

Kritische Betrachtung psychologischer Diagnostik

Bachelor-Thesis
zur Erlangung des Akademischen Grades
Bachelor of Science Rehabilitationspsychologie

Verfasser:	Michael Georg Ortlieb
------------	-----------------------

Matrikelnummer:	20092766
-----------------	----------

Vorgelegt am:	07.09.2012
---------------	------------

Erstgutachter:	Prof. Dr. habil. Gabriele Helga Franke
----------------	--

Zweitgutachter:	Prof. Dr. Mathias Morfeld
-----------------	---------------------------

Inhaltsverzeichnis

1. Zusammenfassung.....	2
2. Einleitung	3
3. Grundlagen von Diagnostik	3
3.1. Begriffsbestimmung	4
3.2. Die psychologische Diagnostik	5
3.3. Die psychiatrische Diagnostik	5
3.4. Ziele und Arten der Diagnostik.....	6
3.5. Testtheoretische Grundlagen.....	8
3.5.1. Gütekriterien	8
3.5.2. Testarten.....	15
3.5.3. Erhebungsstrategien	20
3.6. Die historische Entwicklung der Diagnostik.....	21
4. Kritik an Diagnostik und ihre Folgen	23
4.1. Kritik im historischen Kontext.....	23
4.2. Kritische Betrachtung der Diagnostik heute	26
4.3. Kritik in der Anwendung pädagogischer Diagnostik	31
4.4. Kritik in der Anwendung psychologischer Diagnostik	32
4.5. Kritik in der Anwendung psychiatrischer Diagnostik.....	35
4.6. Auswirkungen einer Diagnose	37
5. Fazit	40
6. Literaturverzeichnis	42
Eidesstattliche Erklärung	45

1. Zusammenfassung

Frage: Die hier vorliegende Arbeit widmet sich dem Thema „Kritische Betrachtung der Anwendung psychologischer Diagnostik“ in welcher folgende Forschungsfragen untersucht wurden: „Gibt es Bereiche in der Diagnostik, die kritisch betrachtet werden können“, „Welche Fehler treten im Rahmen der Diagnostik auf und welche Konsequenzen ergeben sich daraus hinsichtlich ihrer Anwendung für Tester und Klienten?“. Methode: Dieser Arbeit liegt ein umfassendes Fundament an Fachliteratur bestehend aus Zeitschriftenbeiträgen, Lehrbücher und ergänzender Literatur aus der Hochschulbibliothek der Hochschule Magdeburg-Stendal oder aus Datenbanken wie Springerlink, PsyContent, PubMed und Google-Scholar zu Grunde. Zur Untersuchung dieser Forschungsfragen wird zunächst auf die diagnostischen Grundlagen eingegangen, um darauf aufbauend Diagnostik hinsichtlich ihrer Kriterien und Anwendungen kritisch beleuchten zu können. Zur Darstellung der Konsequenzen werden die Prozesse und Folgen von Stigmatisierung genauer vorgestellt. Ergebnisse: Die psychologische Diagnostik als Teilgebiet der Psychologie ist heute eine empirisch belegte und gut geprüfte Wissenschaft, die hinsichtlich ihrer Existenz und Grundlagen nicht mehr in Frage gestellt werden kann. Ihre Teilgebiete, wie bestimmte Testarten, Erhebungsmethoden und die Durchführung von Diagnostik, unterliegen auch heute noch den verschiedensten Fehlerquellen und sind somit kritisch zu hinterfragen. Fehlerquellen sind neben menschlichen Fehlern, zu denen z.B. der Halo-Effekt oder der Milde- und Strengefehler gehören, primär die fehlerhafte Anwendung und die falsche Interpretation von Testergebnissen. Im Bereich der Testtheorie werden einzelne Verfahren, vor allem die testtheoretischen Konstrukte, werden immer noch kritisch Diskutiert. Zu den Konsequenzen, die sich aus der Anwendung von Diagnostik ergeben, gehören der erschwerte Zugang zu Ressourcen und Bildung, die Stigmatisierung und die übermäßige Belastung durch zu viele Testverfahren. Diskussion: Die meisten Kritikpunkte im Zusammenhang mit psychologischer Diagnostik können mit einer genauen Beachtung der Gütekriterien und der jeweiligen Fragestellungen behoben werden. Stigmatisierung kann gesamtgesellschaftlich nur dann effektiv reduziert werden, wenn noch mehr in Krankheitsaufklärung investiert und bestehende Vorurteile abgebaut werden.

Zur Vereinfachung und Vereinheitlichung wird in der gesamten hier vorliegenden Arbeit die männliche Form verwendet.

2. Einleitung

Psychologische Diagnostik ist ein Teilgebiet der Psychologie, welches z.B. der klinischen Psychologie als Instrument dient, um pathologische und normale Verhaltensweisen voneinander abgrenzen zu können. Demnach wird die psychologische Diagnostik nicht nur zur Forschung, sondern auch für die praktische Arbeit von Psychologen verwendet. Ein Blick auf die Internetseite „teste-dich.de“ präsentiert dem User eine Fülle von verschiedenen Testverfahren. Eine Art von Informationsgewinnung, wie sie auch die psychologische Diagnostik verwendet. Jedoch unterscheiden sich diese Testverfahren hinsichtlich ihrer Wissenschaftlichkeit von psychologischen Verfahren. Eine kritische Überprüfung der diagnostischen Kriterien ist im Sinne der Wissenschaftlichkeit daher von immenser Bedeutung. Worin liegen folglich die Unterschiede zwischen professionellen Verfahren und den Tests der Internetseite „teste-dich.de“? Wissenschaftlichkeit im Bereich der psychologischen Diagnostik wird unter anderem hinsichtlich ihrer Gütekriterien bewertet. Im nachfolgenden Kapitel werden entsprechend die wichtigsten Grundlagen dargestellt. Darauf aufbauend wird die psychologische Diagnostik hinsichtlich ihrer Kriterien und Anwendungen kritisch beleuchtet. Da die Diagnostik im Verlauf einer Therapie eine wichtige Rolle spielt und sie demnach diese positiv oder negativ beeinflussen kann, werden zuletzt besonders mögliche negative Konsequenzen für die Klienten berücksichtigt werden.

3. Grundlagen von Diagnostik

Zum grundlegenden Verständnis dieser Arbeit werden im Folgenden die wichtigsten Begriffe zur Diagnostik und Diagnose definiert. Des Weiteren wird der Begriff „Testverfahren“, welcher dem Bereich der psychologischen Diagnostik zugehörig ist, erläutert. Im zweiten Teil des Kapitels zum Thema Grundlagen der Diagnostik werden die für diese Arbeit wichtigen Arten von Diagnostik vorgestellt. Diese sind unter anderen: die psychologische Diagnostik, die psychiatrische Diagnostik und die pädagogische Diagnostik.

3.1. Begriffsbestimmung

Das Wort „Diagnostik“ stammt ursprünglich aus dem griechischen Wort „διάγνωσις“, welches so viel bedeutet wie „gründlich kennenlernen“, „entscheiden“ oder „beschließen“. In der wissenschaftlichen Literatur wird der Begriff Diagnostik folgendermaßen definiert:

Psychodiagnostik ist eine Methodenlehre im Dienste der Angewandten Psychologie. Soweit Menschen die Merkmalsträger sind, besteht ihre Aufgabe darin, interindividuelle Unterschiede im Verhalten und Erleben sowie intraindividuelle Merkmale und Veränderungen einschließlich ihrer jeweils relevanten Bedingungen so zu erfassen, hinlänglich präzise Vorhersagen künftigen Verhaltens und Erlebens sowie deren evtl. Veränderungen in definierten Situationen möglich werden (Amelang & Schmidt-Atzert, 2006, S. 3).

In dieser Arbeit wird Diagnostik in drei relevante Bereiche unterschieden: pädagogische, psychiatrische und psychologische Diagnostik.

Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben (Ingenkamp & Lissman, Lehrbuch der Pädagogischen Diagnostik, 2008, S. 13).

Die psychologische Diagnostik arbeitet zu einem großen Teil mit wissenschaftlichen Testverfahren. Diese sind in der Literatur wie folgt definiert: „Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Raatz, 1998, S. 1).

3.2. Die psychologische Diagnostik

Unter psychologischer Diagnostik versteht man grundlegend das gezielte Sammeln und Erheben von Informationen zur Beantwortung einer psychologischen Fragestellung. Die Diagnostik dient dazu Handlungsgrundlagen zu schaffen, um über den weiteren Behandlungsverlauf von Klienten entscheiden und diesen verbessern zu können. Somit wird eine Prognose den zukünftigen Behandlungsverlauf betreffend möglich. Pospeschill und Spinath (2009) beschreiben die psychologische Diagnostik als diagnostischen Prozess, in welchem für die Diagnostik wichtige Kriterien im Hinblick auf bestimmte Merkmalsträger exploriert werden, die nach ihrer Auswertung eine Diagnose und eine Prognose ermöglichen. „Die Aufgabe der psychologischen Diagnostik [Hervorhebung v. Verf.] besteht in der Erhebung von Differenzen zwischen Merkmalsträgern unter standardisierten Bedingungen zum Zwecke einer Diagnostischen Anwendung“ (S. 14). Damit die psychologische Diagnostik eine Hilfestellung für den Auftraggeber sein kann, muss unter Anwendung von psychologischen Verfahren, eine Fragestellung auf eine Art beantwortet werden, die eine Entscheidung zulässt oder zumindest eine Entscheidungshilfe ist. Im Rahmen dieses Prozesses gibt es mehrere Schritte. Zunächst einmal muss die grundlegende Fragestellung geklärt werden. Mit dieser hängt eine verbindliche Formulierung des Auftrags zusammen. Ist diese dann geklärt, wird diese Fragestellung in eine oder mehrere Hypothesen übersetzt. Diese Hypothese/n gilt es zu prüfen, zu bestätigen oder zu widerlegen. Ist eine Hypothese formuliert, werden konkrete Daten erhoben, welche über die Gültigkeit der Hypothese entscheiden. Die Phase der Operationalisierung spielt im weiteren Verlauf des diagnostischen Prozesses eine wesentliche Rolle. Hierbei wird zunächst ein zur Hypothese geeignetes Verfahren vor allem hinsichtlich der Einhaltung der Gütekriterien, die in einem späteren Kapitel noch genauer erklärt werden, ausgewählt. Danach erfolgt eine genauere Planung der Untersuchung. Der letzte Schritt liegt in der diagnostischen Urteilsbildung. Dieses wiederum muss diagnostisch nutzbar sein, aber auch Prognosen zulassen (Amelang & Schmidt-Atzert, 2006).

3.3. Die psychiatrische Diagnostik

„Im Folgenden [sic.] soll der Begriff der psychiatrischen Diagnostik [Hervorhebung v. Verf.] als Oberbegriff zur Kennzeichnung einer Vielzahl verschiedener Ansätze und Aktivitäten auf sehr unterschiedlichen Ebenen verwendet werden“ (Stieglitz, 2000, S. 2). Ebenen bedeutet in diesem Zusammenhang, Symptom, Syndrom und Diagnose/Störung.

Diese drei Begriffe stehen in einer Hierarchie zueinander. Dabei werden mehrere Symptome zu einem Syndrom zusammengefasst und mehrere Syndrome, sowie weiterführende Daten ergeben die Diagnose. Symptome sind gekennzeichnet durch beobachtbare Verhaltensweisen oder Störungen, die vom Klienten berichtet werden. Die nächst höhere Ebene, die Syndromeebene, wird durch das Zusammenfassen aller beobachtbaren und berichteten Symptome eines Klienten erfasst. „Syndrome werden zu Störungen [Hervorhebung v. Verf.], wenn es Belege dafür gibt, dass die Konstellation von Symptomen als unabhängig voneinander anzusehen sind“ (Stieglitz, 2000, S. 2). Hierbei ist der Begriff Störung nach Stieglitz (2000) mit dem Wort „Diagnose“ gleichzusetzen. Eine Diagnose besteht aus allen erfassten Symptomen und Syndromen und bezieht zusätzlich Daten wie die Anamnese, den zeitlichen Verlauf und Untersuchungsergebnisse ein. Die Bezeichnung und Definition im Bereich der psychiatrischen Diagnostik und entsprechend relevanter Symptome, „(...) orientiert sich bis zum heutigen Tag im Wesentlichen an den klassischen Arbeiten zur Psychopathologie [Hervorhebung v. Verf.] von Karl Jaspers und Kurt Schneider und wurde in den letzten 20 Jahren insbesondere von Scharfetter (1985) systematisiert und weiterentwickelt“ (S. 6). Nach Stieglitz (2000) resultierten aus den erwähnten Arbeiten Dokumentationssysteme, von denen das AMDP-System mit 100 psychischen und 40 somatischen Symptomen im deutschsprachigen Raum, immer noch das am häufigsten eingesetzte System im psychiatrischen Bereich ist.

3.4. Ziele und Arten der Diagnostik

Ziel der Angewandten Psychologie ist es, bestimmte Zustände zu beheben, oder gewünschte Zustände zu erhalten. Um dies zu ermöglichen, werden zunächst Erscheinungsform, Verursachung oder Auslösung ermittelt (=Diagnostik). Testverfahren, die für diesen Zweck eingesetzt werden, bieten die Möglichkeit für die Untersuchung relevante Informationen zu generieren. Interventionen, die aus den diagnostischen Ergebnissen abgeleitet werden, tragen dazu bei, vor allem psychischen Leidensdruck zu minimieren, um einen „Normalzustand“ herstellen zu können. Weiterführende diagnostische Verfahren können den Prozess der Behandlung und Intervention bewerten. Ziel ist es, nicht nur den derzeitigen Stand festzustellen, sondern auch Vorhersagen über künftige Verhaltensweisen des Klienten zu treffen (Jäger & Petermann, 1999). Pospeschill und Spinath (2009) unterscheiden zwei diagnostische Bereiche: die institutionelle und die individuelle Diagnostik. Nach dieser Aufteilung gehört zur institutionellen Diagnostik die Arbeits- und Organisationspsychologie, in welcher diagnostische Verfahren zur Personalauswahl bei der Ausbildung und der Berufswahl eingesetzt werden. Weiterhin zählt auch die pädagogische

Psychologie in diesem Zusammenhang zur institutionellen Diagnostik. Hier geht es um Schulfähigkeits- oder Eignungsfeststellung für bestimmte Schulen oder Hochschulen und um die Feststellung von Leistungsdefiziten und Hochbegabung. Der zweite Bereich ist die individuelle Diagnostik. Hierzu gehören die klinische Psychologie und die klinische Neuropsychologie. Erstere untersucht die Bedingungen und die Ursachen einer psychischen Beeinträchtigung und versucht daraus geeignete Interventionen abzuleiten. Die klinische Neuropsychologie zeigt ein fast gleiches Vorgehen, nur bezieht sie sich auf neurologische Befunde. Ziel ist es wiederum neue Interventions-, Rehabilitations- oder Trainingsmaßnahmen zu finden (Pospeschill & Spinath, 2009). Grundlegend wird in der psychologischen Diagnostik, anders als in der psychiatrischen Diagnostik, mit Testverfahren gearbeitet. Diese unterteilen sich in mehrere Arten und hinsichtlich ihrer Einsatzgebiete, sowie in der Art der Durchführung und der Konstrukte, die sie prüfen sollen. Pospeschill und Spinath (2009) und Kubinger (2006) unterscheiden die Testverfahren zunächst in Leistungs- und Persönlichkeitstests. Pospeschill & Spinath (2009) teilen Testverfahren zusätzlich noch in Verhaltensbeobachtung, projektive Tests und diagnostische Befragung ein (vgl. Kapitel 3.5.3). Testverfahren und ihre Anwendung sind in Deutschland rechtlich nicht explizit geregelt. Psychologisches Diagnostizieren hingegen ist kein freier Begriff und eingebunden in mehrere Gesetze. Die Basis liefert das Grundgesetz. Hier ist der Umgang mit der Würde des Menschen in Artikel 1 GG geregelt. Die ersten drei Artikel des Grundgesetzes finden auch in den Nebengütekriterien, z.B. hinsichtlich der Zumutbarkeit und Ethik, Anwendung. Der Schutz der Privatsphäre eines Patienten ist im §203 des Strafgesetzbuches geregelt. So dürfen z.B. keine persönlichen Daten an Dritte weitergegeben werden. Auch wenn das Psychotherapeutengesetz seit dem 01.01.1999 in Kraft getreten ist, regelt das Heilpraktikergesetz (das Gesetz zur Ausübung von Heilkunde) vom 17.02.1939 in den §§1-3, wer Heilkunde ausüben darf. Die Tätigkeit eines Psychologen und psychologischen Psychotherapeuten fällt ebenfalls in den Bereich der Heilkunde, da nach §1 Abs. 2 jede Arbeit zur Feststellung, Heilung oder Linderung von Krankheit heilkundliche Tätigkeit darstellt. Nach Heyse (2003) unterliegt der Bereich der pädagogischen Diagnostik im Besonderen Schultests keiner rechtlichen Regelung. Der Übergang in den vom Heilpraktikergesetz geschützten Bereich ist jedoch fließend. Wie mit Testergebnissen umgegangen werden soll, wird im Bundesdatenschutzgesetz (BDSG) geregelt. So müssen z.B. personenbezogene Daten anonymisiert werden. Der Verkauf von Tests unterliegt wiederum keinen rechtlichen Regelungen und der Umstand, dass psychologische Testverfahren nur von einer bestimmten Berufsgruppe gekauft werden können, liegt an der freiwilligen Eigenregelung der Verlage (Heyse, 2003).

