

1. Einleitung

1.1 Vorbemerkung zur Rahmendisziplin

Die empirischen Arbeiten, die in dieser Habilitationsschrift zusammengefasst sind, befassen sich mit der Entwicklung psychologisch diagnostischer Instrumente sowie vertiefend mit der Diagnostik autodestruktiver Syndrome. Sie wurden in der Medizinischen Klinik mit Schwerpunkt Psychosomatik der Charité angefertigt. Die inhaltlichen Schwerpunkte der Studien lassen sich allerdings nicht dem engen Themenkanon unterordnen, wie er sich in manch traditionellen Konzepten von Psychosomatik widerspiegelt. Sie stehen vielmehr im Einklang mit der besonderen Ausrichtung der Psychosomatik vor Ort, die dem Konzept einer Integrierten Psychosomatik verpflichtet ist [1]. Damit sind die Arbeiten inhaltlich in jenem breiten Bereich beheimatet, der die Psychosomatik mit ihren benachbarten Disziplinen Medizinische Psychologie, Klinische Psychologie, Psychotherapie und Psychiatrie verbindet.

1.2 Übergeordnete Paradigmen der Psychosomatischen Medizin und ihren Nachbardisziplinen

Globaler Ausgangspunkt aller psychosomatischen Konzepte ist das Zusammenwirken von Seelischem und Körperlichem. Der moderne gesundheitspsychologische Forschungsansatz richtet sich auf Beziehungen zwischen Belastungen und Gesundheit, wobei Gesundheit nicht einfach mit Körperlicher gleichzusetzen ist. Vielmehr beinhaltet der weithin akzeptierte Gesundheitsbegriff gemäß der WHO neben der körperlichen, auch die seelische und die soziale Dimension sowie die Alltagsfunktion [2;3].

Die Beziehungen zwischen Belastung und Gesundheit lassen sich theoretisch entlang zweier unterschiedlicher Paradigmen fassen, dem *pathogenetischen* und dem *salutogenetischen* Modell [4;5].

Das *pathogenetische* Modell geht der Frage nach, ob und unter welchen Bedingungen Personen infolge erlebter Belastung erkranken. Als Belastungen

wurden zum einen kritische Lebensereignisse konzipiert, wobei sich insbesondere für unvorhersehbare und unkontrollierbare negative Ereignisse gesundheitliche Folgen nachweisen ließen [6]. Zum anderen wurde auf chronische Stressoren und ihre gesundheitlichen Auswirkungen abgehoben [7].

Zwei grundsätzliche Pfade lassen sich unterscheiden, über die seelische Belastung in (körperliche) Erkrankung einmünden könnte. Der eine Pfad ist vermittelt über gesundheitsschädigendes Verhalten, welches durch Belastung ausgelöst wird, beispielsweise Substanzabusus [8], Fehlernährung unter Stress [9] oder selbstverletzendes Verhalten [10]. Der zweite Pfad besteht in physiologischen Veränderungen, die sich unter Belastung einstellen, nämlich zum einen länger anhaltende vegetative Reaktionen und zum anderen Änderungen endokrin-immunologischer Prozesse und ihre Folgen [11;12].

Das *salutogenetische* Modell fragt dagegen, wie es Personen trotz belastender Lebensumstände und -ereignisse bewerkstelligen, gesund zu bleiben, d.h. körperliches und seelisches Wohlbefinden aufrechtzuerhalten [13]. Die Kernannahme dieses Modells besagt, dass es Ressourcen gibt, die die Person widerstandsfähig im Umgang mit Belastungen machen. Als Ressourcen gelten neben einer höheren körperlichen Belastbarkeit und Merkmalen des sozialen Netzwerkes vor allem psychologische Merkmale. Solche Charakteristika der Person wurden auch umfassender als „Sense of Coherence“ [14] oder als „seelische Gesundheit“ [15] definiert. Im Einzelnen fallen darunter psychologische Eigenschaften wie die „Hardiness“ [16] oder die Selbstwirksamkeitserwartung [17].

In der Zusammenschau stellen die Beziehungen zwischen Belastung und Gesundheit empirischen Befunden zufolge ein komplexes Zusammenspiel von Variablen aus mehreren Konstruktebenen dar. Studien zu diesen Zusammenhängen beziehen Merkmale aus drei Konstruktebenen ein, nämlich (a) der Ebene der Belastung, (b) der Ebene der körperlichen, psychischen und sozialen Folgen von Belastung (gesundheitsbezogene Lebensqualität) und (c) der Ebene der Ressourcen im Umgang mit der Belastung.

