

PÄDAGOGIK

Karlheinz Ingenkamp
Urban Lissmann

Lehrbuch der Pädagogischen Diagnostik

6. Auflage



BELTZ

Ingenkamp/Lissmann
Lehrbuch der Pädagogischen Diagnostik

Karlheinz Ingenkamp/Urban Lissmann

Lehrbuch der Pädagogischen Diagnostik

6. Auflage

Beltz Verlag · Weinheim und Basel

Dr. Karlheinz Ingenkamp war Universitätsprofessor an der Universität Koblenz-Landau, Campus Landau.

PD Dr. Urban Lissmann ist wissenschaftlicher Mitarbeiter am Zentrum für empirische pädagogische Forschung (zepf) der Universität Koblenz-Landau, Campus Landau.

Das Werk und seine Teile sind urheberrechtlich geschützt. Jede Nutzung in anderen als den gesetzlich zugelassenen Fällen bedarf der vorherigen schriftlichen Einwilligung des Verlages. Hinweis zu § 52a UrhG: Weder das Werk noch seine Teile dürfen ohne eine solche Einwilligung eingescannt und in ein Netzwerk eingestellt werden. Dies gilt auch für Intranets von Schulen und sonstigen Bildungseinrichtungen.

6., neu ausgestattete Auflage 2008

Lektorat: Peter E. Kalb

© 2008 Beltz Verlag · Weinheim und Basel
www.beltz.de
Herstellung: Klaus Kaltenberg
Umschlaggestaltung: Glas AG, Seeheim-Jugenheim
e-book

ISBN 978-3-407-29141-7

Inhaltsverzeichnis

Vorwort	9
1. Gegenstand und Entwicklung Pädagogischer Diagnostik	11
1.1 Was ist Pädagogische Diagnostik?	12
1.2 Zur geschichtlichen Entwicklung Pädagogischer Diagnostik	20
1.3 Zur Modelldiskussion in der Pädagogischen Diagnostik	28
1.3.1 Eigenschafts- versus Verhaltensdiagnostik	28
1.3.2 Ergebnis- versus Prozessdiagnostik	31
1.3.3 Selektions- versus Förderdiagnostik	33
2. Methodenfragen Pädagogischer Diagnostik	39
2.1 Aspekte diagnostischen Handelns	39
2.1.1 Vergleich	40
2.1.2 Analyse	41
2.1.3 Prognose	41
2.1.4 Interpretation	42
2.1.5 Mitteilung und Wirkungskontrolle	42
2.2 Grundfragen des Messens und Skalierens in den Sozialwissenschaften	44
2.2.1 Nominalskala	46
2.2.2 Ordinal- oder Rangskala	47
2.2.3 Intervallskala	48
2.2.4 Verhältnis- oder Proportionalskala	49
2.3 Gütekriterien sozialwissenschaftlicher Messungen	51
2.3.1 Objektivität	51
2.3.2 Zuverlässigkeit	54
2.3.3 Gültigkeit	57
2.3.4 Weitere Gütekriterien.....	60
2.3.5 Qualitätsstandards (DIN 33430)	61
2.4 Normen	63
2.4.1 Prozenränge	64
2.4.2 T-Werte und Abweichungs-Intelligenzquotienten	65
2.4.3 Fähigkeitsorientierte Norm	70
2.5 Verhaltensbeobachtung	74
2.5.1 Bedeutung und Merkmale der Beobachtung	74
2.5.2 Gefahren für die Gültigkeit von Beobachtungsergebnissen	75

2.5.3	Merkmale und Arten wissenschaftlicher Beobachtung	77
2.5.4	Index- und Kategoriensysteme	82
2.5.5	Schätzskaleten (Ratingskaleten)	87
2.5.6	Überprüfung der Beobachtungsgüte	92
2.6	Befragungsmethoden	95
2.6.1	Allgemeine Merkmale und Aufgaben	95
2.6.2	Mündliche Befragung oder Interview	96
2.6.3	Schriftliche Befragung oder Fragebogen	101
2.7	Testmethoden	104
2.7.1	Definition und Klassifikation	104
2.7.2	Testpräsentation	107
2.7.3	Formen der Testaufgabe	110
2.7.4	Klassische Testtheorie	117
2.7.5	Probabilistische Testtheorie (Item-Response-Theorie)	121
3.	Schulleistungsdagnostik	130
3.1	Begriff der Schulleistung und Aspekte der Leistungsdagnostik	130
3.2	Mündliche Prüfungen	137
3.3	Schriftliche Prüfungsarbeiten	142
3.3.1	Untersuchungen zu den Gütekriterien schriftlicher Arbeiten	143
3.3.2	Bemühungen zur Verbesserung der Messqualität traditioneller schriftlicher Arbeiten	151
3.4	Schulleistungstests	155
3.4.1	Definition und Untertypen	155
3.4.2	Bezugsgruppenorientierte Schulleistungstests	158
3.4.3	Kriteriumsorientierte Schulleistungstests	165
3.4.4	Informelle Schulleistungstests	172
3.5	Alternative Beurteilungsverfahren	178
3.5.1	Diagnose- und Beobachtungsbogen	178
3.5.2	Lern- und Zeugnisberichte	185
3.5.3	Portfoliobeurteilung	191
3.6	Vergleich unterschiedlicher Methoden der Leistungsmessung	197
4.	Entwicklungs-, Intelligenz- und Eignungsdagnostik im Bildungswesen	202
4.1	Einschulungsdagnostik	203
4.2	Dagnostik bei Lese- und Rechtschreibschwächen	215
4.3	Intelligenzdagnostik in der Schule	226
4.3.1	Intelligenzforschung und -diagnostik	226
4.3.1.1	Psychometrische Theorien der Intelligenz	227
4.3.1.2	Kognitive Intelligenztheorien	233
4.3.2	Intelligenzdagnostik im Rahmen der Feststellung des sonderpädagogischen Förderbedarfs	236
4.3.3	Dagnostik und Beratung beim Übergang zur Sekundarstufe I ...	244
4.4	Dagnostik im tertiären Bildungsbereich	259

5. Diagnostik sozialer und kognitiv-emotionaler Merkmale	272
5.1 Diagnostik sozialer Beziehungen und Erfahrungen	272
5.1.1 Diagnostik des Sozialklimas in Schulklassen	276
5.1.2 Diagnostik sozialer Verhaltensweisen	279
5.2 Diagnostik von Leistungsmotivation und verwandten Konzepten	286
5.2.1 Leistungsmotivation	287
5.2.2 Kausalattribution	289
5.2.3 Bezugsnormorientierung	290
5.2.4 Selbstkonzept der Fähigkeiten	291
5.2.5 Motivation durch Ziele	293
5.2.6 Implizite und explizite Motive	295
5.2.7 Diagnostische Verfahren	296
5.2.8 Fazit	305
5.3 Diagnostik der Prüfungsangst	307
6. Beratung im Bildungswesen	313
7. Repräsentativerhebungen mit Schulleistungstests	323
7.1 Entwicklung der Forschung	323
7.2 Gründe für die überregionale Lernerfolgsmessung	325
7.3 Entwicklung der Forschung in den USA	327
7.3.1 California Assessment Program	327
7.3.2 National Assessment of Educational Progress	329
7.4 Entwicklung der Forschung in Deutschland	331
7.4.1 Internationale Studien mit deutscher Beteiligung	332
7.4.2 Bundesland übergreifende Schulleistungsprojekte	335
7.4.3 Bundesland interne Schulleistungsprojekte	337
7.5 Ertrag und Verwertung der Ergebnisse	341
7.6 Grenzen und unerwünschte Nebenwirkungen	345
Bezugsnachweis für deutschsprachige Tests	351
Literaturverzeichnis	353
Sachregister	395

Vorwort

Dieses »Lehrbuch der Pädagogischen Diagnostik« war 1985 im deutschen Sprachbereich das erste seiner Art. Im Vorwort zur ersten Auflage hatte ich geschildert, wie mich meine eigenen Erfahrungen als Schullehrer, Schulpsychologe und Hochschullehrer von der Notwendigkeit überzeugten, ein derartiges Lehrbuch zu schreiben, weil »unser Bildungssystem dringend eine systematische Ergänzung seiner traditionellen Beurteilungsverfahren durch objektive Methoden braucht und wir mehr für die wissenschaftliche Absicherung unserer Beurteilungsmethoden tun müssen«.

Dieses Lehrbuch hat viele Jahrgänge von Studenten und Lehrern (im Text sind damit immer beide Geschlechter angesprochen) mit Informationen über subjektive und objektive Verfahren versorgt. In den ersten Jahren musste es gegen die Ausläufer der radikalen Testkritik bestehen. In den 1990er-Jahren herrschte dann vor allem Desinteresse an der Weiterentwicklung der Pädagogischen Diagnostik vor.

Das änderte sich schlagartig mit der PISA-Studie. Minister und Journalisten, die frühere Untersuchungen sowie Publikationen nicht zur Kenntnis genommen hatten, überschlugen sich jetzt in Stellungnahmen. Es wäre besser gewesen, sie hätten sich vorher informiert, dass überregionale Leistungsvergleiche und Vergleiche über die Jahre hinweg auch in Deutschland vorlagen und dass ihre Methoden bei uns auf großen internationalen Kongressen diskutiert worden waren.

Überregionale Leistungsvergleiche, wie sie jetzt fast jedes Bundesland organisiert, haben auf längere Sicht nur dann einen Sinn, wenn Schüler, Eltern und Lehrer einen Gewinn davon haben und in die Gestaltung eingebunden sind. Eltern müssen bei der Konzeption der Untersuchung mitreden, Vertreter der Lehrer sollten zu den Anforderungen Stellung nehmen und die Ergebnisse müssen für den Unterricht nutzbar gemacht werden können. Verordnete Tests, von denen die Wissenschaftler nach Genehmigung der Ministerien nur Mittelwerte und Verteilungen sowie Zusammenhänge mit Oberflächenvariablen mitteilen, helfen einer richtig verstandenen Pädagogischen Diagnostik nur wenig und werden Testaversionen wieder aufleben lassen.