3.5. Testtheoretische Grundlagen

Die psychologische Diagnostik verwendet Testverfahren zur Untersuchung diagnostischer Fragestellung. Im Folgenden wird daher auf die Gütekriterien der psychologischen Testverfahren eingegangen.

3.5.1. Gütekriterien

Zur Qualitätssicherung psychologisch-diagnostischer Verfahren entwickelte die psychologische Testtheorie in den 50er Jahren des vergangenen Jahrhunderts die Gütekriterien. Diese waren eine Vereinbarung zur Wahrung von Standard und Qualität zwischen der Deutschen Gesellschaft für Psychologie, dem Berufsverband deutscher Psychologe und den großen Verlagshäusern, welche Testverfahren publizieren. Rechtsverbindlich wurden die Gütekriterien erst mit der Einführung der Deutschen Industrienorm (DIN) für Testverfahren (Kubinger, 2003). Zu den Gütekriterien gehören Objektivität, Reliabilität, Validität, Skalierung, Normierung, Testökonomie, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit und Fairness (Moosbrugger & Kelava, 2007). Im Folgenden werden die Gütekriterien in Anlehnung an Moosbrugger und Kelava (2007) erläutert:

Die Objektivität kann als grundlegendes Gütekriterium betrachtet werden und wird wie folgt definiert: „Ein Test ist dann objektiv, wenn er dasjenige Merkmal, das er misst, unabhängig von Testleiter, Testauswerter und von der Ergebnisinterpretation misst“ (S. 8). Objektivität als Oberbegriff kann in Durchführungs-, Auswertungs- und Interpretationsobjektivität unterteilt werden. Durchführungsobjektivität ist dann vorhanden, wenn das Ergebnis eines Tests nicht von der Person abhängt, die den Test durchführt oder leitet. Eine Möglichkeit diese Objektivität sicherzustellen ist die Standardisierung. Je genauer die Testinstruktion ist, desto weniger kann der Testleiter die Durchführung durch subjektive Auslegung der Instruktion Einfluss auf das Testergebnis nehmen. Die Standardisierung eines Tests ist dann optimal, wenn nur noch der Proband selbst das Resultat eines Tests beeinflussen kann. Auswertungsobjektivität liegt dann vor, wenn das Testergebnis nicht von der Person abhängt, die den Test auswertet. Die Auswertungsobjektivität ist mit einem offenen Frageformat in einem psychologisch-diagnostischen Testverfahren nur schwer zu erreichen. Multiple-Choice-Fragen ermöglichen dagegen eine höhere Auswertungsobjektivität, da sie weniger Fehler und Interpretation zulassen. Interpretationsobjektivität liegt dann vor, wenn

jeder, der den Test auswertet, aus den vorliegenden Informationen und Ergebnissen die gleichen Schlussfolgerungen ziehen kann.

Reliabilität ist ein weiteres Gütekriterium, welches Aussagen über die Testgenauigkeit ermöglicht und wird wie folgt definiert: „Ein Test ist dann reliabel (zuverlässig), wenn er das Merkmal, das er misst, exakt, d.h. ohne Messfehler, misst“ (S. 11). Die Reliabilität eines psychologischen Tests wird mit Hilfe des Reliabilitätskoeffizienten erfasst. Dieser nimmt einen Wert zwischen eins und null an. Ein Wert von eins bedeutet, dass das Ergebnis frei von jeglichem Messfehler ist, wohingegen ein Reliabilitätskoeffizient von null bedeutet, dass das Ergebnis ausschließlich auf Messfehlern beruht. „Um das Ausmaß der Reliabilität zu bestimmen, wurden im Rahmen der Klassischen Testtheorie mehrere Verfahren entwickelt. So unterscheidet man in vier Vorgehensweisen (vgl. Moosbrugger & Kelava, 2007):

1. Retest-Reliabilität
2. Paralleltest-Reliabilität
3. Testhalbierungs-Reliabilität
4. Innere Konsistenz“ (S. 11-12).

Soll die Reliabilität durch das Verfahren der Retest-Reliabilität bestimmt werden, so wird ein Test dem Probanden zu zwei verschiedenen Zeitpunkten vorgelegt. Dabei liegt die Annahme zugrunde, dass sich das zu messende Merkmal selbst nicht geändert hat. Die Reliabilität ist dann der Wert, der sich aus der Korrelation der beiden Testergebnisse ermitteln lässt. Die Paralleltest-Reliabilität gilt häufig als die beste Möglichkeit, Messfehler zu vermeiden. Einem Probanden wird dabei ein Test in zwei Ausführungen vorgelegt. Beide Tests sollen aus inhaltlich möglichst ähnlichen Items bestehen. Die Korrelation dieser beiden Tests ist dann die Paralleltest-Reliabilität. „Parallel sind zwei Testformen dann, wenn sie trotz nicht identischer Item-Stichproben zu gleichen Mittelwerten und Varianzen der Testwerte führen“ (S. 12). Nicht immer sind Prüfungen die Retest- und die Paralleltest-Reliabilität durchführbar. Oft stehen Probanden nicht zu zwei Terminen zur Verfügung oder es kann aufgrund zu weniger Items keine parallele Testform entwickelt werden. Aus diesem Grund wurde die Testhalbierungsreliabilität entworfen. „In solchen Fällen ist es angebracht, den Test in zwei möglichst parallele Testhälften zu teilen und die Testhalbierungs-Reliabilität (Split-Half-Reliabilität) als Korrelation der beiden Testhälften zu bestimmen“ (S. 12). Die letzte Form der Reliabilität ist die innere Konsistenz. Diese Konsistenzanalyse kann man als Verallgemeinerung der Testhalbierungs-Reliabilität sehen. Hier wird jede Frage bzw. jedes Item als eigener Test bewertet und mit den anderen Items eines Tests verglichen. „Ja stärker die Testteile untereinander positiv korrelieren, desto höher ist die interne Konsistenz des Verfahrens (Cronbach- α -Koeffizient der Reliabilität; (...))“ (S. 13).

Die Validität gibt darüber Auskunft, ob das Merkmal das gemessen wurde, auch jenes Merkmal ist, welches gemessen werden soll. „Ein Test gilt dann als valide (gültig), wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes“ (S. 13). Die Validität gilt als eines der wichtigsten Kriterien bei der Konstruktion von Tests. Objektivität und Reliabilität werden als Grundlage der Validität gesehen. Eine niedrige Reliabilität schließt eine hohe Validität aus. Bei einer hohen Validität wiederum kann von dem Testergebnis auch auf Verhalten außerhalb einer Testsituation geschlossen werden. Validität wird unterschieden in: Inhalts-, Augenschein-, Konstrukt- und Kriteriumsvalidität. „Unter Inhaltsvalidität versteht man, inwieweit ein Test oder ein Testitem das zu messende Merkmal repräsentativ erfasst“ (S. 15). Grundlage ist der Repräsentationsschluss, wobei die einzelnen Items eines Tests aus einem Itempool entnommen werden, welcher das interessierende Merkmal erfasst. Diese werden so ausgewählt, dass sie repräsentativ das interessierende Merkmal wiedergeben. Inhaltsvalidität wird weder in Kennwerten, noch numerisch dargestellt, nur logische und fachliche Überlegungen lassen auf inhaltliche Validität schließen, sofern ein Item einen direkten Ausschnitt aus dem Verhaltensbereich des zu messenden Merkmals darstellt. Die Augenscheinvalidität kann sehr leicht mit Inhaltsvalidität verwechselt werden, da ein Test der Inhaltsvalidität aufweist, auch augenscheinvalide ist. Definiert wird sie wie folgt: „Augenscheinvalidität gibt an, inwieweit der Validitätsanspruch eines Tests, vom bloßen Augenschein her einem Laien gerechtfertigt erscheint“ (S. 15). Die Augenscheinvalidität hat dabei besonders für die Teilnehmer eines Tests eine hohe Bedeutung, da für sie nachvollziehbar wird, was der Test messen möchte. Die Testsituation wird dadurch transparenter. Intelligenztests weisen eine besonders hohe Augenscheinvalidität auf, was zum einen an der Popularität der Tests selbst, zum anderen an dem Bekanntheitsgrad des Intelligenzkonstruktes liegt. Für Laien wird somit nachvollziehbar, dass ein Intelligenztest tatsächlich Intelligenz messen kann. Wissenschaftlich ist die Augenscheinvalidität aber nicht immer zufriedenstellend, da die Validität durch empirische Kennwerte belegt werden muss. „Ein Test weist Konstruktvalidität auf, wenn der Schluss vom Verhalten der Testperson innerhalb der Testsituation auf zugrunde liegende psychologische Persönlichkeitsmerkmale (Konstrukte, latente Variablen, Traits) wie Fähigkeiten, Dispositionen, Charakterzüge, Einstellungen aufgezeigt wurde. Die Enge dieser Beziehung wird aufgrund von testtheoretischen Annahmen und Modellen überprüft“ (S. 16). Damit ist gemeint, dass man von einem Ergebnis in einem Intelligenztest auch auf die zugrundeliegende Intelligenzfähigkeit schließen kann, die der Proband in seinem Alltag abrufen kann. Überprüft man diese Art von Validität, so muss zunächst in struktursuchende und strukturprüfende Ansätze unterschieden werden. Das struktursuchende Vorgehen basiert auf der struktursuchenden deskriptiven Vorgehensweise. Dabei wird die Exploratorische Faktoranalyse zum Einsatz gebracht. Innerhalb eines Items

gibt der Trennschärfe-Koeffizient Auskunft über die Homogenität des jeweiligen Items. Auf diese Art gewonnene Merkmalsdimensionen, erlauben eine erste deskriptive Einordnung der theoretischen Konstrukte, die dem jeweiligen Test zugrunde liegen. Ein Test wird folglich hinsichtlich seines Konstrukts mit Tests verglichen, die ähnliche oder verwandte Konstrukte erfasse. Anschließend werden diese Ergebnisse mit den Resultaten der anderen theoriekonformen Verfahren verglichen. Die Konstruktvalidität wird in konvergente und diskriminante Validität unterschieden. Die konvergente Validität bedeutet das Vergleichen des Testergebnisses mit dem Ergebnis eines anderen Tests, der das gleiche Konstrukt messen soll. Mittels der Korrelation der Ergebnisse kann so ermittelt werden, ob der neue Test auch das zu messende Merkmal misst. Ein Beispiel hierfür ist der Intelligenz-Struktur-Test (IST-2000), ein Verfahren, welches die Intelligenz messen soll. Dabei wird dessen Ergebnis mit dem des Hamburg-Wechsler-Intelligenz-Inventar für Erwachsene (HAWI-E), welcher ebenfalls die Intelligenz messen soll, verglichen. Bei der diskriminanten Validität werden die Ergebnisse von Tests gegenübergestellt, mit dem Unterschied, dass als Vergleich ein Test ausgewählt wird, der explizit ein anderes Konstrukt misst, als der Test, dessen Validität bestimmt werden soll. Ziel dieses Vorgehens ist eine möglichst geringe Korrelation. Dabei soll eine Abgrenzung von den zu messenden Merkmalen und Konstrukten, zu anderen erreicht werden. Um die diskriminante Validität zu prüfen, kann ein Test nicht nur mit konstruktfernen Tests verglichen werden, sondern auch mit solchen, die relativ ähnliche oder dem Konstrukt nahe Merkmale messen. Erst dann kann eine ausreichende Validität aus den Ergebnissen geschlossen werden. Im Rahmen der Konstruktvalidität gibt es neben dem struktursuchenden, auch das strukturprüfende Vorgehen. Ein Verfahren, welches ermöglicht, inferenzstatistische Schlüsse aus einem Test zu ziehen. Dies ist nur möglich, wenn Testmodelle mit latenten Variablen die Grundlage dieses Vorgehens sind. Wurde ein Test mittels exploratorischer Faktorenanalyse überprüft, kann die gefundene Struktur mit der konfirmatorischen Faktorenanalyse kontrolliert werden. Einzelne Dimensionen wiederum können mit Hilfe von IRT-Modellen ebenfalls hinsichtlich einer möglichen Homogenität konfirmatorisch überprüft werden. Die Kriteriumsvalidität ist auch ein Teilbereich des Gütekriteriums der Validität, welche die tatsächliche Verwendbarkeit eines Tests hinsichtlich der Vorhersagbarkeit von Verhalten und Erleben angibt. „Ein Test weist Kriteriumsvalidität auf, wenn vom Verhalten der Testperson innerhalb der Testsituation erfolgreich auf ein Kriterium, nämlich auf ein Verhalten außerhalb der Testsituation geschlossen werden kann. Die Enge dieser Beziehung ist das Ausmaß an Kriteriumsvalidität (Korrelationsschluss)“ (S. 18). Ein Beispiel für Kriteriumsvalidität ist ein Schuleignungstest. Kinder, die bei einem entsprechenden Test besser abschneiden, sollten dann auch in der Schule besser zurechtkommen. Wird dieses „Außenkriterium“ unter der zeitlichen Verfügbarkeit betrachtet, so wird von Übereinstimmungsvalidität gesprochen, wenn

dieses in der Vergangenheit oder Gegenwart bereits vorliegt. Tritt das Außenkriterium wie bei einem Schuleignungstest, der vor Schulantritt durchgeführt wird, erst in der Zukunft auf, so spricht man von Vorhersagevalidität. Bei Übereinstimmungsvalidität ist die Korrelation mit einem bestehenden Kriterium und im Fall der Vorhersagevalidität die Prognose von Bedeutung.

Das Gütekriterium der Skalierung kann am einfachsten mit einem Intelligenztest beschrieben werden. So soll eine Person mit höherer Intelligenz (also eine leistungsfähigere Person) auch ein besseres Ergebnis bei einem Intelligenztest erzielen. Die Forderung der Skalierung bedeutet, dass sich die individuelle Leistungsfähigkeit auch in entsprechenden Testwerten widerspiegeln muss. „Ein Test erfüllt das Gütekriterium der Skalierung, wenn die laut Verrechnungsregel resultierenden Testwerte die empirischen Merkmalsrelationen adäquat abbilden“ (S. 18). Im Wesentlichen hängt die Skalierung von dem zugrundeliegenden Skalenniveau ab. Moosbrugger & Kelava (2007) gehen davon aus, dass eine Messung auf Nominalskalenniveau nicht ausreicht, um die genauen Relationen zwischen den Ergebnissen einzelner Testpersonen zu unterscheiden. Damit zwischen unterschiedlich leistungsfähigen Testpersonen differenziert werden kann, sollte die Messung mindestens auf Ordinalskalenniveau erfolgen. Soll darüber hinaus eine Beurteilung der Größe hinsichtlich der inter- und intraindividuellen Unterschiede möglich sein, sollte mindestens Intervallskalenniveau angewendet werden.

