

Leistungsbeurteilung in der Schule

Herausgegeben von

Kurt Heller

Mit Beiträgen von

Peter Büscher, Mannheim

Walter Fingerhut, Marburg

Anne-Katrin Gaedike, Bonn

Kurt Heller, Bonn

Ralf Horn, Weinheim

Hans-Peter Langfeldt, Heidelberg

Erich Langhorst, Bonn

Horst Nickel, Düsseldorf

Bernhard Rosemann, Bonn

Jürgen Wendeler, Frankfurt/M.

Wilhelm Wiczerkowski, Hamburg

Quelle & Meyer Heidelberg

Das in der Umschlaggraphik verwendete
Faktorenmodell der Schulleistung
wurde Kapitel 2.1. in diesem Buch (Seite 42 f.) entnommen.



© Quelle & Meyer, Heidelberg 1974. Alle Rechte vorbehalten. Jede Vervielfältigung, gleich welcher Art und zu welchem Zweck, ist ohne ausdrückliche Genehmigung des Verlags unzulässig.

Printed in Germany.

Satz und Druck: Pilger-Druckerei, GmbH, Speyer.

Umschlagentwurf: Dieter Hoffmann, Heidelberg.

ISBN 3-494-00798-5

Vorwort

Der vorliegende Sammelband mit 12 zum Teil größeren Originalbeiträgen versucht einen Überblick über die wichtigsten Fragen und Probleme schulischer Leistungsbeurteilung zu geben. Dabei werden sowohl Informationen zum gegenwärtigen Stand der Forschung auf diesem Gebiet vermittelt als auch konkrete Hilfen für die praktische Arbeit der Schülerbeurteilung angeboten, so daß Schulpädagogen und Lehrer aller Schularten bzw. Studierende der Erziehungswissenschaft(en), aber auch Psychologen und Soziologen sowie interessierte Laien und in der Bildungspolitik Engagierte angesprochen sind.

Die Einzelbeiträge zum Generalthema „Leistungsbeurteilung“ sind in vier große Problembereiche gruppiert, denen der Herausgeber ein Einleitungs- und Übersichtsreferat vorangestellt hat. Die zusammenfassende Darstellung soll dem Leser eine erste Orientierung über die verschiedenen Themenkomplexe erlauben, nicht jedoch die Lektüre der Originalarbeiten dieses Readers ersetzen. Weiterhin dienen die jedem Hauptkapitel beigefügten Kurzkomentare sowie ein ausführliches Personen- und Sachregister am Ende des Bandes der schnelleren und gründlichen Einarbeitung in die einschlägigen Begriffe, Probleme und methodischen Möglichkeiten schulischer Leistungsbeurteilung.

Die beiden ersten Originalbeiträge sind dem Beurteilungsgegenstand gewidmet. Während zunächst eine operationale Bestimmung dessen, was man gemeinhin als *Schulleistung* bezeichnet, vorgeschlagen wird, informiert das anschließende Sammelreferat ausführlich über die intellektuellen und außer-intellektuellen Determinanten der Schulleistung. Die Kenntnis dieser Zusammenhänge ist gleichermaßen im Hinblick auf die Förderung und Bewertung von Schülerleistungen bedeutsam.

Die zentrale Thematik des Buches findet ihren Niederschlag in den folgenden Kapiteln über *testtheoretische* Fragen sowie *objektive* und *subjektive Verfahren* der Schülerbeurteilung. Eingehend werden die klassische Testtheorie wie auch moderne testtheoretische Ansätze dargestellt und auf ihre Relevanz für die Leistungsbeurteilung hin kritisch untersucht. Weitere Beiträge beschäftigen sich mit Definitionsproblemen von Schultests und einschlägigen Fragen der Lernzieldefinition. Besonderes Gewicht wird im folgenden auf die Beschreibung der *objektiven* Verfahren zur Leistungsbeurteilung, der sog. standardisierten (formellen) und kriteriums- oder lernzielbezogenen (informellen) Schulleistungstests, gelegt. Erklärtes Ziel der beteiligten Autoren ist es, dem Lehrer nicht nur brauchbare Hinweise für die Anwendung und eventuelle Konstruktion solcher Verfahren zu geben; die mit zahlreichen Beispielen versehenen Anleitungen sollen ihm zugleich die praktische Arbeit in der Schule erleichtern, indem sie ihn zur objektiven,

zuverlässigen und gültigen — somit vielleicht gerechteren — Urteilsfindung befähigen.

Aus dem gleichen Anliegen heraus wird den *subjektiven* Verfahren der Schülerbeurteilung durch Lehrer eine breite Diskussion gewidmet. Neben methodischen Problemen der Beobachtung und Beurteilung des Schülerverhaltens im Unterricht kommen hier u. a. neue varianzanalytische Befunde zur Notengebung sowie umfangreiche experimentelle Untersuchungsergebnisse über die verschiedenen Einflüsse der Aufsatzbeurteilung zur Sprache. Im letzten Beitrag werden schließlich Wege zur Vereinheitlichung der Zensurierung von Schüleraufsätzen aufgezeigt — ein Anliegen, das besonders in der Praxis stehende Schulpädagogen interessieren dürfte.

Dieser thematisch weitgesteckte Rahmen in *einem* Band wäre nicht zu realisieren gewesen ohne die bereitwillige Mitarbeit aller angesprochenen Autoren, denen der Herausgeber darüberhinaus wertvolle Anregungen verdankt. Dank schulde ich ferner den Autoren des Beitrags zu Kap. 2.1. für die freundliche Genehmigung, das dort abgebildete Faktorenmodell der Schulleistung als Einbandgrafik zu verwenden, meinen wissenschaftlichen Mitarbeitern Dipl.-Psych. Renate BONN, Dipl.-Psych. Anne-Katrin GAEDIKE und Dipl.-Psych. Manfred SCHNEIDER für die Erstellung der Register und ihre Mithilfe beim Korrekturlesen sowie dem Quelle & Meyer Verlag für die ansprechende Gestaltung und verlegerische Besorgung des Buches.

Bonn, im September 1973

K. H.

Inhaltsverzeichnis

1.	Einleitung und Übersichtsreferat	13
1.1.	Gegenstand schulischer Leistungsbeurteilungen	14
1.1.1.	Zur Problematik des Leistungsbegriffs in der Pädagogik	14
1.1.2.	Hauptdimensionen der Schulleistung	15
1.1.3.	Bedingungskomponenten der Schulleistung	16
1.2.	Aufgaben und Ziele schulischer Leistungsbeurteilung	18
1.2.1.	Leistungsbeurteilung im Dienste der Unterrichtsorganisation und Bildungsreform	18
1.2.2.	Leistungsbeurteilung als Funktion individueller Beratung	19
1.2.3.	Leistungsbeurteilung als Funktion der Schullaufbahn- bzw. Systemberatung	20
1.3.	Formen und Methoden der Leistungsbeurteilung im Bildungswesen	21
1.3.1.	Testtheoretische Grundlagen	21
1.3.1.1.	Die klassische Testtheorie und ihre Kritik	21
1.3.1.2.	Neuere Modellansätze und ihre Problematik	25
1.3.1.3.	Vorschläge zur Klassifikation von Schultests	26
1.3.2.	Objektive Verfahren schulischer Leistungsbeurteilung	27
1.3.2.1.	Lernzieldefinition als Voraussetzung der Leistungsmessung	27
1.3.2.2.	Informelle Tests (Lernkontrolltests)	28
1.3.2.3.	Standardisierte Schulleistungstests	29
1.3.3.	Subjektive Verfahren schulischer Leistungsbeurteilung	30
1.3.3.1.	Verhaltensbeobachtung und Schülerbeurteilung	30
1.3.3.2.	Leistungsbeurteilung durch Notengebung	31
1.3.3.3.	Probleme der Aufsatzbeurteilung	33
1.4.	Ausblick	34
1.5.	Literaturverzeichnis	35
2.	Schulleistung als pädagogisch-psychologisches Problem	37
2.1.	Empirische Ansätze zur Aufklärung des Konstruktes „Schulleistung“ (<i>H. P. Langfeldt</i> und <i>W. Fingerhut</i>)	38
2.2.	Determinanten der Schulleistung (<i>A.-K. Gaedike</i>)	46
2.2.1.	Kognitive Faktoren der Schulleistung	47
2.2.1.1.	Korrelative Beziehungen zwischen verschiedenen Intelligenztests und der Schulleistung	47
2.2.1.2.	Korrelative Beziehungen zwischen verschiedenen Intelligenzfak- toren und der Schulleistung	54
2.2.2.	Nicht-kognitive Faktoren der Schulleistung	60
2.2.2.1.	Motivation und Arbeitshaltung	60
2.2.2.2.	Persönlichkeitsvariablen	65
2.2.2.2.1.	Angst	66
2.2.2.2.2.	Selbstachtung	69
2.2.2.2.3.	Extraversion	70

2.2.2.3.	Lehrerverhalten	72
2.2.2.3.1.	Setzung „sachfremder“ Leistungsmotivation	72
2.2.2.3.1.	Kognitive Stile	74
2.2.2.3.3.	Unterrichtsatmosphäre	75
2.2.2.3.4.	Direktives Verhalten	77
2.2.2.3.5.	Werthaltungen	79
2.2.2.4.	Sozio-kulturelles Milieu (Familie, Peergroups)	80
2.2.2.4.1.	Äußere, objektive Bedingungen	81
2.2.2.4.2.	Werthaltungen, Normen	82
2.2.3.	Schlußbemerkungen	86
2.2.4.	Literaturverzeichnis	87
3.	Testtheoretische Ansätze der Schulleistungsmessung	94
3.1.	Die klassische Testtheorie als Grundlage standardisierter Schulleistungstests (<i>H.-P. Langfeldt</i>)	96
3.1.0.	Vorbemerkung	96
3.1.1.	Messen und Testen	97
3.1.1.1.	Definition von Messen	97
3.1.1.2.	Skalen für Meßwerte	98
3.1.1.3.	Gütekriterien von Messungen	101
3.1.1.4.	Zusammenfassung	101
3.1.2.	Grundlagen der klassischen Testtheorie	102
3.1.2.1.	Fragestellung	102
3.1.2.2.	Axiome der klassischen Testtheorie	102
3.1.2.3.	Das Reliabilitätskonzept	104
3.1.2.4.	Validitätskonzepte	105
3.1.2.5.	Objektivitätsarten	107
3.1.2.6.	Zusammenhänge zwischen den Gütekriterien	108
3.1.2.7.	Zusammenfassung	109
3.1.3.	Reliabilität von Schultests	109
3.1.3.1.	Retest-Reliabilität	109
3.1.3.2.	Paralleltest-Reliabilität	110
3.1.3.3.	Halbierungs-Reliabilität	111
3.1.3.4.	Die Konsistenz-Reliabilität	111
3.1.3.5.	Das Konzept des Standardmeßfehlers	112
3.1.3.6.	Zusammenfassung	113
3.1.4.	Validität von Schultests	114
3.1.4.1.	Curriculare Validität	114
3.1.4.2.	Kriterienbezogene Validität	115
3.1.4.3.	Konstruktvalidität	116
3.1.4.4.	Zusammenfassung	117
3.1.5.	Objektivität von Schultests	117
3.1.5.1.	Durchführungs- und Auswertungsobjektivität	117
3.1.5.2.	Interpretationsobjektivität	118
3.1.5.3.	Normierung von Testergebnissen	119
3.1.5.4.	Kriterien zur Anwendung unterschiedlicher Normen	122
3.1.5.5.	Zusammenfassung	122
3.1.6.	Formaler Aufbau von Schultests	123
3.1.6.1.	Aufgaben und Aufgabenanalyse	123
3.1.6.2.	Untertests und Testbatterien	125

