daß der eine Graphologe recht gut zur Einschätzung des intellektuellen Niveaus auf Grund der Handschrift in der Lage war, während am anderen Ende der Skala ein Graphologe steht, dessen Urteil in der Mehrzahl der Fälle in der verkehrten Richtung lag. Was besagt demgegenüber der durchschnittliche Validitätskoeffizient von .16? Er verdeckt das wahre Ergebnis der Untersuchung.

Die Frage nach der Reliabilität und der Validität eines Verfahrens ist überhaupt nur sinnvoll, wenn es sich um ein standardisiertes Verfahren handelt, während die explorativen Techniken qualitativ (dem Modus nach) eine außerordentliche Spielbreite haben. Ein Verfahren, das von Mensch zu Mensch von Mal zu Mal wechselt, kann einer statistischen Überprüfung im klassischen Sinne überhaupt nicht unterzogen werden, weil es keine feste und gleichbleibende Gestalt hat. „Every interview is a unique and unreproducible encounter“ (Lopez, 1965, 8). Auch ist bei keinem anderen diagnostischen Verfahren der Verwendungszweck so vielfältig. So vielfältig wie die Verwendungszwecke, sind natürlich auch die Validitätskoeffizienten. Selbst für den Teilbereich der eignungsdiagnostischen Exploration betonen Bolton & Hikey (1969):

„Interviews are not generally predictive; that is, they are not generally valid. Rather their validity must be determined in a given situation, for particular positions, and following specified procedures“ (501).

Ein anderer Grund für die Schwierigkeit der Anwendung der Maßstäbe der klassischen Testtheorie auf projektive Verfahren ist, daß Voraussetzung jeder Validitätsbestimmung ist, daß eine vom zu prüfenden diagnostischen Verfahren unabhängige, ihrerseits reliable und mit dem angezielten psychologischen Konstrukt möglichst identische oder wenigstens hochkorrelierende Kriterienvariable zur Verfügung steht. Daran gebriecht es schon bei projektiven Tests in aller Regel, denn je komplexer das diagnostische Verfahren ist, um so komplexer und weniger klar umgrenzt ist der Komplex der psychologischen Konstrukte, die damit erfaßt werden können, wobei das gleiche diagnostische

Verfahren bei verschiedenen Menschen sogar verschiedene Ausschnitte aus dem Gesamt ihrer personalen Struktur erfassen kann, um so unmöglicher, für den Komplex der erfaßten überdauernden personalen Sachverhalte eine KriterienvARIABLE zu finden oder einen für alle untersuchten Individuen gleichen Komplex von solchen. Ein Vergleich mit Einzelkriterien kann die Leistungsfähigkeit der Exploration prinzipiell nicht in Erscheinung treten lassen. Mit Recht betont Wellek (1958):

„Ein ‚ganzheitliches‘ Verfahren wie die (echte) Exploration kann der Natur der Sache nach nur überlegen, ja sogar nur brauchbar sein, wenn das Erkenntnisziel - der gefragte Gegenstand - gleichfalls ein ganzheitliches ist“ (27).

Daraus wird deutlich, daß die Maßstäbe der klassischen Testtheorie auf Verfahren, die einerseits Ansprüche an die Qualifikation des Untersuchers und Auswerters stellen und andererseits selbst sehr komplex, vielgestaltig, von Fall zu Fall wechselnd sind, entweder überhaupt nicht oder allenfalls unter vielen Vorbehalten und mit nur sehr beschränkter Aussagekraft anwendbar sind.

Gilt das alles schon ganz allgemein für alle in der Durchführung und Auswertung nur wenig standardisierten Verfahren, so gilt es in höchstem Maße für die Exploration, die einer weitgehenden Standardisierung nicht unterworfen werden kann, ohne gleichzeitig denaturiert zu werden. Man muß daher gerade im Falle der Exploration nach anderen Bewährungskriterien Ausschau halten. Maßwerte für Reliabilität und Validität können nur randständige Bedeutung haben und sind wegen der Vielgestaltigkeit der explorativen Techniken und der Vielzahl der Verwendungszwecke ohnehin nur für den speziellen Fall, für den sie ermittelt worden sind, aussagefähig.

6.1 Reliabilität

Für die Bestimmung der Reliabilität ist Voraussetzung, daß „äquivalente“, d.h. vergleichbare Messungen derselben Eigenschaft zur Verfügung stehen. Es stehen vier empirische Methoden zu ihrer Bestimmung zur Verfügung: die Testwiederholung, der Paralleltestvergleich, die Testaufspaltung und die Populationsaufspaltung nach Husen (1949, 62-70, 75), wovon die letztgenannte im Fach unbekannt geblieben ist und deshalb auch keine Anwendung erfahren hat. Alle diese Methoden sind auf explorative Verfahren nicht anwendbar.

Die Testwiederholung verlangt, daß derselbe Test zu einem späteren Zeitpunkt nochmals unter genau den gleichen Bedingungen angewandt wird. Dergleichen wäre nur annäherungsweise bei einem streng standardisierten Interview, aber nicht bei einer aus dem unmittelbaren zwischenmenschlichen Kontakt lebenden Exploration möglich, weil jede Exploration in einem lebendigen Austausch von Fragen und Mitteilungen besteht. Kein Mensch antwortet aber

wie eine Grammophonplatte in der zweiten Exploration genauso wie in der ersten. Jede Änderung der Art und Weise des Antwortens hat aber eine Änderung der Fragenzahl und der Fragenformulierung des Explorators zur Folge. Deshalb kann niemals der gleiche Satz von Fragen in der gleichen Reihenfolge und in der gleichen Formulierung unter gleichen Bedingungen wiederholt werden.

Die Paralleltestmethode setzt voraus, daß ein Paralleltest vorhanden ist, der aus äquivalenten Items besteht. Auf die Exploration angewandt, müßte entweder der gleiche Explorator zwei Sätze äquivalenter Fragen haben oder der gleiche Fragensatz müßte von zwei Exploratoren den gleichen Personen vorgelegt werden. Das erste geht nicht, weil es überhaupt keine Methode gibt zur Überprüfung der Parallelität der beiden Fragensätze. Das zweite geht nicht, weil die gleichen Personen nicht annähernd zum gleichen Zeitpunkt von einem anderen Explorator in der gleichen Weise (hinsichtlich des Zwecks, der Dauer, der Zahl und der Art der behandelten Themen, der Reihenfolge der Fragen, der Fragenformulierung, der Art des zwischenmenschlichen Kontaktes usw.) exploriert werden können. Vielmehr werden verschiedene Exploratoren unvermeidlich persönlichkeitspezifisch verschieden explorieren und die Explorierten auf verschiedene Exploratoren verschieden reagieren, wie schon Nietzsche wußte (Menschliches, Allzumenschliches, Aph. 374) und hernach Klages (1926) weiter ausgeführt hat:

„Mit wieviel Menschen einer in Berührung zu kommen pflegt, über ebensoviel verschiedene Physiognomien verfügt seine Seele. Wir wollen uns an Beispielen verständlich machen. - Die Art und Weise, wie man auf eine und dieselbe Frage Antwort gibt, hängt ganz wesentlich von demjenigen ab, der die Frage stellt. Einen und denselben Vorgang erzählen wir diesem Zuhörer nicht mit genau den nämlichen Worten als jenem“ (19).

Die Exploration des einen ist naturnotwendig anders als die des anderen. Es liegen also gerade keine „Parallelen Tests“ vor, d.h. Tests, die sich gleichen „wie ein Ei dem anderen“ - ganz abgesehen davon, daß sich die testtheoretischen Kennwerte für Parallelität von Tests

$$(M_{X_1} = M_{X_2}, s_{x_1}^2 = s_{x_2}^2, r_{x_1x_2} = r_{x_3x_4}, r_{X_1Y} = r_{X_2Y})$$

für Explorationen jedes Sinnes entbehren und sich übrigens auch gar nicht berechnen lassen.