Das Gütekriterium der Normierung wird synonym auch als Eichung bezeichnet. Das Ziel der Normierung eines Verfahrens besteht darin, möglichst aussagekräftige Werte von Testpersonen zu erhalten. Die in einem Test erzielten Werte einer Testperson werden dann mit Vergleichswerten überprüft. Die Vergleichswerte sollten von Personen stammen, die relevante Merkmale betreffend, wie z.B. Alter, Geschlecht und sozialen Status, eine möglichst hohe Ähnlichkeit zur Testperson aufweisen. „Unter der Normierung (Eichung) eines Tests versteht man das Erstellen eines Bezugssystems, mit dessen Hilfe die Ergebnisse einer Testperson im Vergleich zu den Merkmalsausprägungen anderer Personen eindeutig eingeordnet und interpretiert werden können“ (S. 19). Die Eichung erfolgt für die meisten Testverfahren über Normtabellen. Diese wiederum sollen aus einer möglichst großen und vor allem repräsentativen Stichprobe bestehen. Das Resultat einer Testperson kann dann mit diesen Normtabellen verglichen werden. Dieser Vergleich ermöglicht wiederum Interpretationen, so liegt bei einem Intelligenztest der durchschnittliche Intelligenzquotient (IQ) der Normstichprobe bei 100. Wenn ein Proband ein Ergebnis von über 110 IQ-Punkten erreicht, so ist er nach dieser Normierung als überdurchschnittlich zu bezeichnen. In diesem Zusammenhang sollten die Prozentrangnormen Erwähnung finden. Hier wird der Prozentsatz von Testpersonen der Eichstichprobe ermittelt, die unter oder über

einem bestimmten Ergebnis liegen. So wird der Prozentrang gewonnen, welcher mit der in der Eichstichprobe erzielten prozentualen Häufigkeit eines Ergebnisses kumuliert. Neben dieser Technik gibt es noch weitere Normierungsmöglichkeiten. So kann auch der Abstand eines Testergebnisses zum Mittelwert der zugehörigen Eichstichprobe bestimmt werden. Die resultierende Differenz wird in Einheiten der Standardabweichung angegeben. Auf solchen Standardwerten bauen zum Beispiel IQ- und T-Werte aber auch Centil- und Stanine-Werte auf. Bei der Interpretation dieser Werte, ist darauf zu achten, dass das zugrundeliegende Merkmal normalverteilt ist. Sollte dies nicht der Fall sein, können lediglich Prozentrangwerte für eine Interpretation herangezogen werden. Im Rahmen der Normierung spielt der Geltungsbereich der Normtabelle ebenfalls eine Rolle. Dies bedeutet, dass die Vergleichsstichprobe, welche für die Normierung als Grundlage gilt, möglichst repräsentativ für die Grundgesamtheit sein soll, die ein Test ansprechen oder untersuchen möchte. Als Beispiel eignen sich altersabhängige Tests. So sollte die Vergleichsstichprobe eines Tests für Frauen unter 40 Jahren entsprechend für diese Gruppe repräsentativ sein. Da sich Merkmale/Kriterien eines Tests im Laufe der Zeit ändern können, besteht die Notwendigkeit, Normdaten möglichst aktuell zu halten. Sind die Normdaten veraltet, würde die Aussagekräftigkeit einer Vergleichsstichprobe mit der Zeit nachlassen. So können Lerneffekte oder Änderungen im Leistungsbild der Gesamtbevölkerung durchaus eine Neunormierung erforderlich machen.

Testökonomie erlaubt keine Aussage über die Qualität eines Testverfahrens, sondern ermöglicht die Einschätzung der Wirtschaftlichkeit. Dies beinhaltet Kosten, die durch eine Testung sowohl beim Durchführer, als auch beim Probanden entstehen. Doch lassen sich die Kosten nicht beliebig minimieren, da ansonsten die Qualität darunter leiden würde. So befasst sich Testökonomie mit einer möglichen Kostenreduktion unter der Berücksichtigung anderer Gütekriterien. „Ein Test erfüllt das Gütekriterium der Ökonomie, wenn er, gemessen am diagnostischen Erkenntnisgewinn, relativ wenig Ressourcen wie Zeit, Geld oder andere Formen beansprucht“ (S. 21). Wie bereits aus der Definition ersichtlich wird, gibt es bei der Beurteilung von Ökonomie zwei relevante Größen: zum einen der finanzielle Aufwand, den der Test verursacht und zum anderen der zeitliche Aufwand, der für den Test erforderlich ist. Für die Anschaffung eines Tests ergeben sich nicht nur Kosten durch Anschaffung und Materialverbrauch, sondern auch Kosten für die Anschaffung der nötigen Technik und des Auswertungsprogrammes. Der zeitliche Aufwand entsteht durch die Testzeit, die für die Vorbereitung, Durchführung, Auswertung und Ergebnisrückmeldung aufgewendet werden muss. Daher sollte der Ergebnissgewinn eines Tests größer sein, als die Summe aus Zeitaufwand und Kosten. Da dies aber kein berechenbarer Wert ist, wird meist die Ökonomie an dem Aufwand und den Ertrag eines Testverfahrens geschätzt, in dem diese mit anderen

Verfahren, die die gleichen Merkmalsdimensionen abfragen, verglichen werden. Um den Aufwand gering halten zu können, kann der Testleiter das adaptive Testen nutzen. Zur Durchführung von adaptiven Tests werden nur diese Fragen oder Aufgaben gelöst, die unmittelbar zum Erkenntnisgewinn beitragen. Andere Items, die die Hypothese nicht prüfen, werden ausgelassen. Das Gütekriterium der Ökonomie sollte stets unter Berücksichtigung der Einhaltung anderer Gütekriterien bewertet werden. „So ist eine geringere Ökonomie eines Tests bei einer konkreten Fragestellung insbesondere dann in Kauf zu nehmen, wenn z.B. aus Validitätsgründen der Einsatz gerade dieses Tests sachlich gerechtfertigt ist, weil nur mit ihm die konkrete Fragestellung fachgerecht beantwortbar ist“ (S. 21).

Ein weiteres Gütekriterium ist die Nützlichkeit eines Tests. „Ein Test ist dann nützlich, wenn für das von ihm gemessene Merkmal praktische Relevanz besteht und die auf seiner Grundlage getroffenen Entscheidungen (Maßnahmen) mehr Nutzen als Schaden erwarten lassen“ (S. 21). Ein Test ist daher dann nützlich, wenn ein Merkmal auf Grundlage der Kriteriumsvalidität nützliche Anwendungsoptionen aufweist, oder das zumessende Merkmal nicht mit einem Test ermittelt werden kann, welcher die übrigen Gütekriterien ebenso gut oder besser erfüllen kann.

„Ein Test erfüllt das Kriterium der Zumutbarkeit, wenn er absolut und relativ zu dem aus seiner Anwendung resultierenden Nutzen die zu testende Person in zeitlicher, psychischer sowie körperlicher Hinsicht nicht über Gebühr belastet“ (S. 22). Eine zu testende Person solle hinsichtlich des Zeitaufwandes, sowie der körperlichen und psychischen Belastung soweit möglich geschont werden. Eine verbindliche Unterscheidung in zumutbar und unzumutbar ist meist schwierig. Zum einen ist eine kritische Betrachtung hinsichtlich der Nützlichkeit nötig, zum anderen spielen gesellschaftliche Normen und Werte eine Rolle. So ist es vertretbar einem Menschen, der sich um eine Ausbildung zum Piloten bewirbt, einen schweren und anspruchsvollen Test zu unterziehen. Hier spielt das Ansehen der Ausbildung und die damit assoziierte Anforderung an den Beruf eine Rolle. Es würde auf mehr Unverständnis und Ablehnung stoßen, wenn ein/e Büroangestellte/r einem ebenso anspruchsvollen Test zu absolvieren hätte.

„Ein Testverfahren erfüllt das Gütekriterium der Unverfälschbarkeit, wenn das Verfahren derart konstruiert ist, dass die zu testende Person durch gezieltes Testverhalten die konkreten Ausprägungen ihrer Testwerte nicht steuern bzw. verzerren kann“ (S. 23). Dies ist vor allem bei Tests der Fall, in welchen die Testperson sehr leicht erkennen kann, wie der Test funktioniert. Eine hohe Augenscheinvalidität ermöglicht einer Testperson ein Verfahren schneller zu durchschauen. Dies kann zu sozial angepasstem Antworten führen, welche das Testergebnis verfälschen können. Im Besonderen sind hier Persönlichkeitstests zu nennen,

die von ihrer Anlage eher zur Verzerrung neigen. Ein Leistungstest, wie z.B. ein IQ-Test, der zwar auch eine sehr hohe Augenscheinvalidität besitzt, kann jedoch nur schwer verfälscht werden. Dies ist damit zu begründen, dass eine Testperson sich nicht besser oder intelligenter machen kann, als sie ist.

„Ein Test erfüllt das Gütekriterium der Fairness, wenn die resultierenden Testwerte zu keiner systematischen Benachteiligung bestimmter Personen aufgrund ihrer Zugehörigkeit zu ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppen führen“ (S. 23). Das Gütekriterium der Fairness beinhaltet das Bestreben, Menschen unterschiedlicher Geschlechter oder Hautfarben nicht zu diskriminieren. So sollen Männer und Frauen, aber auch Menschen mit heller oder dunkler Hautfarbe fair und respektvoll behandelt werden. Inwieweit ein Test unfair sein kann, wurde insbesondere in den 70er Jahren des vergangenen Jahrhunderts diskutiert. So wäre ein Item in einem Leistungstest unfair, wenn es z.B. Häckel- oder Strickaufgaben beinhalten würde, da dies Frauen aufgrund der ihnen zugedachten kulturellen Rollenverteilung der letzten Jahrhunderte eher können als Männer. Eine besondere Rolle in diesem Zusammenhang spielen die sogenannten Culture-Fair-Tests. Hier stehen besonders Sprachfähigkeiten im Fokus. Ein Test wäre dann kulturell fair, wenn er von Menschen mit geringen oder keinen sprachlichen Fähigkeiten in einer bestimmten Sprache ebenso gut absolviert werden kann, wie von Menschen mit höheren sprachlichen Fähigkeiten. Ein anderes Beispiel sind hier Tests, die Computeranwendung nötig machen. Besonders ältere Menschen würden so benachteiligt werden, die keine oder wenig Computererfahrung besitzen, im Gegensatz zu jungen Menschen, die bereits mit EDV-Anwendungen aufgewachsen sind. „Die Frage der Fairness betrifft, wenn man die oben genannten Gesichtspunkte bedenkt, auch die Normierung eines Tests und die Interpretation der Testergebnisse. Da es keine Faustregeln zum Umgang mit diesen Gütekriterium gibt, ist jeder Test individuell auf seine Fairness hin zu beurteilen“ (S. 24). Auch eine mögliche Testroutine sollte bei einer Prüfung der Fairness bedacht werden, da dies eine Größe ist, welche das Ergebnis beeinflussen kann.

3.5.2. Testarten

Die psychologisch-diagnostischen Verfahren können nach Wagener (2003) grundlegend in Leistungstests und Persönlichkeitsfragebogen eingeteilt werden. Weiterhin können die Testverfahren in projektive Tests, Verhaltensbeobachtung und diagnostische Befragungen unterteilt werden.

„Als allgemeine Leistungstests [Hervorhebung v. Verf.] werden Verfahren bezeichnet, die allgemeine Grundlagen der kognitiven Leistungsfähigkeit erfassen sollen, z.B. Aufmerksamkeit, Konzentration, Aktiviertheit oder Vigilanz (Wachsamkeit)“ (Pospeschill & Spinath, 2009, S. 82). Im Folgenden werden Leistungstests hinsichtlich psychologischer Konstrukte in Konzentrations-, Eignungs-, Entwicklungs- und Schultests eingeteilt. Inwieweit die Konstrukte Aufmerksamkeit und Konzentration voneinander abgegrenzt werden können, wird in der wissenschaftlichen Literatur unterschiedlich diskutiert, da beide Konstrukte sich gegenseitig bedingen und eine inhaltliche Trennung somit schwierig ist. Aus diesem Grund wird in dieser Arbeit von einer Unterscheidung der beiden Konstrukte abgesehen. Zur Prüfung der Aufmerksamkeit und Konzentration dienen Konzentrations- und Vigilanztests. Ein Konzentrationstest verwendet für die Erfassung allgemeiner Leistungsfähigkeit einfache Aufgaben. Diese einfachen Aufgaben werden dann aber durch Zusatzbedingungen deutlich erschwert. Grundlegend kann man Konzentrationstests in drei Gruppen unterteilen. Die erste Gruppe von Konzentrationstests sind die sogenannten Durchstreichtests. Zu diesen gehört zum Beispiel der D2. Ein Durchstreichtest kann anhand des D2 erklärt werden. Zu der zweiten Gruppe gehören Rechenverfahren, worunter z.B. der Konzentrations-Leistungs-Test (KLT) zählt. Zu einer weiteren Gruppe an Konzentrationstests gehören die Sortierverfahren, wie z.B. der Konzentrations-Verlaufstest (KVT). Weiterhin werden apparative Verfahren, wie z.B. das Wiener Testsystem zu den Konzentrationstests aufgezählt (Kubinger, 2006). Eignungstests sind Verfahren, die vor allem in Bereichen eingesetzt werden, in denen es um die Auswahl geeigneter Personen geht. Ziel ist es, geeignete Bewerber für eine Arbeitsstelle, ein Studium, eine Ausbildung oder allgemein für eine Aufgabe zu finden. Diese Verfahren können z.B. in Assessment-Centern zur Auswahl von Leitungspositionen, zur Auswahl geeigneter Bewerber für den polizeilichen Dienst und bei der Vergabe der Studienplätze für das Medizinstudium eingesetzt werden (Amelang & Schmidt-Atzert, 2006). Entwicklungstests sind Testverfahren, die den Entwicklungsstand eines Kindes mit einem Soll-Wert vergleichen. In diesem Zusammenhang sind vor allem Verhaltensweisen, die von Normverhalten abweichen, von Bedeutung. Ziel von Entwicklungstests ist es, prophylaktisch entwicklungspsychologische Fehlentwicklungen so früh wie möglich zu erkennen, um adäquat auf diese reagieren zu können. Eine spezielle Form der Entwicklungstests sind die Schultests (Rauchfleisch, 2005). „Unter Schultests verstehen wir mit Süllwold (1964) Prüfverfahren, durch die bestimmte Effekte von Unterricht und Erziehung festgestellt werden sollen“ (Rauchfleisch, 2005, S. 145). Ebenfalls gehören zu Schultests auch Testverfahren, die ein zu erwartendes Leistungsniveau abbilden und somit eine Einschätzung ermöglichen sollen, welche Leistungen der Schüler erbringen kann. Dies kann über Essay-Tests erfolgen, in denen sich Schüler über das Freie-Antwortformat zu einer Fragestellung äußern. Ein anderer Weg sind objektive informelle Tests, die ein Lehrer frei von jeglicher Form

durchführen kann. Dabei werden die Schulleistungstests in neun Arten unterteilt: „(...) Einschulungstests, spezielle Schuleignungstests, Mehrfächertests, Lesetests, Rechtschreibtests, Wortschatztests, Rechentests, Fremdsprachentests und sonstige Schulleistungstests“ (Rauchfleisch, 2005, S. 156). Eine weitere Gruppe von Leistungstest sind die Tests zur Prüfung von speziellen Funktionen und Fähigkeiten. Zu dieser Gruppe psychologischer Testverfahren zählen viele unterschiedliche Verfahren, die in vier Arten unterteilt werden: Zum einen in Verfahren zur Prüfung von sensorischen Fähigkeiten, zum anderen Art in Verfahren zur Prüfung motorischer Fähigkeiten. Weiterhin werden in Tests zur Prüfung diverser anderer Bereiche, wie Gedächtnis oder räumliche Vorstellungskraft und in Tests zur Prüfung spezieller Fähigkeiten, zu denen musikalische Fähigkeiten oder technisches Verständnis gehören, unterschieden. Verschiedene Variationen sind nicht nur in den zumessenden Kategorien zu finden, sondern auch in der Darbietungsweise der Test, die zum Teil in Papierform vorhanden sind, zum Teil aber auch über Hilfsmittel, wie z.B. Musikinstrumente, technische Hilfsmittel oder Ähnlichem durchgeführt werden. Besonders die Untersuchung kognitiver Fähigkeiten spielt im Rahmen der klinisch-psychologischen Untersuchung eine bedeutende Rolle (Amelang & Schmidt-Atzert, 2006).

Die nächste Gruppe der Testarten sind die Persönlichkeitstests. Persönlichkeitstests geben Aufschluss über Gefühle, sowie die inneren oder unbewussten Konflikte eines Menschen. Nach Kubinger (2006) basieren heute fast alle Persönlichkeitstests auf dem Modell der Big-Five (vgl. Kapitel 3.6). Die Persönlichkeitstests tragen dazu bei, einen Menschen besser kennen und verstehen zu lernen. Persönlichkeitsfragebögen sind eine der Möglichkeiten, einen Persönlichkeitstest durchzuführen, um Aufschluss über personenbezogene Merkmale zu erhalten. Allgemein werden unter Persönlichkeitsfragebögen Verfahren verstanden, die die Persönlichkeitsstruktur, Interessen und auch spezifische Probleme eines Klienten erfassen und aufzeigen. Diese Verfahren benötigen oft keinen Testleiter, da sie vom Patienten selbst ausgeführt werden können, was eine hohe Ökonomie ermöglicht. Dies ist ein Grund dafür, dass die Persönlichkeitsfragebögen in der Praxis häufig Anwendung finden (Pospeschill & Spinath, 2009).