3.1.6.3.	Ein praktisches Beispiel: Der AST 4	125
3.1.6.4.	Abschließende Definition eines Schultests	128
3.1.7.	Diskussion	129
3.1.7.1.	Kritik an der klassischen Testtheorie	129
3.1.7.2.	Kritik an der Anwendung klassischer Tests	131
3.1.8.	Zusammenfassung	134
3.1.9.	Literaturverzeichnis	135
3.2.	Einige testtheoretische Aspekte kriterienbezogener Leistungs- messung (<i>P. Büscher</i>)	137
3.2.1.	Terminologische Probleme	137
3.2.1.1.	Lernzielorientierte und standardisierte Tests	137
3.2.1.2.	Lernzielorientierte und lehrzielorientierte Tests	137
3.2.1.3.	Normbezogene und kriterienbezogene Tests	138
3.2.1.4.	Kriterien	139
3.2.2.	Testtheoretische Probleme	141
3.2.2.1.	Unzulänglichkeit der klassischen Testtheorie bei kriterienbezoge- nen Tests	142
3.2.2.1.1.	Analyse der Aufgabenschwierigkeit	142
3.2.2.1.2.	Analyse der Aufgabentrennschärfe	143
3.2.2.1.3.	Analyse der Aufgabengültigkeit	144
3.2.2.2.	Verfahren zur Test- bzw. Itemanalyse kriterienbezogener Meß- instrumente	144
3.2.2.2.1.	Der Vortest-Nachtest-Differenzindex D_{pp} nach COX und VARGAS	145
3.2.2.2.2.	Die Itemanalyse von POPHAM	146
3.2.2.2.3.	Der Ü-Koeffizient nach FRICKE	148
3.2.2.2.4.	Das RASCH-Modell	149
3.2.2.2.5.	Reliabilitätsbestimmung nach CARVER	151
3.2.2.2.6.	Reliabilitätsbestimmung nach LIVINGSTON	151
3.2.2.2.7.	Reliabilitätsschätzung durch Skalogrammanalyse	152
3.2.2.2.8.	Reliabilitätsschätzung nach CRONBACH	152
3.2.2.2.9.	Reliabilitätsbestimmung nach JACKSON	153
3.2.2.2.10.	Reliabilitätsbestimmung nach FRICKE	153
3.2.2.2.11.	Klassische Reliabilitätsschätzung	153
3.2.2.2.12.	Validitätsbestimmung	153
3.2.3.	Zusammenfassung	155
3.2.4.	Literaturverzeichnis	156
3.3.	Zur Problematik der Klassifikation von Schultests (<i>B. Rosemann</i>)	158
3.3.1.	Untersuchung der bisherigen Terminologie	158
3.3.1.1.	Standardisierte und nichtstandardisierte Tests	158
3.3.1.2.	Der Begriff des Kriteriums	160
3.3.1.3.	Normbezogene versus lernzielorientierte Tests	161
3.3.1.4.	Normbezogene versus kriteriumsbezogene Tests	161
3.3.2.	Gedanken zu einer pädagogisch begründeten Klassifikation der Schultests	163
3.3.2.1.	Die Unterscheidung zwischen Leistungsfeststellung und Leistungs- bewertung	163
3.3.2.2.	Lernsteuerungstests und Lernkontrolltests	165
3.3.3.	Literaturverzeichnis	166

4.	Objektive Verfahren der Leistungsbeurteilung in der Schule	167
4.1.	Leistungsmessung und Lernzieldefinition (<i>R. Horn</i>) . . .	169
4.1.1.	Voraussetzungen der Leistungsmessung	169
4.1.2.	Operationalisierte Lernziele	170
4.1.3.	Erstellung von Prüfungsaufgaben	176
4.1.4.	Zusammenfassung	180
4.1.5.	Literaturverzeichnis	181
4.2.	Konstruktion und Einsatz von Informellen Tests zur Leistungsbeurteilung (Lernkontrolltests) (<i>B. Rosemann</i>) . . .	182
4.2.1.	Einleitung	182
4.2.2.	Die Formulierung der Items	183
4.2.2.1.	Die Spezifikationstabelle	183
4.2.2.2.	Itemtypen und ihre Konstruktion	186
4.2.2.2.1.	Aufgaben mit gebundenen Antworten	187
	Auswahlantworten	187
	Ordnungsantworten	194
4.2.2.2.2.	Aufgaben mit nicht-gebundenen Antworten	196
	Ergänzungs- bzw. Kurzantworten	196
	Kurzaufsatzantwort (Essay-Test)	198
4.2.2.2.3.	Aufgaben zur Messung komplexer Leistungen	198
	Interpretationsübungen	199
4.2.3.	Entwicklung der Test„vorform“	201
4.2.3.1.	Gruppierung der Items und Erstellung des Testheftes	201
4.2.3.2.	Aufgabenbewertung und Ermittlung der Gesamtleistung	205
4.2.4.	Die Itemanalyse	205
4.2.4.1.	Berechnung des Schwierigkeitsgrades einer Aufgabe	206
4.2.4.2.	Berechnung der Trennschärfe einer Aufgabe	206
4.2.4.3.	Distraktorenanalyse	208
4.2.5.	Itemselektion und -revision	209
4.2.6.	Die Reliabilität des Tests	212
4.2.7.	Die Validität des Tests	215
4.2.8.	Die Normierung	215
4.2.9.	Literaturverzeichnis	220
4.3.	Einsatz standardisierter Schulleistungstests (<i>R. Horn</i>) . . .	222
4.3.0.	Vorbemerkung	222
4.3.1.	Lesetests	223
4.3.2.	Rechtschreibtests	224
4.3.3.	Rechentests	225
4.3.4.	Allgemeine Schulleistungstests	225
4.3.5.	Fremdsprachentests	226
4.3.6.	Tests für verschiedene Fächer	
4.3.7.	Übersicht über die zur Zeit verfügbaren Tests nach Testkategorien und Klassenstufen	228 229
5.	Subjektive Verfahren der Leistungsbeurteilung in der Schule	230
5.1.	Beobachtung und Beurteilung des Schülerverhaltens im Unterricht (<i>E. Langhorst</i>)	233

5.1.1.	Beobachtungsnotwendigkeit und Beurteilungspraxis	233
5.1.2.	Beobachtung	235
5.1.3.	Beschreibung	237
5.1.4.	Beurteilung	241
5.1.5.	Beurteilungsfehler	246
5.1.6.	Literaturverzeichnis	251
5.2.	Leistungsbeurteilung durch Notengebung (<i>W. Fingerhut</i> und <i>H.-P. Langfeldt</i>)	253
5.2.1.	Schulnoten als Urteile	253
5.2.2.	Schulnoten als Meßwerte	254
5.2.2.1.	Verteilungsform von Schulnoten	255
5.2.2.2.	Objektivität von Schulnoten	257
5.2.2.3.	Reliabilität von Schulnoten	259
5.2.2.4.	Validität von Schulnoten	260
5.2.3.	Schulnoten als Variablen	262
5.2.3.1.	Methodische Fragen	262
5.2.3.2.	Schulnoten und Intelligenz	263
5.2.3.3.	Schulnoten und soziale Herkunft	264
5.2.3.4.	Schulnoten und Persönlichkeit	265
5.2.4.	Zusammenfassung	266
5.2.5.	Literaturverzeichnis	267
5.3.	Einflüsse auf die Beurteilung von Schüleraufsätzen — Er- gebnisse einer quasi-experimentellen Versuchsreihe (<i>E. Nickel</i> und <i>W. Wiczercowski</i>)	271
5.3.1.	Einleitung und Problemstellung	271
5.3.2.	Die erste Untersuchung	272
5.3.2.1.	Ausgangsproblem und Fragestellungen	272
5.3.2.2.	Versuchsablauf	273
5.3.2.2.1.	Das Beurteilungsmaterial	273
5.3.2.2.2.	Der Versuchsplan	273
5.3.2.2.3.	Beurteiler und Bewertungsvorgang	275
5.3.2.3.	Ergebnisse	276
5.3.2.3.1.	Beurteilerübereinstimmung und Geschlechtsunterschiede	276
5.3.2.3.2.	Verteilungsform und Streuung der Beurteilungen	277
5.3.2.3.3.	Der Einfluß von Informationsart und Vergleichsreizen auf die Urteile	277
5.3.2.3.4.	Der Einfluß verschiedener Sprachmerkmale	280
5.3.3.	Die zweite Untersuchung	285
5.3.3.1.	Fragestellung	285
5.3.3.2.	Versuchsablauf	286
5.3.3.2.1.	Das Beurteilungsmaterial	286
5.3.3.2.2.	Die Beurteiler und ihre experimentelle Beeinflussung	286
5.3.3.2.3.	Der Versuchsplan	287
5.3.3.3.	Ergebnisse	287
5.3.3.3.1.	Globalanalyse und Reduktion des Versuchsplans	287
5.3.3.3.2.	Effekte der Informationsart auf die Aufsatzbewertung	289
5.3.3.3.3.	Einflüsse der Stichprobenzugehörigkeit	291
5.3.3.3.4.	Wechselwirkungen von Informationsart und Erfahrungshinter- grund der Beurteiler	292