Es hat nicht an Versuchen gefehlt, die Vielzahl der reliabilitätssenkenden Fehler, die bei den explorativen Methoden eine Rolle spielen können, in einer komplizierten Formel zu vereinigen, um auf diese Weise trotz der hohen Komplexität der explorativen Methoden doch noch die Reliabilität berechnen zu können. Für Fälle, in denen die Antworten der Befragten quantifiziert

werden können durch Zuordnung einer Maßzahl (= Rohwert), sieht Fleiss (1970) die Rohwerte als zusammengesetzt an aus folgenden Parametern

- μ = eine additive Konstante,
- α = die Auswirkung des personspezifischen Explorationsstils des Explorators auf die Antworten des Explorierten und dadurch auf die Bewertung der Antworten,
- β = die Kriterien, die der Explorator üblicherweise bei der Bewertung der Antworten anwendet,
- γ = die resultierende Wirkung des Kompromisses, den der Explorator zu schließen hat zwischen den Anforderungen der Exploration und den Anforderungen der Bewertung der Antworten (für eine genaue Auswertung der Exploration ist es wichtig, daß die Antworten des Explorierten vollständig mitgeschrieben werden. Der Exploration hingegen bekommen die durch das Nachschreiben bedingten längeren Sprechpausen schlecht),
- d = die Zufallsverteilung der individuellen Eigenarten der Explorierten,
- e = ein Zufallseinfluß bei der Quantifizierung

und folgenden Interaktionsausdrücken:

- αd = die Interaktion zwischen Explorator und Exploriertem, die sich einerseits in der Weise ausdrücken kann, daß der Explorator im Hinblick auf die Besonderheiten des Explorierten von seinem habituellen Explorationsstil abweicht, und andererseits in der Weise, daß der Explorierte selbst bei gleicher Fragestellung gegenüber einem Explorator anders antwortet als beim anderen,
- βd = die Interaktion, derzufolge der Explorator möglicherweise bei verschiedenen Explorierten verschiedene Bewertungskriterien anwendet,
- γd = die unterschiedliche Gestaltung des Kompromisses im Hinblick auf verschiedene Explorierte.

Daraus ergibt sich, daß der im Falle einer Quantifizierung eines Explorationsbefundes gegebene Wert X sich wie folgt zusammensetzt

$$X = \mu + \alpha + \beta + \gamma + d + (\alpha d) + (\beta d) + (\gamma d) + e.$$

Es ist offensichtlich, daß die einzelnen testtheoretischen Parameter und Interaktionsausdrücke für die konkreten Einzelfälle nur schwer mit Zahlen ausgefüllt werden könnten. Und wenn dies möglich wäre, wäre eine solche Berechnung auch nur für explorative Techniken möglich, die in Zahlenwerte (scores) ausmünden. Und selbst bei solchen explorativen Techniken wäre der dann gewonnene Zahlenwert von höchst fragwürdiger Bedeutung, denn er vereinigt die Leistungen „guter“ und „schlechter“ Explorationsführer und die Antworten ergiebiger und unergiebigiger, selbständiger und beeinflussbarer, aufrichtiger und unaufrichtiger usw. Explorationsteilnehmer. Es ist darum auch nicht verwunderlich, wenn in Sammelreferaten zum Wert der Exploration für die Berufseignungsdiagnostik (Wagner 1949, Ulrich und Trumbo 1965) immer wieder festgestellt wird, daß nur in wenigen Untersuchungen Reliabilitätskoeffizienten mitgeteilt werden. Bei den von Wagner referierten 106 Untersuchun-

gen enthielten nur 25 überhaupt quantitative Angaben. Die mitgeteilten Reliabilitätskoeffizienten lagen in Fällen der Beurteilung umschriebener Merkmale zwischen .23 und .97, für eine pauschale Eignungsbeurteilung zwischen .20 und .85. Mayfield (1964) kommt in einem Sammelreferat, ebenfalls über die Bedeutung der Exploration bei der Bewerberauslese für berufliche Positionen, zu dem Ergebnis, daß allgemeine Beurteilungen der Eignung eines Bewerbers auf der Basis unstrukturierter Explorationen und ohne weitere Vorinformationen eine extrem geringe Inter-Rater-Reliabilität aufwiesen und daß die gleichen Explorationsdaten von verschiedenen Beurteilern in unterschiedlicher, in einigen Fällen sogar in entgegengesetzter Weise interpretiert und gewichtet werden. Dieses Ergebnis sollte indessen nicht als Hinweis auf eine geringe Reliabilität der Exploration genommen werden, sowenig wie es üblich ist, die Tatsache, daß verschiedene Kliniker bei der Beurteilung von MMPI-Profilen nur geringe Übereinstimmung zeigen, gegen die Reliabilität des MMPI als Untersuchungsinstrument ins Feld zu führen. Denn natürlich kann man auch die mit anerkanntermaßen hochreliablen Meßverfahren gewonnenen Befunde diagnostisch falsch verwerten.

Es gibt andere methodische Möglichkeiten, sich von der Reliabilität (wenn man schon bei diesem Begriff bleiben will) der Exploration ein Bild zu machen. U. Lehr (1964) hat einen Versuch unternommen, der als eine der Natur der Exploration gemäße Abwandlung der Testwiederholungsmethode angesehen werden kann.

Das Versuchsmaterial bestand aus den Explorationen von 3 weiblichen Personen. Diese Personen

- waren zur Zeit der Exploration alle gleich alt: 29 Jahre;
- alle hatten die gleiche Schule besucht und im gleichen Jahr Abitur gemacht.
- Ebenso erhielten alle drei eine eingehende Berufsausbildung, gaben jedoch mit dem Zeitpunkt der Heirat die Berufstätigkeit auf;
- alle hatten Kinder.

Aus den Explorationen dieser 3 Frauen wurden je 3 Ausschnitte maschinenschriftlich übertragen, die sich jeweils bezogen

- auf die frühe Kindheit bis zum ersten Schultag,
- auf die Situation der Berufswahl,
- auf die Situation der Partnerwahl.

Die zur gleichen Person gehörenden Explorationsausschnitte waren mit der gleichen Signatur kenntlich gemacht.

Außerdem erhielten die Versuchspersonen die maschinenschriftlichen Übertragungen von Äußerungen der gleichen Personen, die einige Wochen, zum Teil ein Jahr später erhoben worden waren als der Lebenslauf. Diese Äußerungen bezogen sich

- auf soziale Einstellungen,
- auf die Bewertung des bisherigen Lebens und auf die Zukunftsorientierung.

Diese Äußerungen waren nicht signiert.

Diese Materialien wurden einer Gruppe von **30** diagnostisch geschulten Hauptdiplom-Kandidaten übergeben mit der Aufgabe, diese Ausschnitte jenen 3 Personen zuzuordnen, von denen sie Ausschnitte aus der früheren Exploration erhalten hatten. Sowohl die Auszüge aus den Explorationen zum Lebenslauf als auch die später erhobenen Äußerungen waren von allen Informationen gereinigt worden, die als Hinweise auf die Identität der Personen hätten dienen können. Von den 630 vorgenommenen Zuordnungen waren **515** = 83,5% richtig.

Die in der Exploration ersichtlich werdende Verhaltensgestalt kann somit als ein Spiegelbild bestimmter Verhaltenstendenzen, spezifischer Grade, Formen und Richtungen der Aktivität, bestimmter Interessenbevorzugungen und Neigungen angesehen werden. Da der Gegenstand der später aufgenommenen Äußerungen völlig von dem der Exploration des Lebenslaufes abwich, manifestiert sich offensichtlich in den Explorationsausschnitten unabhängig von dem konkreten Inhalt eine durchgängige persönlichkeitspezifische Ausprägung dieser Tendenzen so eindeutig, daß eine Zuordnung von solchen Äußerungen zu anderen, die zeitlich und inhaltlich von jenen deutlichen Abstand zeigen, möglich ist. Thomae (1968) interpretiert dieses Versuchsergebnis folgendermaßen:

„Die Exploration liefert also, wenn sie kunstgerecht und in einer für den Explorierten akzeptablen Motivationslage durchgeführt wurde, kein ad hoc und kein bewußt oder willkürlich zurechtgemachtes Material. Sie ‚entfaltet‘ vielmehr zumindest einen gewissen Ausschnitt aus den Verhaltensweisen des Individuums, den Situationen, so wie sie das Individuum erlebt, und den Zielsetzungen, von denen diese Verhaltensweisen her verstanden werden müssen“ (119).

Sehr gründlich hat sich bezüglich der Zuverlässigkeit der durch Explorationen erhobenen Daten Kinsey vergewissert. Er hatte den einzigartigen Vorteil, ein sehr großes Explorationsmaterial zur Verfügung zu haben (im Laufe der Jahre 1938 bis 1947 hatten er und seine 3 Mitarbeiter 12214 Explorationen durchgeführt), das zu mancherlei Kontrolluntersuchungen Gelegenheit gab.

a) 162 Personen wurden nach Ablauf einer längeren Zeit (18 Monate bis 7 Jahre) einer Zweitbefragung unterzogen. Die Ergebnisse sind in Tab. 13 (122f.) zusammengestellt. Es zeigte sich, daß eine hohe Wiederholungsübereinstimmung bestand hinsichtlich der Angaben darüber, ob eine bestimmte Form sexueller Aktivität vom Befragten jemals ausgeübt worden war. Die Wiederholungskoeffizienten lagen für alle diesbezüglichen Fragen über .90 und in allen bis auf 3 Fälle über .95. Hohe Wiederholungsübereinstimmung besteht ebenfalls für die mehr äußeren biographischen Daten, bei denen die Korrelationskoeffizienten in jedem Fall über .80 lagen, in 6 von 8 Fällen über .90. Deutlich niedriger waren hingegen die Angaben über das Alter bei der ersten Erfahrung mit einzelnen Formen sexueller Betätigung ($r = .5 - .8$). Dennoch sind die Differenzen nicht groß. Sie betragen im allgemeinen 5% oder weniger des arithmetischen Mittels. Am niedrigsten

sind die Koeffizienten bei denjenigen Erlebnissen, die keinen abgrenzbaren Ereignischarakter haben: kindliche sexuelle Spieltätigkeit, unwillkürliche nächtliche Ejakulationen, heterosexuelles Petting. Die Angaben über die durchschnittliche Häufigkeit, mit der die einzelnen Betätigungsformen während eines bestimmten Zeitraumes ausgeübt worden sind, korrelierten miteinander zwischen .58 und .67.