Bei projektiven Verfahren handelt es sich um eine umstrittene Kategorie der psychologischen Testverfahren, da sie nach Kubinger (2006) eine geringe Verrechnungssicherheit aufweisen und ihre Zumutbarkeit fraglich ist. Der Begriff der Projektion lässt sich wiederum auf Sigmund Freud (1856-1939) zurückführen und bedeutet in diesem Zusammenhang: „(...) die Verlegung einer Eigenschaft, die das Ich bedroht und an der eigenen Person nicht wahrgenommen wird, auf eine Person in der Außenwelt“ (Amelang & Schmidt-Atzert, 2006, S. 313). Durch Reize, die auf mehrere Arten gedeutet werden können, werden diese Projektionen hervorgerufen bzw. provoziert. Nach der Theorie von Freud erhält der Tester

dadurch Auskunft über unbewusste und verborgene Elemente der Persönlichkeit eines Klienten (Amelang & Schmidt-Atzert, 2006).

„Die Verhaltensbeobachtung ist eine Methode zur Gewinnung diagnostisch relevanter Daten, die sich immer dann einsetzen lässt, wenn Verhalten zwar wahrnehmbar ist, spezielle formelle Verfahren aber nicht zur Verfügung stehen“ (Amelang & Schmidt-Atzert, 2006, S. 321). Die Arten von Verhaltensbeobachtung werden nach dem Grad der Teilnahme, nach dem Ort und dem Ausmaß der technischen Vermittlung unterschieden. Unter dem Ort der Verhaltensbeobachtung wird zum einen die Laborbeobachtung und zum anderen die Beobachtung im natürlichen Umfeld verstanden. Der Grad der Teilnahme wird in drei abnehmenden Stufen unterteilt: aktive Teilnahme, passive Teilnahme und Nichtteilnahme. Aktive Teilnahme bedeutet, dass der Beobachter direkt in der zu beobachtbaren Situation eingreifen und Einfluss nehmen kann. Während einer Verhaltensbeobachtung mit einer passiven Teilnahme, ist der Beobachter lediglich im Raum und beschränkt sich auf die reine Beobachtung. Anders als bei der aktiven und passiven Teilnahme, wird die Beobachtung bei einer Nichtteilnahme durch den Testleiter über technische Hilfsmittel, wie zum Beispiel Tonband oder Video aufgezeichnet. Dadurch bleibt der Beobachter ungestört und kann den Situationen als neutrale Person folgen. Auch bei Verhaltensbeobachtungen wird versucht, die Gütekriterien weitestgehend einzuhalten. Aufgrund ihrer mangelnden Objektivierbarkeit ist dies nur bedingt möglich. Der Unterschied zwischen allgemeiner und wissenschaftlicher Verhaltensbeobachtung liegt in der Systematisierung von Beobachtung. Die Systematisierung von Verhaltensbeobachtung erfolgt über Segmentierung, der Abbildung von Verhalten, der Verhaltensstichprobe und der Rahmenbedingungen (Pospeschill & Spinath, 2009).

Das Diagnostische Interview wird in Standardisierte Interviews, zu denen auch klinische Interviews gehören und Interviews zur Eignungsdiagnostik untergliedert. Ein Interview ist ein Instrument, mit dessen Hilfe Informationen gewonnen werden sollen. Jedes Interview verwendet zur Erhebung relevanter Informationen das Gespräch. Besonders häufig werden Interviews im Bereich der Marktforschung eingesetzt. „Innerhalb der psychologischen Diagnostik sollte der Begriff `diagnostisches' als Oberbegriff für alle Methoden zur Erhebung von diagnostisch relevanten Informationen mittels Gespräch verstanden werden“ (Amelang & Schmidt-Atzert, 2006, S. 327). Ein wichtiger Begriff im Rahmen der Interviews ist die Anamnese, welche in Anlehnung an die Medizin, als das Erfragen der jeweiligen Vorgeschichte verstanden wird. Exploration ist ebenfalls ein wichtiger Begriff des diagnostischen Interviews und bedeutet die Erkundung des natürlichen Lebensraums eines Probanden (Amelang & Schmidt-Atzert, 2006).

Gruppendiagnostik ist der letzte Bereich der unter Arten der Diagnostik aufgeführt werden soll. Diese wird wiederum in mehrere Arten untergliedert. Die Unterteilung der Gruppendiagnostik richtete sich nach der Art des teilnehmenden Umfeldes. Dazu gehören z.B. Paar-, Familien- und Teamdiagnostik. In den oben genannten Verfahren handelte es ausnahmslos um Verfahren, mit denen psychologische Merkmale einzelner Personen erfasst werden sollten. Dabei fehlte allerdings der Einbezug der sozialen Umwelt eines Menschen.

Angesichts der Bedeutung, die Partnerschaft und Familie, sowie das soziale Umfeld am Arbeitsplatz nicht nur für Staat, Wirtschaft und Gesellschaft, sondern mehr noch für die seelische Gesundheit und Lebenszufriedenheit der darin agierenden Personen aufweist, sind Verfahren von eminenter Wichtigkeit, die sich auf die soziale Interaktion in derartigen Gruppen bzw. der daraus resultierenden spezifischen Phänomene richten (Amelang & Schmidt-Atzert, 2006, S. 344).

Nach Amelang & Schmidt-Atzert (2006) besteht in dieser Art der Diagnostik eine besonders ausgeprägte Verbindung zum Bereich der Interventionen. Da oft erst Störungen im sozialen Miteinander einen Grund für eine Diagnose liefern und normalerweise zu der Intervention Therapie führen, welche am Ende erneut durch eine Diagnostik objektivierbar gemacht werden muss. Die Gruppendiagnostik kann man auch nach den Forschungsansätzen bzw. den Untersuchungsansätzen unterscheiden, in:

1. Kognizierte interpersonelle Merkmale einer oder mehrerer Personen [Hervorhebung v. Verf.]: Vorstellungen, Wahrnehmungen und Emotionen über Beziehungen zu anderen (tatsächlichen oder fiktiven) Personen, mithin interne Repräsentationen.
2. Interaktionelle Merkmale mehrerer Personen in einer sozialen Situation oder in einer zwischenmenschlichen Handlungsepisode [Hervorhebung v. Verf.]: Beobachtung und Analyse konkreter Interaktionen in natürlichen Interaktionen oder in experimentellen Settings (etwa bei der gemeinsamen Lösung einer vorgegebenen Aufgabe).
3. Interpersonelle Eigenschaften einer Person [Hervorhebung v. Verf.]: Personengebundene Interaktionsmerkmale, die von Persönlichkeitsmerkmalen (wie Extraversion oder Dominanz) über Interaktionsstile (z.B. dependenter Interaktionsstil) bis zu konkreten Verhaltensmustern (wie z.B. querulatorisches Verhalten) reichen können (Amelang & Schmidt-Atzert, 2006, S. 344).

Während schon die Paardiagnostik im Vergleich zur Individualdiagnostik neue Ansätze und Theorien benötigt, so gilt dies in einem noch deutlicheren Umfang für die Familiendiagnostik, da hier ein wahres Netz an sozialen Verflechtungen besteht. In diesem Fall sind die bekanntesten Entwickler von Testverfahren im Bereich der Familiendiagnostik diesem

Problem mit der Anwendung von herkömmlichen Methoden begegnet, was zu Methoden wie der Familienskala führte (Amelang & Schmidt-Atzert, 2006).

3.5.3. Erhebungsstrategien

Nach Amelang & Schmidt-Atzert (2006) lassen sich die Erhebungsstrategien in: Gruppen- vs. Einzeltestung, Gestaltung der Testsituation, unimodale vs. multimodale Datenerfassung, einstufige vs. mehrstufige Datenerhebung und Messung von Veränderungen und Unterschieden unterteilen. Unter Einzeltestung wird die Durchführung eines Tests mit nur einem Probanden verstanden. Der Vorteil der Einzeltestung besteht darin, dass eine genauere Verhaltensbeobachtung möglich wird. Gruppentestung bedeutet die Durchführung eines Tests mit mehreren Personen zur gleichen Zeit. Ein Beispiel sind Klassenarbeiten in der Schule, in denen eine ganze Klasse zum selben Zeitpunkt einer Leistungsprüfung unterzogen wird.

Bei der Gestaltung der Testsituation gilt es zunächst die Zielsetzung zu berücksichtigen. Ziel der Gestaltung einer Testsituation ist zum einen die Messung der Leistungsfähigkeit eines Probanden und zum anderen die Selektion von Probanden für eine bestimmte Aufgabe. Im ersten Fall spricht man von Bedingungsselektion. Die Testsituation ist bei der Bedingungsselektion so zu gestalten, dass sie dem Probanden so gut wie möglich entgegenkommt und ihm hilft, seine maximale Leistungsfähigkeit zu zeigen. Dies ist besonders wichtig, wenn der Proband zuvor in einer wichtigen Prüfungssituation versagt hatte und nun geprüft werden soll, ob dies an seiner allgemeinen Leistungsfähigkeit liegt, oder ob er unter optimalen Bedingungen bessere Leistungen erzielen kann. Prüfungsangst ist vielleicht eine der bekanntesten Formen, die eine gesonderte Prüfung nötig machen. So muss davon ausgegangen werden, dass viele Probanden mit einer Testsituation nicht vertraut sind und daher hemmende Effekte auftreten können. So sollte vor einem Leistungstest erst eine Aufwärmphase stattfinden, in welcher sich der Proband in Ruhe mit der Situation anfreunden kann. Besonders bei Probanden, die negative Erfahrungen in Testsituationen gemacht haben, sollte auf eine vertrauenswürdige Atmosphäre während der Durchführung eines Tests geachtet werden. Der Testleiter und dessen Beziehung zum Probanden haben einen großen Einfluss auf die Leistungsfähigkeit während eines Tests. Nachteile einer Bedingungsselektion ergeben sich aus den dort produzierten Ergebnissen, da diese unter Idealbedingungen entstanden sind und daher oft nicht als repräsentativ angesehen werden können. So ist bei einer Selektion von Probanden oder Bewerbern eine

andere Art der Testsituationsgestaltung sinnvoller. Sollen durch den Test repräsentative Ergebnisse erzielt werden, so ist die Testsituation so realitätsnah wie möglich zu gestalten (Amelang & Schmidt-Atzert, 2006).

„Bei ausschließlicher Verwendung nur je eines der bisher besprochenen diagnostischen Verfahren würde man von unimodaler [Hervorhebung v. Verf.] oder multimodaler [Hervorhebung v. Verf.] Datenerfassung sprechen“ (Amelang & Schmidt-Atzert, 2006, S. 371). Im klinischen Alltag findet fast immer die multimodale Diagnostik Anwendung, da meist das Gespräch neben einem Testverfahren eine weitere Methode der Datenerhebung darstellt. Zudem werden besonders in Kliniken gerne Testverfahren der Fremd- und der Eigenbeurteilung kombiniert. Sobald mehrere Methoden der Datenerhebung angewendet werden, wird von multimodaler Diagnostik gesprochen (Amelang & Schmidt-Atzert, 2006).

Unter ein- bzw. mehrstufiger Datenerhebung ist zu verstehen, dass eine diagnostische Untersuchung nicht nur in einer Sitzung, sondern auch in mehreren durchgeführt werden. Meist bestimmen äußere Umstände einer Behandlung, ob ein Klient ein- oder mehrstufig untersucht wird. „Ein Spezialfall von mehrstufiger Datenerhebung, wenn ein- und dieselbe Messung zu wiederholten Zeitpunkten vorgenommen wird, um etwa das Ausmaß der zwischen den Messungen eingetretenen Veränderungen und deren Stabilität abschätzen zu können (...)“ (Amelang & Schmidt-Atzert, 2006, S. 373). Ziel der Messung von Unterschieden ist es, die beiden Werte miteinander zu vergleichen und hinsichtlich ihrer Aussagekraft zu interpretieren (Amelang & Schmidt-Atzert, 2006).

3.6. Die historische Entwicklung der Diagnostik

Historisch gesehen ist die psychologische Diagnostik vielleicht älter als man annehmen mag. So gibt es bereits in der Bibel im Buch der Richter Hinweise auf ein Testverfahren. Dort wurde der Erzählung nach Gideon (ca. 1200 v. Chr.) aufgefordert, Männer anhand verschiedener Kriterien auszuwählen. So fragte er zum Beispiel seine Gefolgsleute, wer sich von ihnen als dumm oder verzagt empfinde, der solle umkehren. Ein ganzer Teil der Männer schien zumindest eine der Eigenschaften bei sich zu entdecken und kehrte um. Zur Auswahl seiner Gefolgsleute ließ er die Männer aus einem Fluss trinken. Diejenigen, die wie Hunde aus diesem Fluss tranken, wurden in die Gefolgschaft aufgenommen (Amelang & Schmidt-Atzert, 2006). Ein weiteres Beispiel für die Historie der Diagnostik entstammt aus alten chinesischen Überlieferungen. Dort wurde jemand, der in ein höheres Amt strebte, einer Reihe an Fragen aus den Bereichen Literatur, Rechnen, Landwirtschaft oder Recht unterzogen. Ein anderer Teil dieser Prüfung bezog sich auf Anwendungsaufgaben wie

Reiten, oder Musizieren. In China war diese Art der Auswahl notwendig, da nicht das Blutrecht die Erbschaft über die Nachfolge regelte. Aus diesem Grund war man auf eine gute Auswahl der Anwärter angewiesen. Grundformen dieser Testung wurden bis ins Jahr 1905 eingesetzt und anschließend von dort nach Europa importiert. Die grundlegende Vorstellung interindividueller Unterschiede und deren wissenschaftliche Untersuchung entstanden erst in den letzten Jahrhunderten. Erste Ansätze lassen sich auf die Evolutionstheorie Charles Darvins (1809-1882) zurückführen, der mit seinen Abhandlungen zur natürlichen Selektion den ersten Grundstein zur Auswahl durch Testverfahren legte (Pospeschill & Spinath, 2009). „Als Vorstufe moderner Diagnostik galten die Bemühungen experimentell arbeitender Psychologen am Ende des 19. Jahrhunderts, psychische Merkmale generell messbar zu - machen“ (Pospeschill & Spinath, 2009, S. 19). Einer der Psychologen, welche über Experimente psychische Merkmale zu erfassen versuchte, war Francis Galton (1822-1911), der den Begriff der Intelligenz als allgemeine Fähigkeit von Menschen untersuchte. Bis zum heutigen Tage wurde die Existenz von Intelligenz nicht mehr infrage gestellt. Die erste psychologische Diagnostik im modernen Verständnis stammt von Alfred Binet (1857-1911). Dieser untersuchte die Leistungsfähigkeit von Schulkindern. Binet, der die einfachen Methoden zur Intelligenz kritisierte, bevorzugte komplexere Indikatoren wie Problemlösen und schlussfolgerndes Denken. Er entwickelte den ersten Vorläufer moderner Intelligenztests. Dabei entdeckte Binet eine Korrelation zwischen Alter und Intelligenzleistung. So entwickelte er verschiedene Schwierigkeitsgrade, die der jeweiligen Altersstufe angemessen waren. Dies bedeutete für ihn, dass eine Aufgabe dann angemessen war, wenn 75% aller Kinder diese Aufgabe lösen konnten. Im Jahre 1912 entwickelte Albert Stern (1878-1966) den Intelligenzquotienten (IQ), ein Wert der Intelligenz unabhängig vom Alter widerspiegeln sollte (Pospeschill & Spinath, 2009). Etwa 25 Jahre nach dem Beginn der Intelligenzdiagnostik begann die jüngste Phase der psychologischen Diagnostik: die Persönlichkeitsdiagnostik. Das erste Verfahren zur Messung von Persönlichkeit war das Personal Data Sheet im Jahre 1918. „Diese Sammlung von 116 Fragen sollte langwierige Psychiaterinterviews während des Ersten Weltkriegs ersetzen“ (Pospeschill & Spinath, 2009, S. 21). Die Fragen des Personal Data Sheet dienten dem herausfiltern von Rekruten, die dem Militärdienst aufgrund ihrer Psyche nicht gewachsen waren. Auf ähnliche Weise erstellte Hans J. Eysenck (1906-1997) auf der Basis klinischer Daten von Soldaten im zweiten Weltkrieg Persönlichkeitsmaße. (Pospeschill & Spinath, 2009) Später entwickelten auch Catell und Norman, die sogenannte lexikalische Methode. Hier wurde aufgrund von Wörterbüchern eine Liste persönlichkeitsbeschreibender Begriffe herausgesucht und unter fünf Oberbegriffen zusammengefasst wurde. Diese Begriffe, dienen bis heute als Grundlage für das Fünf-Faktoren-Modell und die Big-Five. (Pospeschill & Spinath, 2009) Ein paar Jahre nach der Entwicklung des Personal Data Sheets entwickelten

sich die sogenannten projektiven Verfahren. Der bekannteste Vertreter dieses Bereichs ist sicherlich der Rorschach-Test. Hier wurden den Probanden Tintenkleckse gezeigt, die diese dann deuten sollten. (Pospeschill & Spinath, 2009) Bis heute gelten die Bereiche der Intelligenzdiagnostik, Persönlichkeitsdiagnostik und die projektiven Verfahren zu den Anfängen der heute sehr umfassenden Diagnostik (Jäger & Petermann, 1999).