5.3.4.	Die dritte Untersuchung	293
5.3.4.1.	Fragestellung	293
5.3.4.2.	Versuchsablauf	294
5.3.4.2.1.	Das Beurteilungsmaterial	294
5.3.4.2.2.	Beurteilungsstichprobe und Bewertungsablauf	294
5.3.4.3.	Ergebnisse	294
5.3.4.3.1.	Der Einfluß der schulpraktischen Erfahrung	294
5.3.4.3.2.	Der Einfluß verschiedener Sprachkriterien	296
5.3.5.	Diskussion und Ergebnisse	297
5.3.5.1.	Allgemeine Befunde zur experimentellen Situation	297
5.3.5.2.	Diskussion der ersten Untersuchung	299
5.3.5.2.1.	Zur Beeinflussung der Beurteilung durch Vergleichsserien	299
5.3.5.2.2.	Zum Einfluß unterschiedlicher Information über die Ausgangs- situation	299
5.3.5.2.3.	Zum Einfluß verschiedener Sprachvariablen	300
5.3.5.3.	Diskussion der zweiten Untersuchung	301
5.3.5.3.1.	Zum Einfluß unterschiedlicher Informationen über die Leistungen der Schüler	301
5.3.5.3.2.	Die Beurteilung durch Schüler, Studenten und Junglehrer	302
5.3.5.3.3.	Geschlechtsspezifische Einflüsse	303
5.3.5.4.	Diskussion der dritten Untersuchung	303
5.3.5.4.1.	Zum Einfluß unterschiedlicher Information über das Leistungs- verhalten der Schüler	303
5.3.5.4.2.	Zur unterschiedlichen Beurteilung durch Referendare und Lehrer	303
5.3.5.4.3.	Zum Einfluß verschiedener Sprachkriterien	304
5.3.6.	Zusammenfassung und Schlußfolgerungen	304
5.3.7.	Literaturverzeichnis	306
5.4.	Bemühungen um Vereinheitlichung der Aufsatzbeurteilung (<i>J. Wendeler</i>)	309
5.4.1.	Die mangelnde Übereinstimmung von Aufsatzbeurteilungen	309
5.4.2.	Leistungsermittlung und Zensierung	312
5.4.3.	Kriterien der Aufsatzbeurteilung	313
5.4.4.	Erfassung der Urteilsobjektivität	316
5.4.5.	Versuche zur Erhöhung der Urteilsobjektivität	317
5.4.6.	Literaturverzeichnis	322
6.	Autorenverzeichnis	324
7.	Personenregister	327
8.	Sachregister	331

1. Einleitung und Übersichtsreferat

Die schulische Leistungsbeurteilung steckt voller Probleme, wie die Ausführungen in diesem Buch noch verdeutlichen werden. Dies gilt sowohl hinsichtlich der theoretischen Grundlegung als auch im Hinblick auf das Verfahrensangebot (Beurteilungsmethoden). Sofern man Leistungsforderung und Leistungsbewertung als berechtigte Anliegen der Schule ansieht, womit jedoch keiner Verabsolutierung des Leistungsprinzips das Wort geredet wird, muß man sich auch den Problemen, die sich hierbei ergeben, stellen. Andernfalls praktiziert man ‚Vogel-Strauß-Politik‘ und huldigt bekannten Grundsätzen wie „Was ich nicht weiß, macht mich nicht heiß!“ oder „Wasch‘ mir den Pelz, aber mach‘ mich nicht naß!“, was im Grunde — mit Blick auf die Hauptbetroffenen: Schüler und Lehrer (einschließlich Eltern) — das Dilemma nur noch verschärft. Aber selbst, wenn man Modeparolen folgend glaubt, dem Problem der Leistung in unserer Gesellschaft überhaupt aus dem Wege gehen zu können, begibt man sich meines Erachtens nur in neue Schwierigkeiten. Das Terrain, auf dem entsprechende Konflikte dann ausgetragen werden, mag zwar wechseln, etwa von der Leistungsebene auf das Feld ideologischer Auseinandersetzung (statt Leistungserweise werden dann ‚Glaubens‘-Bekenntnisse, ‚Überzeugungs‘-Taten und ähnliche Merkmale die Kriterien abgeben), der Problematik selbst, d. h. dem Individuum und der Gesellschaft gleichermaßen Rechnung tragende Verpflichtungen zu postulieren und daraus resultierende Urteilkriterien zu finden, wird man kaum aus dem Wege gehen können. Siehe ergänzend dazu z. B. die Überlegungen von HENTIGS (1970, S. 140 ff.).

Somit wäre unsere Position abgesteckt: Wir erblicken in der schulischen Forderung nach Lernleistungen eine legitime Forderung. Diese stellt freilich nicht das einzige und vielleicht nicht einmal das wichtigste, gleichwohl ein unentbehrliches, Bildungsziel dar. Erkennt man diese Aufgabenfunktion — neben anderen — an, so muß man auch die bei der Verwirklichung dieses Auftrags in Erscheinung tretenden Probleme artikulieren und wissenschaftlich begründete Wege (Methoden) zu ihrer Lösung aufzeigen. Zweifellos gilt ein solches Postulat zunächst für die pädagogische Funktion der Aktivierung von Schülerleistungen, erst danach erhält es seine volle Berechtigung im Hinblick auf die verschiedenen Ziele der Leistungsbeurteilung (z. B. im Rahmen unterrichtlicher Differenzierungsansätze, schulischer Effizienzkontrollen, der Schullaufbahnberatung usw.), die ja letzten Endes wiederum im Kontext schulischer Bildungsbemühungen zu sehen sind.

Die Beurteilung des Schülerverhaltens unter dem Gesichtspunkt der Lernleistung erweist sich somit als ein Problem, das nicht nur die Schule, sondern auch die Familie, die engagierten Wissenschaftler und Bildungspolitiker

und noch viele andere angeht. Aktualität und Virulenz dieser Problematik wurden jüngst beim Streit um die Zulassungsbestimmungen und Kriterien für die Auswahl der Studienbewerber besonders gefragter Studienfächer erneut offenbar. Bedeutet dies nicht eine Herausforderung an uns alle?

1.1. Gegenstand schulischer Leistungsbeurteilungen

1.1.1. Zur Problematik des Leistungsbegriffs in der Pädagogik

Mit dem Begriff der *Schulleistung* sei hier das gesamte Leistungsverhalten, soweit es im Kontext schulischer Bildungsbemühungen virulent wird, angesprochen. Dabei verdienen der dynamische Aspekt (Lernprozeß) und der statische Aspekt (Lernprodukt) gleichermaßen Beachtung — entgegen manchen Begriffsvorstellungen, die nur das Ergebnis der Schülerleistung und nicht auch deren Bedingungsgefüge im Auge haben. Diesen mehr oder weniger operational definierbaren Kategorien können schließlich ethische (Leistungspflicht) und ökonomische Überlegungen (Leistungsnotwendigkeit) — als Sinndimensionen menschlicher Leistung überhaupt — hinzugefügt werden.

Daraus resultiert nach FURCK (1964, S. 118 ff.) eine doppelte Aufgabenfunktion der Schulleistung, nämlich die Persönlichkeitsbildung des Schülers und die Sicherung des volkswirtschaftlichen Leistungspotentials. „Das Problem der Leistung in der Schule spitzt sich so zu der Frage nach dem rechten Verhältnis von individueller Bildsamkeit und ihr angemessener Anforderung zu“ (loc. cit.). Damit ist im weiteren Sinne das pädagogisch-didaktische Anliegen der Leistungsförderung angesprochen. Sicherlich lassen sich noch andere Aspekte und möglicherweise ganz neue Perspektiven zum Schulleistungsproblem aufweisen, die bislang artikulierten, realisierbaren Ansätze (z. B. von FURCK, v. HENTIG oder GAUDE & TESCHNER u. a.) kulminieren letzten Endes doch immer wieder um die hier skizzierten Problem-bereiche. Im Hinblick auf die zentrale Thematik dieses Buches erübrigt sich vorläufig eine weiterführende Diskussion; die eine oder andere Grundsatzfrage wird im Rahmen der einzelnen Beiträge erneut aufgegriffen und im entsprechenden Kontext zu diskutieren sein.

Unsere Definitionsformel „Schulleistung meint das gesamte Leistungsverhalten im Kontext der Schule“ schließt prinzipiell das Schüler- (bzw. Eltern-) und Lehrerverhalten ein, wiewgleich die Vokabel ‚Schulleistung‘ häufig nur die Leistung des Schülers indiziert. Ein solchermaßen eingeeengter Leistungsbegriff ist jedoch nur akzentuierend, d. h. im Hinblick auf jeweils thematisierte Fragestellungen, angebracht. Die Gefahr, hierbei den Gesamtkomplex interdependenter Zusammenhänge aus dem Auge zu verlieren, ist freilich nicht von der Hand zu weisen. Nicht zufällig haben wir den Beitrag von A.-K. GAEDIKE zum Thema *Determinanten* der Schulleistung nebst

dem Diskussionsbeitrag von LANGFELDT & FINGERHUT zur *Faktorenstruktur* der Schulleistung an den Anfang gestellt. Möglichen Mißverständnissen sollte hiermit von vornherein begegnet werden. Wir müssen davon ausgehen, daß Schulleistungen mehrdimensional bedingt sind und in hohem Maße von persönlichkeits- und sozialpsychologischen Faktoren abhängige Variablen(bündel) darstellen. Vor diesem Hintergrund sind praktisch alle Beiträge dieses Sammelbandes zu sehen, auch dann, wenn im einen oder anderen Fall diese Implikationen nicht *expressis verbis* zum Ausdruck kommen.

1.1.2. Hauptdimensionen der Schulleistung

Ausgehend von den üblichen Indikatoren der Schulleistung versuchen LANGFELDT & FINGERHUT in ihrem Beitrag eine empirische Beschreibung dieses Phänomens. Die zu diesem Zweck faktoranalytisch analysierten Schulzeugnisse teils eigener, teils fremder Provenienz deuten übereinstimmend an, „daß die Schulleistung — so wie sie durch Schulnoten erfaßt wird — bei weitem nicht so differenziert ist, wie es Zeugnisse mit zehn und mehr Einzelnoten nahelegen“. Für die Schülerleistungen auf der Sekundarstufe konnten lediglich drei Faktoren nachgewiesen werden: ein Fremdsprachenfaktor, ein mathematisch-naturwissenschaftlicher Faktor und ein sachkundlicher Faktor (sensu DENIG & WEIS 1970). Für die Primarstufe (Grundschüler und lernbehinderte Sonderschüler) ergaben sich sogar nur zwei Faktoren: ein allgemeiner Schulleistungsfaktor (vor allem auf die Fachzeugnisse in Deutsch und Rechnen bezogen) und ein Faktor der schulischen Disziplin (mit entsprechenden Ladungen auf den sog. Kopfnoten). Die Ergebnisse von FUNKE (1972) und ZIMMERMANN (1968) bestätigen — unabhängig voneinander — diese gegenüber dem wesentlich differenzierteren Zeugnisbild stark reduzierte Faktorenstruktur der tatsächlichen Schülerleistung. Im Hinblick auf die Schulleistungsbeurteilung herkömmlicher Art (Zensurierung) ergäbe sich hieraus die Konsequenz, künftig nicht mehr als zwei oder drei Schulnoten zu vergeben. Zumindest unter dem Gesichtspunkt streng leistungsbezogener Beurteilung sind mehr als drei Einzelnoten irreführend und überflüssig zugleich.