- b) Eine weitere Überprüfungsmöglichkeit ergibt sich aus dem Vergleich der Angaben von Ehepartnern. Kinsey hat solche Vergleiche für 231 Ehepaare vorgenommen (Tab. 14, S. 126). Verglichen wurden deren Angaben zu insgesamt 32 Einzelpunkten. Bei 3/4 der erfragten Einzelpunkte liegen die Übereinstimmungskoeffizienten über .70, bei der Hälfte über .80 und bei 1/4 über .90. In der Hälfte der Einzelpunkte besteht zwischen den Angaben der beiden Ehepartner eine nahezu vollständige Übereinstimmung in 90 bis 100% der Berichte. Diese Übereinstimmung ist erstaunlich, wenn man in Betracht zieht, daß zwischen der Befragung der beiden Ehepartner Zeiträume von 2 bis 6 Jahren (und mehr) lagen.
- c) Die Explorationen für den Kinsey-Bericht sind überwiegend von 3 Wissenschaftlern ausgeführt worden. Das gab die Möglichkeit, die Frage zu untersuchen, ob verschiedene Exploratoren gleiche Resultate zu erreichen pflegen. Es wurden zu diesem Zweck verglichen die Ergebnisse, die sie bei gleichartigen (parallelen) Gruppen von Befragten gewonnen hatten. Die Gruppen waren homogen unter den Gesichtspunkten des Geschlechts, der Rasse, des Familienstandes, des Alters, des Bildungsgrades, der Stadt-/Land-Zugehörigkeit, der Konfession. Es wurden Vergleiche durchgeführt für die Gruppen, in denen jeder Untersucher mehr als 300 Fälle exploriert hatte. In Tab. 16 (S. 134) sind 75 Vergleichswerte wiedergegeben. 35 davon
- „are so similar that the differences are immaterial - closer than any person could calculate about his own history“ (135),

wobei wiederum die Zahlen über die Häufigkeit, mit der die einzelnen Formen sexueller Betätigung praktiziert worden sind, eine etwas geringere Übereinstimmung aufwiesen als die Zahlen über die Verbreitung. Das Ergebnis dieser Vergleiche ist:

„There seems no reason to doubt that any other group of investigators could duplicate these results if their scientific objectivity and their methods in interviewing were comparable to those used in the present study“ (135).

- d) Kinsey, der persönlich im Laufe eines knappen Jahrzehnts 7036 Explorationen durchgeführt hatte, hatte noch eine weitere Möglichkeit zur Überprüfung der Reliabilität seiner Methode im Laufe der Zeit: Er verglich seine Explorationsergebnisse aus den ersten 4 Jahren (1938-1942) mit denen aus den letzten 4 Jahren (1943-1946) der Erhebungszeit, bildete unter den Gesichtspunkten des Geschlechts, der Rasse, des Familienstandes, des Al-

ters und des Bildungsgrades homogene Gruppen und verglich die Mittelwerte aller Gruppen, die mehr als 300 Personen umfaßten, aus der ersten Hälfte der Erhebungszeit mit gleichartigen Gruppen aus der zweiten Hälfte. Die Ergebnisse sind zusammengestellt in den Tabellen 21 (S. 142), 22 (S. 144) und 23 (S. 146). Die Mittelwerte aus beiden Erhebungszeiträumen sind bei allen Gruppen nahezu identisch.

„The comparisons in Tables **21-23** seem to indicate that methods of securing subjects, proficiency in interviewing, skill in using the code in which the data are recorded, and calculations and judgments which the data undergo in their statistical treatment, can be maintained at such uniform levels as many persons would have considered impossible in a case history study which is liable to error from so many sources, and which deals with as taboo a subject as sex“ (**147**).

6.2 Validität

Die Validitätsüberprüfung kann sich auf drei Stadien der Anwendung eines Erhebungsverfahrens beziehen: die Datenerhebung, die Auswertung der erhobenen Befunde (Kategorisierung, Quantifizierung, Evaluation, Scoring) und ihre Verarbeitung für eine Diagnose bzw. Prognose.

6.2.1 Die Validität der Datenerhebung

Validität der erhobenen Daten ist das Maß, in dem die von den Befragten gemachten Angaben mit einem (objektiven und seinerseits reliablen) Kriterium für den explorativ erhobenen Sachverhalt übereinstimmen.

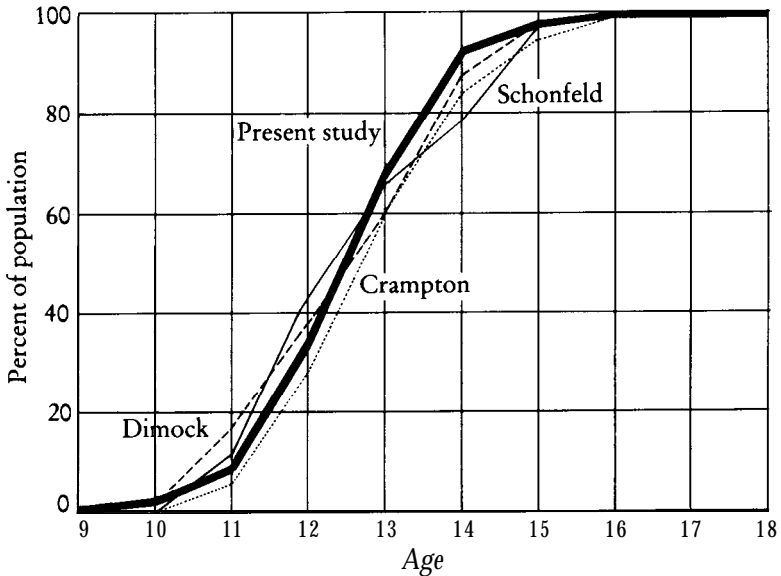
Die Leistungsfähigkeit der explorativen Verfahren bei der Datenerhebung ist selbstverständlich unterschiedlich für verschiedene Arten von Daten.

Die Leistungsfähigkeit der Exploration für die Erhebung von Daten der körperlichen Reifung in der Pubeszenz wurde von Kinsey (1948) untersucht durch einen Vergleich der explorativ erhobenen Daten mit somatoskopischen Beobachtungen. Es sei hier als Beispiel herausgegriffen das von den befragten erwachsenen Männern angegebene Alter des ersten Auftretens der Pubesbehaarung. Die sich daraus ergebende Altersverteilung wurde verglichen mit den somatoskopisch ermittelten Altersverteilungen des gleichen Merkmals, die in jenen Zeiträumen erhoben worden sind, in denen die von Kinsey befragten Männer die Pubeszenzentwicklung durchliefen. Eine frappierende Übereinstimmung ergab sich sowohl hinsichtlich der Mittelwerte (Tab. 15, S. 130) als auch bezüglich der Verteilungen (Abb. 15, S. 131).

Tabelle 1: Vergleich der Mittelwerte des ersten Auftretens von Pubesbehaarung nach den Befragungsergebnissen von Kinsey (1948) mit den Ergebnissen somatoskopischer Beobachtungen von Crampton (1908), Dimock (1937) und Schonfeld (1943)

	Crampton	Dimock	Schonfeld	Kinsey
N	3835	1406	1475	2511
M	13,44	13,08		13,45
Z			13,17	13,43
s	1,51			0,03

Auszug aus Kinsey u.a. (1948, Tab. 15, S. 130).



Wiedergabe von Kinseys (1948) Abb. 15 (S. 131).

Abb. 1: Vergleich der Altersverteilung des ersten Auftretens von Pubesbehaarung nach den Befragungsergebnissen von Kinsey (1948) mit den Ergebnissen somatoskopischer Beobachtungen von Crampton (1908), Dimock (1937) und Schonfeld (1943)

Diese gute Übereinstimmung ist deshalb so erstaunlich, weil das Alter des ersten Auftretens der Pubesbehaarung sicher zu den nicht gerade leicht erinnerbaren Ereignissen der persönlichen Entwicklung gehört.