Ich bin seit vielen Jahren emeritiert und habe mich einem anderen Sachgebiet zugewandt. Die Bearbeitung dieses Lehrbuchs hat dankenswerterweise einer meiner früheren Studenten, Diplomanden und Doktoranden übernommen, Herr Privatdozent Dr. Urban Lissmann. Er hat mit fundierter Kenntnis neuere Forschungsergebnisse eingearbeitet und die von mir angestrebten Intentionen doch erhalten.

Karlheinz Ingenkamp

1. Gegenstand und Entwicklung Pädagogischer Diagnostik

Der Physiklehrer ließ den 16-jährigen Rolf ein Experiment vor der Klasse an den Geräten erklären. Rolf beschrieb die Anordnung in wenigen Sätzen recht sicher, als die Tür aufging und die Sekretärin eine Nachricht hineinreichte. Nachdem der Physiklehrer sie überflogen hatte, wandte er sich Rolf zu, entschuldigte sich und bat ihn weiter zu erklären. Rolf stand da wie erstarrt. Es half ihm nicht, dass der Lehrer den letzten Satz sinngemäß wiederholte. Mechanisch bewegte er seine Lippen, folgte mit den Augen der Anordnung der Geräte und sprach erst dann, langsam sicherer werdend, weiter und kam zu einem sinnvollen Ende. Der Physiklehrer überlegte: Hatte Rolf alles nur auswendig gelernt und nicht verstanden? Reagierte er überängstlich auf jedes ungewohnte Ereignis? Wie sollte er sich Klarheit verschaffen und Rolf behandeln?

Andreas hatte den in seinem Haus wohnenden Austauschschüler Henry wieder mit ins Jugendheim gebracht. Aber wie an den anderen Tagen vergnügte sich Andreas im Hobbyraum, während Henry mit Marko im Leseraum an einer Zeitung Übersetzungsübungen machte. Der Heimleiter, der dies seit Tagen beobachtete, wusste, dass Marko in der 7. Klasse von der Realschule hatte abgehen müssen und jetzt Mitte der 9. Klasse ein durchschnittlicher Hauptschüler war. Aber wie er Henry ausfragte, sich Redewendungen erklären ließ, wie er Henry über deutschen Satzgebrauch informierte, das zeigte nicht nur Interesse, das ließ auch Begabung erkennen. Hatte der Junge in der Schule nicht den richtigen Draht zum Lehrer gehabt? Wie begabt war er? Welche Förderung und Unterstützung bekam er im Elternhaus? Mit wem sollte man reden und was für Fakten hatte man zur Hand?

Wie immer der pädagogische Alltag aussehen mag, immer wieder stellen sich uns die Fragen: Wie ist diese Beobachtung zu bewerten? Zufall? Anzeichen der »echten« Leistungsfähigkeit? Wie kann man das Verhalten erklären? Was sollte man tun? Wie würde sich dieser, wie jener pädagogische Eingriff auswirken? Alle diese Fragen fallen in den Aufgabenbereich Pädagogischer Diagnostik. Man kann die Fragen vorwissenschaftlich und naiv angehen, oder man kann darüber nachdenken, Verfahrensweisen zur Kenntnis nehmen, Instrumente studieren.

In diesem Kapitel soll im ersten Teil erörtert werden, was Pädagogische Diagnostik ist, im zweiten Teil, wie sie sich historisch entwickelt hat und was sie jetzt leisten will, und im dritten Teil geht es um die Modelle ihres Vorgehens. Die anderen Kapitel sind dann den Methoden und Anwendungsbereichen gewidmet.

1.1 Was ist Pädagogische Diagnostik?

Der Sache nach ist Pädagogische Diagnostik so alt wie pädagogisches Handeln. Wer planmäßig lehrte, hat auch versucht, die Erfolge seiner Bemühungen zu erkunden. Das geschah über viele Jahrtausende pädagogischer Tätigkeit hinweg mit Methoden, die nach heutigem Verständnis als vorwissenschaftlich gelten. Erst in den letzten zwei Jahrhunderten wurden zunehmend wissenschaftlich kontrollierte Verfahren eingesetzt.

Der Begriff »Pädagogische Diagnostik« wurde 1968 von mir in Anlehnung an Medizinische und Psychologische Diagnostik im Rahmen eines Forschungsprojekts vorgeschlagen und von meinem Mitarbeiter Ingo Hartmann 1969 im Titel eines unveröffentlichten Projektberichts benutzt (vgl. Lüking 1976).

Klauers Feststellung »Die Pädagogische Diagnostik ist aus der Psychologischen Diagnostik herausgewachsen« (Klauer 1982b, XI) trifft so nicht zu. Pädagogische Diagnostik war nach ihren Aufgaben, Zielen und Handlungsfeldern immer eigenständig. Sie hat zwar ihre Methoden und auch viele Denkweisen der Psychologischen Diagnostik entlehnt, als sie begann, sich auf wissenschaftlich abgesicherte Verfahren zu besinnen. Aber auch die Psychologische Diagnostik hat einige Jahrzehnte vorher viele ihrer Methoden und Modelle der Medizin und Biologie entlehnt, ohne deswegen aus diesen Disziplinen »herausgewachsen« zu sein.

In den 70er- und 80er-Jahren des letzten Jahrhunderts war Pädagogische Diagnostik noch eher ein – heftig umstrittenes und unklares – Programm als eine etablierte wissenschaftliche Teildisziplin, wie z.B. die Psychologische Diagnostik im Rahmen der Psychologie. Daher ist es nicht verwunderlich, dass es auch recht unterschiedliche Definitionen zur Pädagogischen Diagnostik gibt. Die Bund-Länder-Kommission erklärt im Bildungsgesamtplan: »Unter Pädagogischer Diagnostik werden alle Maßnahmen zur Aufhellung von Problemen und Prozessen sowie zur Messung des Lehr- und Lernerfolges und der Bildungsmöglichkeiten des Einzelnen im pädagogischen Bereich verstanden, insbesondere solche, die der individuellen Entscheidung über die Wahl der anzustrebenden Qualifikationen der Schullaufbahn, des Ausbildungsganges im Tertiären Bereich und der Berufsausbildung sowie der Weiterbildung dienen« (Bund-Länder-Kommission für Bildungsplanung 1973, 75).

Was hier im Vordergrund steht, die Hilfe bei der Schullaufbahnberatung, rückt bei Reulecke und Rollet mehr zur Peripherie, wenn sie schreiben: »Diagnostik in schulischen Entscheidungssituationen hat den Zweck, Informationen zur Optimierung des pädagogischen Handelns zu gewinnen. Entsprechend unterscheidet man zwischen Pädagogischer Diagnostik im engeren Sinn, die die Planung und Kontrolle von Lehr- und Lernprozessen zum Gegenstand hat und Pädagogischer Diagnostik im weiteren Sinn, die alle diagnostischen Aufgaben im Rahmen der Bildungsberatung umfasst« (Reulecke & Rollett 1976, 177).

Unter der Überschrift »Pädagogische Diagnostik ist untrennbar mit Curriculumentwicklung und -evaluation verbunden« erklärt Mauermann: »Pädagogische Diagnostik setzt Lernergebnisse in Beziehung zu Lernvoraussetzungen, schätzt die An-

gemessenheit von Lernzielen oder lernorganisatorischen Maßnahmen auf der Basis von festgestellten Lernbedingungen ab« (Mauermann 1977, 47).

Klauer wirft bisherigen Definitionen vor, dass sie versuchten, die Pädagogische Diagnostik durch Angabe ihrer Aufgabenfelder zu kennzeichnen und behauptet: »Die Klassifikation der pädagogisch-diagnostischen Aufgaben kann überhaupt nicht ein für alle Mal geleistet werden. Sie gehört deshalb nicht in ihre Definition. So verbleibt nur ein relativ formaler Ansatz zur Definition: Pädagogische Diagnostik ist das Insgesamt von Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen« (Klauer 1982b, 5).

Bei dieser allgemeinen Formulierung muss Klauer dann durch Zusatzerklärungen Pädagogische Diagnostik von Forschung abgrenzen. Er tut das, indem er »die diagnostische Erkenntnisbemühung nicht auf die Entdeckung *allgemeiner* Zusammenhänge gerichtet [sieht], sondern auf die nähere Kategorisierung oder Einordnung eines *Einzelfalls*«, der immer im Dienste einer aktuellen pädagogischen Entscheidung steht.

Die Unterscheidung zur Forschung ist wichtig und der Abgrenzung kann weitgehend gefolgt werden. Der Verzicht auf Aufgaben ist aber unnötig und nur zu begründen, wenn man wie Klauer fast karikierend auf Aufgabendetails und ihre jeweils modische Ausprägung zurückgreift. Die zentralen und grundsätzlich bedeutsamen Aufgaben ändern sich im pädagogischen Bereich nicht so schnell und eine Definition für eine Teildisziplin ist fast immer nur auf begrenzte Zeitdauer angelegt. Der interessierte Leser findet weitere Definitionen in der entsprechenden Literatur (Barkey 1976; Döscher, Kuhr & Ziegenspeck 1977; Hopf 1975; Ingenkamp 1977, 1988, 1999c; Jäger, Frey, Wosnitza & Flor 2001; Kutscher, Raeithel & Stevens 1973; Schwarzer & Schwarzer 1979; Sumaski & Lüttge 1976). Die folgende Definition wird für angemessen gehalten:

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.