Auf allen drei Konstruktebenen ist für die empirische Erhebung der Einsatz psychologisch diagnostischer Instrumente erforderlich. Eine zentrale

Zielsetzung der hier vorgelegten Arbeiten, lag darin, psychologisch diagnostische Instrumente zu entwickeln bzw. für den deutschen Sprachraum zu adaptieren, auf ihre psychometrischen Qualitäten zu testen sowie Referenzwerte für ausgewählte Krankengruppen zu ermitteln. Insgesamt wurden Instrumente zu Konstrukte aus allen drei genannten Merkmalsebenen, Belastungen, Ressourcen und gesundheitsbezogene Lebensqualität, berücksichtigt. In der vorliegenden Übersicht werden exemplarisch die Instrumente zu Stress und zu Depression herausgegriffen und es wird in einer Arbeit zur psychosomatischen Therapieevaluation verdeutlicht, welche besonderen Anforderungen an die Messung dort gestellt werden. In einer klinischen Vertiefung werden Syndrome autodestruktiven Verhaltens und Instrumente zu ihrer Diagnostik behandelt.

Zur Einführung wird vorab ein kurzer Überblick über psychologische Diagnostik in der Psychosomatischen Medizin und ihren angrenzenden Disziplinen folgen. Der daran anschließende Abschnitt informiert über Syndrome pathologisch autodestruktiven Verhaltens.

1.3 Psychologische Diagnostik in der Psychosomatischen Medizin und ihren Nachbardisziplinen

Klinisch psychologische, medizinpsychologische und gesundheitspsychologische Diagnostik bezieht sich in der Regel auf nicht direkt beobachtbare, sondern so genannte latente Merkmale [18]. Psychologische Diagnostik erhebt Information, die es uns erlauben soll, auf solche latenten Merkmale zu schließen. Ihr kommt ein breites Spektrum an Funktionen zu, die in der Psychosomatischen Medizin wie auch der Gesundheitspsychologie und damit in allen medizinischen Anwendungsfeldern potentiell relevant sind.

Im Folgenden werden die grundlegenden Funktionen, zu unterscheidende Datenquellen, Datenebenen und Arten methodischer Zugänge psychologischer Diagnostik dargestellt. Die Instrumentenentwicklung folgte historisch gesehen bestimmten Strömungen, die zum Teil durch die wissenschaftlich-theoretische Fortentwicklung, zum Teil durch klinische Problemstellungen sowie ökonomische Notwendigkeiten bestimmt war. Diese Strömungen werden

skizziert und es wird ein Einblick in die probabilistische Testtheorie und deren Möglichkeiten als modernem methodischen Ansatz in der psychometrischen Diagnostik gegeben.

1.3.1 Funktionen psychologischer Diagnostik

Der psychologischen Diagnostik in der Psychosomatischen Medizin und ihren Nachbardisziplinen kommen grundsätzlich 5 Funktionen zu: *Deskription*, *Klassifikation*, *Erklärung*, *Prognose* und *Evaluation* [18;19].

Die systematische *Deskription* einschließlich der Quantifizierung eines Phänomens, etwa eines Symptoms oder eines Verhaltensmusters, ist insofern eine grundlegende Funktion psychologischer Diagnostik, als sie die Ausgangslage für alle weiteren Funktionen bildet. Unterschieden wird zwischen der Beschreibung von Ist-Zuständen und von Veränderungen.

Die *Klassifikation* meint die Zuordnung von Fällen zu einer Klasse in einem gegebenen Klassifikationssystem, etwa der Klassifikation psychischer Störungen und Verhaltensstörungen in der ICD-10 Teil F.

Sofern die psychologische Diagnostik auf die *Erklärung* eines Merkmals abzielt, kann sie sich auf die Messung von prädisponierenden, auslösenden oder aufrechterhaltenden Bedingungen dieses Merkmals richten. Dies kann die Messung von akut auslösenden Belastungsfaktoren ebenso beinhalten wie die von biographischen und prädisponierenden Persönlichkeitsfaktoren oder aufrechterhaltenden Verhaltens- und Verstärkungsmustern.

Die *Prognose* richtet sich auf die Vorhersage von Verläufen psychischer oder Verhaltensmerkmalen. Besonders relevant ist die Prognose hinsichtlich der Verläufe mit oder ohne Intervention.

Die *Evaluation* bezieht sich auf die Bewertung von Interventionen oder ganzen Versorgungsmodellen und -systemen. Die Bewertung basiert auf Prozess- und Ergebnismessungen. In diesen Kontext gehören Wirksamkeitsnachweise und die Qualitätssicherung [20;21].

1.3.2 Datenquellen und Datenebenen

Hinsichtlich der Datenquellen lassen sich Selbstbeurteilungen der Person von Fremdbeurteilungen durch andere Personen, etwa durch die behandelnden Klinker oder durch Angehörige, abgrenzen [22;23]. Die entsprechenden Methoden werden auch als Selbstbeurteilungsverfahren und Fremdbeurteilungsverfahren bezeichnet. Ferner kommen in der Funktions- und Leistungsdiagnostik apparative Messungen zum Einsatz, auf die hier nicht näher eingegangen wird.