Bezüglich des Alkoholkonsums führte Schmidt (1973) einen Vergleich von Befragungsergebnissen mit Verkaufsdaten (Kanada) durch, wobei er eine sehr gute Übereinstimmung bei niedrigen und bei mittleren Konsumquanten feststellte. Hingegen beträgt die am stärksten konsumierende Gruppe (= Personen, die 11 und mehr Flaschen alkoholischer Getränke kaufen) auf Grund der Befragungsergebnisse nur 1,8%, nach den Verkaufsdaten aber 6,54%. Besonders jene Alkoholkonsumenten, die täglich alkoholische Getränke einkaufen, machen besonders niedrige Angaben zum Konsum. Auch Boland und Roizen (1973) fanden sehr gute Übereinstimmungen zwischen Befragungsergebnissen und Verkaufsdaten bei niedrigen Konsummengen. Die Übereinstimmung wurde um so geringer, je höher die Verkaufsdaten waren. Aus diesen Ergebnissen ist zunächst einmal zu schließen, daß starke Trinker dazu tendieren, zu niedrige Konsummengen anzugeben - sei es, daß sie tatsächlich keine Kontrolle über die von ihnen genossenen hohen Alkoholmengen haben, sei es, daß sie ihren hohen Alkoholkonsum zu verschleiern trachten. Ähnliche Verhältnisse sind sicher überall zu erwarten, wo es um das Bekenntnis sozial mißbilligten Verhaltens geht. Hier zeigt sich einer der Gründe dafür, daß es gewiß auch für die explorativen Methoden eine differentielle Validität gibt: ihre Validität ist, selbst bei gleicher Thematik der Befragung, nicht für alle Menschen gleich hoch. Sie ist niedriger für sozial mißbilligte Verhaltensweisen. Dieser Tendenz gilt es, schon während der Befragung durch die bewährten Techniken der „Exploration verheimlichter Sachverhalte“ entgegenzuarbeiten. Der durch die Wirksamkeit dieser Tendenz auftretende Fehler dürfte aber für die Praxis nicht gravierend sein, denn es kommt dadurch eher ein systematischer Fehler zustande (eine speculation à la baisse bei Mengenangaben für sozial mißbilligtes Verhalten, die um so stärker ist, je höher die tatsächlichen Häufigkeiten bzw. Mengen sind), so daß die Rangfolge in den Konsumangaben und damit die hohe Reliabilität erhalten bleiben, wie Boland und Roizen (1973) für den Alkoholgenuß hervorheben.

6.2.2 Validität der Bewertung

Explorationsbefunde müssen für eine ganze Reihe von Fragestellungen in irgendeiner Form bewertet werden: sie müssen quantifiziert oder einer Auswertungskategorie zugeordnet oder irgendwie interpretiert werden.

Eine Untersuchung hierzu ist im Arbeitskreis von Thomae von Fisseni (1974) durchgeführt worden. 237 Frauen in Altersheimen wurden über das Leben im Heim befragt, und zwar einerseits in einer unstrukturierten Exploration, die einem lockeren Themenkatalog folgte, der den Tageslauf der Befragten umschrieb, und andererseits mit einem standardisierten Fragebogen, der mit 50 gezielten Fragen den gleichen Themenkreis erfaßte. Die Auswerter ordneten, unabhängig voneinander, das Explorationsmaterial dem gleichen Ja/Nein-Fra-

gebogen zu, der von den Probanden ausgefüllt worden war. Die Inter-Rater-Reliabilität betrug .73; die Übereinstimmung mit den originalen Fragebogenbeantwortungen betrug 87.

6.2.3 Validität der diagnostischen Verwertung

Die erhobenen und bewerteten Befunde bilden die Basis für die Diagnose. Die diagnostische Validität ist ein Maß für die Richtigkeit der Diagnose oder der Vorhersage. Die bewerteten Befunde sind die Prädiktoren. Voraussetzung für die Richtigkeit der Diagnosenstellung sind einerseits die Erarbeitung einer empirisch begründeten klaren Symptomatologie der zu diagnostizierenden psychologischen Sachverhalte. Eine solche ergibt sich aus empirischen Untersuchungen der Häufigkeit des Vorkommens der als Symptome aufzufassenden Befunde bei Vorliegen des zu diagnostizierenden Sachverhaltes. In mustergültiger Weise hat Kunkel (1977) dieses Vorgehen für die Prognose der Rückfallwahrscheinlichkeit in das Delikt des Führens eines Kraftfahrzeuges in alkoholbedingt fahruntüchtigem Zustand vorgeführt. Man kann diesen Teil der Validität als „symptomatische Bedeutsamkeit“ (Wohlfahrt, 1939) oder als „diagnostische Validität“ bezeichnen. Sodann ist (weitere) Voraussetzung der richtigen Diagnosenstellung, daß der Diagnostiker die „symptomatische Bedeutsamkeit“ der erhobenen Befunde kennt, d.h. sie richtig gewichtet und sie im „semantischen Umfeld“ (K. Bühler, 1933) richtig interpretiert.

In einem unveröffentlichten Versuch von U. Lehr (Thomae 1968, 117f.) waren 46 Studenten der Psychologie Ausschnitte aus der Exploration eines 38jährigen Kaufmanns, dem wegen Betrugs und Unterschlagung die FE entzogen worden war, vom Tonband dargeboten worden, und zwar a) über Kindheit und Jugend des Probanden und b) zu den Straftaten und seinen Auseinandersetzungen mit dem Straßenverkehrsamt. Die Beurteiler hatten die Aufgabe, den Eindruck wiederzugeben, den sie daraus von der Persönlichkeit des Probanden erhielten. Die von mindestens der Hälfte der Beurteiler übereinstimmend gegebenen Charakterisierungen wurden verglichen mit den Aussagen über diese Persönlichkeit, wie sie auf Grund der psychologischen Untersuchung, bei der u.a. angewendet worden waren: Kieler Determinationsgerät, Hawie, Rorschach, TAT, graphologische Analyse, gewonnen wurden. Der Vergleich der Aussagen, die die Beurteiler auf Grund der Exploration über Kindheit und Jugend überwiegend übereinstimmend machten, mit den Befunden aus der persönlichkeitsdiagnostischen Untersuchung ergab bei 24 Aussagen völlige Übereinstimmung, bei 16 Aussagen war eine Bestätigung aus den anderen diagnostischen Verfahren nur teilweise zu erhalten, bei 2 Aussagen widersprachen sich der Eindruck aus der Exploration und die anderen diagnostischen Befunde. Im Vergleich der Aussagen, die die Beurteiler auf Grund der Exploration zur Sache (= seine bisherige Verkehrsbewährung) machten, mit den

Befunden aus der Persönlichkeitsuntersuchung ergaben sich bei 18 Aussagen Übereinstimmungen, für 19 Aussagen waren entsprechende Befunde aus den anderen diagnostischen Verfahren nicht beizubringen, bei 6 Aussagen bestand ein Widerspruch zu den Befunden aus den anderen diagnostischen Verfahren. Thomae 1968 interpretiert das Ergebnis dieses Versuches wie folgt:

„Die in der Exploration gegebene Selbstdarstellung des Probanden ist nicht nur das Resultat von Rationalisierungen, von Vereinfachungen und Verfälschungen. Insbesondere dort, wo konkrete Verhaltensweisen in konkreten Situationen geschildert werden, wird eine psychische Realität erkennbar, die andere Verfahren niemals in dieser Eindeutigkeit aufzuweisen vermögen . . . Soweit aus solchen Schilderungen Verhaltenstendenzen erschlossen werden, lassen sich diese wenigstens zum erheblichen Teil auch durch ‚objektivere‘ Verfahren nachweisen. Bei anderen dürfte die Exploration auf Tendenzen verweisen, die den übrigen Verfahren gar nicht zugänglich werden.

Darüber hinaus zeigt das Ergebnis, daß das Maß der Übereinstimmung zwischen der Verwertung der Explorationsdaten durch Beurteiler und den Testbefunden von dem jeweiligen Inhalt der Mitteilung abhängt“ (118).

Die von Thomae entwickelten Prinzipien einer Analyse der formalen Qualitäten des Tageslaufes wurden in einer (unveröffentlichten) Untersuchung von Olbrich (1967) auf ihren Aussagewert hin überprüft (Thomae 1968). Das Material bildeten 40 in der Exploration gegebene Schilderungen von Tagesläufen an einem „typischen Werktag“ von 60- bis 75jährigen Männern und Frauen (erhoben im Rahmen der VW-Alters-Studie). Zwei Beurteiler stufte, unabhängig voneinander, die Schilderung des Verhaltens der befragten Personen in 8 formalen Kategorien des Verhaltens (Thomae 1968, Tab. 5, S. 130) auf einer 9stufigen Skala ein, ohne sonst irgendeine Kenntnis von den untersuchten Personen zu haben. Zwischen den Werten der 8 hier verwendeten Ausgangsskalen und den übrigen in dieser Untersuchung erhobenen Daten (Gesamtzahl 563) über die Erzähler dieser Tagesläufe ergaben sich bei Anwendung der Rangreihenkorrelation mehr als 300 signifikante bis sehr signifikante Korrelationen. Die Ergebnisse von Olbrich belegen, daß eine methodisch fundierte Erhebung und Auswertung von Daten über Verhaltensweisen in mittleren biographischen Einheiten (Tageslauf) symptomatische Bedeutsamkeit besitzt sowohl für Varianten des Verhaltens in kurz-zeitigen Einheiten (z.B. Leistung beim Determinationsgerät nach Mierke) wie bei größeren Einheiten der Biographie (z.B. Formen der Auseinandersetzung mit einer beruflichen oder familiären Situation).