Unter diagnostischer Tätigkeit wird dabei ein Vorgehen verstanden, in dem (mit oder ohne diagnostische Instrumente) unter Beachtung wissenschaftlicher Gütekriterien beobachtet und befragt wird, die Beobachtungs- und Befragungsergebnisse interpretiert und mitgeteilt werden, um ein Verhalten zu beschreiben und/oder die Gründe für dieses Verhalten zu erläutern und/oder künftiges Verhalten vorherzusagen.

Die Hauptaufgabe der Pädagogischen Diagnostik besteht also darin, für den Lernenden richtige Entscheidungen zu treffen. Die Entscheidungen beziehen sich auf Förderungs-, Platzierungs- und Selektionsmaßnahmen. Pädagogische Diagnostik folgt dabei dem Optimierungsgrundsatz und verwendet wissenschaftliche Methoden. Es sollte auch nicht übersehen werden, dass Pädagogische Diagnostik sich auf den Gesamtbereich pädagogischen Handelns bezieht, auch wenn die diagnostischen Tätigkeiten im Bildungswesen durch Umfang und gesellschaftliches Gewicht dieser Institutionen häufig in den Vordergrund treten.

Diagnostische Tätigkeiten werden auch ausgeübt, wenn es nicht darum geht Lernprozesse bei Individuen oder gemeinsam Lernenden zu verbessern und Hilfen für individuelle Entscheidungen zu erhalten, sondern allgemeinere Erkenntnisse z.B. über die Angemessenheit bestimmter didaktischer Vorgehensweisen, Medien usw. für Schüler mit bestimmten Merkmalen zu gewinnen. Aber das dient dann der Lehr-Lernforschung oder der pädagogischen Forschung, die ja bei empirischen Untersuchungen fast ständig diagnostische Methoden benutzt, ohne deshalb der Pädagogischen Diagnostik zugeordnet werden zu können.¹

Auch die *Evaluation von Curricula* ist m.E. nicht Aufgabe der Pädagogischen Diagnostik. Schon 1974 definierte Wulf: »Evaluation zielt auf die Beschreibung und Bewertung von Reformen im Elementar-, Primär-, Sekundär-, Tertiär- und Weiterbildungsbereich des Bildungswesens« (Wulf 1974, 204). Evaluation in diesem Sinne erfasst Merkmale bei einzelnen Schülern, um Aussagen über ein Curriculum, eine Form der Lernorganisation usw. treffen zu können. Der Schüler wird nur als »Merkmalsträger« gebraucht. Seine persönliche Identität interessiert nicht. In der Pädagogischen Diagnostik werden aber Daten des Lernenden erhoben, um gerade über ihn selbst etwas auszusagen, didaktische Maßnahmen seinen individuellen Voraussetzungen anzupassen. Die Messvorgänge können identisch sein, aber die Aussagen zielen einmal auf allgemeine Tatbestände, das andere Mal auf den individuellen Schüler oder die gemeinsam lernenden Schüler, um aktuelle pädagogische Maßnahmen für diese bestimmten Personen zu verbessern.

Hier sehe ich das wichtigste Unterscheidungskriterium zwischen Pädagogischer Diagnostik und pädagogischer Forschung. Krapp und Prell sind zwar der Auffassung: »Die Evaluation überschreitet den üblichen Aufgabenkatalog der Diagnostik«, aber sie weisen einer »Pädagogischen Diagnostik im Rahmen der Evaluation« Aufgaben der Treatment- oder Systemevaluation zu (Krapp & Prell 1982, 845f.). In der

1 Nach Auffassung des Bearbeiters sind die Grenzen zwischen Pädagogischer Diagnostik und pädagogischer Forschung fließend. Zum Beispiel liefern Repräsentativerhebungen mit Schulleistungstests Erkenntnisse (vgl. Kapitel 7), die zwar primär auf die Schulsystem- und nicht die Individualebene bezogen, aber dennoch pädagogisch diagnostisch relevant sind, weil durch sie Voraussetzungen, Verlauf und Folgen individuellen Lernens erforscht werden. Mit dieser erweiterten Auffassung lässt sich der Gegenstandsbereich der Pädagogischen Diagnostik veränderten heutigen Bedürfnissen anpassen – somit eine Möglichkeit, ihren »Bedeutungsverlust« (im Sinne von Jäger et al. 2001, 849) zu kompensieren.

Beschreibung wird jedoch deutlich, dass die Autoren hier nicht ohne den Begriff der Forschung auskommen.²

Zur Veranschaulichung der in der pädagogischen Beurteilung wirksamen Faktoren wird auf eine schematische Darstellung in der Abbildung 1.1-1 (S. 16) verwiesen. Diese Schemadarstellung ist im Vergleich zu anderen Modelldarstellungen (z.B. Pritz 1981; Viebahn 1980) sehr vereinfacht worden, um wesentliche Faktoren deutlicher hervortreten zu lassen.

Aus dem Schema ist zunächst zu erkennen, dass Merkmale der Beurteilungssituation *und* institutionelle Rahmenbedingungen auf Beurteiler und Beurteilte einwirken. Die Merkmale der Beurteilungssituation können sehr vielfältig sein und schließen die Qualität des vorangegangenen Unterrichts ebenso ein wie das Ziel des Beurteilungsvorganges, die beide wiederum Niveau und Form der Beurteilung beeinflussen. Auch die institutionellen Rahmenbedingungen wirken auf die Beurteilung ein, indem sie Planung und Erwartung gleichermaßen tangieren, z.B. dadurch, dass die Richtlinien bestimmte Prüfungsformen fordern und die hohe Klassenfrequenz mehr individuelle und informelle Beurteilungsvorgänge nicht zulässt. Diese Merkmale und Rahmenbedingungen wirken vor allem auf die Erwartungen und Vorplanungen des Beurteilers und des Beurteilten ein.

Die Merkmale der Beteiligten sind hier zur Vereinfachung fast identisch benannt und nach Kompetenzen, Orientierungen, sozial-emotionalen Aspekten und Erfahrungen gegliedert. Auch der Beurteilte braucht beispielsweise Prognosekompetenz, denn er muss nicht nur aus formellen Ankündigungen des Beurteilers, sondern auch aus seinen vorausgegangenen Handlungen erschließen, welche Beurteilung ihn erwartet und wie der Beurteiler voraussichtlich auf bestimmte Handlungen reagieren wird. In Kompetenzen, Orientierungen und sozial-emotionalen Aspekten wirken relativ überdauernde Einstellungen und situative Erfahrungen in einem ständig neu anzupassenden System zusammen. Wenn diese »Rahmenfaktoren« vorwiegend auch auf die Erwartungs- und Vorplanungsphase einwirken, so beeinflussen sie die nachfolgende Handlungsphase nicht nur über diese Erwartung, sondern auch direkt.

2 Dieser Auffassung kann sich der Bearbeiter nicht anschließen. Aus der in der letzten Fußnote angesprochenen erweiterten Sicht kann auch die *Evaluation von Curricula* zu einer Angelegenheit der Pädagogischen Diagnostik werden. Wenn Wottawa unter Evaluation das »Bewerten von Sachverhalten« versteht, wenn »es also im Gegensatz zur klassischen empirischen Forschung nicht um das bloße Feststellen von Fakten, sondern auch um deren wertende Interpretation und, darauf gestützt, die Optimierung von Entscheidungen« geht (Wottawa 2001, 152), dann werden Schüler zwar als Merkmalsträger gebraucht, um Aussagen über das Curriculum zu treffen, aber die Evaluation kann außerdem dazu beitragen, individuelle Lernbedingungen zu verbessern. Definitionsgemäß lässt sich dann von einer »Evaluation mit Anteilen Pädagogischer Diagnostik« sprechen (Jäger et al. 2001, 850). Dies wurde schon früher von Krapp und Prell so gesehen: »Die Evaluation überschreitet den üblichen Aufgabenkatalog der Diagnostik ...«. Aber daraus zu schließen: »Pädagogische Diagnostik im Rahmen der Evaluation hat also primär die Aufgabe, die Effektivität eines oder mehrerer Treatments zu bestimmen und gegebenenfalls miteinander zu vergleichen« (Krapp & Prell 1982, 845f.) greift zu kurz, weil die Zweckbestimmung »um individuelles Lernen zu optimieren« fehlt.