In der Forschung überwiegt der Einsatz der Selbstbeurteilungsverfahren [24]. Sie basieren auf Selbstbeobachtung und Reflexion und liefern Selbstbeschreibungen und -einschätzungen [18]. Indiziert sind sie vor allem in Merkmalsbereichen, die dem Individuum selbst am besten zugänglich sind und die von außen schwerlich beobachtet oder erschlossen werden können, etwa bei der Erhebung von Befindlichkeit, Schmerzen, Lebenszufriedenheit oder persönlichen Einstellungen, Bewertungen und Überzeugungen. Sie sind jedoch auch in solchen Merkmalsbereichen sinnvoll, die prinzipiell per Fremdbeobachtung erfasst werden können, wie z.B. der Diagnostik von Angst oder Depression. Ihre Vorteile liegen in dem enormen inhaltlichen Spektrum verfügbarer Verfahren, in der Ökonomie ihres Einsatzes für den Untersucher, ihrer Objektivität (gemeint ist damit die Standardisierung ihrer Durchführung, Auswertung und Interpretation) und ihrer Vergleichbarkeit mit Referenzwerten.

Fremdbeurteilungsverfahren kommen vor allem bei solchen Merkmalen zum Einsatz, über die die Person nicht in der Lage oder nicht willens ist, offen Auskunft zu geben. Dies betrifft z.B. heimlich selbstschädigendes Verhalten, wie es im Falle sog. artifizieller Störungen auftritt.

Zwischen Selbst- und Fremdbeurteilungsverfahren lässt sich allerdings keine absolute Trennung ziehen, da auch Fremdbeurteilungen häufig Items beinhalten, die auf Selbstaussagen der Patienten basieren. Es sei angemerkt, dass sich auch bei inhaltlich korrespondierenden Verfahren keine vollständigen

Übereinstimmungen zwischen Selbst- und Fremdbeurteilung ergeben. Aus empirischen Studien zu dieser Frage wird gefolgert,

- dass die beiden Verfahrensarten teils unterschiedliche Phänomene abbilden,
- dass bei der Messung psychischer Beeinträchtigungen Fremdbeurteilungsverfahren besser differenzieren als Selbstbeurteilungsverfahren,
- dass es für beide Verfahrensarten spezifische Urteilsfehler gibt und
- dass die Übereinstimmung zwischen Fremd- und Selbstbeurteilung mit fortschreitendem Behandlungsverlauf größer wird [22].

Hinsichtlich der Datenebenen wird theoretisch zwischen biologischen, psychologischen und sozialen Daten unterschieden. Gelegentlich finden sich auch Unterscheidungen in subjektive und objektive Daten, wobei Einschätzungen als subjektiv gelten und eindeutig bestimmbare Fakten als objektiv. Da aber auch so genannte Fakten, wie zum Beispiel Laborwerte oder (krankheitsbedingt ausfallende) Arbeitsstunden – oft nicht exakt und zuverlässig quantifizierbar sind, sondern mit einem Messfehler behaftet sein können, kann diese Unterscheidung irreführend sein.

Da für die Diagnostik komplexer Zusammenhänge keine der genannten Datenquellen und auch keine der Ebenen Alleingültigkeit beanspruchen kann, gilt Multimodalität, also die Einbeziehung unterschiedlicher Datenebenen und Datenquellen, als optimal. Dies gilt verstärkt dort, wo mit einer eingeschränkten Validität der Daten einer bestimmten Quelle gerechnet werden muss, wie auch in der Evaluation von Interventionen, die Ziele auf mehreren Ebenen verfolgen und daher an subjektiven wie auch objektivierbaren Kriterien gemessen werden sollten [25;26].

1.3.3 Qualitative und quantitative Verfahren

Psychologisch diagnostische Instrumente lassen sich grob in qualitative und quantitative Verfahren einteilen.

Qualitative Verfahren zielen darauf ab, die Individualität psychologischer Merkmale möglichst ohne a priori festgelegte Beschränkungen (durch Vorgabe von Itemformulierungen oder bestimmten Skalen) zu erfassen. Sie bevorzugen daher offene Verfahren der Erhebung, also Verfahren mit möglichst geringer Vorstrukturierung der Antwortmöglichkeiten, und inhaltsanalytische Auswertungsmethoden.

Quantitative Verfahren zielen darauf ab, individuelle Ausprägungen vorab definierter Merkmale zu erfassen und einem Messwert, also einer numerischen Größe zuzuordnen. In diesem Zusammenhang ist auch von psychometrischer Diagnostik die Rede. Der Begriff „Psychometrie“ betont also die Messbarkeit psychischer Merkmale [18].

Beide methodischen Zugänge haben ihre jeweilige wissenschaftliche Relevanz, indem sie unterschiedliche Funktionen in der Erkenntnisgenese erfüllen [27;28]. Die Wahl der Methoden sollte sich nach dem Stand der Theorieentwicklung und der empirischen Befundlage richten. Qualitative Methoden gelten vor allem dort als indiziert, wo theoretische Ansätze erst zu entwickeln oder weiterzuentwickeln sind und Hypothesen generiert werden. Quantitative Methoden gelten als indiziert, wenn Hypothesen zu prüfen sind.