Die Bedeutung der Exploration für die Eignungsdiagnostik ist sehr schwer konkret nachweisbar, weil in der Praxis in der Regel das Eignungsurteil nicht allein auf Grund von Explorationsbefunden abgegeben wird, sondern auch die äußere Erscheinung des Bewerbers, die Verhaltensbeobachtung, Informationen aus Zeugnissen und Akten, meistens auch Testergebnisse in das abschließende Eignungsurteil eingehen. Bei Bewährungskontrollen der in der Ober-

gutachterstelle zur Beurteilung der Kraftfahreignung des Landes Nordrhein-Westfalen (Leiter: Undeutsch) verwendeten Untersuchungsverfahren war es möglich, einige ausschließlich aus der Exploration gewonnene Daten auf ihren prognostischen Wert hin zu untersuchen. Welzel (1976 und 1982) verfolgte an einer Untersuchungsgruppe von 420 (meist mehrfach) wegen Trunkenheit am Steuer vorbestraften Probanden, die in der Obergutachterstelle untersucht worden waren, deren Verkehrsbewährung in den auf die Untersuchung folgenden Jahren. Von den ausschließlich explorativ erhobenen Daten stehen das „Alter zum Zeitpunkt der ersten Heirat“ und der „größte Alkoholkonsum bei einem Trinkanlaß im letzten Jahr vor der Untersuchung“ in gesicherter Beziehung zu einem erneuten Rückfall in ein Trunkenheitsdelikt beim Führen von Kraftfahrzeugen innerhalb von 3 Jahren nach Wiedererteilung der Fahrerlaubnis ($\Phi = .36$). Lediglich einige aus der Gesamtheit der Untersuchungsergebnisse abgeleitete Persönlichkeitseigenschaften und durch Aktenanalyse erhobene Prädiktoren erreichen höhere oder ähnlich hohe Validitätskoeffizienten.

Nicht mehr exakt faßbar, aber doch immerhin deutlich erkennbar ist die Rolle der Exploration in der Eignungsdiagnostik hervorgetreten in einer Bewährungsuntersuchung von Trankell (1956). Das Material der Bewährungskontrolle waren Bewerber um Einstellung als Flugzeugführer bei der schwedischen Luftwaffe. Die Eignung wurde von den Psychologen auf einer 9stufigen Skala angegeben, wobei sowohl die Testergebnisse als auch der in der Exploration gewonnene Eindruck berücksichtigt wurden. Als Kriterium diente die

Tabelle 2: Korrelationen der Untersuchungsvariablen mit der Endbeurteilung nach Abschluß der fliegerischen Grundausbildung

Variable	1945	1946	1947	1948
Techn. Verständnis		.42	.28	.28
Flugtechn. Orientierung			.25	.05
Allgemeine Flugorientierung		.24	.34	.05
Sterzinger, Qualität,	.09	.18	.27	.09
Sterzinger, Quantität	.00	.22	.31	.17
Formale Intelligenz	.25	.31	.47	.27
Introversion	.16	.01	.06	.21
Sensibilität		.04	.16	.13
Körperliche Gewandtheit			.30	.10
Eignungsurteil	.39	.08	.57	.56
N		96	96	100

Bewertung, die die angenommenen Kandidaten am Ende der fliegerischen Grundausbildung erhielten. Die Ergebnisse sind in Tab. 2 wiedergegeben: Korrelationen der Untersuchungsvariablen mit der Endbeurteilung nach Abschluß der fliegerischen Grundausbildung.

Es zeigt sich, daß die von den Psychologen vorgenommenen Einstufungen mit dem Kriterium - mit Ausnahme des Aufnahmejahrganges 1946 - höher korrelieren als die Testwerte. Die aus der Exploration gewonnenen Zusatzinformationen und der auf Grund der Exploration gebildete Gesamteindruck ermöglichen eine genauere Prognose als irgendeiner der Tests. Eine Ausnahme bildet der Aufnahmejahrgang 1946. Eine Erklärung dafür findet sich, wenn man die Validitätskoeffizienten für die beteiligten Psychologen einzeln berechnet (Tab. 3).

Tabelle 3 : Korrelationen der Eignungsurteile der einzelnen Psychologen mit der Endbeurteilung nach Abschluß der fliegerischen Grundausbildung

	Explorator	N	Eignungsurteil
1946	{ Psychologe A	69	.21
	{ Psychologe B	27	.00
1947	{ Psychologe C	69	.33
	{ Psychologe D	27	.50
1948	{ Psychologe E	39	.36
	{ Psychologe F	61	.49

(Aus: Trankell 1956, Tab. 19, S. 87)

Diese Zusammenstellung offenbart beträchtliche Unterschiede in der prognostizierenden Befähigung der beteiligten Psychologen. Die Eignungsprognosen des Psychologen B haben keinerlei prognostischen Wert, während die Psychologen F und D sehr hohe Validitätskoeffizienten erzielten. Die Leistungsunterschiede zwischen den beteiligten Exploratoren traten noch deutlicher in Erscheinung, wenn als Validitätskriterium das Alternativmerkmal „fliegerische Grundausbildung wegen mangelhaften Ausbildungserfolges abgebrochen ./ erfolgreich abgeschlossen“ verwendet wird. Die individuellen Validitätskoeffizienten der beteiligten Psychologen lagen hier zwischen .40 und der (nicht unbeträchtlichen) negativen Korrelation von -.21 (Tab. 18, S. 87). Die Auslese im Jahre 1946 wurde von den zwei Psychologen A und B vorgenommen, die dazu am wenigsten befähigt waren.

Ein ähnliches Ergebnis erbrachte die Bewährungskontrolle der von der Obergutachterstelle des Landes Nordrhein-Westfalen zur Beurteilung der Kraftfahreignung erstellten psychologischen Gutachten (Deters, 1978). 878 Probanden, die in den Jahren 1966-1972 begutachtet worden waren, wurden über einen Zeitraum von 3.5-7.0 Jahren bezüglich ihrer späteren Verkehrsbewährung weiter verfolgt. In den Untersuchungen der Obergutachterstelle nimmt die Exploration eine zentrale Stellung ein. Sie erstreckt sich sowohl auf die Persönlichkeit des Probanden wie auch auf seine bisherige Verkehrsbewährung und bei Trunkenheitstätern selbstverständlich auch auf die Entwicklung ihrer Trinkgewohnheiten und ihrer Einstellungen zum Problemkreis „Alkohol und Verkehr“. Die durchschnittliche Dauer einer Exploration beträgt hier 2;15 Stunden. Bei der Prognose künftiger Verkehrsbewährung haben die aus der Exploration gewonnenen Befunde ein sehr starkes Gewicht. Es wurde nach verschiedenen statistischen Verfahren die prospektive Validität der in die Begutachtung eingegangenen Variablen ermittelt. Es wurden auch multiple Konstriktionskoeffizienten berechnet, um die optimale Kombination von Variablen für die Prognosestellung kennenzulernen. U. a. wurde eine Trennung der Ergebnisse nach Kombinationen von Prädiktorvariablen mit und ohne Einbeziehung der Variablen „Gutachtenergebnis“ vorgenommen. Es ergab sich, daß die Sechser-Kombinationen, bei denen das Gutachtenergebnis einbezogen worden war,

„fast durchweg numerisch höhere Vorhersageverbesserungen erzielen als die Sechser-Kombinationen, bei denen die Variable ‚Gutachtenergebnis‘ ausgespart worden war“ (114).

Es ergab sich weiter, daß bei allen als optimal gefundenen Prädiktorkombinationen der jeweils beste Einzelprädiktor die Variable „Gutachtenergebnis“ ist (S. 16 und Tab. 6/B 1).

Dieses immer wieder zu beobachtende diagnostische Plus, das die Exploration liefert, hat zwei Gründe. Einerseits erfaßt die Exploration inhaltliche Aspekte der Persönlichkeit, die andere Verfahren niemals in dieser Eindeutigkeit aufzuweisen vermögen. Zum anderen ist sie ein ganzheitliches Verfahren, das es ermöglicht, die in der Exploration und die mit anderweitigen Untersuchungsverfahren gewonnenen Befunde aufeinander zu beziehen und in einer im Hinblick auf die konkrete Persönlichkeit adäquaten Weise zu gewichten. Wellek (1958) hatte schon auf Grund theoretischer Überlegungen darauf hingewiesen, daß die Exploration, je mehr sie „die Persönlichkeit als Ganzes im Auge hat“,

„desto mehr leistet die Methode und desto eher wird sie zum ‚Rückgrat‘ der Diagnose, die dann eben das Ganze, nicht ausgegliederte Sektoren betrifft“ (24).