Sofern man neben den Lehrerurteilen in Form von Zensuren noch weitere Daten, z. B. Schulleistungstests, Intelligenztests, biographische Informationen usw., in die Faktorenanalyse miteinbezieht, gewinnt man ein höher strukturiertes Modell der Schulleistung. Wie FINGERHUT & LANGFELDT (1971) unter Bezug auf ein entsprechendes Datenmaterial von 1756 Viertkläßkinder nachweisen konnten, ergeben sich hierbei folgende vier Faktoren: 1. „Schulische Leistungsfähigkeit“ mit Ladungen auf den Zensuren der Kernfächer (Deutsch, Rechnen, Heimatkunde), sämtlichen Skalen des AST 4 (Allgemeiner Schulleistungstest von FIPPINGER) und den Untertests

3 + 4 des LPS (Leistungsprüfsystem von W. HORN); 2. „Genereller Notenfaktor“ mit Ladungen auf den Fachleistungs- und Kopfnoten sowie den Variablen Alter und Geschlecht; 3. „Test-Intelligenz“ mit Ladungen auf sämtlichen Skalen des LPS; 4. „Rechenleistungen“ mit Ladungen auf der Rechennote, den Rechenskalen des AST 4 und den LPS-Subtests 3 + 4, 14 und 15 (Arbeitsprobe).

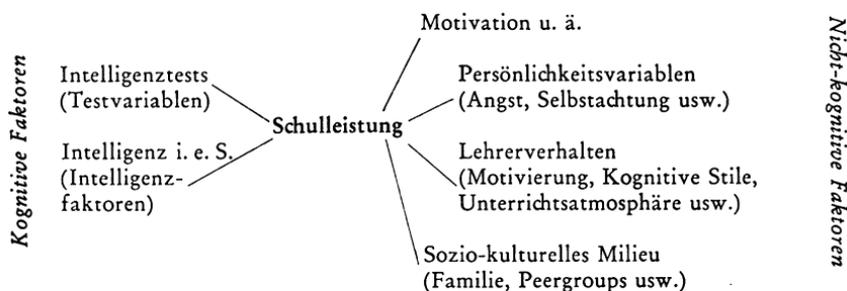
Diese Ergebnisse lassen einen gewissen Zusammenhang zwischen einigen Intelligenzfaktoren (Reasoning, Accuracy, Number) und der (Grund-) Schulleistung, besonders hinsichtlich des allgemeinen Leistungsstandards und der Rechenleistung, erkennen. Die gerade in der jüngeren Literatur häufig vertretene Auffassung, wonach zwischen Intelligenz und Schulleistung keine oder nur schwache Zusammenhänge bestehen, muß demnach korrigiert bzw. folgendermaßen präzisiert werden. „Es kann ... nur gesagt werden, daß Schulnoten und Intelligenz relativ unabhängig voneinander sind“, nicht aber Schulleistung und Intelligenz. Da Schulnoten allein nach den Befunden von LANGFELDT & FINGERHUT kein adäquates Abbild der Schulleistung geben, können diese „nur in mäßigem Umfange etwas über die *tatsächlichen* Bedingungen der Schulleistung aussagen“. Dieses Ergebnis erhellt indirekt die Bedeutung kognitiver Faktoren im Hinblick auf die Konstituierung von Schulleistungen, ohne hiermit den Einfluß nicht-kognitiver Determinanten als gering einzuschätzen.

Eine modifizierte Faktorenanalyse der Daten von FINGERHUT & LANGFELDT ergab schließlich fünf abgrenzbare Cluster: 1. Intelligenztestleistungen (LPS), 2. Schultestleistungen (AST 4), 3. Fachleistungsnoten (Deutsch, Rechnen, Heimatkunde), 4. Kopfnoten (Betragen, Fleiß, Aufmerksamkeit, Ordnung), 5. Merkmale biographischer Art (Alter, Geschlecht, soziale Herkunft). Aus der Distanz der Cluster zueinander postulieren die Autoren jetzt ein zweidimensionales Konstrukt „Schulleistung“, demzufolge sich Schulleistungen im gegenwärtigen Bildungssystem als Funktion von „*tatsächlicher Leistung* aufgrund von Begabung“ (Faktor I) und „Anpassung an dieses System“ (Faktor II) erweisen. Siehe dazu Abb. 1 auf Seite 42 ff. Zur Stützung ihres Interpretationsmodells, das die Autoren als — vorläufigen — Diskussionsbeitrag gewertet wissen wollen, können sie auf faktorenanalytische Untersuchungsbefunde von SEITZ und LÖSER aus den Jahren 1969 bis 1971 verweisen.

1.1.3. Bedingungskomponenten der Schulleistung

Der folgende Beitrag von A.-K. GAEDIKE beschäftigt sich nun mit den einzelnen Determinanten der Schulleistung, wozu die Autorin die wichtigsten Untersuchungen der letzten zwei Dezennien gesammelt hat. Nachstehendes Schema, das wir in Anlehnung an eine frühere Publikation (vgl. HELLER

1970, S. 67) bringen, möge dem Leser die Orientierung durch den Dschungel einer kaum mehr zu sichtenden Zahl von Einzelbefunden zu diesem Themenkomplex erleichtern. Zugleich sind hiermit die Hauptpunkte des Sammelreferats skizziert.



Wenn im folgenden von *Determinanten* die Rede ist, so gilt es zu beachten, daß sich dieser Begriff auf ^{wechselseitige} *korrelative* Beziehungen beschränkt, d. h. nicht ohne weiteres etwas über Ursache-Wirkungs-Zusammenhänge aussagt. Streng genommen kann bei Beziehungen dieser Art nicht exakt „zwischen ‚verursachenden‘ Determinanten und ‚enthaltenen‘ Faktoren“ unterschieden werden. Während der zweite Aspekt bereits Gegenstand der vorhergehenden Betrachtungen war, sollen im Mittelpunkt des GAEDIKESchen Beitrags vorrangig die *Bedingungskomponenten* (erster Bedeutungsaspekt von ‚Determinante‘) der Schulleistung untersucht werden. Für die Interpretation der hier einschlägigen Befunde gelten freilich die angeführten (methodologischen) Kautelen ungemindert.

Zunächst werden die *kognitiven* Determinanten der Schulleistung untersucht, also die Frage, „ob ‚gute‘ Schüler auch zugleich ‚intelligente‘ Schüler sind und umgekehrt“. Die meisten der von GAEDIKE gesammelten Untersuchungsergebnisse zur Korrelation zwischen (allgemeinen) Intelligenztest- und Schulleistungswerten liegen im Bereich von $r = 0.40$ und $r = 0.70$, gelegentlich auch darüber oder darunter. Dabei korrelieren „schulnahe“ Intelligenztests (Verbaltests) im allgemeinen mit der Schulleistung höher als „schulferne“ (Handlungstests). Ferner korrelieren Intelligenztests höher mit Schulleistungstests als mit anderen Indikatoren, z. B. Schulzensuren. Entsprechende Zusammenhänge fallen bei jüngeren Schülern (Grundschule) deutlicher aus als bei älteren Schülern (Sekundarstufe) und dort wiederum deutlicher bei Hauptschülern als bei Realschülern und Gymnasiasten, was analog auch — durchgängig — für den Vergleich von Mädchen und Jungen gilt.

Allerdings tragen nicht alle kognitiven Faktoren in gleichem Maße zum Schulerfolg bei. So bestimmen vor allem verbale Fähigkeiten, Faktoren des

logischen und schlußfolgernden Denkens (Reasoning) sowie Zahlenverständnis, teilweise auch technische Fähigkeiten (Space) und Wahrnehmungsfaktoren (Perceptual Speed, Accuracy) den Bildungserfolg, gemessen an den traditionellen Formen der Schulleistung (siehe noch HELLER 1970, S. 127 ff.). „Es kommt also nicht (nur) darauf an, intelligent oder kreativ zu sein, sondern (von den Bildungsinstitutionen) bevorzugt werden Schüler, die in ganz bestimmter Weise intelligent sind.“

Ausführlicher werden dann die nicht-kognitiven Determinanten der Schulleistung erörtert. Hierbei gewinnen die Lern- und Leistungsmotivation sowie Faktoren der Arbeitshaltung des Schülers, Persönlichkeitsvariablen wie Ängstlichkeit, Selbstachtung, Extraversion vs. Introversion, aber auch Merkmale des Lehrerverhaltens (z. B. Setzung „sachfremder“ Leistungsmotivation, Entwicklung und Pflege kognitiver Stile, Unterrichtsatmosphäre, direktives Verhalten, Werthaltungen und Einstellungen) sowie äußere und innere Bedingungen des sozio-kulturellen Hintergrundes (Familie, Peer-groups) mehr oder weniger starken Einfluß auf das Leistungsverhalten des Schülers. Die aufgezählten Determinanten bilden ein Bezugsgeflecht vielseitiger und mehrschichtiger Einflußgrößen. Dabei scheinen die Bezugsmuster einem Variationsspielraum ausgesetzt zu sein, der größer ist als der, dem die Phänomene der Schulleistung selbst unterliegen.

Somit liegt nach dem resümierenden Urteil der Autorin „noch ein weites Feld für entsprechende Forschungen vor uns“, bis alle Beziehungen aufgedeckt und deren Kenntnis für eine optimale Schulleistungsförderung eingebracht werden können. Immerhin dürfte eine Reihe konkreter Hinweise am Ende jedes Abschnittes diese Aufgabe des Schulpädagogen schon jetzt spürbar erleichtern.

1.2. Aufgaben und Ziele schulischer Leistungsbeurteilung

1.2.1. Leistungsbeurteilung im Dienste der Unterrichtsorganisation und Bildungsreform

Diese Funktion der Leistungsbeurteilung steht in unmittelbarem Zusammenhang mit schulpädagogischen bzw. unterrichtsdidaktischen Problemen. „Nicht die Beurteilung und Benotung eines Schülers, die fast schon zum Selbstzweck geworden ist, sondern die Bewertung des Unterrichtserfolges und die Erfassung von Lernschwierigkeiten ist u. E. die vornehmste Aufgabe pädagogisch-psychologischer Prüfmethode, ausgerichtet auf das Ziel, rechtzeitiges und wirkungsvolles pädagogisches Handeln zu ermöglichen und zu initiieren“ (ZIELINSKI 1973, S. 58). Unter dieser Prämisse dienen schuli-

sche Leistungskontrollen vorab zur Überprüfung aufgestellter Lernziele und zur Diagnose des Lehr-Lernprozesses. Die so gewonnenen Informationen bieten unverzichtbare Hilfen für die Curriculumentwicklung und -revision, die Lernzielbestimmung, die didaktische Planung und Analyse des Unterrichtsverlaufs u. dgl. m. Die Leistungsbeurteilung ist somit notwendiger Bestandteil einer optimalen Unterrichtsgestaltung.