1.3.4 Gütekriterien

Die Entwicklung eines psychologisch diagnostischen Verfahrens muss berücksichtigen, ob das Instrument die Anforderungen an die Gütekriterien erfüllt. Neben den primären Gütekriterien der Objektivität, Reliabilität und Validität eines Verfahrens interessieren vor allem die Akzeptanz durch die Patienten, die Ökonomie und die Normierung.

Obwohl dies allgemein akzeptierter Standard ist, finden sich in der Literatur häufig zwei Mängel: Insbesondere in älteren Publikationen werden oft noch Instrumente eingeführt, ohne dass Reliabilitäten mitgeteilt werden [29] oder es werden keine Reliabilitäten für die einzelnen Skalen eines Verfahrens oder für unterschiedliche Referenzstichproben mitgeteilt, obgleich sich diese Werte erheblich unterscheiden können. Zweitens wird bei mehrdimensionalen

Verfahren, deren Struktur ursprünglich einmal faktorenanalytisch ermittelt wurde, nicht mehr geprüft, ob sich diese Struktur in anderen Populationen überhaupt replizieren lässt.

1.3.5 Trends und aktuelle Aufgaben für die Entwicklung klinisch psychometrischer Verfahren

Die klinisch-psychologische Diagnostik hat eine kaum zu überblickende Fülle unterschiedlicher Verfahren hervorgebracht. In der Entwicklung dieser Methoden von den Anfängen bis heute sind verschiedene sich teilweise überschneidende Trends auszumachen.

Zunächst stand das Ziel im Vordergrund, überhaupt standardisierte psychometrische Instrumente zu entwickeln, die den Anforderungen an die Gütekriterien, Objektivität, Reliabilität und Validität, entsprechen, und die klinische Beobachtungen und Einschätzungen aus Erstuntersuchung, Exploration und Anamnese objektivieren und validieren, ggfs. auch korrigieren helfen sollten. Klinisch relevante Beispiele dafür sind Fragebögen zu Depression, Angst oder Körperbeschwerden [18].

Im Weiteren lassen sich zwei parallele Trends erkennen: Zum einen wurden übergreifende Inventare und Testbatterien entwickelt, die es sich zur Aufgabe gemacht hatten, alle relevanten Subdimensionen eines breiten übergeordneten Konstruktes abzudecken. Übergeordnete Konstrukte waren beispielsweise Persönlichkeit (z.B. der Persönlichkeitsfragebogen MMPI-2 [30]), Psychopathologie (z.B. die Symptom-Checklist SCL-90-R [31]) oder gesundheitsbezogene Lebensqualität (z.B. der WHOQOL100 [32]). Dies ging damit einher, dass die entsprechenden Verfahren oft einen beträchtlichen Gesamtumfang annahmen. Gleichzeitig wurden für viele eingeschränkte Fragestellungen spezialisierte Instrumente entwickelt, die zwar in ihrem jeweiligen Umfang überschaubar blieben, deren Geltungsbereich jedoch zwangsläufig enger war (z.B. Selbstwirksamkeitserwartungen in Bezug auf das Rauchen [33]).

Gegenwärtig lassen sich mehrere Trends ausmachen:

Erstens werden die Bemühungen fortgesetzt, stark ausdifferenzierte Einzelinstrumente für spezifische Problemstellungen zu entwickeln.

Zweitens werden mit dem großen Aufwand von Konsensus-Konferenzen internationale Vereinheitlichungen vorgenommen, wie dies von den Klassifikationssystemen der WHO bekannt ist. Prominente Beispiele hierfür sind das Lebensqualitätsinventar WHOQOL und seine Kurzversion WHOQOL-Bref, das mittlerweile in vielen Sprachen validiert ist [32] oder der SF-36 [34], eines der meisteingesetzten Verfahren zur Lebensqualitätsmessung in der Medizin überhaupt.

Drittens werden etablierte Verfahren, die sich im praktischen Einsatz bewährt haben, ökonomisiert, d.h. anhand der geltenden empirischen Gütekriterien gekürzt und neu validiert. Dieser Trend folgt der Notwendigkeit, die Untersuchungsbelastung von Patienten und Probanden wie auch die Kosten für die Erhebung und Auswertung gering zu halten. Ein hohes Maß an Ökonomie ist vor allem dann relevant, wenn komplexe Zusammenhänge zwischen verschiedenen gesundheitspsychologisch relevanten Variablen überprüft werden sollen [35;36], da dies mit einem hohen Bearbeitungsaufwand verbunden ist. Für die klinisch-diagnostische Routine liegt ein weiteres Ziel der Ökonomisierung darin, möglichst zeitnah über Befunde zu verfügen, die weitere Behandlungsentscheidungen fundieren helfen sowie der Rückmeldung an die Patienten dienen.