Noch lebendiger und anschaulicher schildert den unersetzlichen Wert der Exploration Walther (1941) aus den Erfahrungen der ganzheitlich-charakterolo-

gisch ausgerichteten Offiziersbewerber-Eignungsuntersuchungen in der deutschen Wehrmachtpsychologie:

„Die Eignung der Aussprache zur Lösung der verschiedenartigsten diagnostischen Probleme ist nicht von ungefähr. Es würde eine Abhandlung für sich erfordern, das im einzelnen darzulegen. In diesem Zusammenhang mag es genügen, daran zu erinnern, daß wir in der Aussprache die einzigartige Möglichkeit besitzen, jede Reaktion des Prüflings in ihrer Bedeutung für das Ganze der Persönlichkeit genau abzuschätzen. Seine Äußerungen sind nicht (wie etwa schriftliche) von ihm abgelöst, sondern bleiben im Verband nicht nur der zugehörigen Ausdruckserscheinungen und Verhaltensweisen, sondern auch der situativen Faktoren. Schließlich lassen sich die tieferen inneren Zusammenhänge, in denen die Reaktionen und Äußerungen des Prüflings stehen, grundsätzlich bis an die Grenze des Möglichen erkunden. Aus diesem Grunde ist die Exploration die *via regia* der psychologischen Diagnostik und wird es immer bleiben. Das heißt mit anderen Worten auch: es wird gar nicht möglich sein, eine ‚bessere‘ Methode zu finden“ (24).

Auf der anderen Seite gibt es auch niederschmetternde Validitätskoeffizienten aus dem Bereich der Eignungsdiagnostik. Solche finden sich bei Eysenck (1951, 1952) zusammengestellt, aber auch in allen Sammelreferaten über die Leistungsfähigkeit explorativer Techniken in der Eignungsdiagnostik (Wagner 1949, Mayfield 1964, Ulrich & Trumbo 1965, Moffatt 1969, Wright 1969, Triebe 1976). Wegen der vielfältigen Verwendungsmöglichkeit explorativer Techniken ist es ganz selbstverständlich, daß es auch eine Vielzahl, und zwar sehr unterschiedlicher Validitätskoeffizienten geben muß. Niedrige Koeffizienten besagen zunächst einmal, daß Explorationen - im Gegensatz zu voll standardisierten Tests - auch ganz unzulänglich durchgeführt und ausgewertet werden können. Wellek (1958) sagt zu den von Eysenck berichteten niedrigen Validitätskoeffizienten:

„Wie schlecht muß da exploriert worden sein, wenn keinerlei Bewährungserfolg erzielt werden konnte!“ (26).

Auch Thomae (1968) meint, Eysencks

„diesbezüglichen Befunde sprechen zunächst einmal für einen nicht sonderlich hohen Standard der Verwendung von Explorationstechnik und der Techniken und Ziele der Auswertung der auf diese Weise gewonnenen Daten“ (117).

Zum anderen stellt sich das - für alle Validitätsuntersuchungen leidige - Problem der Kriterienvariablen im Falle eines ganzheitlichen Verfahrens, wie es die Exploration in typischer Weise ist, mit besonderer Schärfe. Hierzu hatte Wellek schon 1958 ausgeführt:

„Eine Exploration kann selbst im günstigsten Falle das Ergebnis eines Tests nicht besser voraussagen als dieser selbst. Ist also der der Bewährungskontrolle zugrunde gelegte Maßstab ein Test oder doch testartig, dann ist es ein methodischer Kurzschluß, zu sagen, in der Bewährung komme ein Test besser heraus als eine Exploration, denn das

liegt in der Sache selbst. Ein ‚ganzheitliches‘ Verfahren wie die (echte) Exploration kann der Natur der Sache nach nur überlegen, ja sogar nur brauchbar sein, wenn das Erkenntnisziel - der gefragte Gegenstand - gleichfalls ein ganzheitliches ist“ (27).

„Allgemeiner gesagt: eine nicht qualitativ ausgerichtete Bewährungskontrolle ist zur Validierung eines ganzheitlichen diagnostischen Verfahrens nicht geeignet“ (28).

Wahrscheinlich in noch viel höherem Maße ist ein anderer Grund für die unterschiedliche Höhe der gefundenen Validitätskoeffizienten maßgebend: Lopez (1965) weist darauf hin, daß die Eignungsdiagnose ein mehrstufiger Vorgang ist:

„This decision-making function of the employment interviewer presupposes three separate and prior steps: description, evaluation, and prediction. The selection interviewer must first elicit sufficient information from the applicant (description) to compare with a set of preestablished job specifications (evaluation) to enable him to draw a conclusion about the probable future behavior of the interviewee in a specific set of circumstances (prediction). On the basis of this prediction he then makes a decision ...“ (10f.).

Für die diagnostische Valenz der Exploration ist von ausschlaggebender Bedeutung, daß für die Diagnose eine klare und vollständige Symptomatik der zu diagnostizierenden Sachverhalte von der Forschung erarbeitet worden ist, der nicht nur zu entnehmen ist, welche Befunde überhaupt symptomatische Bedeutsamkeit besitzen, sondern vor allem auch, welches Gewicht den einzelnen Symptomen beizumessen ist. Das Fehlen einer solchen klaren Symptomatik ist z.B. der Grund dafür, weshalb selbst bei standardisierten Tests und Persönlichkeitsfragebögen in der Klinischen Psychologie von verschiedenen Auswertern unterschiedliche klinische Diagnosen gestellt werden. Für die Eignungsdiagnostik haben Triebe, Fischer und Ulich (1973) mit Recht hervorgehoben: Eine Analyse der Arbeitsanforderungen

„ist der Ausgangspunkt jeder den Ansprüchen wissenschaftlicher Objektivität genügenden Eignungsdiagnostik. Ihr kommt die Schlüsselfunktion zu; von ihrer Angemessenheit und Genauigkeit hängt letzten Ende ab, ob sich die eignungsdiagnostischen Verfahren - mit den resultierenden Prognosen - bewähren“ (27).

Niedrige Validitätskoeffizienten können ihren Grund ganz einfach haben in unzutreffenden Vorstellungen von der Anforderungsstruktur der betreffenden Stelle oder Position auf Seiten des Entscheidungsträgers. Die Strukturierung der eignungsdiagnostischen Exploration muß von einer Analyse der Arbeitsanforderungen her vorgenommen werden, während in der Praxis, wie die vorliegende Literatur zeigt, weit eher ein dem Explorator mehr oder weniger bewußtes „Stereotyp des guten Bewerbers“ den Orientierungsrahmen für die Gesprächsführung und für die Beurteilung abgibt.

„Insofern könnte man - etwas überspitzt - sagen, daß mit Hilfe unstrukturierter Interviews ausgewählte Bewerber zwar wahrscheinlich meist die Anforderungen ihres

Interviewers erfüllen werden, daß es aber weitgehend vom Zufall abhängen wird, in welchem Maße dessen Anforderungen mit denen der für den Bewerber vorgesehenen Tätigkeit übereinstimmen“ (Triebe 1976, 40).

Aus den Sammelreferaten von Ulrich und Trumbo (1965) und von Wright (1969) ergibt sich eine Überlegenheit strukturierter Explorations hinsichtlich ihrer Reliabilität und Validität. Dies gilt aber wiederum nur, wenn sich die Voraussage auf etwas Spezialisiertes, Eingeengtes bezieht. Allgemein formuliert Wellek (1958):

„Je enger die Fragestellung, zumal in Richtung auf konkrete Leistungen und Erfolge, umschrieben ist, je weniger läßt sich von der bloßen Exploration, je mehr von entsprechend zugeschnittenen Tests und sogar Fragebögen erwarten. Und umgekehrt: je allgemeiner aufs Persönlichkeitsganze gehend die Fragestellung, je mehr läßt sich von der Exploration und von der unmittelbaren oder auch mittelbaren ausdrucksmäßigen Kenntnisnahme, je weniger von Tests und erst recht von Fragebögen erwarten“ (25).

Auf Grund einer Übersicht über die vorliegende anglo-amerikanische Literatur zur Leistungsfähigkeit der Exploration im Kontext der Eignungsdiagnostik kommt Triebe (1976) zu dem Ergebnis:

„Im Hinblick auf die Möglichkeiten des Interviews scheint-wenn auch wohl z.T. aus recht verschiedenen Gründen - unter Praktikern und Wissenschaftlern fast einheitlicher Optimismus zu herrschen“ (8).