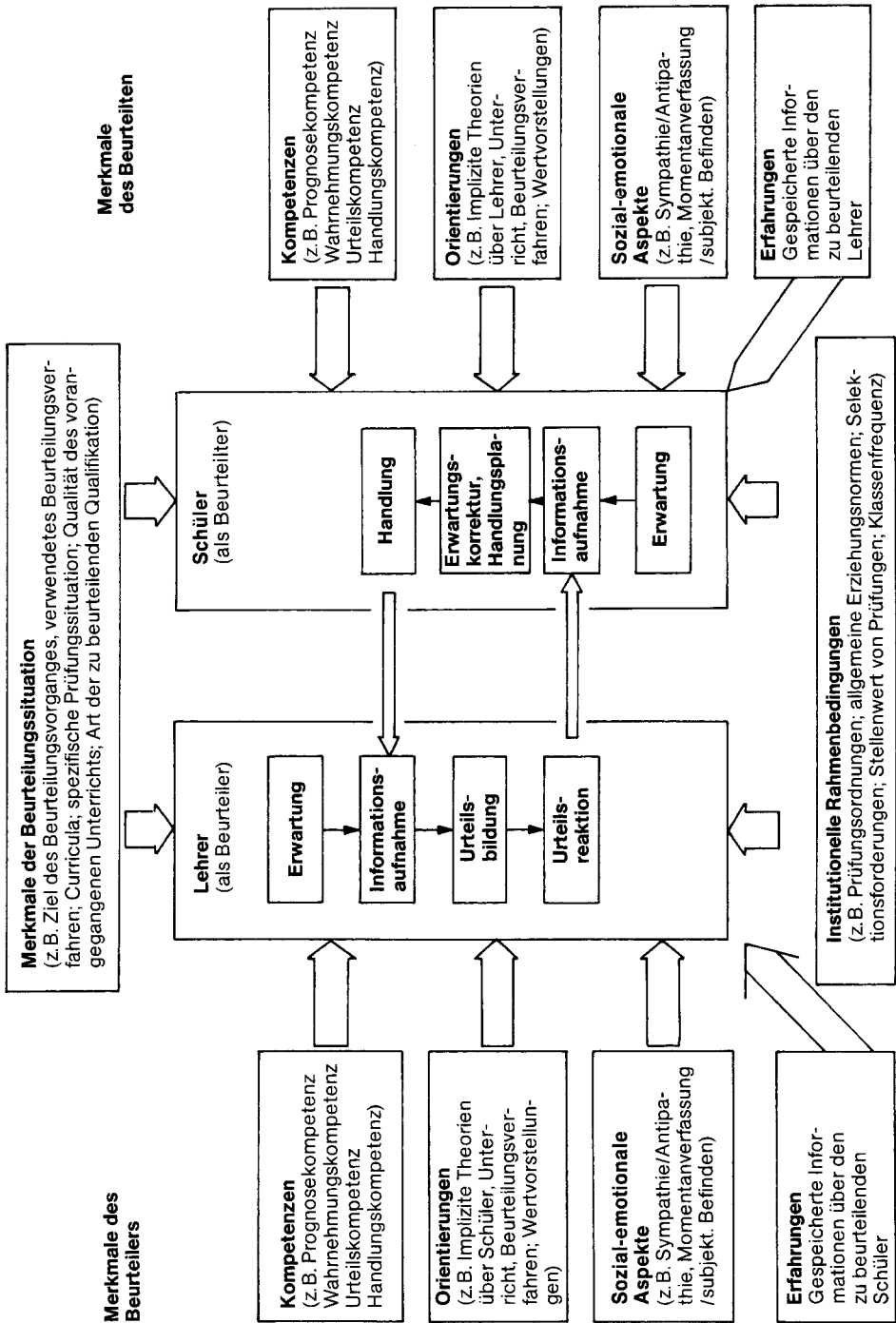


Abb. 1.1-1: Schematische Darstellung der pädagogischen Beurteilung als Interaktion zwischen Beurteiler und Beurteilten

Mit »Informationsaufnahme« ist hier neutral eine Phase bezeichnet, die von einer eher beiläufigen Wahrnehmung in der Unterrichtssituation bis zur standardisierten Testsituation sehr unterschiedliche Gestaltungen erfahren kann. Die hier aufgenommenen Informationen werden zu einem Urteil verarbeitet, das in einer formellen Nachricht dem Schüler mitgeteilt werden kann, das sich aber oft auch nur in einer Geste des Beurteilers ausdrückt, die der Beurteilte wiederum richtig interpretieren, auf seine Erwartungen zurückbeziehen und danach seine Handlungsplanung korrigieren muss.

Die Darstellung in Abbildung 1.1-1 bildet den Urteilsvorgang nicht ab, sondern gibt ein stark vereinfachtes Schema zur Veranschaulichung. Das kann wiederum durch das Schema in der Abbildung 1.1-2 (S. 18) demonstriert werden. Hier ist einerseits auf die Wiedergabe von Determinanten verzichtet worden, andererseits ist der Prozess der Urteilsbildung stärker differenziert worden. Wir beobachten eine Schülerhandlung bereits teilweise selektiv, gesteuert durch unsere Erwartungen. Das wirkt sich umso stärker aus, je weniger standardisiert (Beobachtungsbogen, Test) die Beurteilungssituation ist. Unser Beobachtungsergebnis überprüfen wir mehr oder weniger reflektiert an unserer Erwartung. Je selektiver wir beobachtet haben, desto geringer ist die Chance, eine Unstimmigkeit zu entdecken. Es gibt (selbst-)sichere Beobachter, die durch die Selektivität ihrer Wahrnehmung nur immer ihr eigenes (Vor-)Urteil bestätigen. Wenn wir eine Inkongruenz finden, können wir neue Überprüfungssequenzen in Gang setzen, bis hin zum Neubeginn planmäßiger diagnostischer Bemühungen. Wir können aber auch – früher oder später – aufgeben, zu der Phase kommen, die hier »Resignation« genannt wurde, oft aber auch als »Oberflächlichkeit« oder anders bezeichnet werden könnte, sofern wir nicht die später beschriebenen Experten in Beratungsdiensten einschalten.

Wenn für uns hinreichend erklärt ist, was wir beobachtet haben, dann ordnen wir es nach unseren Normen ein und kommen zu einem Urteil. Dieses kann mehr oder weniger formell mitgeteilt werden und führt gleichzeitig zu Konsequenzen für weitere pädagogische Aktionen.

Auch dieses Schema ist vereinfacht, denn ob wir z.B. »resignieren« hängt natürlich auch wieder von persönlichen Einstellungen, beruflichen Erfahrungen und Kenntnissen über den Schüler, aber auch von anderen Rahmenfaktoren, von der aktuellen Schülerhandlung, von Handlungen anderer Schüler und weiteren situativen Faktoren ab. Der Versuch, hier ein noch differenzierteres Modell darzubieten, ist zum Scheitern verurteilt. Beurteilungsvorgänge sind im Allgemeinen so umfassend, so differenziert, so netzartig in Wechselbeziehungen verbunden, dass sie zweidimensional gar nicht mehr anschaulich dargestellt werden können. Außerdem sind unsere empirisch gesicherten Theorien nicht umfassend genug, um eine wesentlich detailliertere Darstellung zu rechtfertigen.

Eine kurze Bemerkung zum Unterschied von Pädagogischer und *Psychologischer Diagnostik* sei noch angefügt. Pädagogische Diagnostik war nie eine externe Agentur, die auftragsgemäß Klienten prüfte, ihre Merkmale kennzeichnete und sie mit diesen Kennzeichnungen an den Auftraggeber zurückgab. Pädagogische Diagnostik kann

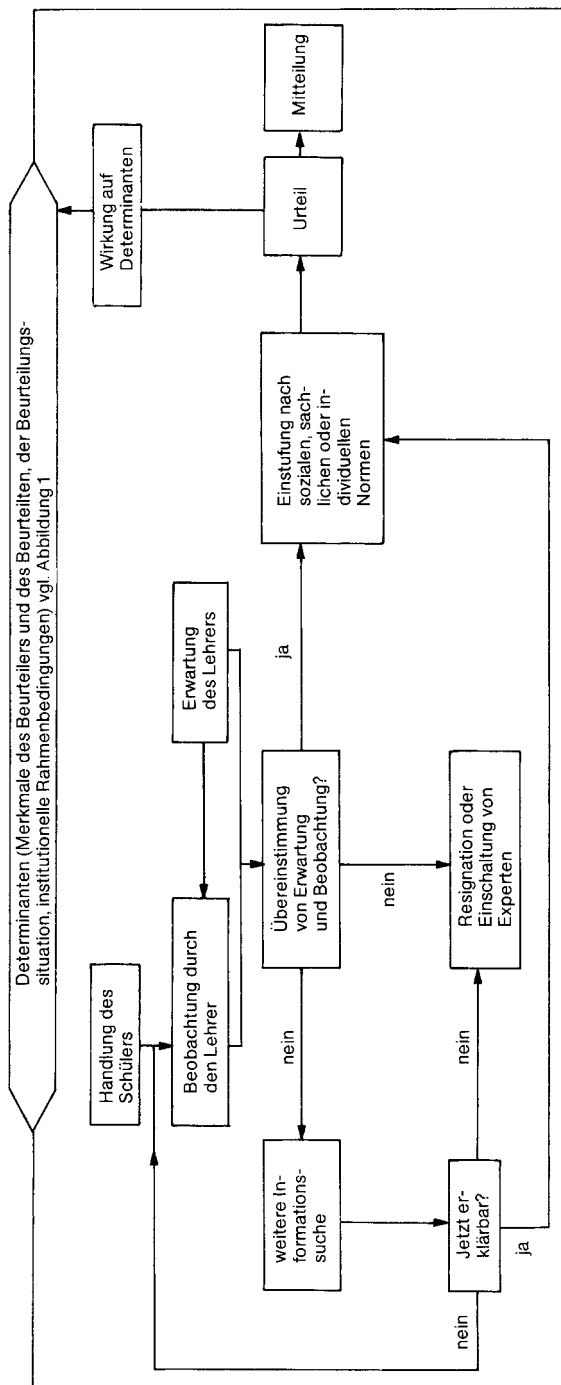


Abb. 1.1-2: Schematische Darstellung der Urteilsbildung in der Pädagogischen Diagnostik

nur Dienstleistung innerhalb von Erziehung und Unterricht sein. Für die Pädagogik ist nicht kennzeichnend, dass sie für das Erkennen und Verstehen eines Gegenstandsbereichs zuständig ist, sondern dass sie Theorie einer Praxis ist und dieser Aufgabe ist die Pädagogische Diagnostik untergeordnet.

Nach Lukesch sind die Pädagogische und *Psychologische Diagnostik* nicht deckungsgleich (Lukesch 1998). Als spezieller Beitrag der Pädagogischen Diagnostik zur Psychologischen Diagnostik gilt Klauers Theorie der kriteriumsorientierten Messung (Klauer 1987; vgl. auch Kapitel 3.4.3).