Diese Trends lassen sich auch als Aufgaben für die künftige Entwicklung und Weiterentwicklung psychometrischer Instrumente formulieren. Die erste Aufgabe lautet also, Instrumente für relevante klinische oder gesundheitspsychologische Konstrukte zu entwickeln, für die noch keine diagnostischen Erhebungsmethoden vorliegen. Die zweite ist, international vergleichbare Instrumente zu entwickeln, bzw. sofern im englischsprachigen Raum bewährte Instrumente vorliegen, diese auch für den deutschen Sprachraum zu adaptieren. Die dritte ist, die klinische Diagnostik, insbesondere sofern sie Bestandteil von UntersuchungsROUTINEN in der Patientenversorgung sein soll, zu ökonomisieren. Möglichkeiten der Ökonomisierung bieten sich nach den Annahmen der Klassischen Testtheorie, insbesondere aber nach den Annahmen der Probabilistischen Testtheorie.

1.3.6 Klassische Testtheorie und Probabilistische Testtheorie

Der Konstruktion eines jeden psychologischen Testverfahrens liegt eine Testtheorie zugrunde. Testtheorien beziehen sich auf den Zusammenhang zwischen dem zu erfassenden, nicht direkt beobachtbaren Merkmal und dem konkreten Testverhalten des Patienten [37]. Klassische Testtheorie und probabilistische Testtheorie (gelegentlich auch als „moderne“ Testtheorie bezeichnet) dienen beide dem Ziel, die Merkmalsausprägung des Patienten zu bestimmen. Ihre zugrunde liegenden Modellannahmen unterscheiden sich allerdings deutlich.

Die ganz überwiegende Mehrzahl standardisierter psychometrischer Instrumente im klinischen Bereich gründen in ihrer Konstruktion auf den Annahmen der Klassischen Testtheorie (KTT). Vereinfacht besagt diese, dass sich ein empirisch erhobener Messwert aus einem wahren Wert und einem Fehleranteil zusammensetzt, wobei der Fehleranteil in unsystematischen Einflüssen auf die Messung, wie zum Beispiel Unaufmerksamkeit in der Bearbeitung oder Lärmbelästigung in der Untersuchungsumgebung, begründet liegt [38]. Dies hat Implikationen dafür, wie bei einer Messung der Testwert des Patienten, also zum Beispiel ein Score in einem Depressionsfragebogen, berechnet wird und wie die Messgenauigkeit (Reliabilität) definiert und überprüft wird. Der Zusammenhang zwischen dem zugrunde liegenden Merkmal des Patienten – auch als *latent trait* bezeichnet – und einem konkreten Messwert des Patienten ist der Klassischen Testtheorie folgend ein deterministischer Zusammenhang.

KTT-basierte Fragebögen präsentieren allen Patienten jeweils alle Fragen eines Merkmalsbereiches zur Bearbeitung, so dass auch solche Items, die im Einzelfall irrelevant sein können, bearbeitet werden müssen. Mit der Länge einer Skala steigt in der Regel ihre Messgenauigkeit. Solche Verfahren bilden daher stets einen Kompromiss zwischen Ökonomie und Messgenauigkeit. Üblicherweise wird nach diesem Ansatz aus den Items einer Skala ein Summenscore oder Mittelwert gebildet, in den die Werte aller Items in gleicher Gewichtung eingehen. Der Score soll besagen, an welcher Position im

Ausprägungsbereich des Merkmals der „wahre Wert“ des Patienten liegt. Für die Einzelfalldiagnostik wird dieser Wert mit den Werten einer konkreten Normierungsstichprobe verglichen. Will man die Bestimmung des „wahren Wertes“ statistisch absichern, lassen sich Konfidenzintervalle berechnen. Die Messgenauigkeit wird üblicherweise über die interne Konsistenz (Cronbachs Alpha), Testhälftenkorrelationen, Paralleltestkorrelation oder Retest-Korrelationen bestimmt.

Ein hierzu kontrastierender Ansatz sind die probabilistischen Testtheorien, die in der Literatur meist unter der Bezeichnung Item-Response-Theorie (IRT) zu finden sind [39;40]. Probabilistische Testtheorien "gehen davon aus, dass zwischen dem Testergebnis und der zu messenden latenten Dimension (...) ein Wahrscheinlichkeitszusammenhang besteht, nicht ein deterministischer, der bloß durch das Hinzukommen von Fehlern verwischt ist" ([41], S. 796). Der Zusammenhang zwischen einer Itembeantwortung und einem latenten Merkmal ist insofern probabilistisch, als die Wahrscheinlichkeit einer Antwort auf ein Item (die Wahrscheinlichkeit der beobachteten Reaktionen) von latenten Merkmalen abhängt. Diese Merkmale sind sowohl bestimmte Itemmerkmale als auch die zu messenden Merkmale der Person [42-44].