4. Kritik an Diagnostik und ihre Folgen

Im folgenden Abschnitt möchte ich zunächst auf die Kritik an der Diagnostik in den 70er und 80er Jahren eingehen und diese ausführlich erörtern. Anschließend wird ein Vergleich zur heutigen Situation gezogen. Im weiteren Verlauf werden mehrere Bereiche der Diagnostik vorgestellt und hinsichtlich der Anwendung kritisch beleuchtet.

4.1. Kritik im historischen Kontext

Heute ist die Kritik an psychologischer Diagnostik fast vollständig verschwunden. Aktuell ist die Diagnostik ein wichtiger und etablierter Teilbereich der Psychologie und erfüllt nach Westmeyer (2004) wichtige Aufgaben in den Bereichen der Klassifikation, der Erklärung und Entscheidungshilfe. Eine psychologische Praxis ohne Diagnostik ist heute kaum mehr vorstellbar. Jedoch wurde die psychologische Diagnostik vor allem in den letzten Jahrzehnten von mehreren Seiten stark kritisiert und hinsichtlich ihrer Sinnhaftigkeit diskutiert. Dies gilt besonders für die deutschsprachigen Länder mit Ausnahme der DDR. Die Welle der Kritik an Diagnostik war nicht zuletzt dem geschuldet, dass die Diagnostik lange Zeit als das einzige Fachgebiet der Psychologie galt. „Mit seinem Buch Personality and Assessment [Hervorhebung v. Verf.] hatte Mischel (1968) nicht nur die eigenschaftsorientierte Persönlichkeitsforschung, sondern auch die auf diesem Ansatz aufbauende psychologische Diagnostik herausgefordert“ (Westmeyer, 2004, S. 11). Das Buch stellte somit den Beginn der psychologischen Krise dar. Einige Jahre zuvor wurde bereits von mehreren Verhaltenstherapeuten der Begriff der Verhaltensdiagnose als weitere Alternative zur psychologischen Diagnostik eingeführt. Diese und andere kritische Sichtweisen stießen im deutschsprachigen Raum auf enorme Resonanz, welche maßgeblich zur Krise in der psychologischen Diagnostik beigetragen hat. Besonders in den 70er Jahren wurde im

deutschsprachigen Raum, die Kritik am medizinischen Modell psychischer Störungen laut. Dieses Modell dominierte lange Zeit nicht nur die Psychiatrie, sondern auch die klinische Psychologie. Darin wird davon ausgegangen, dass diagnostische Kategorien zeitlich überdauernd sind und zwischen einzelnen Störungsbildern sehr differenziert unterschieden werden kann. Weiterhin beschreibt dieses Modell, dass ohne Therapie das Krankmachende weiter fortwirke und eine Heilung alleine nicht möglich sei. So hatte die klinisch-diagnostische Forschung als Hauptaufgabe die einzelnen Kategorien herauszufinden und zu identifizieren. Die klinisch-psychologische Praxis ordnete die Personen in das von der Forschung entwickelte Klassifikationssystem ein. Genau dieser Ansatz war Ausgangspunkt der Kritik. Nach Westmeyer (2004) wurde Ende der 70er Jahre das Motto „psychisch krank durch Etikettierung“ geprägt. Dieses Motto bewirkte eine intensive Auseinandersetzung mit der Sinnhaftigkeit und den Auswirkungen einer Diagnose. Westmeyer (2004) hebt in diesen Zusammenhang insbesondere den Prozess der Etikettierung eines Menschen und die damit verbundenen Schwierigkeiten und eventuell zusätzlich auftretende Störungen hervor. Kritik an der psychologischen Diagnostik erfolgte auch sehr stark durch die gesellschaftstheoretische Perspektive. So diente die traditionelle Diagnostik in vielen Bereichen als Selektion. Die Auslese an Personen weckte seitens der Kritiker die Befürchtung, dass bei einer solchen immer Menschen zurück bleiben oder auf der Strecke bleiben müssen. In diesem Fall wird von einer zwangsläufigen Benachteiligung ausgegangen. Diagnostik würde dazu beitragen, bestimmten Menschen den Zugang zu Ressourcen oder Chancen zu verwehren. „Dahinter standen, das war für die Proponenten dieser Perspektive klar, mehr oder weniger unverhüllt die Verwertungsinteressen des Kapitals. Einer psychologischen Diagnostik, die sich diesen Interessen nicht entgegenstellte, wurde die Existenzberechtigung abgesprochen“ (Westmeyer, 2004, S. 11). Für viele passte die Diagnostik einfach nicht in ihr Welt- und Menschenbild. So wurde der Mensch selbst für veränderbar gehalten, die Diagnostik sprach aber von stabilen Eigenschaften und genetischer Determinierung. Die von der Diagnostik postulierten Eigenschaften passten nicht zu dem Glauben und den Einstellungen der 70er Jahre. Diese Kritik verunsicherte viele Diagnostiker. So wurde in der Schweiz diskutiert, ob eine Fahrtauglichkeitsprüfung im Hinblick auf die starke Kritik an Diagnostik noch zu rechtfertigen sei. Weiterhin wurde auch der Bereich der Bildungspsychologie infrage gestellt. Befürworter der Diagnostik hoben die Bedeutung des genetischen Einflusses auf Intelligenz hervor und versuchten diese empirisch zu belegen. Dies war der Auftakt zur später aufkommenden IQ-Kontroverse. Dabei spielten Intelligenztests eine zentrale Rolle. Hinterfragt wurde nicht nur deren Wissenschaftlichkeit, sondern auch die dahinter liegende Theorie. Die Proponenten stützten sich in ihren Aussagen auf die Ergebnisse und Resultate von Intelligenztests. Die Kontrahenten dagegen stellten die Tests als solche bereits in Frage und so konnte auch keine gemeinsame

Grundlage in dieser Diskussion gefunden werden. Zusätzlich hatten die Gegner der Intelligenztests auch die damalige politische Stimmung hinter sich, wodurch die Proponenten zusätzlich unter Druck gerieten. Eysenks Abhandlungen über die Ungleichheit der Menschen (1975) wurden besonders in der damaligen Zeit als Provokation aufgefasst - ein Effekt, den der Autor wohl bewusst hervorrufen wollte. In einem Symposium der schweizerischen Gesellschaft für Psychologie wurden mehrere Themen unter der Überschrift: „Krise der Diagnostik“ diskutiert. Diese Diskussion wurde in der Schweizerischen Zeitschrift für Psychologie aufgegriffen und weitergeführt. Zentrale Diskussionspunkte waren in diesem Zusammenhang die Diagnostik, welche die persönliche Freiheit und Autonomie beeinträchtigt, sodass die zur Verfügung stehenden Mittel zu einem großen Teil ungenügend seien und zu falschen Ergebnissen führen würden. Nachfolgend wurde darüber diskutiert, ob der Versuch der Diagnostiker, ein Wesen, das sich in einem ständigen Wandel befände, mit konstanten und prognostizierbaren Fähigkeiten und Eigenschaften zu beschreiben, nicht im Prinzip verfehlt sei. Doch war das Ziel dieser Diskussion nicht die Eliminierung der Diagnostik, sondern ihre Verbesserung. Zweifel an den zentralen methodischen Hilfsmitteln in dem Bereich der eigenschaftsorientierten Persönlichkeitsforschung, sowie der Diagnostik förderten zusätzlich die damalige Krisenstimmung. Es wurde nicht nur die Faktorenanalyse, sondern auch die gesamte klassische Testtheorie infrage gestellt. Diese hatte zuvor jahrzehntelang die unangefochtene Basis der Konstruktion psychologischer Tests dargestellt. In dieser Zeit gab die Diagnostik ein komplett uneinheitliches Bild ab. An die pädagogische Diagnostik wurde niemals zuvor solch widersprüchliche Anforderungen gestellt (Ingenkamp, 1981). „Von der Forderung nach methodischer Perfektionierung der schulischen Beurteilung unter Einsatz verschiedener Testarten, Schätzskalen, Kriterienkataloge, verstärkter Häufigkeit der Überprüfung in enger Anlehnung an die curricularen Schritte bis zur Forderung nach Abschaffung jeder Überprüfung, die einen Vergleich von Schülern einschließt, reicht die Skale“ (Ingenkamp, 1981, S. 20). Doch einige Jahre später verebte die Krise, welche nie wirklich existentiell für die Psychologie oder eines ihrer Teilbereiche war. Nach der Schaffung offener Streitpunkte und Fragen wurde nun rein pragmatisch nach Lösungen gesucht. Die Krise hatte nie die Notwendigkeit aufgeworfen, die Diagnostik umfassend zu verändern oder zu revolutionieren. Diese aber führte zu einer konstanten Weiterentwicklung der Diagnostik. Das Ende der Krise wurde von den Kritikern der Diagnostik selbst ausgelöst, da diese gemeinsam mit Befürwortern der Diagnostik nach Lösungen suchten. So verebte die Kritik an den Grundlagen der Diagnostik und machte dem modernen Bild der Diagnostik als anerkannte und empirisch belegte Wissenschaft Platz (Westmeyer, 2004).

4.2. Kritische Betrachtung der Diagnostik heute

Die Kritik an den Grundlagen der Diagnostik aus den 70er Jahren ist einer Diskussion einzelner Teilgebiete gewichen. So wird bis heute z.B. über das Konstrukt und die Testung von Intelligenz diskutiert. Die Anwendung von Diagnostik zählt in bestimmten Bereichen auch dazu. Einzelne Methoden der Diagnostik werden auch heute noch diskutiert. So gibt es weiterhin Uneinigkeit betreffend der Anwendung projektiver Verfahren. Im folgenden Teil dieser Arbeit werden einige dieser Bereiche aufgegriffen und kritisch hinterfragt.

An dieser Stelle werden die Konzentrationstests erneut aufgegriffen und hinsichtlich ihrer Fehlerquelle und Risiken diskutiert. Konzentrationstests, zu denen auch die Durchstreichtests, wie der D2 gehören, sollen nun erläutert werden. Ein Problem der Durchstreichtests sind die freien Bereiche, die, wenn man es analysiert, keinen Rückschluss auf die Konzentration erlauben. Hierbei bleibt die Frage offen, inwieweit die unbearbeiteten Abschnitte überhaupt bemerkt oder beachtet wurden. Um dieses Problem zu begegnen, wurde das vollständige Markierungsprinzip entwickelt. Bei diesem Prinzip werden auch die nichtrelevanten Abschnitte mit anderen Zeichen markiert. Kritisch betrachtet wird bei dem D2 auch die Differenzierung zwischen den Buchstaben d und p, welche besonders für Legastheniker eine zusätzliche Herausforderung darstellt (Kubinger, 2006). Doch auch Rechentests als zweite Form der Konzentrationstests sind nicht ohne mögliche Fehlerquellen. Ein Rechentest prüft eher die Rechenfähigkeit eines Probanden, als die Konzentrationsleistung. Daher ist ein Rechentest bereits hinsichtlich seiner Validität kritisch zu betrachten. Bei einem Rechentest wird davon ausgegangen, dass alle Probanden den gleichen Automatisierungsgrad im Rechnen erreicht haben. Dies ist aber eine Annahme, die hinsichtlich der sehr unterschiedlichen Bildungsgrade und Schulbiografien sehr kritisch angesehen wird. Ein Rechentest kann nur dann realistisch die Konzentration messen, wenn sich die Probanden nur darin unterscheiden. Doch spätestens seit dem Auftauchen von Dyskalkulie kann nicht mehr ausgegangen werden, dass sich die Probanden hinsichtlich der Konzentration voneinander abgrenzen (Rauchfleisch, 2005).

Eignungstests sind im Gegensatz zu Leistungstests Kriterien näher konstruiert. Diese versuchen demnach ein bestimmtes Kriterium möglichst genau zu erfassen. Besonders häufig werden Eignungstests in Assessment-Center zur Personalauswahl und bei anderen Vorauswahlen von Bewerbern eingesetzt. Besonders für das Medizinstudium und für die Ausbildung bei der Polizei müssen Bewerber Eignungstests durchlaufen. Besonders bei

bekanntesten Eignungstests besteht die Möglichkeit diese zu trainieren. In diesem Falle würde ein Test keine Auskunft mehr über die Studientauglichkeit, wie im Falle des TMS, mehr geben, sondern Kriterien, wie Einsatz und Lernbereitschaft ermitteln. Ein besonderes Risiko bei der Anwendung von Eignungstests ist die Verwirtschafterlichung der Vorbereitung. So bieten erfahrene Testleiter ihre Dienste gegen Geld an und werben damit, die Personen optimal auf den jeweiligen Eignungstest zu vorzubereiten. In diesem Fall würde nicht mehr die eigene Motivation, sondern die Verfügbarkeit von finanziellen Mitteln darüber entscheiden, wer die begehrten Ausbildungs- und Studienplätze oder ähnliches bekommt. Eine genaue Untersuchung, in wie weit sich die Vorbereitungen auswirken, ist deswegen kaum zu realisieren, da die angebotenen Vorbereitungen in Umfang und Art sehr stark schwanken. So gibt es Kurse die nur einstündig sind und nur grob durchgesprochen wird, was die Klienten erwartet. Auf der anderen Seite sind Kursangebote auf dem Markt, die über ein Jahr dauern und in denen sehr exakt geübt und vorbereitet wird. Doch grundlegend kann man sagen, dass die Bereitschaft sich intensiv vorzubereiten, eine durchaus gewünschte Eigenschaft darstellt, die sogar von den Testanbietern gefördert wird (Hänsgen & Hofer, 1996).

„Das Hauptproblem bei den Entwicklungstests liegt darin, dass es bekanntlich bei Kindern außerordentlich schwierig ist, mit hinlänglicher Verlässlichkeit ein bestimmtes Persönlichkeitsmerkmal zu erfassen und zu Verhaltensvorhersagen aufgrund von Testuntersuchungen an Kleinkindern zu gelangen“ (Rauchfleisch, 2005, S. 138). Besonders bei Kindern ist die intraindividuelle Leistungsfähigkeit stark schwankend. Rauchfleisch (2005) stellte fest, dass ein Kind auch wenn es völlig normal ist, stark in seiner Leistungsfähigkeit variiert. So ist eine Aufgabe an einem Tag kein Problem und das Kind absolviert diese mit Leichtigkeit und am nächsten Tag kann es diese Aufgabe nicht mehr lösen. Zudem spielt die affektive Beziehung zwischen dem Kind und den Eltern eine maßgebliche Rolle und beeinflusst stark die Fähigkeit des Kindes, sich in der Testsituation zu äußern. Rauchfleisch (2005) zieht dabei folgendes Fazit: Entwicklungstests können nur mit großer Vorsicht an Kindern durchgeführt werden und es ist besser, je länger der Tester das Kind und seine Fähigkeiten kennt. Zudem gilt, je jünger die Kinder sind, umso ungenauer und irrtümlicher können die Ergebnisse von Entwicklungstests ausfallen.

Eines der zentralen Probleme der Schultests ist die Schwierigkeit, Kriterien die besonders valide, aber auch reliabel sind zu finden. Wobei speziell im Bereich der Eignungsdiagnostik, zu der auch die Schultests gehören, eine einmalige Testung nicht als ausreichend angesehen wird. So würde die Diagnostik, wenn sie den entsprechenden Forderungen nachkäme, zu einem Begleitinstrument. Ein Instrument, das während der gesamten Zeit einer Ausbildung durchgeführt wird, könnte die konstante Leistungsentwicklung

wiederspiegeln. Diese Art jedoch ist bislang nicht weit verbreitet. So werden für Menschen lebenswichtige Entscheidungen anhand der Ergebnisse einzelner Tests getroffen (Rauchfleisch, 2005). Rauchfleisch (2005) erwähnt dabei, dass besonders verhaltensgestörte Kinder von Entscheidungen aufgrund der Tests negativ betroffen sind. So untersuchte der Autor Schüler, die die Sonderschule mit einem IQ von 120 durchliefen. Er kam zu dem Ergebnis, dass es nicht die mangelnde Leistungsfähigkeit sei, sondern die Verhaltensstörungen, die zu schlechten Testresultaten führten. So zeichnet sich im pädagogischen Bereich die Tendenz ab, die Beurteilung von Leistungsfähigkeit objektive Dritte überlassen. Nach Rauchfleisch (2005) ist dieser Wunsch auch vollständig nachzuvollziehen, da sich ein Lehrer bei der Beurteilung nicht von Sympathie und Antipathie leiten lassen möchte. Doch genau das spiegelt die Gefahr wieder, die Rauchfleisch (2005) in Schult- und Leistungstests sieht. Geht man davon aus, dass besonders im entwicklungspsychologischen Bereich die Testbeurteilung die weiteren Wege der Klienten maßgeblich beeinflussen kann und zum Teil Karrieren deswegen unmöglich werden, dann wird die Kritik besonders an der einmaligen Testung prekär. Im Weiteren kritisiert Rauchfleisch (2005) auch die Mitteilung von Testresultaten, die nach seiner Meinung nach nicht ausreichend mit den Klienten und deren Eltern besprochen wird.