Unter dem Aspekt der (inneren) Bildungsreform wäre hier die Wechselwirkung von Curriculumentwicklung und Leistungsbeurteilung hervorzuheben (vgl. Symposium des Europarates in Berlin 1971). So erfordert die Überprüfung festgelegter Curricula, z. B. im Hinblick auf schulpädagogische und lernpsychologische Möglichkeiten, jeweils bestimmte Methoden ihrer Bewertung. Theoretisch bedeutet dies, daß *alle* kognitiven und nicht-kognitiven Determinanten der konkret geforderten Schülerleistung in einem solchen Bewertungssystem Berücksichtigung finden müssen. Praktisch ist man dieser Idealforderung bislang allenfalls bezüglich der Kontrolle der kognitiven Variablen nahegekommen, wohingegen nicht-kognitive Variablen nur gelegentlich Beachtung fanden (vgl. Kap. 2.2. in diesem Buch). Andererseits bewirkt die Entwicklung bestimmter Beurteilungsmethoden oft auch Modifikationen und Ergänzungen schulischer Curricula, deren Beschränkung auf kognitive Inhalte keine Notwendigkeit darstellt. Daraus läßt sich die Forderung ableiten, die Diskussion über schultestmethodische und curriculare Neuerungen nicht isoliert, sondern in wechselseitiger Abstimmung voranzutreiben. Nur so kann man der Gefahr unerwünschter Stabilisierungseffekte wirksam begegnen.

1.2.2. Leistungsbeurteilung als Funktion individueller Beratung

Dieser Aspekt der Leistungsbeurteilung zielt auf die „Auto-Evaluation“, die Orientierung des Schülers (bzw. seiner Eltern) über seine eigenen Schulleistungen, seine Lernfortschritte (intraindividueller Vergleich) und seine Position innerhalb der Klassen- oder Altersgruppe (interindividueller Vergleich). Dazu ist zweierlei notwendig: erstens die genaue Kenntnis der Lernanforderungen, was eindeutig definierte Lernziele erfordert, und zweitens die Transparenz des Beurteilungssystems, dessen unerläßliche Bestandteile objektive, zuverlässige und gültige Bewertungskriterien bilden. Ferner müssen dem Schüler Notwendigkeit, Zweck und Formen der Leistungsbeurteilung einsichtig gemacht werden.

Während im vorigen Abschnitt unterrichtsdidaktische und somit lehrerzentrierte bzw. gruppenimmanente Interessen (z. B. der Schulklasse bzw. Lerngruppe) im Vordergrund der Betrachtung standen, kommt unter dem Gesichtspunkt schülerzentrierter Aufgabenfunktionen der Individualdiagnose besondere Bedeutung zu. Hier sind diagnostizierte Begabungs-Leistungs-

diskrepanzen (z. B. Phänomene des sog. Underachievement versus Overachievement, d. h. unter- versus übererwartungsgemäße Schulleistungen), vorübergehende oder andauernde Lern- und Leistungshemmungen (z. B. Konzentrationsstörungen, sozio-kulturelle und sensorische Deprivationserscheinungen wie Sprachbarrieren oder Hör- und Sehfehler), aber auch partielle versus allgemeinere Leistungsschwächen (z. B. Legasthenie, Rechenschwäche usw. versus Lernbehinderung[en] im Sinne der Sonderschulbedürftigkeit) sehr oft unerläßliche Voraussetzung für die Bestimmung angemessener (schul)pädagogischer Maßnahmen bzw. therapeutischer Behandlungsansätze. Darüberhinaus will — und sollte — jeder Schüler, auch der ‚unauffällige‘, kontinuierlich über den persönlichen Stand und Fortschritt und erst recht über eventuelle Rückschritte im Lernprozeß informiert werden. Ob diese Information durch Tests, sog. Diagnosebögen oder Zensuren erfolgt, ist prinzipiell von zweitrangiger Bedeutung, sofern die verwendeten Methoden der Leistungsbeurteilung hinreichend objektiv, verläßlich und gültig sind. Daß diese Anforderungskriterien in der Praxis der Schülerbeurteilung — besonders bei den subjektiven Verfahren (z. B. der Notengebung) — häufig nicht erfüllt sind, betrifft eine Reihe von Problemen, die im folgenden Kapitel 1.3. bzw. 1.3.3. zur Sprache kommen.

1.2.3. Leistungsbeurteilung als Funktion der Schullaufbahn- bzw. Systemberatung

Die bisher erörterten Funktionen schulischer Leistungsbeurteilung sind nicht unabhängig von der Struktur des jeweiligen Schul- und Bildungssystems, wie umgekehrt eine gewisse feed-back-Wirkung auf die Bildungsinnovation zu erwarten ist. Von unterschiedlichen Aufgabenschwerpunkten und entsprechend modifiziert eingesetzten Beurteilungstechniken, etwa im dreigliedrigen versus Gesamtschulsystem, abgesehen kulminieren sämtliche Beurteilungs- und Beratungsansätze letztlich um die Differenzierungsproblematik. An dieser Frage und ihrer Lösung führt m. E. kein Weg vorbei, wenigstens ist bislang kein funktionierendes Schulsystem, das diese Problematik ohne Schaden ausklammern könnte, bekannt.

Die Schullaufbahnberatung im traditionellen (vertikal gegliederten) Schulsystem orientiert sich vornehmlich an den Anforderungskriterien der einzelnen Schularten (Hauptschule, Realschule, Gymnasium), für die — über kürzere oder längere Perioden — in etwa konsistente Lern-Leistungsbedingungen bzw. invariante Schülermerkmale postuliert werden. Der Gefahr systemstabilisierender Tendenzen versucht man hier durch regelmäßige Leistungskontrollen mit entsprechenden Übergangsmöglichkeiten (Prinzip der Durchlässigkeit) zu begegnen. Eine solche Laufbahnberatung müßte sich

allerdings weniger am Selektions- als vielmehr am Klassifikationsmodell orientieren (vgl. HELLER 1970).

In der Gesamtschule verlagern sich analoge Wahl- und Entscheidungsprozesse mehr nach ‚innen‘, d. h. in den Unterricht und das didaktische Handeln. Hier werden dann Kriterien benötigt, die angemessene Gruppierungen im Niveauunterricht, Wechsel zwischen Fachleistungskursen u. ä. erlauben bzw. sinnvolle Wahlpflichtkombinationen im Hinblick auf die Berechtigungsfunktion der Schulabschlußqualifikation garantieren. Dazu bedarf es wiederum objektiverer Lernleistungskontrollen.

Die angeführten Beispiele zeigen, daß es bislang und auch wohl in absehbarer Zukunft kein Bildungssystem gibt, das auf die Funktion der Leistungsbeurteilung in dieser oder jener Form verzichten könnte. Die unterschiedlichen Methoden schulischer Leistungsbeurteilung und ihre Einsatzmöglichkeiten im Dienste der hier getrennt aufgewiesenen, in Wirklichkeit jedoch verzahnten, Aufgabenfunktionen stehen deshalb im Mittelpunkt der folgenden Ausführungen. Dabei sollen theoretische und praktische Probleme gleichermaßen Beachtung finden.

1.3. Formen und Methoden der Leistungsbeurteilung im Bildungswesen

1.3.1. Testtheoretische Grundlagen

1.3.1.1. Die klassische Testtheorie und ihre Kritik

Die klassische Testtheorie, auch Meßfehlertheorie genannt, bildet nach wie vor die Hauptgrundlage standardisierter (formeller und informeller) Schulleistungstests. Jeder, der sich solcher Meßverfahren im Rahmen der Schülerbeurteilung bedient, sollte deshalb — schon um unkritischer Anwendung und Fehleinschätzungen vorzubeugen — die theoretischen Voraussetzungen testdiagnostischen Vorgehens in etwa kennen. Da die klassische Testtheorie unabhängig vom Inhalt der einzelnen Verfahren (z. B. Intelligenztest, Leistungstest usw.) gilt, maß man diesem Konzept in der pädagogisch-psychologischen Diagnostik bis in die jüngste Vergangenheit praktisch uneingeschränkte Bedeutung bei. Obwohl heute die Relativierung dieses Anspruchs unbestritten ist, finden sich zunehmend Tendenzen, die Theorie als solche überhaupt abzulehnen. Abgesehen davon, daß man damit das Kind mit dem Bad ausschüttet, entarten solche Versuche nicht selten in absurde ideologische Verstrickungen, wie ein kürzlich erschienener Artikel in der Deutschen Schule eindrucksvoll demonstriert (vgl. NEANDER 1973). Damit ist m. E. weder der theoretischen Neubesinnung noch den praktischen Bedürfnissen in irgendeiner Weise gedient.

Das Wort „Test“ bedeutet ursprünglich ‚Prüfung‘, ‚Stichprobe‘ oder ‚Zeugnis‘, wobei sich der Begriff in der Testdiagnostik vor allem auf standardisierte Verhaltensstichproben bezieht, d. h. auf Prüfungen, die unter genau festgelegten Bedingungen durchgeführt werden. Solche Prüfverfahren oder *Tests* messen bestimmte Merkmalsausprägungen (z. B. Intelligenz, Schulleistung, Ängstlichkeit usw.), indem sie *interindividuelle* Unterschiede (Unterschiede zwischen den einzelnen Individuen) oder *intraindividuelle* Unterschiede (Unterschiede in bezug auf dieselbe Person, z. B. im Verlauf der Ontogenese) erfassen. Bei den sog. standardisierten Schulleistungstests steht der erste Aspekt im Vordergrund, d. h. der Vergleich eines individuellen Meßwertes im Test mit der Gruppennorm, die als Testleistungsmaßstab in Form von Alters- und/oder Schul- bzw. Klassennormen operational definiert ist.

In dem testtheoretischen Beitrag von LANGFELDT (s. Kap. 3.1.) werden zunächst die Begriffe „Messen“ (Testen) und „Meßskala“ sowie die wichtigsten Anforderungskriterien in bezug auf Messungen, auch „Gütekriterien“ genannt, erläutert. *Messen* (Testen) wird hier als ein Vorgang des Vergleichens anhand eines Maßstabs (Testnormen) aufgefaßt, wobei die Meßwerte durch unterschiedliche Skalenniveaus repräsentiert sein können. Die wichtigsten Skalentypen sind: 1. die *Nominal-* oder *Klassifikations-skala* (unterste Ebene des Messens), die nur einem Kriterium, dem der Äquivalenz, genügt; 2. die *Ordinal-* oder *Rangskala* (nächsthöhere Ebene), die zwei Kriterien, dem der Äquivalenz und dem der rangmäßigen Beziehung, genügt; 3. die *Intervallskala*, die drei Kriterien, der Äquivalenz, der Rangbeziehung und der Intervallkonstanz, genügt; 4. die *Rational-* oder *Verhältnisskala* (oberste Ebene des Messens), die zusätzlich zu den aufgeführten Kriterien einen absoluten Nullpunkt aufweist, also vier Kriterien genügt. Je höher das repräsentierte Skalenniveau ist, desto besser und vielfältiger sind die Verarbeitungsmöglichkeiten entsprechender Kennwerte und damit ihr Informationsgehalt. Schulleistungstests messen in der Regel auf Rangskalenniveau, selten auf Intervallskalenniveau (z. B. Intelligenztests).