Probabilistische Testtheorien anzuwenden bedeutet, diese beiden Merkmale mithilfe komplexer statistischer Modelle möglichst genau zu bestimmen, um sie für die Konstruktion eines Testes zu nutzen. Ausgangspunkt sind die Wahrscheinlichkeitsverteilungen für unterschiedliche Antwortoptionen eines Items (daher der Name „Item Response Theorie“) bezogen auf die Ausprägung des zu messenden latenten Merkmals [45]. Das Kernstück stellt die Modellierung des Antwortverhaltens durch eine mathematische kurvilineare Funktion dar, die *item response function* genannt wird und sich auch als *item response curve* visualisieren lässt [43].

Für die Entwicklung eines Testverfahrens müssen diese Wahrscheinlichkeitsfunktionen zunächst ermittelt werden. Die Items werden also im ersten Schritt einer Kalibrierungsstichprobe zur Bearbeitung vorgelegt. Deren Antworten erlauben eine Schätzung der Itemparameter. Darauf aufbauend kann im zweiten Schritt ein Instrument entwickelt werden, das für den Einzelfall eine Schätzung des latenten Traits anhand des beobachteten

Antwortverhaltens der Person und der bekannten Itemparameter vornimmt [43;46;47].

Computerisierte Erhebungsmethoden bieten hierbei den Vorteil, die einzelnen Items nicht mehr in immer gleich bleibender Abfolge vorzugeben, sondern sie an das Antwortverhalten der Patienten anzupassen [46-50]. Nach jeder Antwort des Patienten kann die Merkmalschätzung auf der Grundlage der bekannten Itemparameter neu berechnet werden, wobei die Präzision dieser Schätzung sukzessive zunimmt. Jede neue Merkmalschätzung ist die Grundlage für die dann folgenden Items. Der Testalgorithmus stoppt, wenn die Merkmalschätzung die vorab festgelegte Messpräzision erreicht hat.

Abbildung 1 verdeutlicht diesen Prozess. Zu Beginn wird ein Item präsentiert, das den Merkmalsbereich in seiner gesamten Spannweite abdecken soll. In der Abbildung erfüllt das Item mit seiner polynominalen Verteilung von fünf Antwortalternativen diese Forderung. Die Ausprägung des Merkmals ist auf der x-Achse wiedergegeben. Die Wahrscheinlichkeit, mit der bei einer gegebenen Merkmalsausprägung die jeweilige Antwortalternative angekreuzt wird, ist auf der y-Achse abgetragen. Gibt die Person etwa bei Frage A die Antwortalternative 4 (rote Kurve) an, so ist die Wahrscheinlichkeit, dass ihre Merkmalsausprägung im unteren Bereich der Population liegt, eher gering. Demnach wäre für dieses Merkmal eine Ausprägung zwischen den Prozenträngen 60 bis 80 zu erwarten. Diese stellen die so genannten Item-Response-Thresholds oder Item-Reaktions-Schwellen dar.

Die nächste Frage wird adaptiv so ausgewählt, dass sie eine Präzisierung und Absicherung für genau diesen – nunmehr eingegrenzten – Bereich leistet. Im Idealfall unseres Beispiels liegt ihr Antwortbereich um den Prozentrang von 70. Ihre Antwortverteilung wäre damit leicht nach rechts verschoben, wie in der Abbildung für das Item B ersichtlich wird. Wird auch Item B mit der Antwortalternative 4 beantwortet (rote Kurve), engt sich die wahrscheinliche Ausprägung des Merkmals weiter im oberen Bereich ein (Prozentränge zwischen 78 bis 90). Die nächste Frage sollte wiederum diesen Bereich möglichst gut differenzieren. Die Antworten für die ersten drei Items können in der Folge rechnerisch miteinander in Beziehung gesetzt werden. Dabei wird ermittelt, in welcher Antwortspanne das zu messende Merkmal liegt. Die

Wahrscheinlichkeit hierfür (von der die Breite der Antwortspanne abhängt) wird in dem für das Instrument spezifischen Algorithmus festgelegt.