Während sich zwischen Pädagogischer und Psychologischer Diagnostik noch Unterschiede erkennen lassen, verschwinden sie bei der Gegenüberstellung von Pädagogischer und Pädagogisch-psychologischer Diagnostik. *Pädagogisch-psychologische Diagnostik* hat ihre Wurzeln in der pädagogischen Praxis (Tent & Stelzl 1993) und macht Aussagen zu menschlichem Erleben und Verhalten. Gegenstand der Pädagogischen Diagnostik sind definitionsgemäß die Voraussetzungen, Lernprozesse und Ergebnisse von Lernenden, d.h. es geht letztlich um »Aussagen über menschliches Verhalten und Erleben in Lehr-, Lern- und Erziehungssituationen«. Deshalb sollte man nicht versuchen, die beiden Begriffe künstlich zu trennen, sondern lieber synonym verwenden (Leutner 2001, 521). Für die Gleichsetzung beider Begriffe spricht sich implizit auch Jäger aus, wenn er die Pädagogisch-psychologische Diagnostik mit meiner Definition der Pädagogischen Diagnostik aus dem Jahr 1999 wiedergibt (Jäger 2003a, 313). Es hat den Anschein, dass Erziehungswissenschaftler lieber den Begriff Pädagogische Diagnostik verwenden, während Psychologen den Begriff Pädagogisch-psychologische Diagnostik bevorzugen. Auf diese Weise entsteht jedoch eine neue Verunsicherung und »Fachegoismus« dominiert über interdisziplinäres Denken.

Zusammenfassung

Pädagogische Diagnostik war in vorwissenschaftlicher Ausprägung schon immer Bestandteil des pädagogischen Handelns. In der wissenschaftlichen Ausgestaltung ändern sich prinzipiell nicht Inhalt und Umfang der Beurteilungsmaßnahmen, sondern nur ihre theoretische und methodische Fundierung.

Pädagogische Diagnostik unterscheidet sich von pädagogischer Forschung durch den Bezug auf aktuelle pädagogische Maßnahmen bei einzelnen Lernenden oder einer Gruppe von Lernenden.

In der Pädagogischen Diagnostik werden Lernvorgänge einschließlich ihrer Voraussetzungen, Rahmenbedingungen und Ergebnisse untersucht, um aktuelle Lernvorgänge zu optimieren oder Lernerfolge für die Gesellschaft zu optimieren.

Die im Rahmen der Pädagogischen Diagnostik ablaufenden Urteilsvorgänge sind in jedem Fall durch vielfältige institutionelle, personale und situative Faktoren beeinflusste, sehr umfassende, differenzierte und vernetzte Prozesse, die beim gegenwärtigen Forschungsstand erst ansatzweise abgebildet werden können.

Weiterführende Literatur

Einen aktuellen Überblick über die Pädagogische Diagnostik findet man bei:

Ingenkamp, K. (1999): Pädagogische Diagnostik. In: Jäger, R.S. & Petermann, F. (Hrsg.): Psychologische Diagnostik. Ein Lehrbuch (4. Aufl.) (S. 495–510). Weinheim: Psychologie Verlags Union.

Jäger, R.S., Frey, A., Wosnitza, M. & Flor, D. (2001): Pädagogische Diagnostik. In: Roth, L. (Hrsg.): Pädagogik. Handbuch für Studium und Praxis (S. 848–872). München: Oldenbourg.

Über neuere Entwicklungen zur Pädagogischen Diagnostik informierten die im Beltz Verlag erschienenen Jahrbücher der Pädagogischen Diagnostik. Unter dem Titel »Tests und Trends« lagen von 1981 bis 1997 insgesamt 11 Jahrbücher vor. Die Reihe »Tests und Trends« wird seit 2000 im Verlag Hogrefe weitergeführt (siehe Hasselhorn, Schneider & Marx 2000).

1.2 Zur geschichtlichen Entwicklung Pädagogischer Diagnostik

Aus den Definitionen konnte schon entnommen werden, dass in den Aufgaben der Pädagogischen Diagnostik zwischen

1. der Diagnostik zur Verbesserung des Lernens und
2. der Diagnostik zur Erteilung von Qualifikationen

eine Spannung besteht. Diese Spannung führt häufig zu Rollenkonflikten, wenn beide Aufgaben – wie in Deutschland – in einer Hand, der des Lehrers, liegen. Abgrenzung und Berechtigung beider Aufgabenschwerpunkte werden uns in diesem Buch noch wiederholt beschäftigen. In diesem Teilkapitel werden ein *stärker pädagogisch* und ein *stärker gesellschaftlich* orientierter Aufgabenschwerpunkt unterschieden und festgehalten, dass beide Aufgaben eine – vielfach verwobene – eigene geschichtliche Entwicklung durchgemacht haben, die zur Veranschaulichung hier getrennt skizziert werden soll.

Pädagogische Diagnostik zur **Verbesserung des Lernens** ist unentbehrlicher Bestandteil jedes planmäßigen Lehrvorganges. Wer einen anderen im weitesten Sinne des Wortes unterrichtet, achtet auch darauf, wie der Lernende auf den Unterricht reagiert. Diese Beobachtung des Lernerfolges geschah über viele Jahrhunderte hinweg weitgehend intuitiv und ohne wissenschaftliche Reflexion. Der »geborene« Erzieher erkannte besser als andere, wann Schwierigkeiten im Lernprozess auftraten und reagierte flexibler darauf. Je informeller die Unterrichtssituation war, desto eher konnte auch auf formalisierte Qualitätsmerkmale (die Gütekriterien, vgl. Kapitel 2.3) des diagnostischen Vorganges verzichtet werden. Wer als Hauslehrer die Kinder einer Familie in Latein unterrichtete, brauchte sich nicht an einem methodisch fixierten Maß für Objektivität und Gültigkeit für seine Beobachtungen und Rückmeldungen zu orientieren. Solange die Kinder sich nicht ungerecht und schlecht beurteilt fühlten und solange vor allem die Eltern den Eindruck hatten, der Unterricht sei erfolg-

reich und für die Kinder von erzieherischem Wert, waren die heute erarbeiteten wissenschaftlichen Kriterien für diagnostisches und didaktisches Vorgehen irrelevant. Auch heute ist die Unterrichtssituation entscheidend dafür, in welchem Maße formalisierte Gütekriterien zur Beurteilung des Lehrerverhaltens herangezogen werden. Wenn ein Jugendlicher sich von seinem Klavierlehrer falsch beurteilt und unangemessen behandelt fühlt, dann wird er – sofern die Eltern sich seiner Einschätzung anschließen – im Allgemeinen den Lehrer wechseln. Bei so informellen Unterrichtsveranstaltungen ist die Sicherung diagnostischer Gütekriterien durch Ausbildungsregelungen, Prüfungsvorschriften und Aufsichtsbehörden nicht üblich. Wenn im heutigen Schulunterricht nicht mehr so informell vorgegangen werden kann und bei Beurteilungen die Gütekriterien der Objektivität, Zuverlässigkeit und Gültigkeit beachtet werden müssen, so liegt das nicht daran, dass der Unterricht heute inhumaner, leistungsorientierter oder unpersönlicher ist, sondern daran, dass er Pflichtunterricht und »Massen«-Unterricht ist, der zu Berechtigungen führt. Das macht eine stärkere Regelung notwendig, auch um Willkür und Protektion zu verringern.

Unabhängig von den Rahmenbedingungen des Unterrichts sollte die Diagnose bei der Verbesserung des Lernens vor allem folgenden Aufgaben dienen:

1. der Selbst- und Fremdkorrektur falscher Lernergebnisse,
2. dem Erkennen von Lerndefiziten,
3. der Bestätigung erfolgreicher Lernschritte,
4. der Planung nachfolgender Lernschritte,
5. der Motivierung durch Hinweise auf Lernerfolge und Steuerung des Schwierigkeitsgrads der nächsten Lernschritte,
6. der Verbesserung der Lernbedingungen.

Diese Maßnahmen wurden bei informellem Unterricht mehr intuitiv ergriffen. Im stärker formalisierten Unterricht unserer Zeit kann beobachtet werden, dass diese Handlungen planmäßig in didaktische Systeme eingebaut werden. Das ging mit der Verfeinerung diagnostischer Methoden Hand in Hand. Um die Entwicklung anzudeuten, kann hier nur auf wenige Beispiele verwiesen werden.

Man hat bei Gruppenunterricht schon immer versucht, die Schüler so zu gruppieren, dass sie möglichst viele Lernvoraussetzungen für den Unterricht gemeinsam haben. Auf welche Gruppierungsmerkmale man zurückgriff, hing vom Stand der didaktischen Theoriebildung und von der Qualität der verfügbaren diagnostischen Verfahren ab. Jahrhunderte hindurch hat man sich mit sehr groben Einstufungen des Unterrichtserfolgs oder mit relativ stabilen Merkmalen wie Alter oder »Allgemeinbegabung« begnügt. Die Fortschritte der pädagogisch-psychologischen Forschung und der diagnostischen Techniken haben dann zu immer mehr verfeinerten didaktischen Vorgehensweisen geführt, um individuelle Lernvoraussetzungen und Lernangebote möglichst optimal aufeinander abzustimmen. Unter der Sammelbezeichnung »Adaptiver Unterricht« (Schwarzer & Steinhagen 1975) wurden damals so unterschiedliche didaktische Maßnahmen wie Programmierter Unterricht, Ziel-

erreichendes Lernen (mastery learning), Aptitude-Treatment-Interaktion, Computerunterstütztes Lernen und anderes mehr zusammengefasst. Allen diesen Modellen ist die enge Verzahnung von diagnostischen Schritten und darauf aufbauenden didaktischen Eingriffen gemeinsam.