Wie bereits erwähnt, erfreuen sich die Persönlichkeitsfragebogen auch deswegen so großer Beliebtheit, da sie von den Probanden alleine ausgefüllt werden können und niemand nötig ist, der bei diesen Tests anwesend sein muss. Im Allgemeinen sind diese Tests so aufgebaut, dass den Klienten eine Reihe von Fragen vorgelegt wird, die dann zu beantworten sind. Doch genau in diesem eigenständigen Beantworten liegt eine der großen Gefahren dieser Testart. Ein Test, der von dem Klienten selbst ausgefüllt wird, unterliegt starken Übertreibungs- und Verfälschungseffekten. Ein weiteres Problem sind die Begrifflichkeiten selbst, die in den jeweiligen Testformen verwendet werden. Die Probanden gebrauchen diese Begriffe alle unterschiedlich, was wiederum zu einem Verfälschungseffekt führen kann. Rauchfleisch (2005) sieht Persönlichkeitsfragebögen hinsichtlich ihrer ethischen Richtlinien kritisch. So geht der Testleiter davon aus, dass die Fragen, die dem Probanden gestellt werden, ihm keine seelischen und moralischen Verletzungen zuführen. Doch auch vermeintlich „harmlosen“ Fragen können durch psychopathologische Sensibilisierung anders und damit als unethisch wahrgenommen werden.

Nicht unproblematisch ist auch die Tatsache, dass wir bei der Auswertung solcher Skalen die verschiedenen Symptome addieren [Hervorhebung v. Verf.] und daraus einen Gesamtwert für ein bestimmtes Syndrom errechnen. Dabei werden zum Teil Symptome gleichwertig behandelt, die im Erleben und im Alltag von Patienten völlig verschiedene Relevanz besitzen (beispielsweise werden Suizidversuche, Versündigungsideen und

verschiedene bei Depression auftretende körperliche Störungen, ohne dass diese Symptome unterschiedlich gewichtet würden, einfach addiert) (S. 171).

Weiterhin ist darauf zu achten, dass ein solcher Fragebogen die subjektive Sicht eines Klienten widerspiegelt. Daher sollte nicht unabhängig von Individualbefunden oder wissenschaftlichen Untersuchungen davon gesprochen werden, dass sich der Proband durch dieses oder jenes Persönlichkeitsmerkmal auszeichne. Nach Rauchfleisch (2005) sind Persönlichkeitsfragebögen, obgleich sie sehr ökonomisch sind, sehr unpersönliche Verfahren, da der Klient sie alleine ausfüllt. Somit ist diese Art der Testverfahren bei vielen Klienten nicht sehr gemocht und es sollte drauf geachtet werden, in einer Testbatterie eine Mischung aus verschiedenen Verfahren auszuwählen, um diesen Umstand zu relativieren.

Die Theorie hinter einem projektiven Verfahren wird durch seine Kriterien und Konstrukte bestimmt. Somit fällt auch eine Einteilung in eine der Testarten schwer, da nicht immer klar zu unterscheiden ist, ob dieses Verfahren nun eher in den Bereich der Persönlichkeitsdiagnostik fällt, oder eher der Leistungsdiagnostik zugeordnet werden kann. Dies liegt an der divergenten Darstellung der Begrifflichkeiten von Projektion und auch an der Unklarheit darüber, welche Variante gemeint ist. Amelang & Schmidt-Atzert (2006) nennen ein Beispiel für ein Problem, welches bei einem projektiven Verfahren auftauchen kann: Wenn ein Klient ein Tiger sein möchte, der beißen kann, ist nun die Konsequenz dieses Wunsches, dass er aggressiv ist? Oder ist er nicht aggressiv und möchte es gerne sein? Ein Tiger, der in der Versuchsperson die Aggressivität anspricht und nicht etwa den Aspekt der Ästhetik liefert zwar einen Hinweis darauf, dass Aggressivität bei dem Klienten eine Rolle spielt, aber es wird nicht deutlich welche.

Da man dem Testverhalten selbst nicht ansehen kann, wie es zustande gekommen ist und auf welcher Ebene es demzufolge interpretiert werden kann, bedarf es zusätzlicher Informationen etwa der folgenden Art: Aufgrund empirischer Untersuchungen muss in Erfahrung gebracht werden oder ist bereits bekannt, dass

- Ein bestimmter Test immer auf einer Ebene „funktioniert“
- Ein bestimmter Proband immer auf einer Ebene „funktioniert“
- Ein bestimmter Proband in dem bestimmten Test auf einer bestimmten Ebene „funktioniert“ (Amelang & Schmidt-Atzert, 2006, S. 313-314).

So kann festgehalten werden, dass ein Test bei den Klienten auf verschiedenen Ebenen funktioniert. Nach Amelang & Schmidt-Atzert (2006) ist dies für projektive Verfahren sehr wahrscheinlich. Eine niedrige Validität ist die Folge. Doch ungeachtet dieser Schwierigkeiten haben projektive Verfahren bis heute eine große Faszination. Als gutes Beispiel dient der

Rorschach-Test. Ein Problem von einem Rorschach-Test ist die nicht festgelegte Zahl der Antworten. Die Auswertung variiert daher bei gleichem Ergebnis von Auswerter zu Auswerter. Daher ist die Auswertungsobjektivität des Rorschach-Tests nicht gegeben. Dies gilt ebenso für die Reliabilität, da die Tafeln alle verschieden sind. Amelang & Schmidt-Atzert (2006) kommen zu dem Ergebnis, dass der Rorschach-Test kaum als objektives, reliables und/oder valides Verfahren angesehen werden kann.

Bei Beobachtungsverfahren wird im Allgemeinen die Objektivität kritisiert. Interessant ist dabei die Erkenntnis, dass die Genauigkeit der Wiedergabe mit dem Wissen einer möglichen Kontrolle korreliert. Im Weiteren können Fehlerquellen, wie z.B. der Halo-Effekt, der logische Fehler und der Milde- und Strengefehler auftreten. Ein Mildefehler entsteht immer dann, wenn ein Beobachter aus Angst zu schlechten Ergebnissen zu kommen, einen zu hohen Mittelwert erzeugt. Im Gegensatz zum Mildefehler wird von einem Strengefehler gesprochen, wenn der Beobachter negativ bewertet, um einen geringeren Mittelwert zu erzeugen (Amelang & Schmidt-Atzert, 2006).

Auch die Erhebungsmethoden sollen nun einem genauen Blick unterworfen und kritisch beleuchtet werden. Zunächst ist die Gruppentestung im Vergleich zur Einzeltestung hinsichtlich der Aufwendung von Zeit sehr ökonomisch. So kann in der gleichen Zeit nicht nur eine Person, sondern eine große Anzahl an Probanden getestet werden. Dies hat den Vorteil, dass viele Ergebnisse, die unter ähnlichen Umständen erfasst wurden, erzielt werden können. Der einzige Grund, warum heute immer noch Einzeltests durchgeführt werden, liegt daran, dass viele Tests keine andere Durchführung zulassen. Zudem können durch Einzeltests eine deutlich genauere Verhaltensbeobachtung erreicht werden. Ein weiteres Problem bei Gruppentestung ist die mögliche Benachteiligung einzelner durch z.B. eine Lese-Rechtschreibschwäche eines Probanden. Weitere Nachteile der Gruppentestung sind die möglichen Störungen durch Dritte und die Gefahr des Abschreibens. Eine Möglichkeit dem Abschreiben zu begegnen ist die Entwicklung von Pseudoparalleltests, bei denen sich die Aufgaben zwar gleichen, diese aber in einer unterschiedlichen Reihenfolge angeordnet sind (Pospeschill & Spinath, 2009).

Auch die Art der Fragestellung sollte hinsichtlich ihrer Nützlichkeit für psychologische Testverfahren hinterfragt werden. So benötigen z.B. offene Fragen sowohl in der Beantwortung, als auch in der Auswertung mehr Zeit. Offene Fragen sind zudem in ihrer Auswertungsobjektivität eingeschränkt. Bei einem gebundenen Antwortformat hingegen können z.B. Rateeffekten auftreten, die das Ergebnis verfälschen. Diagnostische Entscheidungen zur Generierung von Informationen werden von dem Testleiter getroffen. Dieser aber unterliegen wie alle Menschen den allgemeinen Urteilsfehlern. Zu den

allgemeinen Urteilsfehlern gehören z.B. falsche Schlussfolgerungen, oder der Halo-Effekt (Pospeschill & Spinath, 2009).

4.3. Kritik in der Anwendung pädagogischer Diagnostik

Diagnostik im Rahmen der Pädagogik ist grundlegend auf Intervention ausgelegt. Zu ihr gehören zwei große Bereiche: Schullaufbahnberatung und individuumszentrierte Schülerhilfe. Intervention im Rahmen der Schullaufbahnberatung beschränkt sich primär auf Bedingungsselektion. Dagegen wird bei individueller Schülerhilfe unter Intervention eher eine Beratung verstanden. Der Einsatz von diagnostischen Verfahren wird in diesem Bereich durch die Anpassung von Lernvoraussetzungen und schulischen Anforderungen begründet (Ingenkamp & Lissman, 2008).

Bis heute wird im Besonderen den Lehrern vorgeworfen, dass ihre Schulleistungsbeurteilungen sehr inobjektiv und unzuverlässig seien. Besonders die Vorhersagequalität von Urteilen über Schüler im Einzelfall sei ungenügend. „Zumindest im Hinblick auf die mangelnde Vorhersagegültigkeit besteht ohne Zweifel die ernstzunehmende Gefahr, daß für Schüler unangemessene Übergangsentscheidungen getroffen werden, sofern solche Entscheidungen von den bisherigen Leistungen eines Schülers abhängig gemacht werden“ (Rheinberg, 1986, S. 28). Doch nicht nur die Schulleistungsbeurteilung wird im Bereich der pädagogischen Diagnostik kritisiert. Ein anderer Bereich ist die Anwendung von motivationspsychologischen Konzepten im Schulunterricht. Hier geht die Gefahr vom Bezugssystem aus. Dies bedeutet, dass ein Schüler zu einem bestimmten Zeitpunkt an der Leistungsfähigkeit seiner Mitschüler gemessen wird. Daher wird in einem solchen Fall von einem normorientierten Schulleistungstest gesprochen. Die individuelle Entwicklung eines Schülers wird dabei nicht berücksichtigt, sofern die Mitschüler den gleichen Wissenszuwachs erfahren haben. Es kommt ebenfalls darauf an, dass ein Schüler lernt, sich nicht nur unter sozialen, sondern auch unter individuellen Bezugsnormen zu beurteilen. Besonders in Hinblick auf die leistungsschwachen Schüler ist es sehr unvorteilhaft, wenn ihre Leistungen alleine durch den sozialen Vergleich erfolgen, da sie dann immer schlechter abschneiden, als die leistungsstärkeren Schüler. Dies wiederum hat negative Auswirkungen auf deren Motivation. In einem Feldexperiment von Williams, Pollack, & Ferguson (1975) zeigte sich, dass leistungsschwächere Schüler im Unterschied zu einer Schülergruppe, deren Leistungen individuell zurückgemeldet wurde, ein Leistungsvergleich mit einer Klasse, sich negativ auf die Motivation auswirkte. Doch eine allgemeine Kritik auf

die pädagogische Diagnostik und deren Leistungsrückmeldungen können nicht abgeleitet werden, da Lehrer neben den rein sozialen Kriterien auch individuelle Bezugsnormen als Rückmeldung nutzen. Besonders zur Interpretation der Leistungsfähigkeit von Klassen oder Schülern, treten Probleme durch die Freiheit eines Lehrers auf, vom eigentlichen Lehrplan abweichen zu können. Auch Noten selbst sind, wenn man die immensen Schwankungen der Noten für eine Arbeit betrachtet, ein sehr inexaktes Instrument, um Leistungen widerzuspiegeln. Dies sind Aspekte, die eine exakte und genaue Diagnostik im pädagogischen Bereich erschweren. „Damit besteht die Gefahr, dass Unterrichtsmethoden oder Lehrstiele fälschlich als gleichwertig angesehen werden, weil die eingesetzten Meßinstrumente entweder etwas anderes als die unterrichteten Inhalte messen und/oder zu unsensibel für Unterschiede im Lernzuwachs verschiedener instruierter Lerngruppen sind (Rheinberg, 1986, S. 30). Ein weiterer Kritikpunkt ist die selektive Wahrnehmung von Lehrpersonen und deren unterschiedlichen Erwartungshaltungen. So sind Lehrer oft von dem Halo-Effekt betroffen und warten bei Schülern, die sie für begabt halten, länger auf eine Antwort und geben auch mehr Hilfestellung bei der Lösung (Rheinberg, 1986).

4.4. Kritik in der Anwendung psychologischer Diagnostik

„Wenn Diagnostik letztlich das diagnostizierte Subjekt erst hervorbringt, das sie objektiv zu erfassen beansprucht, sollten wir dann Diagnostik nicht grundsätzlich ablehnen?“ (Schrödter, 2006, S. 22) Nach Schrödter (2006) sei die psychologische Diagnostik als Mittel ethisch sehr problematisch. Dies sei vom positiven Hintergrund ihres Einsatzes und der aus Diagnostik entstehenden positiven Effekte unabhängig. Er begründet dies mit Machteffekten, die in der Diagnostik eine Rolle spielen. Daraus ergibt sich die Notwendigkeit diagnostische Mittel nur dann einzusetzen, wenn diese frei von Machteffekten sind. Der Kontext einer diagnostischen Situation bestimmt auch die Art und die Ausprägung der jeweilig vorherrschenden Machteffekte. Es ist somit wichtig, sich von einer polarisierenden Haltung gegenüber Diagnostik zu verabschieden. Als eine mögliche Macht nennt Schrödter (2006) die Normierungspraxis. Um diese beurteilen zu können, muss zunächst klargestellt werden, welchen Sinn und Zweck die Diagnostik in der jeweiligen Situation haben soll. Grundlegend wird das Diagnose- und Klassifikationssystem hinsichtlich der Bildung einer gemeinsamen Sprache, Erleichterung einer Homogenisierung von Gruppen und hinsichtlich der Systematisierung des Wissens einer Disziplin unterteilt. Ein Klassifikationssystem stellt eine gemeinsame Sprache zwischen den Professionen und den Mitarbeitern dar. Die Begriffe eines solchen Systems sind operationalisiert und damit kritisierbar. Weiterhin ist eine Funktion der Klassifikationssysteme das Beschreiben und die Homogenisierung von