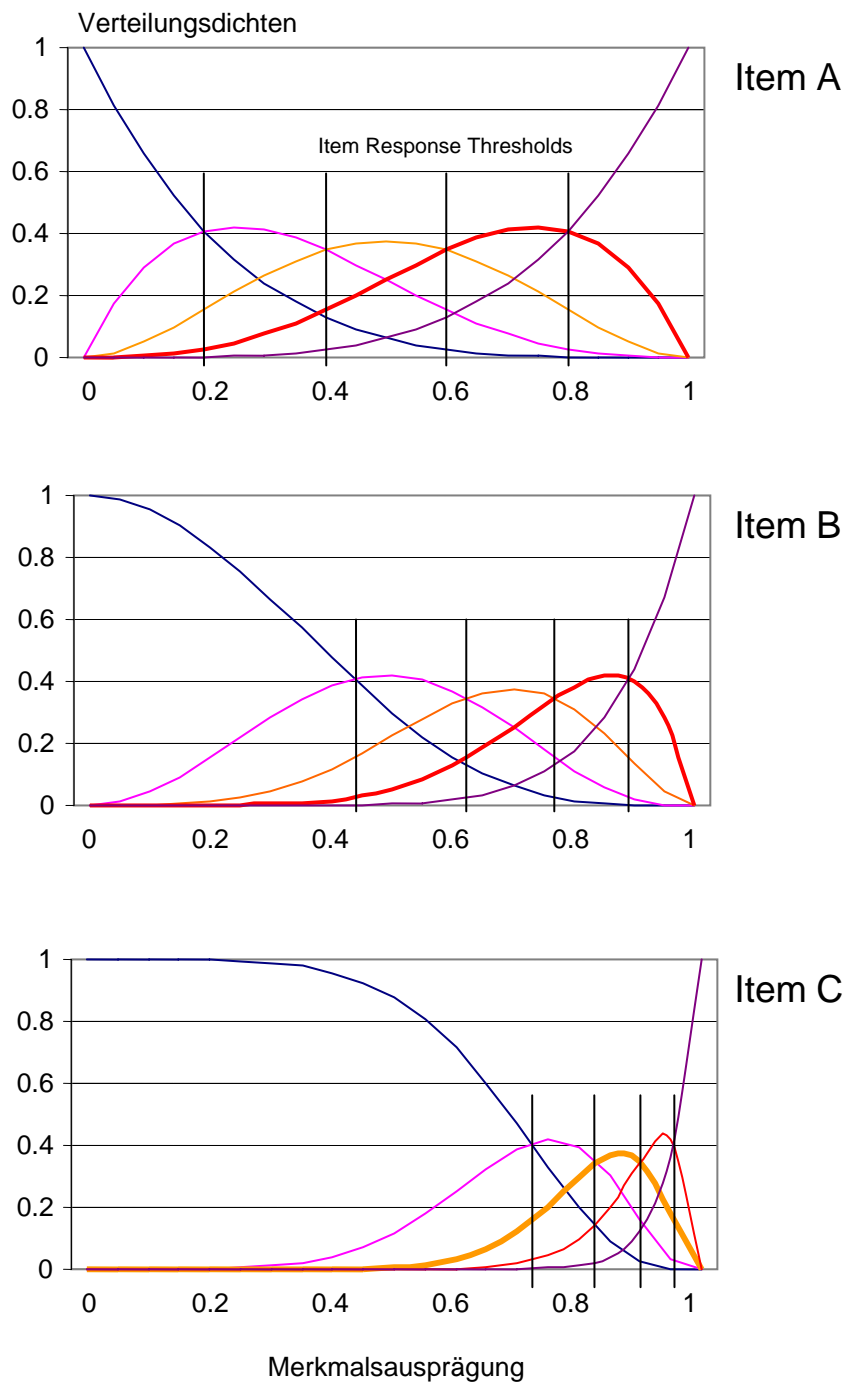


Abbildung 1: Zusammenhang zwischen Merkmalsausprägung und Antwortverteilungen für drei exemplarische Items

In der Abfolge der Items soll also jedes folgende Item den vorher (mit einer festgelegten Wahrscheinlichkeit) ermittelten Ausprägungsbereich des

Merkmals weiter präzisieren. Benötigt werden deshalb Items mit unterschiedlichen Antwortverteilungen, um einen anfangs eher breiten Bereich in verschiedene Richtungen einengen zu können. Die Anzahl der tatsächlich vorgegebenen Items hängt dabei stark vom Individualfall ab. In empirischen Untersuchungen konnten mit IRT-basierten computer-adaptiven Tests erhebliche Itemseinsparungen gegenüber herkömmlichen Verfahren bei gleich bleibend hoher Messgenauigkeit realisiert werden [46].

Der wesentliche Vorteil dieses Vorgehens liegt in einer erheblichen Ökonomisierung der Messung bei größtmöglichem Erhalt der Messpräzision. Ein weiterer Vorteil liegt in der Stichprobenunabhängigkeit sowohl der Itemparameter als auch der Personparameter eines Verfahrens, so dass beispielsweise die Messgenauigkeit eines IRT-basierten Verfahrens – anders als in der KTT – nicht für jede neue Stichprobe neu ermittelt werden muss. Darüber hinaus bietet die IRT weitere Optionen, wie beispielsweise die Messgenauigkeit in extremen Bereichen der Merkmalsausprägung zu kontrollieren, Items aus verschiedenen Erhebungen auf einer gemeinsamen Skala zu vereinen, oder Personen mit inkonsistentem Antwortmuster im Sinne von Falschantworten oder zufälligem Ankreuzen zu ermitteln [39;43;45].