Diese Möglichkeiten gilt es zu nutzen. Für die vielfältigen Anwendungsbereiche explorativer Techniken gilt es,

die relevanten Themen der Exploration empirisch zu ermitteln,
 Fragetechniken zu entwickeln,
 Auswertungstechniken zu erarbeiten,
 die diagnostische Relevanz von Explorationsdaten zu ermitteln,
 Psychologen auf Grund dieser Erkenntnisse in Explorationstechnik und
 -auswertung auszubilden,

damit der Kreis von Psychologen, der mit diesem methodischen Instrumentarium erfolgreich zu arbeiten versteht, vergrößert wird, denn dieses Erhebungsverfahren hat einzigartige und daher unverzichtbare Vorteile, auf die namentlich Thomae immer wieder hingewiesen hat:

„Es ist somit nicht eine methodische Voreingenommenheit oder Borniertheit, die uns die besondere Bedeutung der Exploration für die systematische Beobachtung menschlichen Verhaltens in verschiedenen biographischen Einheiten hervorheben läßt. Vielmehr ist es die Einsicht, daß nur das Individuum selbst Zeuge seines Verhaltens im natürlichen Ablauf seines Lebens ist. Da wir keine Zeiten herbeisehnen dürften, in denen eine Dauerbeobachtung durch Fremde staatlich oder wissenschaftlich sanktioniert wird, können wir auf die Aussagen dieses Zeugen nicht verzichten“ (1968, 111).

Literatur

- Anger, H. 1969. Befragung und Erhebung. Handbuch der Psychologie, Band 7/1: Sozialpsychologie, Göttingen: Hogrefe, 567-617.
- Antons, K. & Schulz, W. 1976. Normales Trinken und Suchtentwicklung. Band 1, Göttingen: Hogrefe.
- Arnold, W. & Pauli, R. 1957⁶. Psychologisches Praktikum. Band 1, Stuttgart: Fischer.
- Assessment Staff 1948. Assessment of men: selection of personnel for the office of strategic services. New York: Rinehart.
- Baade, W., Lipmann, O. & Stern, W. 1909. Fragment eines biographischen Themas. Zeitschrift für Angewandte Psychologie, 3, 191-215.
- Bain, R. K. 1960. The researcher's role: a case study. In: Adams, R. N. and Preuss, H. H. (eds): Human Organisation research. Field relations and techniques, Homewood: Dorsey Press, 140-152.
- Beck, W. 1942. Begegnung und Erkenntnis. Zeitschrift für Angewandte Psychologie, 62, 328-369.
- Bellows, R. 1954. Psychology of Personnel in Business and Industry. Englewood Cliffs: Prentice-Hall.
- Bellows, R. M. & Estep, M. F. 1954. Employment psychology: The interview. New York: Holt, Rinehart & Winston.
- Bingham, W. van Dyke & Moore, B. V., with the collaboration of Gustad, J. W., 1959. How to interview. New York: Harper & Brothers.
- Blakeney, R. N. & McNaughton, J. F. 1971. Effects of temporal placement of unfavorable information on decision making during the selection interview. Journal of applied Psychology, 55, 138-142.
- Böcher, W. 1968. Befragungsstil und Verdeckungsmöglichkeiten im eignungsdiagnostischen Gespräch. Fahreignung und Verkehrssicherheit, Mitteilungsblatt des MPI Stuttgart, 17, 20-25.
- Boland & Roizen 1973 zit. nach Antons u. Schulz 1976.
- Bolton, D. L. & Hickey, M. E. 1969. Effect of interviews on teacher selection decisions. Journal of Applied Psychology, 53, 501-505.
- Bühler, K. 1933. Ausdruckstheorie. Jena: Fischer.
- Cattell, R. B. 1957. Personality and motivation: structure and measurement. Yonkers-on-Hudson, New York: World Book.
- Cosper, R. 1969. Interviewer bias in a study of drinking practices. Quarterly Journal of Studies on Alcohol, 30, 152-157.
- Crampton, C. W. 1908. Psychological age a fundamental principle. Child Development, 15, 3-52.
- Deters, H. 1978. Beiträge zur Prognose der Eignung zum Führen von Kraftfahrzeugen. Unveröff. vervielf. Forschungsbericht der Arbeits- und Forschungsgemeinschaft für Straßenverkehr und Verkehrssicherheit. Köln.

- Dimock, H. S. 1937. Rediscovering the adolescent. A study of personality development in adolescent boys. New York: Association Press.
- Ekman, G. 1947. Reliabilitet och konstans. Uppsala: Hugo Gebers Förlag.
- Erbslöh, E. 1972. Interview. Stuttgart: Teubner Studienskripten.
- Eysenck, H.-J. 1951. Uses and abuses of psychology. London: Penguin Books; deutsche Übersetzung: Wege und Abwege der Psychologie. Hamburg 1956.
- Fear, R. A. 1953. The evaluation interview. New York: McGraw Hill.
- Fisseni, H.-J. 1974. Zur Zuverlässigkeit von Interviews. Archiv für Psychologie, 126, 71-84.
- Fleiss, J. L. 1970. Estimating the reliability of interview data. Psychometrika, 35, 143-162.
- Friedrichs, H. 1967. Die aussagepsychologische Exploration. Handbuch der Psychologie, Band 11, Göttingen: Hogrefe, 3-25.
- Froehlich, C. P. 1958. The completeness and accuracy of counseling interview reports. Journal of General Psychology, 58, 81-96.
- Froehlich, R. E. & Bishop, F. M. 1973. Die Gesprächsführung des Arztes. Berlin: Springer.
- Gross, H. 1893. Handbuch für Untersuchungsrichter, Polizeibeamte und Gendarmen usw. als System der Kriminalistik. Graz: Leuschner & Lubensky.
- Gross, H. 1898. Kriminal-Psychologie. Leipzig: Vogel.
- Husén, T. 1949. Om innebörden av psykologiska mätningar. Lund-Köbenhavn: Gleerup och Munksgaard.
- Inbau, F. E. & Reid, J. E. 1967², 1974 reprint. Criminal interrogation and confessions. Baltimore: The Williams & Wilkins Company.
- Jung, C. G. 1937. Die Beziehungen der Psychotherapie zur Seelsorge. Zürich: Rascher. Wiederabgedruckt in: Gesammelte Werke, 11. Bd., 2. Aufl. 1979, S. 355-376.
- Kahn, R. L. & Cannell, C. F. 1966¹⁰. The dynamics of interviewing. New York: John Wiley & Sons.
- Kemmler, L. & Echelmeyer, L. 1978. Anamnese-Erhebung. Handbuch der Psychologie, Band 8/2, Göttingen: Hogrefe, 1628-1648.
- Kephart, N. 1952. The employment interview in industry. New York: McGraw-Hill.
- Kinsey, A. C., Pomeroy, W. B. & Martin, C. E. 1948. Sexual behavior of the human male. Philadelphia and London: Saunders; deutsche Übersetzung: Das sexuelle Verhalten des Mannes. Berlin und Frankfurt 1964.
- Kinsey, A. C., Pomeroy, W. B., Martin, C. E. & Gebhard, P. H. 1953. Sexual behavior in the human female. Philadelphia and London: Saunders; deutsche Übersetzung: Das sexuelle Verhalten der Frau. Berlin und Frankfurt 1963.
- Klages, L. 1926. Zur Ausdruckslehre und Charakterkunde. Heidelberg: Kampmann.
- Klebersberg, D. v. 1970. Fahrverhalten. Kleine Fachbuchreihe, Band 8, Wien: Kuratorium für Verkehrssicherheit.

- Kreipe, K. 1936. Zur Methode der Exploration. In: Abhandlungen zur Wehrpsychologie, 1. Folge, Zeitschrift für Angewandte Psychologie und Charakterkunde, Beiheft 72, 104-114.
- Kröber, W. 1942. über die Hauptaussprache (Exploration). Wehrpsychologische Mitteilungen, 4/1, 26-39.
- Kunkel, E. 1977. Biographische Daten und Rückfallprognose bei Trunkenheitstätern im Straßenverkehr. Köln: Verlag TÜV Rheinland.
- Kunkel, E. 1978. Die Bedeutung der biographischen Daten für die Fahreignungsprognose. Köln: Verlag TÜV Rheinland.
- Ledvinka, J. 1971. Rate of interviewer and the language elaboration of black interviewees. Journal of Social Issues, 27, 185-197.
- Ledvinka, J. 1972. Rate of employment interviewer and reasons given by black job-seekers for leaving their jobs. Proceedings of the Annual Convention of the American Psychological Association, 7, 441-442.
- Lehr, U. 1964. Diagnostische Erfahrungen aus explorativen Untersuchungen bei Erwachsenen. Psychologische Rundschau, 14, 97-106.
- Lewin, K. 1936. Principles of Topological Psychology, New York-London: McGraw-Hill. Deutsche Übersetzung: Grundzüge der topologischen Psychologie. Bern-Stuttgart-Wien: Huber 1969.
- Lienert, G. A. 1969³. Testaufbau und Testanalyse. Weinheim: Beltz.
- Lopez, F. M. 1965. Personnel interviewing. New York: McGraw-Hill.
- Magnusson, D. 1959. A study of ratings based on T. A. T. Stockholm: The Swedish Council for Personnel Administration.
- Margis, P. 1911. Das Problem und die Methoden der Psychographie. Zeitschrift für Angewandte Psychologie, 5, 409-451.
- Mayfield, E. C. 1964. The selection interview: A re-evaluation of published research. Personnel Psychology, 17, 239-260.
- McFarlane, J. W. 1938. Studies in child guidance. I. Methodology or data collection and Organisation. Society for Research in Child Development, 3, 1-254.
- Meier, R. D. 1972. The effectiveness of modeling procedures and instruction for teaching verbal employment interview behaviors to high school seniors. Dissertation Abstracts. 33, (6-B), 2817.
- Metzger, W. 1942. Psychologie und Menschenkenntnis. Die Erziehung. Leipzig: Quelle & Meyer.
- Michel, L. 1969. Empirische Untersuchungen zur Frage der Übereinstimmung und Gültigkeiten von Beurteilungen des intellektuellen Niveaus aus der Handschrift. Archiv für die gesamte Psychologie, 121, 31-54.
- Mierke, K. 1944. Psychologische Diagnostik. In: N. Ach (Hrsg.): Lehrbuch der Psychologie, 3. Bd., Bamberg: Buchner, 1-79.