Unabhängig vom jeweiligen Skalenniveau wird der Interpretationsspielraum noch durch die sog. *Testgütekriterien* (Objektivität, Reliabilität, Validität) beeinflusst. *Objektivität* meint hier die Unabhängigkeit der Testergebnisse von der Person des Testleiters sowohl in bezug auf die Testanweisung (Instruktion) als auch in bezug auf die Testauswertung und Testinterpretation. Die Durchführungsbestimmungen eines Tests müssen so präzise festgelegt sein, daß Intersubjektivität gewährleistet ist. Diese Forderung ist in der Praxis der Testdurchführung nicht immer leicht zu erfüllen. Die *Reliabilität* oder Zuverlässigkeit bezieht sich auf die Meßgenauigkeit. Ein Schulleistungstest ist beispielsweise dann reliabel (zuverlässig), wenn die Ergebnisse unabhängig vom Zeitpunkt der Messung zustandekommen, d. h.

wenn die wiederholte Anwendung des betr. Tests bei derselben Klasse oder einzelnen Individuen in etwa zum gleichen Resultat führt. Es gibt verschiedene Aspekte der Reliabilität und dementsprechend unterschiedliche Methoden ihrer Kontrolle, die alle von LANGFELDT besprochen werden. Die *Validität* oder Gültigkeit bezieht sich auf die Genauigkeit, mit der ein Test das mißt, was er messen soll. So sagt die Validität eines Rechentests etwas darüber aus, wie genau tatsächlich Rechenkenntnisse (und nicht etwa Konzentrationsfähigkeit oder andere Merkmale) erfaßt werden. Während sich also die Reliabilität auf die Zuverlässigkeit des Meßinstrumentes (Tests) selbst und damit auf die formale Meßgenauigkeit bezieht, informiert uns die Validität darüber, „welche psychodiagnostischen Schlußfolgerungen die numerischen Resultate eines Tests zulassen und welchen Grad an Sicherheit solche Schlußfolgerungen aufweisen“ (Michel 1964, S. 47). Damit ist die diagnostische Valenz oder Treffsicherheit eines Testverfahrens angesprochen. Was endlich den Zusammenhang zwischen Reliabilität und Validität betrifft, so gilt: Die Reliabilität ist die notwendige, aber keine hinreichende, Voraussetzung für die Validität eines Tests. Sehr oft erweist sich die Validierung eines Schultests als das schwierigste Unterfangen überhaupt, wobei man sich nicht selten mit der Bestimmung der curricularen Gültigkeit begnügt. Seltener kommen hier die Methoden der Kriterienvalidierung (z. B. zur Ermittlung der Übereinstimmungsvalidität von Testleistung und Schulnoten als „Außenkriterien“) oder der Konstruktvalidierung (analog zur Validierung „neuer“ Intelligenztests; vgl. HELLER 1973, S. 75) zum Einsatz. Bei der Konstruktion eines Schulleistungstests kommt man den Anforderungskriterien der Reliabilität und Validität dadurch entgegen, daß man schon frühzeitig — in der sog. Aufgaben- oder Itemanalyse — *Schwierigkeit* und *Trennschärfe* jeder einzelnen Testaufgabe ermittelt, um von vornherein die unbrauchbaren Aufgaben (Items) auszusondern.

Damit nun die Ergebnisse eines Schulleistungstests vernünftig interpretiert werden können, sind zwei weitere Voraussetzungen unerlässlich: die Normierung des Tests und die Ermittlung des sog. (Standard-)Meßfehlers. In den *Testnormen* (z. B. PR = Prozentränge auf Ordinalskalenniveau versus Z, C, Abweichungs-IQ u. ä. Standardwert-Normen auf Intervallskalenniveau bzw. flächentransformierte T-Standard-Äquivalent-Normen) liegen hierfür relativierte Testwerte, d. h. auf die jeweilige Alters- oder Klassen-
gruppe bezogene Vergleichsmaßstäbe vor. So besagt eine Schülerleistung auf dem 75. PR, daß der betreffende Schüler in dem untersuchten Schulfach bessere Leistungen erzielt als 75 % seiner Bezugsgruppe (und schlechtere Leistungen als 25 %), während eine Leistung auf dem 50. PR durchschnittliche Leistungsfähigkeit indiziert. Nun wäre es naiv anzunehmen, daß die ermittelte (beobachtete) Testleistung von PR = 75 exakt die tatsächliche (wahre) Leistungsfähigkeit des Schülers wiedergibt; vielmehr muß davon

ausgegangen werden, daß in diesem Testergebnis — wie in jedem Meßwert — eine zunächst unbekannte Fehlerquote steckt, die zu Lasten der Irreliabilität bzw. mangelnden Objektivität einer Untersuchung geht. Diese Annahme trifft im Kern das *Meßfehlerkonzept*, das wichtigste Axiom der klassischen Testtheorie. Danach setzt sich jeder *beobachtete Wert* (Meßwert) aus dem „wahren Wert“ (einem zeitlich konstanten Parameter) und einem „Fehlerwert“ (Meßfehler) zusammen. Beide werden als unkorrelierte Größen postuliert, wobei der Meßfehler als Zufallsvariable auftreten soll. Mit dessen Hilfe werden dann — auf der Basis wahrscheinlichkeitstheoretischer Überlegungen — die sog. *Vertrauensintervalle* oder *Konfidenzintervalle* ermittelt, d. h. jene Bereiche, in denen mit einer bestimmten Wahrscheinlichkeit der wahre Wert erwartet wird. Auf diese Weise läßt sich die Interpretation der durch Tests erfaßten Schülerleistungen auf eine (meßtheoretisch) gesicherte Grundlage stellen — ein Vorteil gegenüber anderen Formen der Schülerbeurteilung (z. B. Zensurierung), der sowohl dem interindividuellen Vergleich als auch intraindividuellen Untersuchungszielen (z. B. der Erfassung von Lernleistungsfortschritten oder sog. Profilanalysen) zugute kommt (s. noch HELLER 1973, S. 66 ff. u. 157 ff.).

Die *Kritik der klassischen Testtheorie* setzt an der Axiomatik an, d. h. an Widersprüchen zwischen einigen testtheoretischen Voraussetzungen und empirischen Befunden dazu. Die wichtigsten Voraussetzungen der klassischen Testtheorie lassen sich in folgenden drei Axiomen zusammenfassen (nach HELLER 1973, loc. cit.). 1. *Existenzaxiom*: Zu jedem beobachteten (gemessenen) Wert existiert ein „wahrer“ Wert im Sinne einer bestimmten individuellen Merkmalsausprägung (z. B. Höhe der Schulleistung). Diese wird als Konstante — wenigstens über einen gewissen Zeitraum hinweg — angenommen. 2. *Fehleraxiom*: Der Meßfehler einer Messung ist eine Zufallsvariable. Für diese gilt, daß die Summe bzw. das arithmetische Mittel der Fehlerwerte den Wert Null ergibt. 3. *Verknüpfungsaxiom*: Der beobachtete Wert (Meßwert) setzt sich additiv aus wahren Wert und Fehlerwert zusammen. Daraus kann eine Reihe von Sätzen abgeleitet werden, worauf wir nicht mehr eingehen wollen.

Die auf der Grundlage der klassischen Meßfehlertheorie konstruierten Schulleistungstests lassen vor allem folgende Phänomene ungeklärt: erstens die Beobachtung, daß — entgegen obiger Annahme — die Fehlerwerte nicht unabhängig von den wahren Werten auftreten, d. h. „daß unterschiedliche wahre Werte auch unterschiedliche Fehlerwerte bedingen“ können und diese nicht allein von der (mangelnden) Reliabilität des Tests, sondern auch von der jeweiligen Untersuchungspopulation beeinflusst werden; zweitens die (prinzipielle) Variabilität der wahren Werte, also die Tatsache, daß Merkmalsschwankungen (z. B. Intelligenz- oder Schulleistungszuwachs durch gezielte Fördermaßnahmen) auftreten können; drit-

tens die damit verbundene Schwierigkeit, meßtheoretischen Anforderungen (vor allem der Reliabilität und Validität) *und* berechtigten schulpädagogischen Anliegen (Lernleistungsförderung) gleichermaßen zu genügen. Inwiefern diese Probleme durch die „modernen“ testtheoretischen Ansätze einer Lösung zugeführt werden können, soll im folgenden behandelt werden.

1.3.1.2. Neuere Modellansätze und ihre Problematik

Nach einigen terminologischen Vorklärungen befaßt sich BÜSCHER (s. Kap. 3.2.) mit der durch das RASCH-Modell eingeleiteten neuen Entwicklung einer *psychologischen* Meßtheorie, die vor allem am Problem der Populationsabhängigkeit klassischer Leistungsmessung ansetzt. „Für gewöhnlich werden die Eigenschaften eines psychologischen Tests durch individuelle Unterschiede innerhalb einer bestimmten Population definiert, und die Beurteilung jeder einzelnen Person ist mit der Referenzpopulation verknüpft. In vielen Fällen ist man jedoch daran interessiert, Individuen *per se* zu vergleichen, ohne sich dabei auf eine Population beziehen zu müssen. Die Frage etwa, ob sich ein Kind während eines oder mehrerer Jahre verbessert hat, kann unmöglich mit Hilfe verschiedener Tests beantwortet werden, die an Populationen unterschiedlich alter Kinder geeicht wurden. Im Rahmen eines neuen testtheoretischen Ansatzes versuchte RASCH (1960) derartige Probleme zu lösen, um das Studium einzelner Personen oder einzelner Testitems (Testaufgaben) unabhängig von Referenzpopulationen zu ermöglichen“ (STENE 1968, S. 229).