Populationsgruppen zu erleichtern. Zudem systematisiert ein Klassifikationssystem auch das Wissen einer Disziplin. Daraus folgt, dass Klassifikationssysteme besonders in sozialpädagogischer Sicht eine Basis für Entscheidungen sind. Doch können sie interventionspraktische Entscheidungsprobleme nicht selbst lösen. Ferner kann die psychologische Diagnostik aus Gesichtspunkten der Stigmatisierung kritisch betrachtet werden. In diesem Fall soll durch das Beispiel der Schwulenbewegung die Gefahr von psychologischer Diagnostik vorgestellt werden. Innerhalb der Schwulenbewegung besteht die Hoffnung auf die baldige Entdeckung eines Gens, welche für die Homosexualität verantwortlich gemacht werden kann. Von der psychologischen Forschung wird demnach erwartet, dass sie das entsprechende Gen bald gefunden haben. Doch was wäre die Folge? Sollte jemand dieses Gen nicht aufweisen, würde diese Person erst recht als wiedernatürlich etikettiert werden. So zeigen sich deutlich die Tendenz und die Gefahr, Lebensentwürfe aus Diagnosen und Diagnostik abzuleiten. „Als ‚Wahrheitsspiele‘ geben Diagnosesysteme also häufig die Rahmung vor, welcher definiert, was überhaupt als wahr und falsch gelten kann“ (Schrödter, 2006, S. 22). Dabei ist ein häufiges Argument, dass diagnostische Kategorien lediglich Konstrukte sind, die historisch kontingent sind und sich genauso auch anders entwickelt haben könnten. Besonders kritische Stimmen behaupten, dass eine Krankheit oder eine Diagnose etwas ist, dass nur erfunden wurde und somit eine solche Störung gar nicht existiert. Ein Beispiel ist auch die Ungenauigkeit solcher Klassifikationen. Ein Testverfahren untersucht Probanden hinsichtlich eventuell bestehender Lernstörungen. Kinder und Erwachsene sollen zu diesem Zweck getrennt untersucht werden. Diese Begriffe lassen jedoch keine genaue Abgrenzung zu, da nicht eindeutig definiert werden kann, ab wann ein Proband als erwachsen oder als Kind gilt. Dieses sprachliche Problem lässt sich ebenso auf andere Bereiche übertagen. Auch bei einer psychischen Krankheit ist nicht eindeutig definiert, ab wann genau eine Störung vorliegt. So kann der Diagnostik trotz aller Bemühungen ein bestimmtes Maß an Willkür vorgeworfen werden. Schon bei der Abgrenzung alltäglicher Begriffe wird dieses Problem deutlich. So kann nicht eindeutig definiert werden, zu welcher Minute der Wechsel zwischen Tag und Nacht stattfindet. Doch wäre es nach Schrödter (2006) ein Fehler von der Schwierigkeit der Abgrenzung sofort auf ein willkürliches Konstrukt zu schließen. Willkürlich wäre ein Konstrukt nur dann, wenn man es durch ein beliebig Anderes ersetzen könnte und genau das ist nicht möglich. „Die normativen Vorannahmen, auf die Diagnosesysteme gründen, dürfen niemals unhinterfragt bleiben“ (Schrödter, 2006, S. 23). Auch die sprachliche Grundlage kann bereits ein Kritikpunkt sein. So wäre jemand, der eine Sprache nicht gut beherrscht, in einer Testsituation alleine schon aus diesem Grund benachteiligt. Die meisten Tests, die im Rahmen meiner Praktika durchgeführt wurden, waren alle sprachgebunden. Als Beispiel soll der Intelligenz-Struktur-Test (IST-2000) aufgegriffen werden. Diesen Intelligenztest habe ich

während eines Praktikums in einer Suchtklinik selbst mit einem Probanden mit Migrationshintergrund durchgeführt. Während der Durchführung wurde deutlich, wie schnell er an die Grenzen seiner sprachlichen Fähigkeiten stieß. Das Ergebnis des Tests war ein IQ-Wert, der unter dem Durchschnitt und nur knapp über dem Wert der geistigen Behinderung lag. Für mich stellt sich in diesem Zusammenhang die Frage, ob das Ergebnis des Probanden seinen wahren Intelligenzwert widerspiegelte oder ob dieses durch seine Sprachbarriere zustande kam. Der IST-2000 liefert somit ein gutes Beispiel für die kritische Sichtweise auf Intelligenzdiagnostik. Die Intelligenzdiagnostik soll Auskunft über die maximale Leistungsfähigkeit eines Probanden geben. Wie aus dem Beispiel ersichtlich wird, trifft dies nicht zwangsläufig auf alle Probanden zu. Im Weiteren ist zu erwähnen, dass nicht nur bei Probanden mit Migrationshintergrund, sondern auch bei Probanden mit einer geringen Schulbildung oder einer Lese-, Rechtschreibschwäche, Schwierigkeiten während der Durchführung von Tests auftreten können. In beiden Fällen hat ein Konstrukt, das nicht zu Intelligenz gehört, einen großen und maßgeblichen Einfluss auf die mögliche Leistung in einem Intelligenztest. Beim Durchsehen der Rechenaufgaben der verschiedenen Leistungstests, ist mir aufgefallen, dass das Rechnen mit Hochzahlen Bestandteil des Rechenteils wie z.B. im IST-200 ist. Doch nur Probanden, die eine höhere Schulbildung haben, sind effektiv in der Lage, diese Aufgaben zu lösen. Für mich ist dies eine klare Benachteiligung von Schülern aus bildungsarmen Hintergründen und stelle mir daher die Frage, wie genau Intelligenz wirklich gemessen werden kann. Als weiteren Kritikpunkt aus meinen Erfahrungen in den Praktika, sehe ich in der Zumutbarkeit und Ökonomie von Tests. Die Patienten bekamen in den Rehabilitationseinrichtungen meiner Praktika zu Beginn und zum Ende der Therapie eine Test- und Fragebogenbatterie. Seitens der Patienten wurde der enorm hohe Zeitaufwand beklagt, den diese Verfahren in ihren Klinikalltag verursachten. Teilweise berichteten die Patienten davon, das gesamte Wochenende für die Beantwortung der Testbatterie benötigt zu haben und somit zu wenig Zeit vorhanden war, sich in den Klinikalltag einzuleben. In diesem Zusammenhang stellte sich mir die Frage, inwieweit diese Fülle an Tests wirklich sinnvoll ist und ob die Anzahl der Fragebogen nicht etwas reduziert werden kann. Eine Betreuerin in der Rehabilitationsklinik bestätigte, dass die wichtigsten Informationen aus den Testverfahren SCL-90R und dem BDI-II gewonnen wurden. Gründe, warum auf die weiteren Testverfahren nicht verzichtet wurde, haben sich mir nicht erschlossen. Meinem Erachten nach sollte hinsichtlich der Zumutbarkeit und Ökonomie der Einsatz von psychologischen Tests stärker in der klinischen Praxis abgewogen werden.

4.5. Kritik in der Anwendung psychiatrischer Diagnostik

Bei jedem Menschen wird das Urteilsvermögen von allgemeinen Urteilsfehlern geleitet. Dies trifft in gleichem Maße auch auf professionelle Diagnostiker zu. Auch diese werden in ihren Entscheidungen von Urteilsfehlern bestimmt. „Hierbei kann es sich zum Beispiel um Fehler aufgrund falscher Schlussfolgerungen handeln (z.B. logischer Fehler: Annahme, daß bestimmte pathologische Phänomene immer zusammen auftreten müssen) oder den sog. Halo-Effekt (ein besonders Merkmal beeinflusst die Wahrnehmung anderer Merkmale) (Stieglitz, 2000, S. 16). Wichtig ist dabei aber, dass Risiken und Fehler nicht nur durch den Diagnostiker entstehen, sondern auch durch den Klienten. Stieglitz (2000) sieht primär zwei Fehlerquellen:

- Unwissentliche Fehler [Hervorhebung v. Verf.] bedingt durch Erinnerungsfehler, Selbstbeurteilungs- und Beobachtungsfehler
- Absichtliche Verfälschungen [Hervorhebung v. Verf.] wie zum Beispiel Simulation (absichtliche Vortäuschung eines Symptoms), Dissimulation (absichtliche Negierung eines Symptoms), Bagatellisierung (absichtliches Herunterspielen eines Symptoms) oder Aggravation (absichtliches Herausstellen eines Symptoms) (S. 17).

Weiterhin liegt eine Quelle für Fehler und Risiken bei den sogenannten Varianzquellen, wobei sich diese primär auf den Diagnostiker beziehen:

Varianzquelle	Kennzeichen
Subjektvarianz	Ein Patient wird zu zwei verschiedenen Zeitpunkten untersucht, in dem er sich in verschiedenen Krankheitszuständen befindet.
Situationsvarianz	Ein Patient wird zu zwei Zeitpunkten untersucht, in denen er sich in verschiedenen Phasen und seiner Erkrankung befindet.
Informationsvarianz	Verschiedenen Untersuchern stehen unterschiedliche Informationen zum Patienten und seiner Erkrankung zur Verfügung.
Beobachtungsvarianz	Verschiedene Untersucher kommen zu unterschiedlichen Urteilen und Bewertungen über Vorhandensein und Relevanz der vorliegenden Symptome.
Kriterienvarianz	Verschiedene Untersucher verwenden unterschiedliche diagnostische Kriterien für die Diagnose derselben Störung.

Tabelle 1 (Spitzer & Fleiss, 1974)

Die oben genannten Risiken spielen in der Anwendung der psychiatrischen Diagnostik eine konstante Rolle und müssen daher kritisch betrachtet werden. Meiner Meinung nach sind

diese Fehlerquellen zwar durchaus relevant, erlauben aber noch keine Kritik an der Anwendung von psychiatrischer Diagnostik. Kritik kann dann angebracht werden, sobald ein Diagnostiker diese Fehlerquellen ignoriert. Diese Fehlerquellen lassen sich direkt in die psychologische Diagnostik übertragen, da dort dieselben Fehlerquellen und Varianzen auftreten können. Jede Art von Diagnostik oder Diagnose hat eine Auswirkung auf den Klienten. Diese Auswirkung soll nun näher betrachtet werden. Bei Diagnostik stehen eher Testverfahren und die damit verbundenen Belastungen im Vordergrund. Bei einer Diagnose ist eher der Prozess der Stigmatisierung und das Leben mit einer Diagnose von Bedeutung (vgl. Kapitel 4.6) (Schomerus, 2010).

Die Personalauswahl gehört sicherlich zu den ältesten Verfahren im Rahmen der Psychologie. Wie im geschichtlichen Teil bereits erwähnt (vgl. Kapitel 3.6), fanden erste Tests zur Personalauswahl bereits vor mehreren tausend Jahren statt. In der Geschichte von Gideon ging es um die Auswahl geeigneter Gefolgsleute. Dieser Bereich ist auch der Ursprung der modernen Personalauswahl und begann beim Militär. Dort hatte man Anfang des 20. Jahrhunderts die Notwendigkeit erkannt, Anwärter für die Offizierslaufbahn einem Auswahlverfahren zu unterziehen, um deren Tauglichkeit festzustellen. Zu diesem Zweck versuchten Psychologen ein Testverfahren zu entwickeln, mit welchem man die Eigenschaften, die man idealerweise in einem Offizier sah, abprüfen konnte. Bereits wenige Jahre nach dessen Einführung, wurde dieser Test kritisch diskutiert. Was man Anfang des Jahrhunderts noch für erstrebenswert gehalten wurde, war wenige Jahre darauf schon nicht mehr aktuell. Zudem stellte man berechtigt die Frage, ob die geprüften Charaktereigenschaften wirklich notwendig für eine Laufbahn als Offizier waren. Bereits zeitgleich wurde auch eine Kraftfahrer- und Funkerprüfung etabliert (Petri, 2004). Heute sind die Auswahlmethoden richtigen Assessmentcentern gewichen, welche in einer regelrechten Testbatterie versuchen umfassend die Fähigkeiten von Personen vor allem von Führungskräften zu prüfen. Kritisch können folgende Punkte bei Assessmentcentern und einmaligen Auswahlverfahren betrachtet werden: Zum einen unterliegt jeder Mensch einer schwankenden Tagesform. Diese aktuelle Form kann bei einmaligen Testungen dazu führen, dass nicht der Beste oder Leistungsfähigste gewählt wird, sondern der mit der besten Tagesform. Des Weiteren haben Menschen mit Prüfungsangst deutlich weniger Chancen in Testsituation, als Personen ohne Prüfungsängste. Diese können folglich ihre maximale Leistungsfähigkeit nicht zeigen. Das führt dazu, dass Ergebnisse erzielt werden, die dazu beitragen, dass nicht der Beste ausgewählt wird. Ein Assessment-Center kann daher nur teilweise seine Zielstellung erfüllen (Petri, 2004).

Die Diagnostik im Rahmen der Rehabilitation kann nach den bisher gemachten Erfahrungen primär anhand des Gütekriteriums der Zumutbarkeit kritisch diskutiert werden. Immer wieder

wurde seitens der Patienten geäußert, dass die zeitliche Belastung durch die Testbatterie viele zu hoch war. Die Interpretation der Ergebnisse eines Testverfahrens variiert von Auswerter zu Auswerter. Meiner Erfahrung nach wurden die Ergebnisse nur als Tendenzen, oder Hinweise gewertet. So wie mit Diagnostik in der Praxis umgegangen wird, wird er auch von der Fachliteratur gefordert. Im Rahmen der Therapie kommt psychologische Diagnostik im Sinne der Testverfahren nur sehr wenig zum Einsatz. Meist hat man es hier mit der klassischen psychiatrischen Diagnosestellung zu tun. Nur ein Teilbereich der psychologischen Diagnostik bleibt für mich kritisch: der Einsatz und die Verwendung von projektiven Tests. Bereits wurde mittels der Fachliteratur die mangelnde Objektivität dargelegt. Für mich stellt sich in diesem Zusammenhang aber die Frage der Sinnhaftigkeit dieser Verfahren. Letzten Endes können sie nur einen Hinweis auf zu grundlegende Konflikte geben. Um einen solchen Konflikt zu explorieren, halte ich persönlich die Gesprächsführung oder das diagnostische Interview für Verfahren, die sich weit besser eignen. In keinem von mir besuchten Praktikumsbetrieb wurden projektive Verfahren eingesetzt. Besonders kritisch sehe ich die hohe Bedeutung des Rorschach-Tests in den Vereinigten Staaten, in denen dieser Test sogar über die mögliche Schuldhaftigkeit und Schuldfähigkeit einer Person entscheiden könne.

4.6. Auswirkungen einer Diagnose

Auf der akutstationären Abteilung meines letzten Praktikums nahm ich an mehreren Gruppentherapiesitzungen teil. Das Thema einer dieser Gruppen war Partnerschaft. Innerhalb der Gruppendynamik einer Therapiestunde stellten sich zwei gegensätzliche Meinungen heraus. So standen die Klienten, die primär an einer Depression erkrankt waren, dem Thema der Partnerschaft deutlich positiver gegenüber, als die Klienten, die unter einer schweren Persönlichkeitsstörung litten. Besonders ist mir während der Therapiestunde eine junge Frau aufgefallen, bei der zuvor eine Borderline-Persönlichkeitsstörung festgestellt wurde. Sie beklagte, dass die Diagnose der Borderline-Persönlichkeitsstörung dazu führe, dass sie aufgrund der daraus resultierenden Etikettierung als psychisch unnormal in der Gesellschaft gelte. Diese Etikettierung hätte zur Folge, dass sie im sozialen Leben zum Beispiel auch bei der Partnerwahl benachteiligt werden würde. Dieser Satz führte dazu, dass die Patienten über das Bild ihrer Krankheiten in ihrem Umfeld diskutierten. Ihre Einstellung gegenüber dem gesellschaftlich-kulturellen Umgang mit Diagnose sorgte für eine lebendige Diskussion während der Therapiestunde. Dabei wurden primär das soziale Leben mit einer Diagnose und der gesellschaftliche Umgang mit psychisch Erkrankten zum Thema. Es stellte sich heraus, dass nicht nur die junge Frau, die an den Folgen einer Borderline-

Persönlichkeitsstörung litt, sondern auch viele ihrer Mitpatienten Erfahrungen mit Stigmatisierungen aus ihrer sozialen Umgebung gemacht haben. So erlebten viele Rehabilitationspatienten von psychischen Krankheiten unter anderen Stigmatisierungen aus Vorurteilen, Ausgrenzungen, Beleidigungen und Diskriminierungen. Aufgrund der mangelnden Kenntnis und Erfahrung mit dem Umgang von psychischen Erkrankungen, waren die Menschen aus dem sozialen Umfeld der Patienten nicht in der Lage, adäquat auf diese zu reagieren. Eine umfassende bundesweite Aufklärung könnte dazu beitragen, den Stigmatisierungsprozessen entgegen zu wirken. Aus diesen Erfahrungen heraus, wurde mir bewusst, welche Auswirkungen Stigmatisierungen auf einen Patienten haben können. Bei Stigmatisierung geht es nicht um die Symptome einer Krankheit selbst, sondern mehr um das subjektive Empfinden von Betroffenen, die Reaktionen des Umfeldes auf Menschen mit psychischen Krankheiten und um falsche Vorstellungen, die sich aus den psychischen Erkrankungen selbst ergeben. „Die verbreiteten Mythen halten sich hartnäckig in der Öffentlichkeit, in der Medienwelt, sogar professionelle im medizinischen oder Sozial-Bereich tätige Personen entgehen ihnen nicht“ (Meise & Ibelshäuser, 2006, S. 17).

Psychische Erkrankungen sind auch heute noch unverändert mit einem Stigma behaftet (...). Weitreichende Programme zum Stigmaabbau (...) haben zwar zu einem nachweisbaren Wissenszuwachs geführt, der jedoch vielfach nicht die erhoffte Verbesserung der öffentlichen Einstellung gegenüber den betroffenen Menschen zur Folge hatte (...). Dies spiegelt sich auch im konkreten Stigmaerleben psychisch erkrankter Menschen wieder. Etwa die Hälfte der psychiatrieerfahrener Menschen sehen sich über Ländergrenzen hinweg im Alltag mit Ablehnung und negativer Diskriminierung konfrontiert (...)" (Schulze, Janeiro, & Kiss, 2010, S. 276).