Im Prinzip bedingen und fördern sich die Individualisierung und Verfeinerung didaktischer Maßnahmen und die steigenden Anforderungen an die Gütekriterien diagnostischer Instrumente gegenseitig. Adaptiver Unterricht ist ohne ein reichhaltiges Arsenal möglichst genauer diagnostischer Verfahrensweisen nicht zu verwirklichen. Dieser Grundsatz wird auch nicht dadurch aufgehoben, dass man bei der häufigen und den Lernprozess unmittelbar begleitenden Diagnose aus praktischen Gründen oft auf leicht verfügbare und mehr informelle Techniken zurückgreifen muss. Wir werden das später noch als *Häufigkeits-Genauigkeits-Dilemma* erörtern. Im Prinzip ist moderne, wissenschaftlich fundierte Didaktik ohne ein reichhaltiges Instrumentarium möglichst objektiver Verfahren der Pädagogischen Diagnostik zum Scheitern verurteilt.

Dieses Instrumentarium ist in den letzten hundert Jahren entwickelt worden, wie den geschichtlich orientierten Darstellungen entnommen werden kann (Ayres 1918; Freeman 1939; Goodenough 1969; Ingenkamp 1990; Judges 1969; Laux 1990; Marsolek 1978; Resnick 1982; neuerdings auch Gregory 2002). Sie zeigen, dass die Vorläufer der Schulleistungstests um 1864 in den »scale books« des Engländers George Fishers gesehen werden können und der Amerikaner J.M. Rice bereits 1894 seine Rechtschreibliste zur Untersuchung des günstigen didaktischen Wegs einsetzte. 1908 publizierte der Thorndike-Schüler Stone den ersten Arithmetiktest. Nachdem die Vereinigung der amerikanischen Schulräte 1914 objektive Verfahren befürwortete, setzte die Flut von Schulleistungstests in den USA ein. Bis in die Gegenwart ist z.B. mit diagnostischen Tests und Kriteriumstests immer ein reichhaltiges Angebot zur Verbesserung des Unterrichts angeboten worden (z.B. Phye 1997).

Die häufige Gleichsetzung von Tests mit Ausleseverfahren ist falsch, wie uns auch die deutschen Erfahrungen bestätigen, wo z.T. rigorose Selektionen traditionell mithilfe von Zensuren vorgenommen werden. Eines der frühen amerikanischen Handbücher (Lindquist 1950) widmet folgenden vier Funktionen des »Educational Measurement« je ein Kapitel: Zur Förderung des Lernens, zur Verbesserung des Unterrichts, zur Schülerberatung und zur Platzierung im Bildungssystem. Heute liegt der Schwerpunkt mehr auf »Educational Assessment«, aber die Ziele sind vergleichbar geblieben: ... Sammeln und Bewerten von Informationen zum Zweck (a) der Diagnose von Stärken und Schwächen des Lernens; (b) der Planung und Verbesserung des Unterrichts und (c) der Evaluation von Leistung einschließlich Platzierungsentscheidungen (Phye 1997, 10). Das beweist, welches Gewicht die Förderung des Lernens und Lehrens in einer wissenschaftlich begründeten Pädagogischen Diagnostik einnimmt.

In viel stärkerem Maße sind diagnostische Tätigkeiten zur **Erteilung von Qualifikationen** von den jeweiligen gesellschaftlichen Strukturen und Forderungen abhängig. Solange die Übernahme von Führungsrollen in der Gesellschaft nicht von er-

worbenen persönlichen Qualifikationen abhing, sondern von mit der Geburt gegebenen Merkmalen (Adel, Stände), vom Ämterkauf oder religiöser Erweckung, brauchte man keine besonderen diagnostischen Verfahren zur Erteilung von Qualifikationen. Als frühestes Beispiel für Verfahren zur Diagnose persönlicher Lernerfolge als Voraussetzung für die Übernahme von Ämtern werden in der Literatur die chinesischen Examina für den öffentlichen Dienst mehr als 1.000 Jahre vor Christi Geburt erwähnt (DuBois 1966). Eine Rekrutierung der Beamten durch Leistungsnachweise war notwendig, weil der Geburtsadel zahlenmäßig zu gering war, um die Ämter zu besetzen.

In Europa war erst nach 1700 eine gesellschaftliche Situation gegeben, in der der absolutistische Staat in seinem Wunsch, die Macht des Geburtsadels zu schwächen und das aufstrebende Bürgertum in seinem Bestreben, Führungspositionen zu übernehmen, ein gemeinsames Interesse hatten, die durch zunehmende Bürokratisierung und Militarisierung gewachsene Zahl von Ämtern mithilfe von Examina zu vergeben. In den meisten europäischen Staaten wurden zwischen 1790 und 1870 Prüfungen für den Öffentlichen Dienst eingeführt. Holmes und Lauwerys (1969) schrieben dazu: Es sollte festgehalten werden, dass eine gesellschaftliche Folge der Einführung von Examen die größere Chancengleichheit war, die sie ermöglichten.

In den letzten beiden Jahrhunderten sind mit zunehmender Demokratisierung die individuellen Lernerfolge immer wichtiger für den persönlichen Aufstieg geworden. Vor allem das Bürgertum hat auf dem Wege über Schul- und Hochschulprüfungen die leitenden Positionen besetzt. Geburtsrechte und Ämterkauf wurden ausgeschaltet. Mit der Pädagogischen Diagnostik zur Feststellung von Qualifikationen war bis in die zweite Hälfte unseres Jahrhunderts hinein ein eminentes gesellschaftspolitisches Reformpotenzial zur Liberalisierung individuellen Aufstiegs verbunden.

Das Instrumentarium zur Kennzeichnung der erworbenen persönlichen Qualifikation entsprach den relativ groben und allgemeinen Anforderungen bei der »Sortierung« von Bewerbern. In vielen Fällen brauchte man nur zu wissen, ob bestimmte Lehrgänge erfolgreich abgeschlossen waren oder nicht. In den anderen Fällen genügte eine Grobeinteilung in vier bis sechs Stufen. Diese Funktion übernahmen die Zensuren, die im *Zeugnis* zusammengefasst wurden. Aus geschichtlichen Darstellungen wissen wir, dass das Zeugnis im Mittelalter als Benefizienzeugnis eine ganz andere Aufgabe hatte (Breitschuh 1979; Dohse 1967; Ingenkamp 1995; Ziegenspeck 1973, 1999). Es wurde nur an bedürftige Schüler als Nachweis zur Erlangung von Stipendien und ähnlichen Unterstützungen ausgegeben und enthielt mehr Informationen über Anwesenheit und Führung als über Schulleistungen. Mit der Durchsetzung der Jahrgangsklasse um 1850 (Ingenkamp 1969b) und der Einführung der allgemeinen vierjährigen Grundschule 1920, durch die die schulische Begutachtung der dort gezeigten Lernerfolge, Begabungen und Interessen für den Übergang zu Sekundarschulen mitentscheidend wurde, gewann das Zeugnis dann eine außerordentliche Bedeutung, die in den Regelungen für Numerus-Clausus-Fächer einen Höhepunkt erreicht hat.

Für die Pädagogische Diagnostik war es eine verhängnisvolle Entwicklung, dass die Erteilung von Qualifikationen und Berechtigungen überwiegend der Schule übertragen wurde und dort alle anderen diagnostischen Aufgaben überwuchert hat.

Die Zensur, die für die Grobentscheidungen des Berechtigungswesens ausreichte, wurde auch das vorherrschende Verfahren zur Kennzeichnung von Lernerfolgen, obwohl sie viel zu wenige Informationen zur Verbesserung von Lernprozessen enthielt. Aber sie war einfach zu handhaben und hatte bereits einen durch Gewohnheit gefestigten Platz, bevor differenzierte und genauere Verfahren zur Verfügung standen. Als Instrument des Berechtigungswesens sind Zensur und Zeugnis in Deutschland bis in die Gegenwart hinein erstaunlich wenig kritisiert und schon gar nicht in amtlichen Regelungen durch Testverfahren ergänzt worden, wenn man die Zulassung zum und Prüfungen im Medizinstudium außer Acht lässt (vgl. Kapitel 4.4).

Die Kritik an der Zensurenggebung und die Entwicklung von *Schultests* erfolgten in Deutschland vor allem unter dem Einfluss messmethodischer Weiterentwicklungen, didaktischer Neuansätze und des Ausbaus der Lehr-Lernforschung. Frühere, vereinzelte Einwände hat Zielinski 1961 referiert. In internationalen Konferenzen um 1930 wurden kritische Analysen der Beurteilungspraxis angeregt. Sie schlugen sich in den Arbeiten von Hartog und Rhodes 1936 in England, Kandel 1936 in den USA und – durch den Nationalsozialismus beendet – von Bobertag 1933 in Deutschland nieder (vgl. Ingenkamp 1975, 1990). Die erste »Welle« der Beschäftigung mit Pädagogischer Diagnostik hat in den angelsächsischen Ländern und in Schweden zu einer Intensivierung der methodischen und pädagogischen Diskussion und zu einem verstärkten Einsatz objektiver Verfahren geführt. In den Ländern unter deutscher Herrschaft waren objektive diagnostische Methoden unerwünscht und der Subjektivismus und die weltanschauliche Gesinnungsforschung belasteten die Praxis der Zensurenggebung noch mehr.