Am Beispiel *kriterienbezogener* Leistungstests behandelt nun BÜSCHER alle einschlägigen Probleme moderner testtheoretischer Provenienz. Ausführlich werden hierbei neue Verfahrensansätze zur Test- bzw. Itemanalyse erörtert. Deren praktische Relevanz im Hinblick auf die Schülerbeurteilung ist jedoch — vorerst noch — ungeklärt. Symptomatisch hierfür ist die abschließende Bewertung des RASCH-Modells durch BÜSCHER. „Die praktische Anwendung des RASCH-Modells auf lehrzielorientierte Tests ist problematisch, da zum einen eine große Anzahl von Probanden erforderlich ist, zum andern der Konstruktions- und Rechenaufwand enorm ist und schließlich das Modell nicht mehr angewandt werden kann, wenn alle Probanden das Lehrziel erreicht (bzw. nicht erreicht) haben. Das bedeutet, wenn überhaupt, dann ist das RASCH-Modell nur für normbezogene Messung sinnvoll.“

Damit, so könnte es scheinen, sind wir wieder an den Anfang unserer Problemdiskussion verwiesen. Dem ist nicht so. Vielmehr stehen wir mitten in einer Umorientierungsphase, in der neue Denkmodelle entwickelt (vgl. FISCHER 1968, FRICKE 1972) und die klassische Testtheorie relativiert (nicht verworfen) werden. Soviel steht jetzt schon fest, daß kriterien-

bezogene Messungen genauso wie normbezogene „ihre eigene bedeutsame pädagogische Funktion“ haben. „Keineswegs wird das eine Meßverfahren das andere verdrängen.“

1.3.1.3. Vorschläge zur Klassifikation von Schultests

So unterschiedlich wie die testtheoretischen Ansätze ist die in der Literatur verwendete Terminologie im Bereich der Schulleistungsmessung. Nach einer kritischen Analyse in sich oft widersprüchlicher Konzepte unterbreitet ROSEMANN (s. Kap. 3.3.) eigene Vorschläge zu einer *pädagogisch* begründeten Klassifikation von Schultests. Während sich die konventionellen Bezeichnungen vorwiegend an meßtheoretischen Kriterien orientieren, begründet ROSEMANN sein Klassifikationskonzept auf den *Funktionen* schulischer Leistungsbeurteilung. Dabei ist die Unterscheidung von „Leistungsfeststellung“ und „Leistungsbewertung“ von Bedeutung. „Im ersteren Falle verschafft man sich lediglich Information darüber, was die Schüler im Verlaufe des Unterrichtsgeschehens gelernt bzw. nicht gelernt haben. Diese Informationen per se kann der Lehrer in vielfältiger Weise verwenden. Im zweiten Falle geht man einen Schritt weiter, man will die festgestellten Leistungen der Schüler bewerten, wobei verschiedene Bezugspunkte für die Bewertung in Betracht kommen können.“ Für Testverfahren, die im Dienste der erstgenannten Funktionseinheit stehen, schlägt der Autor die Sammelbezeichnung „Lernsteuerungstests“ vor, für Verfahren der zweiten Kategorie die Bezeichnung „Lernkontrolltests“. Ohne Zweifel werden mit dieser Einteilung zentrale Aufgabenfunktionen der Leistungsdiagnostik in der Schule getroffen, nämlich die „Lenkung und Steuerung des Lernprozesses“ und die „Bewertung der Ergebnisse dieses Vorganges“. Während die erste Zielsetzung dem „permanenten Informationsaustausch zwischen Lernenden und Lehrenden“ dient und somit eine optimale Instruktion ermöglichen soll, liegt der entscheidende Vorteil der Lernkontrolltests in der „Objektivierung des Bewertungsvorganges“, also der im Vergleich zur nicht-testgebundenen subjektiven Leistungsbeurteilung — Benotung nach schriftlicher (z. B. Klassenarbeiten) oder mündlicher Prüfung, Aufsatzbeurteilung usw. — größeren Zuverlässigkeit der Urteilsfindung.

Anhand des ROSEMANNSchen Ordnungskonzeptes lassen sich nunmehr die verschiedenen Testformen folgendermaßen zusammenfassen: in *Lernsteuerungstests* (nach bisheriger Benennung die [informellen] kriteriumsbezogenen Schulleistungstests ohne Benotung und — gelegentlich auch — normbezogene Tests für kleinere Lerneinheiten) und *Lernkontrolltests* (die sog. standardisierten Schulleistungstests, informelle normbezogene Tests sowie [informelle] kriteriumsbezogene Tests mit Benotung). ROSEMANN selbst betont

den Charakter seines Ordnungskonzeptes als Diskussionsgrundlage, weshalb wir auch bewußt zunächst darauf verzichteten, die einzelnen Beitragsautoren quasi auf eine einheitliche Terminologie in diesem Band zu verpflichten. Die Vorschläge ROSEMANNS scheinen mir jedoch einer gründlichen Überlegung wert, vor allem im Hinblick auf die praktischen Bedürfnisse der Testanwendung in der Schule.

1.3.2. Objektive Verfahren schulischer Leistungsbeurteilung

1.3.2.1. Lernzieldefinition als Voraussetzung der Leistungsmessung

Anders als beim Programmierten Unterricht (PU), wo man sich von Anfang an vor die Notwendigkeit gestellt sah, *operationalisierte* Lehr-/Lernziele zu formulieren, sind die Probleme einer exakten Unterrichtsplanung im Sinne präziser Zieldefinitionen in der traditionellen Unterrichtslehre verhältnismäßig spät Gegenstand der Diskussion geworden. Virulent wurden solche Problemfragen eigentlich erst in dem Moment, wo das Bemühen um Überprüfung der Unterrichtsziele — analog zum PU — einsetzte, sei es im Rahmen curricularer Innovationen, bei der Erprobung neuer Schulmodelle (Gesamtschule) und damit zusammenhängenden Fragen der Unterrichtsdifferenzierung oder auch, um didaktische und schulpädagogische bzw. therapeutische Fördermaßnahmen auf leistungsdiagnostischer Grundlage zu sichern. Dazu bedarf es in jedem Falle operationalisierter Lernziele, d. h.: Lernleistungsprüfungen sind ohne (vorausgehende) Lernzielbestimmung weder pädagogisch sinnvoll noch meßtechnisch (via Kriteriumstests) durchführbar.

Operationalisierung von Lernzielen meint hier (nach HORN in Kap. 4.1.) die genaue Festlegung der vom Schüler am Ende einer Unterrichtseinheit geforderten Verhaltensweisen (Operationen). „Diese Verhaltensweisen müssen direkt beobachtbar sein. Daher findet man bei Lernzielen häufig Formulierungen wie ‚Der Schüler soll . . . nennen (aufschreiben, lösen usw.) können‘.“

Bei der Festlegung von Lernzielen orientiert man sich heute vielfach an der BLOOMschen Taxonomie, die 6 verschiedene, hierarchisch geordnete, Komplexitätsstufen enthält: Wissen, Verstehen, Anwendung, Analyse, Synthese, Evaluation. Die Kategorien 2 bis 6 betreffen dabei mehr „intellektuelle Fähigkeiten“ und repräsentieren Formen zunehmender Komplexität des kognitiven Verhaltens, die auch als „Strategien des Problemlösens“ (*Problemlösungsstrategien* = relativ inhaltsunabhängige „Grundmuster“) aufgefaßt werden können. Für die praktische Anwendung des BLOOMschen Klassifikationsmodells bei der Unterrichtsplanung gibt HORN wertvolle

Hinweise und erläutert diese an einem Beispiel aus dem Naturkundeunterricht, wobei die Darstellung seines eigenen Lehrplan-Analyseschemas besondere Beachtung verdient. Ausführlich wird dann auf die Konstruktion von Prüfungsaufgaben eingegangen. Auch hierzu bietet HORN eine Reihe praktischer Beispiele. Insgesamt wird somit deutlich, daß die Messung schulischer Lernleistungen vor allen Methodenfragen zunächst von exakten Lernzieldefinitionen abhängt, wobei der Operationalisierung der Lernziele vorrangige Bedeutung zukommt.

1.3.2.2. *Informelle Tests (Lernkontrolltests)*

Hauptanliegen des folgenden Beitrags von ROSEMANN (s. Kap. 4.2.) ist die Information über Lernkontrolltests, aufgezeigt am Beispiel sog. informeller Leistungsmessung. Dabei soll der Leser mit den Konstruktionstechniken und einigen Problemen informeller Tests soweit vertraut gemacht werden, daß er in der Lage ist, „den einen oder anderen Test selbst zu entwickeln bzw. bestehende Tests kritisch zu beurteilen“.

Ein den Ausführungen vorangestelltes Ablaufdiagramm vermittelt einen schnellen Überblick über die einzelnen Arbeitsschritte. Nach der *Operationalisierung der Lernziele* (siehe oben) und der Erstellung einer *Spezifikationstabelle* (in der nicht nur die Anzahl der Testaufgaben oder Items festgelegt, sondern auch eine Entscheidung darüber getroffen wird, welche Lernziele überprüft und somit in den Test aufgenommen werden sollen) erfolgt die eigentliche *Aufgabenkonstruktion*. Hierzu werden vom Autor alle relevanten Item- und Antworttypen besprochen und an Hand einschlägiger Beispiele das Vorgehen veranschaulicht. Ausführlich erörtert ROSEMANN dann die einzelnen Schritte zur Entwicklung der sog. *Testvorform* (Aufgabengruppierung, Erstellung des Testaufgabenheftes einschließlich der Instruktion, Festlegung des Bewertungsschlüssels usw.) sowie die Vorbereitung und Durchführung der *Itemanalyse* (Schwierigkeits- und Trennschärfenanalyse jeder einzelnen Testaufgabe sowie Distraktorenanalyse, Aufgabenselektion bzw. -revision). Schließlich werden die Methoden zur Überprüfung der *Reliabilität* und *Validität* sowie Verfahren zur *Normierung* des Tests dargestellt, wobei eine Reihe von Illustrationsbeispielen wiederum die praktische Arbeit der Testkonstruktion und zugleich das Verständnis testtheoretischer Grundlagen — nunmehr am konkreten Objekt — wesentlich erleichtern dürfte. Der Weg von der Lernzieldefinition bis hin zur *Testendform* ist somit lückenlos beschrieben.

Die Lektüre dieses Beitrags sollte jeden, der sich mit Fragen praktischer Schulleistungsmessung zu beschäftigen hat, verhältnismäßig rasch und umfassend über entsprechende Möglichkeiten (informeller Tests) informieren.

Selbst derjenige Leser, der noch über keine oder nur wenige Kenntnisse und Erfahrungen auf diesem Gebiet verfügt, wird sich hierbei m. E. ohne größere Mühe zurechtfinden.

1.3.2.3. Standardisierte Schulleistungstests

Im nächsten Beitrag (s. Kap. 4.3.) gibt HORN einen Überblick über die wichtigsten zur Zeit verfügbaren sog. standardisierten Schulleistungstests und deren Anwendungsmöglichkeiten im Rahmen der Schülerbeurteilung. Die Unterschiede zwischen *informellen* und *standardisierten* (formellen) Schulleistungstests sind freilich nur gradueller Natur, wie die Ausführungen ROSEMANNs in Kap. 3.3. deutlich gemacht haben. Demnach unterscheiden sich beide Verfahrensansätze vor allem bezüglich der vergleichsweise niedrigeren vs. höheren Reliabilität, des engeren vs. weiteren Anwendungsbereiches, der spezielleren vs. allgemeineren Testinhalte, der Normenfunktion, d. h. verfügbarer Testnormen für bestimmte Klassenstufen (derselben Schulart) vs. für verschiedene Klassen und Schularten sowie hinsichtlich einiger Konstruktionsaspekte, z. B. vorwiegend durch Lehrer vs. Testexperten erstellt (s. S. 158 f.). Die Bezeichnung „standardisierter“ Schulleistungstest bezieht sich also keineswegs, wie oft angenommen wird, ausschließlich auf das Normenkriterium.