Allerdings ist der primäre Aufwand der Testentwicklung im Vergleich zu herkömmlichen Verfahren aus zwei Gründen höher. Zum einen werden größere Zahlen an Items benötigt, um auch in den extremeren Bereichen des Merkmalskontinuums genügend differenzierende Items zur Verfügung zu haben. Zweitens werden größere Patientstichproben benötigt, um Itemverteilungen zu ermitteln, den Algorithmus zu entwickeln und die Modelle zu überprüfen. Dies erklärt, warum probabilistische Modelle und computer-adaptive Algorithmen bislang noch wenig genutzt werden (siehe auch [51]).

Anwendungen dieser Methodik sind bislang überwiegend im Bereich der Leistungsmessung zu finden. Die Entwicklung klinisch-psychologischer Instrumente, die auf der IRT basieren und computer-adaptives Testen nutzen, hat dagegen erst begonnen [45;46;52].

1.3.7 Klinische Vertiefung: Diagnostik autodestruktiver Syndrome

Das Störungsspektrum autodestruktiver Syndrome ist im Vergleich zu anderen Phänomenen gestörten Erlebens und Verhaltens – etwa Angst, Depression oder gestörtes Essverhalten – besonders wenig empirisch erforscht. Dies liegt nicht zuletzt daran, dass bislang weder hinsichtlich der Diagnostik noch hinsichtlich der klassifikatorischen Einordnung autodestruktiver Syndrome Konsens besteht [53-55].

Autodestruktives Verhalten, das nicht im Rahmen von hirnorganischen Veränderungen, Intelligenzminderung oder psychotischen Erkrankungen auftritt, findet in den gängigen diagnostischen Klassifikationssystemen psychischer und Verhaltensstörungen der Weltgesundheitsorganisation (ICD-10) und der American Psychiatric Association (DSM-IV TR) mehrfach Berücksichtigung, einmal als eigenständige Kategorie in Form der „artificialen Störung“, dann als diagnostisches Kriterium der so genannten „Borderline-Persönlichkeitsstörung“, ferner in den Zusatzcodierungen weiterer Gründe für die Inanspruchnahme medizinischer Leistungen (X- und Z-Codierungen für vorsätzliche Selbstschädigung in der ICD).

Pathologische Selbstschädigung kann sowohl heimlich als auch offen imponieren mittels direkter, indirekter oder kombinierter Manipulationsmethoden seitens der Patienten [54;56-58]. Heimliche Selbstschädigung wird meist als artifizielle Störung begriffen, offene Selbstschädigung eher als Symptom einer Persönlichkeitsstörung. Überschneidungen sind allerdings häufig, so dass hier eher von Störungstypen als von abgegrenzten Kategorien auszugehen ist [54;59].

Während offene Selbstschädigung vor allem in aversiven emotionalen Zuständen auftritt, lässt sich heimliche Selbstschädigung am ehesten als eine Störung im Umgang mit Krankheit und der Krankenrolle interpretieren. Autodestruktives Verhalten scheint in jedem Fall komplexe Funktionen zu erfüllen [60-62]. Es ist davon auszugehen, dass meist prolongierte Krankheitsverläufe vorherrschen, durch Betreiben der Patienten, teils aber

auch durch Zutun des medizinischen Systems bedingt. Dies geht mit ineffektiven Therapieverläufen und mit einem Fortschreiben der erhöhten Inanspruchnahme medizinischer Ressourcen durch diese Patientengruppierung einher [63-66]. In dieser Hinsicht weist die artifizielle Störung deutliche Überschneidungen mit der Somatisierungsstörung auf.

Im deutschsprachigen Raum finden sich bis dato keine gesicherten epidemiologischen Daten zu dem Störungsspektrum autodestruktiven Verhaltens. Um solche Daten zu erheben und Vergleichbarkeit zwischen verschiedenen Studien zu ermöglichen, bedarf es geeigneter Instrumente, die den allgemein gültigen psychometrischen Gütekriterien entsprechen.

Für den Kliniker stellen sich diagnostische und therapeutische Probleme. Für offen autodestruktives Verhalten gibt es keinen eigenständigen und befriedigenden Diagnoseschlüssel. Heimliche Autodestruktion ist gerade infolge der Verheimlichung und der damit verbundenen Täuschungstendenzen der Patienten schwierig zu diagnostizieren. Therapiekonzepte müssen auf Störungswissen gründen, welches wiederum auf der Basis eines klassifikatorischen und diagnostischen Konsenses der Fachwelt gewonnen werden sollte. Dies ist für autodestruktive Syndrome bislang nicht der Fall. Die Formulierung von Therapiekonzepten ist somit erschwert, wenngleich im Rahmen der Behandlung der Borderline-Persönlichkeitsstörung (DSM-IV 301.83) bereits Ansätze zu direkt selbstschädigendem Verhalten entwickelt wurden [67].

Angesichts des geringen empirisch fundierten Wissensstandes zu dem Störungsspektrum und der oft sehr langwierigen und unbefriedigenden Behandlungsverläufe sind psychometrisch überprüfte diagnostische Instrumente vonnöten.