Nach dem Krieg gab es in den ersten eineinhalb Jahrzehnten in Theorie und Praxis Pädagogischer Diagnostik kaum Neuansätze, wenn man von vereinzelten Testkonstruktionen absah, die durch die US-Militärregierung gefördert wurden, aber ohne breiteres Echo blieben (Marsolek 1978). Erst bei den Schulreifetests, von denen 1960 bereits neun Gruppentests verfügbar waren, erfolgte eine stärkere Anwendung. 1960 wurden für den Schulbereich insgesamt 74 Individual- und Gruppentests angeboten, aber ein großer Teil genügte nicht methodischen Gütekriterien. Wichtige Markierungspunkte in dieser Zeit sind die erste Anleitung zur Testentwicklung nach dem Krieg (Belser 1959), Lienerts klassisches Werk über »Testaufbau und Testanalyse« aus dem Jahr 1961 (Lienert & Raatz 1998) und zwei Publikationen über Schultests (Ingenkamp 1962; Samstag & Baus 1962). Hauptanwendungsbereiche objektiver Verfahren waren im nächsten Jahrzehnt Einschulungstests (vgl. Kapitel 4.1) und Tests zur Diagnose der Lese- und Rechtschreibschwäche (vgl. Kapitel 4.2). Tests zur Übergangsauslese spielten eine vergleichsweise geringe Rolle.

Eine internationale Konferenz in Berlin 1967 gab neue Impulse vor allem für informelle Tests (Kapitel 3.4.4) und Kriteriumstests (Kapitel 3.4.3) (Ingenkamp 1985; Ingenkamp & Marsolek 1968). 1970 lagen für den gesamten deutschsprachigen

Raum 122 Schulleistungs-, Einschulungs-, Intelligenz- und Eignungstests, Konzentrationstests und Fragebogen vor (Ingenkamp 1985). Das ist ein dürftiges Testangebot, wenn man bedenkt, dass es auf alle 13 Klassenstufen, alle Unterrichtsfächer und sonstigen Lernzielbereiche bezogen ist. Von diesem Angebot wurde nur in den beiden oben genannten Bereichen und in Modellversuchsschulen in nennenswertem Umfang Gebrauch gemacht. In den Unterrichtsfächern der Sekundarstufe, vor allem im Gymnasium, wurden nur selten Tests eingesetzt. Andererseits haben objektive Verfahren in die Lehr-Lernforschung Eingang gefunden und es gibt praktisch keine wichtige Untersuchung des folgenden Jahrzehnts, die auf sie verzichtet hätte.

Bis 1980 nahm die Zahl der für die Schule angebotenen Tests und Fragebogen zwar auf 222 zu (Ingenkamp 1985), aber ihre Anwendungshäufigkeit sank vor allem wegen der so genannten »Anti-Test-Bewegung« (Zeuch 1973) rapide ab. Die *Testkritik* der 1970er-Jahre wurde hauptsächlich aus folgenden Quellen gespeist:

1. Eine verbreitete *politische* Kritik an der Wachstumsgesellschaft setzte ein und zeigte zunehmend antitechnische und antiwissenschaftliche Züge. In ihren radikalen Formen äußerte sich diese Kritik als neomarxistische *Kapitalismuskritik*.
2. Mit der üblichen Dekade Verspätung erreichte die amerikanische *Anti-Test-Welle* auch Deutschland. In den USA war dies ein Teil der dortigen Staats- und Gesellschaftskritik. In Deutschland wurden bei der starken Fixierung unserer empirisch orientierten Sozialwissenschaftler an amerikanischer Literatur Verfahren in den »Stellvertreterkrieg« einbezogen, die für unsere Gesellschaft eine ganz andere Bedeutung hatten.
3. Die Kritische Theorie der Frankfurter Schule griff auch auf Psychologie und Pädagogik über und wandte sich vor allem gegen die *quantitativen* Vorgehensweisen der empirisch orientierten Sozialwissenschaften.
4. Im allgemeinen Trend gegen *technologische* Vorgehensweisen, gegen quantifizierende Methoden erstarkten auch wieder die vorübergehend zurückgedrängten Kräfte der deutschen geisteswissenschaftlichen Psychologie und Pädagogik und betonten erneut den Vorrang von Subjektivität und subjektivem Verstehen.

Die Kritik war vor allem eine methodische Kritik an Tests und eine Kritik an möglichen Nebenwirkungen von Tests, wie die Vermessung des Menschen, Systemstabilisierung, Steigerung von Angst- und Konkurrenzdenken, Gesellschaftskritik (ausführlicher bei Ingenkamp 1981). Bei der gesellschaftlich orientierten Kritik kommt es oft zu einer Art »Stellvertreterkrieg«, bei dem die Verfahren Pädagogischer Diagnostik attackiert, tatsächlich aber bestehende Gesellschaftsstrukturen gemeint werden. Von der Anti-Test-Bewegung spricht heute kaum noch jemand und mancher Testkritiker hat inzwischen seine Überzeugung geändert.³

Bis 1990 erhöhte sich das Testangebot noch einmal um 12 auf 234, was hauptsächlich auf die Zunahme publizierter Fragebogen im nicht-kognitiven Bereich zurückzuführen war (Lissmann 1992, 553).

3 Vgl. z.B. Grubitzsch 1978 mit 1991 (Grubitzsch 1991b; Grubitzsch & Rexilius 1978).

Als Beurteilungsalternativen, die die Vorteile objektiver Instrumente mit traditionellen Beurteilungsformen verbinden wollten, sind in den 1970er-Jahren vorwiegend an Gesamtschulen und Orientierungsstufen Diagnosebogen entwickelt worden, die sich allerdings nicht durchgesetzt haben und heute in abgespeckter Form als Beobachtungsbogen eingesetzt werden (Kapitel 3.5.1). Die ziffernlosen Zeugnisse und die Zeugnisberichte, die in den letzten Jahren vorwiegend in den ersten Klassenstufen die Ziffernzeugnisse ersetzen, lassen keine messmethodischen Bemühung mehr erkennen, sondern vertrauen ganz auf Ermutigung und pädagogisches Engagement (Kapitel 3.5.2). Die Portfoliobeurteilung steht im Dienste der neuen Anforderungen an Lehren und Lernen (offener Unterricht, Schlüsselqualifikationen, vgl. Kapitel 3.5.3), muss sich aber in Deutschland zunächst einmal bewähren. Engagement allein reicht bei keiner dieser Beurteilungsalternativen aus, um die diagnostische Kompetenz der Lehrer zu erhöhen. Bedauerlicherweise werden unsere Lehrer weder in der Ausbildung noch in der Fort- und Weiterbildung hinreichend auf die Beobachtung der Schüler oder das Schreiben von Zeugnisberichten vorbereitet.

Zwischen 1970 und 2000 sind im allgemeinbildenden Schulwesen in Pädagogischer Diagnostik kleine Fortschritte gemacht worden. Neue Verfahren wurden konstruiert, wie z.B. ein Beurteilungsbogen zur Schulfähigkeit, der nach der klassischen Testtheorie entwickelt wurde, oder curriculare Tests von Lehrern, die vom methodischen Niveau her mit informellen Tests zu vergleichen sind. Außerdem wurden zahlreiche diagnostische Verfahren als Forschungsinstrumente entwickelt, die es eigentlich wert sind, im Unterricht verwendet oder sogar veröffentlicht zu werden. Wie die Sprachstandserhebung bei Ausländern oder Aspekte des Verständnisses sprachlicher Strukturen stammen sie aus der Lernforschung oder der Evaluation von Fördermaßnahmen. Beispiele mit Quellenangaben finden sich in der Literatur bei Probst; dort werden sie als »graue Tests« bezeichnet (Probst 1992). Zwar sind in der Testtheorie, Methodik von Kriteriumstests (Kapitel 3.4.3) und in den soziometrischen Verfahren (Kapitel 5.1) bedeutende wissenschaftliche Fortschritte erzielt worden, aber in der Schulpraxis hat sich das nicht ausgewirkt. Ein zweiter Höhepunkt der Testanwendung, wie zwischen 1962 und 1975, zeichnete sich lediglich im tertiären Bereich ab, als objektive Verfahren ergänzend zur Diagnose der Studierfähigkeit herangezogen wurden (vgl. Kapitel 4.4). Mit der politischen Entscheidung, den Test nicht mehr als Hauptkriterium für die Zulassung zum Medizinstudium zu verwenden, fand der Boom ein jähes Ende.

Auch außerhalb des Bildungswesens, z.B. in der Freien Jugendarbeit und in Betrieben, haben moderne Diagnosemethoden nur vereinzelt Anwendung gefunden. Das Qualifizierungsproblem der diagnostisch Tätigen setzt sowohl der Optimierung des Lernens als auch der gesellschaftlich notwendigen Beratung und Lenkung enge Grenzen. Das wenige, das über die geschichtliche Entwicklung der Pädagogischen Diagnostik in der DDR zu sagen ist, wird an anderer Stelle ausgeführt (Döbert & Geissler 2000; Ettrich & Guthke 1991; Ingenkamp 1999c, 498).

Dennoch kann auf eine Verbesserung der Situation der Pädagogischen Diagnostik gehofft werden. Was mit einer Konferenz internationaler Experten auf Initiative

des Zentrums für empirische pädagogische Forschung begonnen hatte (Ingenkamp & Schreiber 1989), scheint heute Früchte zu tragen. Die Repräsentativerhebungen mit Schulleistungstests, z.B. PISA (vgl. Kapitel 7.4), waren der Anlass, ab 2004 bundesweit Bildungsstandards zu etablieren und deren Einhaltung mit überregionalen Tests zu kontrollieren. Die Tests sollen einmal flächendeckend an den Nahtstellen des Schulsystems eingesetzt werden (vgl. Kapitel 7.5). Ob der Pädagogischen Diagnostik deshalb ein neuer Höhepunkt bevorsteht, wird sich noch zeigen müssen, da außer der Qualifizierung der Lehrerschaft ihre Einstellung zu einer »verordneten« Diagnostik sowie die Finanzierungsfrage der Reform entscheidende Einflussfaktoren darstellen.