- Moffatt, G. W. 1969. The selection interview. A review. *Personnel Practice Bulletin*, **25**, 15-23.
- Olbrich, M. 1967. Formale Analyse der Verhaltensstruktur auf Grund des Tageslaufes. Vordiplomarbeit Bonn.
- Oldfield, R. C. 1951⁴. The psychology of the interview. London: Methuen.
- Orne, M. T., Thackray, R. I. & Paskewitz, D. A. 1972. On the detection of deception. In: N. S. Greenfield and R. A. Sternbach (eds): *Handbook of Psychophysiology*. New York: Holt, Rinehart and Winston, 743-785.
- Parkinson, C. N. 1957. *Parkinsons Law*. Boston: Houghton Mifflin Comp. Deutsche Übersetzung: *Parkinsons Gesetz*. Stuttgart 1958.
- Pfahler, G. 1939. Das Gespräch als Methode erbcharakterologischer Rassenforschung. Bericht über den XVI. Kongreß der Deutschen Gesellschaft für Psychologie, **119-122**.
- Pongratz, L. 1957. Das psychologische Explorationsgespräch. *Psychologische Rundschau*, **8**, 195-205.
- Rechetnick, J. & Barkus, Ph. 1966. The anatomy of a Workshop. *Personnel Review*, **27**, 199-205.
- Richardson, S. A., Dohrenwend, B. S. & Klein, D. 1965. *Interviewing*. New York: Basic-Books.
- Roethlisberger, F. J. & Dickson, W. J. 1939. *Management and the worker*. New York: Wiley.
- Rowe, P. M. 1963. Individual differences in selection decisions. *Journal of Applied Psychology*, **47**, 304-307.
- Salber, W. 1960. Qualitative Methoden der Persönlichkeitsforschung. In: *Handbuch der Psychologie*, Band 4, Göttingen: Hogrefe, 30-58.
- Salber, W. 1969. Strukturen der Verhaltens- und Erlebensbeschreibung. In: *Enzyklopädie der Geisteswissenschaftlichen Arbeitsmethoden*, München und Wien: Oldenbourg, 52.
- Scheuch, E. 1962. Das Interview in der Sozialforschung. In: R. König (Hrsg.): *Handbuch der empirischen Sozialforschung*, Stuttgart: Enke.
- Schmidt 1973. zit. nach Antons u. Schulz 1976.
- Schonfeld, W. A. 1943. Primary and secondary sexual characteristics. Study of their development in males from birth through maturity with biometric study of penis and testes. *American Journal for the Diseases of the Child*, **65**, 535-549.
- Schraml, W. 1964. Das psychodiagnostische Gespräch (Exploration und Anamnese). *Handbuch der Psychologie*, Band 6, Göttingen: Hogrefe, 868-897.
- Schuller, A. & Rosenmeier, H. P. 1973. *Medizinstudium und Sozialwissenschaften*. München: Urban & Schwarzenberg.
- Soudijn, K. A., Mellenbergh, G. J. & Hartemink, B. G. 1970. Evaluatie van een handleiding voor de interviewer. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, **25**, 618-626.

- Spiegel, B. 1965. Die Aufgaben der Psychologie in der werbewissenschaftlichen Forschung. *Werbewissenschaftliches Referatenblatt*, Nr. 3.
- Spriegel, W. R. & James, V. A. 1958. Trends in recruitment and selection practices. *Personnel*, 35, 42-48.
- Steinkamp, S. W. 1966. Some characteristics of effective interviewers, *Journal of Applied Psychology*, 50, 487-492.
- Stern, W. 1900. *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth.
- Stern, W. 1902. Zur Psychologie der Aussage. *Zeitschrift für Strafrechtswissenschaft*, **22**, 315-370.
- Taft, R. 1955. The ability to judge people. *Psychological Bulletin*, 52, 1-23.
- The McGraw-Hill course in effective interviewing. New York: McGraw-Hill 1973.
- Thomae, H. 1968. *Das Individuum und seine Welt*. Göttingen: Hogrefe.
- Thomae, H. & Schmidt, H. D. 1967. Psychologische Aspekte der Schuldfähigkeit. *Handbuch der Psychologie*, Band 11, Göttingen: Hogrefe, 326-396.
- Thoms, K. 1975. Anamnese. In: S. Keil (Hrsg.): *Familien- und Lebensberatung*. Ein Handbuch. Stuttgart: Huber.
- Thorne, F. C. 1970. Psychological „twenty questions“: A method for teaching diagnostic interviewing. *Journal of Clinical Psychology*, 26, 331-334.
- Trankell, A. 1956. Rekryteringen av piloter i svenska flygvapnet. *Tidskrift i militär hälsovård*, Jg. 81, 61-90.
- Trankell, A. 1959. The psychologist as an instrument of prediction. *Journal of Applied Psychology*, 43, 170-175.
- Triebe, J. K. 1976. *Das Interview im Kontext der Eignungsdiagnostik*. Bern, Stuttgart, Wien: Huber.
- Triebe, J. K., Fischer, K. & Ulich, E. 1973. Problemstudie zur Informations- und Entscheidungsfindung bei der Auswahl von Bewerbern für den öffentlichen Dienst. In: Studienkommission für die Reform des öffentlichen Dienstrechts, Band 10, 15-104, Baden-Baden.
- Ulrich, L. & Trumbo, D. 1965. The selection interview since 1949. *Psychological Bulletin*, 63, 100-116.
- Undeutsch, U. 1954. Die Entwicklung der gerichtropsychologischen Gutachter Tätigkeit. Bericht über den XIX. Kongreß der Deutschen Gesellschaft für Psychologie, Göttingen, 132-154.
- Undeutsch, U. 1965². Forensische Psychologie. In: *Handwörterbuch der Kriminologie*, hrsg. von R. Sieverts, Band I, Berlin: De Gruyter, 205-231.
- Undeutsch, U. 1967. Beurteilung der Glaubhaftigkeit von Zeugenaussagen. *Handbuch der Psychologie*, Band 11, Göttingen: Hogrefe, 26-181.
- Vernon, Ph. E. 1957. *Personality tests and assessment*. London: Methuen.

- Wagner, R. 1949. The employment interview: A critical summary. *Personnel Psychology*, 2, 17-46.
- Walther, E. H. 1941. Die Exploration als Mittel zur Handlungsuntersuchung. *Wehrpsychologische Mitteilungen*, 3/2, 23-28.
- Wellek, A. 1958. Exploration und ganzheitliches Verfahren. *Psychologische Rundschau*, 9, 24-28.
- Welzel, U. 1976. Die Rückfallprognose bei Trunkenheitstätern. Faktor Mensch im Verkehr, Heft 25, Darmstadt: Tetzlaff.
- Welzel, U. 1982. Differentielle Prognosekriterien zur Beurteilung der Rückfallwahrscheinlichkeit in das Delikt „Trunkenheit am Steuer“. In: Winkler, W. (Hrsg.): *Verkehrspsychologische Beiträge I. Faktor Mensch im Verkehr*, Heft 32. Braunschweig: Rot-Gelb-Grün Verlag, im Druck.
- White, W. F. 1955. *Street corner Society*. Rev. Ed., Chicago: University of Chicago Press.
- Wiley, L. & Jenkins, W. S. 1964. Selecting competent raters. *Journal of Applied Psychology*, 48, 215-217.
- Wohlfart, E. 1939. Die Bildpostkartenwahl. *Wehrpsychologische Mitteilungen*, H. 5 und 7.
- Wolfe, L. A. 1970. A comparison of the effectiveness of teaching basic interviewing skills by three group methods: Modeling with role playing, role playing, and discussion. *Dissertation Abstracts*, 31 (3-A), 1029-1030.
- Wright, O. R. 1969. Summary of research on the selection interview since 1964. *Personnel Psychology*, 22, 391-413.