In vielen Fällen wird die Stigmatisierung sogar als zweite Krankheit bezeichnet. Begründet wird dies mit den Vulnerabilitäts-Stress-Paradigma und den damit verbundenen sozialen Konsequenzen. So können mit dem Stigma verbundene Diskriminierungen als Stressoren wirksam werden und so sogar zu einer Verschlimmerung beitragen. Viele Betroffene leiden folglich nicht nur an den Symptomen ihrer Erkrankung, sondern auch unter den Auswirkungen von Stigmatisierung. Besonders Diskriminierung und Ausgrenzung werden von den Klienten als sehr schmerzhaft erlebt. Um Stigmatisierung besser verstehen zu können, macht es Sinn, diesen Prozess entsprechend seines Verlaufs in vier Phasen aufzuteilen. Der Prozess der Stigmatisierung beginnt mit der Etikettierung (labeling). „(...) Am Anfang steht das Labeling [Hervorhebung v. Verf.], die Etikettierung einer Normabweichung, etwa durch eine psychiatrische Diagnose“ (Schomerus, 2010, S. 253). Die Folgen des Labelings bestehen meist aus einer Reaktion aus Ärger oder Angst. Das Umfeld begegnet dem Betroffenen oft mit mehreren Stereotypen, wie z.B. Unberechenbarkeit oder

Aggressivität. Dadurch kommt es zur Ausgrenzung des Betroffenen, die Abwertung und Diskriminierung zur Folge haben. So kann also festgestellt werden, dass Stigmatisierung zunächst ein von außen kommender Prozess ist, der sich in der Umgebung des Betroffenen abspielt. Diese sehen sich dann „(...) mit Benachteiligung im Berufsleben, mit Ablehnung in persönlichen Kontakten (...)“ (Schomerus, 2010, S. 253) und anderem gegenüber. Zusätzlich kommen dann auch Benachteiligungen im Hinblick auf strukturelle Aspekte hinzu. Ein Nachteil entsteht z.B. bei der Verteilung von Ressourcen im Gesundheitswesen. Doch neben den Stigmatisierungsprozessen, die aus der sozialen Umwelt auf den Klienten einwirken, gibt es auch solche Prozesse, die im Betroffenen selbst ablaufen. So „(...) rührt ein großer Teil der Belastung durch Stigma aus der Antizipation negativer Reaktionen, was zu Geheimhaltung der Erkrankung und zur Vermeidung potentiell als bedrohlich eingeschätzter Situationen führt (...)“ (Schomerus, 2010, S. 254). Neben der Geheimhaltung der Erkrankung findet häufig während eines Stigmatisierungsprozesses auch Selbststigmatisierung statt. Patienten sind unweigerlich auch mit der eigenen Haltung gegenüber einer Diagnose konfrontiert. In einem solchen Fall wendet der psychisch Kranke bestehende Vorurteile und negative Haltungen gegen sich selbst an. Er zählt sich selbst zur Gruppe der Ausgestoßenen. Inzwischen ist eindeutig, dass diese Prozesse der Selbststigmatisierung einen ungünstigen Einfluss auf den Krankheitsverlauf nehmen, da sie auf der einen Seite schon die Inanspruchnahme von Hilfen erschweren, auf der anderen Seite werden auch Selbstwertgefühl und Selbstwirksamkeit beeinträchtigt. Diese sind wichtige Effekte für den Verlauf einer Therapie (Schomerus, 2010). Hier zeigt sich, dass sich Menschen mit einer Borderline-Persönlichkeitsstörung einer subjektiv höheren Stigmatisierung und Diskriminierung gegenüber sehen als Menschen mit beispielsweise einer Schizophrenie (Schulze, Janeiro, & Kiss, 2010). Ein hoffnungsvoller Aspekt dabei ist der Umstand, dass besonders jüngere Menschen sich gegenüber psychischen Erkrankungen offener zeigen und etwas weniger zu Vorurteilen neigen als Ältere (Sulzenbacher, Schmid, Kämmler, De Col, & Meise, 2002). Die Folgen von Stigmatisierung sind schwer und vielfältig. Sie reichen von einer erschwerten Behandlung, bis hin zu Behandlungsvermeidung. Behandlungsabbrüche gehören ebenfalls dazu. Dies trägt dazu bei, dass psychische Krankheiten als unheilbar angesehen werden. Doch deutlich wurde gezeigt, dass der Glaube an die Wirksamkeit von Behandlung sehr starke positive Effekte bringen würde. Im Umgang mit Stigmatisierung zeigen die Patienten unterschiedliche Ansätze, die zu einem gewissen Grad mit der Erkrankung korrelieren (Meise & Ibelshäuser, 2006). So wurde deutlich, dass Menschen, die unter der Borderline-Persönlichkeitsstörung, leiden häufiger Situationen und Kontakte meiden, in welchen die Betroffenen befürchten, dass sie stigmatisiert werden könnten. Sie beschrieben es als besonders schlimm, wenn sie nur über die Krankheit definiert werden und eigene Leistungen nicht mehr gesehen werden.

Betroffene, die unter Schizophrenie leiden, beklagten sich häufig, dass man ihre Krankheit zu wenig ernst nehme und man sie nur als zu „faul“ erachte. Häufig seien sie mit Vorwürfen, sich eine schöne Zeit zu machen, konfrontiert. Im Gegensatz zu Menschen mit Borderline-Persönlichkeitsstörung, empfanden Menschen mit Schizophrenie die Zwangsmaßnahmen als sehr traumatisierend. Dazu zählten Dinge wie Zwangseinweisung, oder Zwangsbehandlung. Das Ausmaß der Auswirkungen von Stigmatisierung macht deutlich, dass es sich lohnt, sich diesen Mechanismen entgegen zu stellen. Es konnte gezeigt werden, dass besonders die informative Vorstellung der Krankheit durch einen Betroffenen in Zusammenarbeit mit einem Experten die besten Resultate im Abbau von Vorurteilen erbrachte (Schulze, Janeiro, & Kiss, 2010). Doch woher kommen die Vorurteile, die den Betroffenen stigmatisieren? Im Beispiel der Schizophrenie existiert das weit verbreitete Bild, dass diese Menschen eine gespaltenen Persönlichkeit hätten. Eine Ursache lässt sich in Emil Kräpelins Arbeit zu dem Thema der Dementia Praecox finden, in der die psychische Krankheit unweigerlich zum seelischen Siechtum führe. Zwar gibt es durchaus solch chronisch verlaufende Krankheiten, doch sind diese eher selten und in den meisten Fällen ist es möglich mit richtiger Intervention den Krankheitsverlauf positiv zu beeinflussen. Dass man die Schizophrenie für eine Krankheit hält, bei der die Persönlichkeit gespalten wird, was nach der International-Classification-of-Disfunction (ICD-10) eher der dissoziativen Identitätsstörung entspricht, kommt durch die fehlerhafte Rückübersetzung des Wortes Schizophrenie. Doch hat sich dieser Fehler bis heute gehalten und trägt dazu bei, dass die Betroffenen unter vielleicht stärkerer Stigmatisierung zu leiden haben als nötig wäre (Miller, 2011).

5. Fazit

Aus der hier vorliegende Arbeit zum Thema „Kritische Betrachtung der Anwendung psychologischer Diagnostik“ kann folgendes geschlussfolgert werden: Die fundamentale Kritik an psychologischer Diagnostik ist kaum mehr haltbar. Lediglich die Anwendung und die Interpretation von Testergebnissen können und müssen kritisch hinterfragt werden. Im Bereich der Leistungstests gibt es vereinzelt Diskussionen über die zugrundeliegenden Konstrukte. Die psychologische Diagnostik hat sich im Laufe des letzten Jahrhunderts sehr verändert. Dies lag vor allem daran, dass sie einer ständigen kritischen Betrachtung unterzogen wurde. So wurden beispielsweise die Gütekriterien entwickelt, um sich an das naturwissenschaftliche Verständnis anpassen zu können. Die Entwicklung hin zur Naturwissenschaftlichkeit hatte sich nicht nur positiv auf die Diagnostik ausgewirkt. So wurde zunehmend aus den Augen verloren, dass es sich bei der Psychologie um eine

Wissenschaft über Menschen, die sich inter- und intraindividuell unterscheiden, handelt. Der Ansatz die methodischen Grundlagen zu verbessern und zu modernisieren ist dennoch sehr zu begrüßen, da deren Überprüfung zu einer Weiterentwicklung der Diagnostik führte. Die psychologische Diagnostik konnte sich trotz schwerer Kritik in der heutigen Psychologie durchsetzen. Das Ergebnis ist einer der am besten fundierten Bereiche in der Psychologie. Durch die psychologische Diagnostik können allgemein gültige Aussagen getroffen werden, die für den Entscheidungsprozess innerhalb einer Therapie von Bedeutung sind. Weiterhin hat sich gezeigt, dass einheitliche Testverfahren trotz aller Unterschiede in menschlichen Eigenschaften ihre Berechtigung haben. Die diagnostischen Testverfahren sollten dennoch nur als Hilfestellung genutzt werden. Wird ein einzelnes Testverfahren als alleinige Entscheidungsgrundlage verwendet, so kann es dazu kommen, dass der Therapieerfolg der Behandlung einer psychischen Erkrankung ausbleibt. Eine Diagnose sollte sich demnach nicht nur auf ein Ergebnis oder ein Verfahren stützen. Effekte, wie Übertreibung, sozialangepasstes Antworten und andere Verfälschungseffekte können die Folge sein. Aus meiner Erfahrung während meiner Praktika, kann ich bestätigen, dass mit Diagnostik zumeist sehr verantwortungsvoll umgegangen wurde. So stützten die Kollegen ihre Aussagen nicht nur auf die Ergebnisse einzelner Tests, sondern bezogen in ihre Entscheidungen immer eigene Beobachtungen und andere Erhebungstechniken mit ein. So stellten die Testverfahren immer nur eine Informationsquelle über die Erlebniswelten des Patienten und zur Sicherung eines Verdachtes dar. Heute fällt aufgrund ihrer Nützlichkeit bei der Therapieplanung eine kritische Betrachtung von Diagnostik schwer. So kann festgehalten werden, dass die psychologische Diagnostik ein wissenschaftliches und empirisch gut belegtes Instrument zur Generierung von Informationen ist. Eine Ausnahme sind die projektiven Verfahren, die sich besonders durch die mangelnde Objektivität auszeichnen. Hier stellt sich die Frage, ob man nicht ganz auf solche Tests verzichten sollte, da sie letzten Endes keine zusätzlichen Informationen liefern, die man nicht auch durch Verhaltensbeobachtung, oder Gespräche schneller und besser explorieren könnte. Die Stigmatisierung von Menschen mit psychischen Störungen ist nicht nur Folge der mangelnden Kenntnis der sozialen Umgebung des Patienten, sondern auch Produkt der Etikettierung durch Diagnostik. Stigmatisierungsprozesse die das Leben der Patienten mit einer Diagnose beeinträchtigen, sind in der heutigen Gesellschaft weiterhin üblich. So haben Betroffene unter Stigmatisierungsprozessen, wie Ablehnung, Diskriminierung und Ausgrenzung zu leiden. Eine umfassende bundesweite Aufklärung über psychische Störungen würde Stigmatisierungsprozessen in unserer Gesellschaft entgegen wirken und somit zu einer Minimierung des psychischen Leidens von Patienten beitragen. Eine Gleichberechtigung psychischer und somatischer Krankheiten hinsichtlich ihrer Akzeptanz in der Gesellschaft wäre wünschenswert.

6. Literaturverzeichnis

- Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer Verlag.
- Freyberger, H. J., & Stieglitz, R.-D. (2006). Leitlinien zur Diagnostik in der Psychiatrie und Psychotherapie. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 54 (1), 23-33.
- Hänsgen, K.-D., & Hofer, R. (1996). Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium in der Schweiz. In K.-D. Hänsgen, & N. Ischi, *Eignungsprüfung für das Medizinstudium. Kriterien und Testverfahren* (S. 82-90). Bern: Zentrum für Testentwicklung und Diagnostik am psychologischen Institut der Universität Freiburg.
- Heyse, H. (2003). Rechtsgrundlagen - Deutschland. In K.-D. Kubinger & R.-S. Jäger, *Schlüsselbegriffe der psychologischen Diagnostik* (S. 362-367). Weinheim: Beltz Verlag.
- Ingenkamp, K. (1981). Testkritik ohne Alternative: Eine kritische Darstellung der Argumentation radikaler Schultestkritik in der deutschen Fachliteratur. In R. S. Jäger, K. Ingenkamp, & G. Stark, *Tests und Trends 1981. Jahrbuch der Pädagogischen Diagnostik* (S. 71-101). Weinheim: Beltz.
- Ingenkamp, K., & Lissman, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz Verlag.
- Jäger, R. S., & Petermann, F. (1999). *Psychologische Diagnostik*. Weinheim: Beltz Verlag.
- Kubinger, K. D. (2003). Gütekriterien. In K. D. Kubinger, & R. S. Jäger, *Schlüsselbegriffe der psychologischen Diagnostik* (S. 195-204). Weinheim: Beltz.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe Verlag.
- Lauber, C., Nordt, C., Falcató, L., & Rössler, W. (2002). Behandlungsvorstellungen der Bevölkerung zu Depression und Schizophrenie. *Neuropsychiatrie*, 16 (1+2), 99-102.
- Lienert, G. A., & Ratz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz Verlag.

- Meise, U., & Ibelshäuser, A. (2006). Diagnose: Stigma. Wechselwirkungen von psychischer Gesundheit und Stigmatisierung. *ProMente Austria*, 3, 17-18.
- Miller, K. (2011). *Die Stigmatisierung von Menschen mit psychischen Erkrankungen durch einen stationären Aufenthalt aus Sicht der Betroffenen*. Dissertation Freiburg i. Brsg.: Klinik für Psychiatrie und Psychotherapie der Albert-Ludwigs-Universität Freiburg i. Br.
- Moosbrugger, H., & Kelava, A. (2007). *Testtheorien und Fragenbogenkonstruktion*. Heidelberg: Springer Verlag.
- Petri, S. (2004). Personalauswahl zwischen Psychotechnik und Charakteranalyse. Die Kompetenzverschiebung der deutschen Militärpsychologie von 1914 bis 1942. *Zeitschrift für Psychologie*, 212 (4), 200-211.
- Pospeschill, M., & Spinath, F. (2009). *Psychologische Diagnostik*. München: Ernst Reinhardt Verlag.
- Rauchfleisch, U. (2005). *Testpsychologie*. Göttingen: Vandebhoeck & Ruprecht GmbH.
- Rheinberg, F. (1986). Gefahren pädagogischer Diagnostik. In K.-H. Klauer, *Handbuch der Pädagogischen Diagnostik* (S. 27-38). Düsseldorf: Schwann Verlag.
- Schomerus, G. (2010). Das Stigma psychischer Krankheit - Ein Gegenstand der Psychotherapie? *Zeitschrift für Psychiatrie, Psychologie und Psychiatrie*, 58 (4), 253-255.
- Schrödter, M. (2006). Die Herrschaft machtvoller Diagnostik verhindern! *Sozialextra*, 22-24.
- Schulze, B., Janeiro, M., & Kiss, H. (2010). Das kommt ganz drauf an... Strategien zur Stigmabewältigung von Menschen mit Schizophrenie und Borderline-Persönlichkeitsstörung. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 48 (4), 275-285.
- Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry*, 125, 341-347.
- Stieglitz, R.-D. (2000). *Diagnostik und Klassifikation psychischer Störungen*. Göttingen: Hogrefe Verlag.

Sulzenbacher, H., Schmid, R., Kämmler, G., De Col, C., & Meise, U. (2002). Schizophrenie ... "bedeutet für mich gespaltene Persönlichkeit". Ein Programm gegen Stigmatisierung von Schizophrenie in Schulen. *Neuropsychiatrie*, 16 (1), 93-98.

Westmeyer, H. (2004). Die sogenannte Krise der psychologischen Diagnostik. *Diagnostica*, 50 (1), 10-16.

Williams, R. G., Pollack, U. J., & Ferguson, N. A. (1975). Differential effects of two grading systems on student performance. *Journal of Educational Ppsychology*, 67, 253-258.

Eidesstattliche Erklärung

Vorname: Michael Georg **Nachname:** Ortlieb **Matrikelnummer:** 20092766

Hiermit versichere ich, Michael Ortlieb, dass ich die vorliegende Arbeit selbständig angefertigt und ohne fremde Hilfe verfasst, keine anderen als die angegebenen Hilfsmittel benutzt und wörtlich oder inhaltlich entlehnte Stellen als solche kenntlich gemacht habe. Diese Arbeit wurde noch keiner Prüfungsbehörde in gleicher oder ähnlicher Form vorgelegt.

Stendal, den 05.09.2012