Zusammenfassung

Unter den **Aufgaben der Pädagogischen Diagnostik** können die mehr didaktisch orientierten zur Optimierung des Lernens und die mehr gesellschaftlich orientierten zur Steuerung des Bildungsnachwuchses und der Erteilung von Qualifikationen unterschieden werden. Die Methoden des Berechtigungswesens haben dabei in Deutschland die Methoden zur Optimierung des Lernens im letzten Jahrhundert völlig überwuchert und eine Weiterentwicklung didaktisch orientierter Pädagogischer Diagnostik stark behindert. Der **Zifferzensur** wurden immer wieder neue Aufgaben übertragen, ohne dass ihre Eignung dafür überprüft worden war. Da der Zugang zu weiterführender Bildung und damit mittelbar auch zu leitenden Positionen immer stärker vom vorangegangenen Schulerfolg abhängig wurde, gewann die Zifferzensur immer mehr an Gewicht.

Die Entscheidung über die diagnostischen Vorgehensweisen bei der Selektion und der Erteilung von Qualifikationen ist eine politische. Bei dieser Entscheidung muss berücksichtigt werden, dass eine demokratische Gesellschaft sozialen Aufstieg an persönlich erworbene Merkmale binden muss, die in politischen Entscheidungen zu definieren sind. Bei diesen Entscheidungen ist davon auszugehen, dass das Berechtigungswesen zwar eine legitime, aber keineswegs die einzige Aufgabe Pädagogischer Diagnostik ist.

Die Bemühungen um eine Objektivierung der pädagogischen Beurteilung wurden in Deutschland durch politische Verhältnisse und durch das traditionelle Übergewicht der geisteswissenschaftlichen Pädagogik behindert. Erst seit Mitte der 1950er-Jahre wurden **Tests** in einigen Bereichen des Bildungswesens eingesetzt. Aber das Angebot war gering, die Anwendung nahm ab Mitte der 1970er-Jahre deutlich ab. Neueichungen und Neuentwicklungen erfolgten erst wieder in den 1990er-Jahren. In der Lehr-Lernforschung, im tertiären Bereich und neuerdings auch an den Nahtstellen des Schulsystems ist ein Vordringen objektiver Verfahren zu beobachten.

Ohne vielfältige und möglichst genaue Instrumente zur Erfassung des Lernerfolgs und ohne eine entsprechende Ausbildung der Lehrer müssen aber sowohl die didaktischen Verbesserungen zur Optimierung des Lernens als auch gezieltere Bildungsplanung und bessere Bildungsberatung Stückwerk bleiben.

Weiterführende Literatur

Über die Entwicklung objektiver Verfahren in drei Epochen geben einen Überblick:

Ingenkamp, K. (1990): Pädagogische Diagnostik in Deutschland 1885–1932. Weinheim: Deutscher Studienverlag.

- Ingenkamp, K. (1985): 1955–1985: Drei Dekaden Pädagogischer Diagnostik in Deutschland: Entwicklungen, Kontroversen und Perspektiven. In: Jäger, R.S., Horn, R. & Ingenkamp, K. (Hrsg.): Tests und Trends 4. Jahrbuch der Pädagogischen Diagnostik (S. 13–40). Weinheim: Beltz.
- Laux, H. (1990): Pädagogische Diagnostik im Nationalsozialismus 1933–1945. Geschichte der Pädagogischen Diagnostik, Band 2. Weinheim: Deutscher Studien Verlag.

1.3 Zur Modelldiskussion in der Pädagogischen Diagnostik

Die Pädagogische Diagnostik ist nicht nur eng mit ihren Nachbarwissenschaften der Fachdidaktik, Psychologie und Soziologie verbunden, sondern auch so sehr von politischen und kulturellen Strömungen abhängig, dass man keine abgegrenzte, eigenständige und kontinuierliche Entwicklung dieser Teildisziplin erwarten kann.

Im Folgenden soll gezeigt werden, wie abhängig die Pädagogische Diagnostik in der Diskussion um die *Modelle ihres Vorgehens* von den Nachbarwissenschaften und allgemeinen pädagogischen Erwägungen ist. In der Literatur wurden immer wieder vor allem folgende Positionen kontrastierend behandelt:

1. Eigenschafts- versus Verhaltensdiagnostik
2. Ergebnis- versus Prozessdiagnostik
3. Selektions- versus Förderdiagnostik
4. Norm- versus Kriterienorientierung

Die ersten drei Kontroversen sollen so weit erörtert werden, dass ein Verständnis für unterschiedliche Positionen geschaffen wird. Auf die Spannung zwischen norm- und kriteriumsorientierter Messung wird im Kapitel über Schulleistungsdiagnostik eingegangen (Kapitel 3.4.2 und 3.4.3).

1.3.1 Eigenschafts- versus Verhaltensdiagnostik

Diese Kontroverse ist in den 70er- und frühen 80er-Jahren des 20. Jahrhunderts in der Psychologie entstanden. Die Psychologie ist in den vergangenen Jahrzehnten innerhalb der Persönlichkeitspsychologie und der Psychodiagnostik weitgehend dem Denkschema des Alltagshandelns gefolgt. Wenn man einen Menschen auch nur ein wenig kennt, dann registriert man sein Verhalten nicht mehr in jeder Situation der Begegnung wie das Verhalten eines unbekanntem Menschen, von dem man noch nicht weiß, wie er reagieren wird. Von dem bekannten Menschen, den man im Geschäft, im Bus, auf dem Rummelplatz als freundlichen Menschen mit guten Umgangsformen erlebt hat, erwartet man auch auf dem Fußballplatz, dass er sich weitgehend ähnlich verhalten wird, selbst wenn er die Sympathie des anderen für die gegnerische Mannschaft erkennt. Wenn er das nicht tut, werden die eigenen Verhaltenserwartungen mit Einschränkungen versehen, wie z.B. »ein freundlicher, umgänglicher Mensch, solange man nicht beim Fußball, in der Religion o.Ä. anderer

Meinung ist«. Das heißt, die Verhaltensgeneralisierung wird mit spezifischen Einschränkungen versehen, aber man geht noch nicht so weit, das Verhalten des Menschen als nicht vorhersehbar, als »unberechenbar« anzusehen. Wenn man so weit kommt, entsteht im Alltagsleben meist der Eindruck des Krankhaften.

Wenn man Menschen sehr gut kennt, vertraut man seiner Verhaltenserwartung noch viel mehr. Man ist vielleicht zum Schwur bereit, dass dieser friedfertige, entgegenkommende Mensch eine bestimmte Gewalttat nicht begangen haben kann. Bei einem anderen sagt man vielleicht, dass man ihm unter diesen Bedingungen traut, in dieser oder jener Versuchungssituation aber nicht für ihn »die Hand ins Feuer legen würde«. Die Stärke relativ stabiler Merkmale, Eigenschaften, Charakterzüge wird an der Versuchungs- oder Ablenkungskraft bestimmter Situationen relativiert, bevor man zu Vorhersagen kommt. Menschen, die Mitmenschen in Grenzsituationen, z.B. in Krieg und Gefangenschaft, erlebt haben, sind dabei vielleicht etwas skeptischer. Aber im Prinzip tendieren alle dazu, das Verhalten der Mitmenschen möglichst als vorhersehbar einzustufen bzw. die mitmenschlichen Begegnungen in gewisser Weise »berechenbar« zu halten. Dann müssen Menschen aber auch etwas haben, was zu Verhaltenskontinuität führt.

Nicht viel anders sind auch die Annahmen der Differentiellen Psychologie. »Hinter« den individuell unterschiedlichen Verhaltensweisen und Erlebnisverarbeitungen vermutet man eine begrenzte Anzahl von Eigenschaften, Persönlichkeitsmerkmalen, Verhaltensdispositionen, Wesenszügen o.Ä., die dieses Verhalten steuern und ihm über verschiedene Situationen hinweg eine gewisse Stabilität verleihen. Es wird etwas »konstruiert«, was »in uns« für eine Konstanz des Verhaltens sorgt. Daher spricht die Psychologie auch von »Konstrukten«, denn natürlich gibt es keine materielle Lokalisierung dieser Dispositionen.

Auch die Erziehung will diese überdauernden Konstrukte erreichen. Schon Herbart (1776–1841) hat zwischen »Regierung« und »Zucht« unterschieden, wobei sich »Regierung« auf die Erziehung zu Ordnung und Disziplin, »Zucht« jedoch auf die Erziehung des »sittlichen Kerns des Subjekts, auf Motivation und Wille« beziehen (zitiert nach Kron 2001, 201). Immer wieder wird in ähnlichen Definitionen betont, dass Erziehung nicht nur aktuelles Verhalten beeinflussen will, das sei Dressur, sondern die das Verhalten steuernden Kräfte, und zwar langfristig.

Die Differentielle Psychologie hat – vereinfacht gesagt – versucht, jene Eigenschaften zu diagnostizieren, die das Verhalten steuern. Da Menschen sich unterschiedlich verhalten, wurde gefolgert, dass sie auch in ihren »Eigenschaften« differieren. Die gemessenen Eigenschaften erlaubten eine umso einfachere Prognose, je weniger Veränderung sie unterworfen waren. Wäre z.B. das Konstrukt »Intelligenz« die entscheidende Disposition für Schul- und Lebenserfolg, so wäre es von bestechender Einfachheit, die Intelligenz so früh wie möglich zu diagnostizieren und den Menschen in Umwelten einzugliedern, wo die Intelligenz sich weiter entwickeln und nützlich eingesetzt werden kann. Da man nun bald gemerkt hat, dass mit *einer* Disposition allein nur ein Teil des späteren Verhaltens zu erklären war, hat man Eigenschaftsbündel in verschiedenen Kombinationen zur Erklärung herangezogen.