Unter *inhaltlichem* Gesichtspunkt lassen sich die standardisierten Schulleistungstests folgendermaßen gruppieren: 1. *fächerübergreifende* Tests oder sog. Omnibusverfahren (z. B. Allgemeiner Schulleistungstest für 2. Klassen AST 2 bzw. für 3. Klassen AST 3 usw.); 2. *fachspezifische* Tests (z. B. Erdkundetest Deutschland ETD 5—7, Geschichtstest Neuzeit GTN 8—10, Naturlehretest NLT 9 usw. oder Diagnostischer Englisch-Leistungstest ELT 6—7, Französischer Wortschatztest FWS 9—12 usw.); 3. *lernbereichsorientierte* Tests, etwa Lesetests (z. B. Lesetest LT 2), Rechtschreibtests (z. B. Diagnostischer Rechtschreibtest DRT 3), Rechentests (z. B. Bruchrechentest BRT 6). Im Hinblick auf den *Anwendungsbereich* bzw. die Zielgruppe oder *Untersuchungspopulation* könnte man schließlich Tests für verschiedene Schulstufen und Schularten unterscheiden: 1. Tests für den *Primar-* vs. *Sekundarstufenbereich* bzw. die einzelnen *Schulklassen*; 2. Tests für die *Schultypen* des Gymnasiums, der Realschule und Hauptschule vs. Grundschule — seltener Gesamtschulen, die fast ausschließlich informelle Tests zur Lernleistungskontrolle verwenden (s. GAUDE u. TESCHNER 1970).

Neben zahlreichen Test- und Aufgabenbeispielen sowie praktischen Hinweisen für den schulischen Einsatz standardisierter Leistungstests bringt HORN am Ende seines Beitrags eine Gesamtübersicht der wichtigsten in der BRD lieferbaren Schultests (Stand 1973). Darüberhinaus kann sich jeder

Leser leicht selbst an Hand der beigelegten Verlagsanschriften durch Anforderung von Prospektmaterial bzw. Gesamtverzeichnissen über die neuesten Testangebote informieren.

1.3.3. Subjektive Verfahren schulischer Leistungsbeurteilung

Zu den „subjektiven“ Verfahren rechnen wir alle nicht-messenden Methoden der Schülerbeurteilung, d. h. Verfahrensweisen, die im Sinne der klassischen Meßfehlertheorie (vgl. Meß- bzw. Testgütekriterien) — erfahrungsgemäß — nicht objektiv sind und einen relativ geringen Grad an Zuverlässigkeit und Gültigkeit aufweisen. Im einzelnen fallen hierunter alle Formen konventioneller Notengebung, mündliche Prüfung, Aufsatzbeurteilung u. ä. Aber auch die in der Schulpraxis bislang viel zu wenig beachteten Methoden der (wissenschaftlichen) Verhaltensbeobachtung und einer Reihe von Beurteilungstechniken i. e. S. (vgl. HELLER et al. 1974, Kap. 2.1.3.) sind der Kategorie der subjektiven Verfahren zuzuordnen. Damit befaßt sich nun der folgende Beitrag von LANGHORST (s. Kap. 5.1.).

1.3.3.1. Verhaltensbeobachtung und Schülerbeurteilung

Über die Notwendigkeit der Verhaltensbeobachtung bzw. einzelner Techniken der Beurteilung i. e. S. im Unterricht sollte es eigentlich keine Diskussion mehr geben. Ohne den Einsatz dieser Methode(n) müßten wir oft auf wichtige Informationen verzichten, z. B. über das Schüler- und Lehrerverhalten im Interaktionsspiel des Unterrichtsablaufs, damit zugleich auf nicht via Tests erfäßbare Daten in bezug auf die Unterrichtsplanung, die Curriculumentwicklung, Lernzielbestimmung usw. STAKE formulierte dies in einem anderen Zusammenhang sehr deutlich, wenn er sagt (zit. nach ROSENSHINE 1973, S. 201): „Ohne Informationen über die Lehrmethoden ist es weder möglich, das Wesen eines Curriculums zu verstehen, noch zu wissen, was als nächstes ausprobiert werden muß. In manchen Evaluationsuntersuchungen werden die mit Hilfe der Unterrichtsbeobachtung gewonnenen Daten die wichtigsten und wertvollsten sein.“

Nach einem kurzen Überblick über gängige Verfahren der Schülerbeobachtung und -beurteilung im Unterricht diskutiert LANGHORST eine Reihe von Möglichkeiten, diese Verfahrensansätze zu systematisieren und somit auf eine objektivere Basis zu stellen. Dabei gilt es zunächst, die *Hauptphasen der Verhaltensbeobachtung*, nämlich die *Beobachtung i. e. S.* (auf den Beobachtungsgegenstand zentrierte, Unwesentliches selegierende Wahrnehmung), die *Beschreibung* oder *Deskription* (Protokollierung des beobachteten Verhaltens) und die eigentliche *Beurteilung* (Interpretation der Beob-

achtungsdaten), strikt auseinanderzuhalten. Nur so kann die Zahl der subjektiven (unkontrollierbaren) Einflüsse auf ein Minimum reduziert und die Aussagekraft der Ergebnisse erhöht werden.

Im zweiten Teil seines Beitrags geht der Autor ausführlicher auf die Möglichkeiten der Schülerbeurteilung mit Hilfe von *Schätzskalen* (rating scales) ein. Unterrichtsrelevante Formen dieses Vorgehens werden an Hand von praktischen Beispielen demonstriert und kritisch auf ihre Verwendungsmöglichkeiten hin untersucht. In diesem Zusammenhang nimmt die Erörterung der häufigsten *Beurteilungsfehler* einen wichtigen Platz ein, entscheidet doch ihre Kenntnis letzten Endes darüber, „ob der Lehrer die Vielzahl der subjektiven Urteilstendenzen . . . beim Umgang mit seinen Schülern in etwa kontrollieren und steuern kann . . . Entsprechende Anstrengungen sind schon deshalb lohnenswert, weil — erwiesenermaßen — im Fall einer gerechten Verhaltensbeobachtung sich der Schüler vom Lehrer besser verstanden fühlt.“

1.3.3.2. Leistungsbeurteilung durch Notengebung

„Schulnoten sind Maßzahlen für erbrachte Leistungen der Schüler, die der Lehrer nach seinen Erfahrungen und Einschätzungen auf der Notenskala einstuft. Noten kommen also aufgrund eines Urteilsprozesses des Lehrers zustande und sind mit all den Mängeln behaftet, die man bei Urteilsprozessen nachgewiesen hat“ (FINGERHUT u. LANGFELDT in Kap. 5.2. dieses Buches). Dabei sind situative Einflüsse genauso beteiligt wie sozial- und persönlichkeitspsychologische Momente (vgl. Beurteilungsfehler). Daraus können wir folgern, daß Schulzensuren prinzipiell schlechte Indikatoren für Schulleistungen sind. Siehe noch INGENKAMP (1971).

Diese Vermutung läßt sich durch zahlreiche empirische Belege stützen. So kommen FINGERHUT und LANGFELDT in dem zitierten Beitrag, der die wichtigsten (neueren) Untersuchungen zu diesem Thema berücksichtigt, zu dem eindeutigen Ergebnis, daß die Lehrerzensuren herkömmlicher Art den meßtheoretischen Forderungen der Objektivität, Reliabilität und Validität nicht oder höchst unvollkommen genügen. „Dies kann den Lehrern jedoch nicht zum Vorwurf gemacht werden. In jeder Unterrichtsstunde wird von ihnen eine Vielzahl schneller und unabhängiger Entscheidungen und Beurteilungen verlangt, die sie nur bewältigen können, wenn sie bestimmte Urteils- und Verhaltensstrategien entwickeln. Diese notwendige Bildung von Stereotypen verhindert aber exakte Urteile.“ Daraus kann m. E. nur die Folgerung abgeleitet werden, dem Lehrer die nötigen Hilfen in Form objektiver und zuverlässiger Methoden (z. B. formelle und informelle Tests) an die Hand zu geben, um auf diese Weise die subjektiven Urteile abzusichern. Nur unter dieser Voraussetzung wäre es — wenn überhaupt —

sinnvoll, Schulnoten die fast alles im Bildungsgang entscheidende Funktion zuzuerkennen.

Die Absicherung subjektiver Lehrerurteile sollte also *mittelbar* durch Tests u. ä. erfolgen und nicht durch unkritische Angleichung an testtheoretische Modelle versucht werden, da ein solches Unterfangen mancherlei pädagogischen und erzieherischen Absichten widersprechen würde. So führt beispielsweise ZIELINSKI (1973, S. 11) nicht weniger als 10 verschiedene Funktionen der Zensurierung auf. Sowohl die Intention der Schule, Lernleistungen zu fördern und damit Schülerleistungen (positiv) zu verändern als auch die Verquickung von pädagogischer Anreizfunktion und reiner Meßfunktion der Schulnoten lassen alle Bemühungen, konsistente (zuverlässige) Leistungsmessungen via Zensuren zu erzielen, praktisch im Ansatz scheitern. Entweder müßte man auf berechnete schulpädagogische Anliegen hier verzichten, um — nicht weniger berechtigten — meßtheoretischen Anforderungen zu genügen, oder man sollte inkompatible Forderungen fallen lassen und schieflieh zwischen pädagogischen und Meßfunktionen bei der Notengebung trennen (was theoretisch möglich ist) bzw. ganz auf Noten verzichten (was vielen praktisch unmöglich erscheint). Zweifellos wäre die letzte Alternative die konsequenteste Haltung. Da wir aber ihre Realisierungschance momentan als gering erachten, bleibt wohl als *vorläufiger* Weg aus dem Dilemma nur die erste Alternative übrig, deren Gefahrenmomente (Vermischung der Funktionen, Beurteilungsfehler usw.) durch Testhilfen soweit als möglich kompensiert werden müssen. Hierzu werden die formellen und die informellen Schulleistungstests einschließlich sog. Standardarbeiten (vgl. WENDELER 1969) gleichermaßen beitragen können.

Prinzipiell wäre damit auch der Weg gewiesen, *einheitlichere Beurteilungsmaßstäbe* zu erzielen. Nach den Ergebnissen von FINGERHUT und LANGFELDT sowie zahlreicher anderer — in diesem Buch zitierter — Forscher werden Schulnoten fast durchweg fachspezifisch erteilt, wobei sich drei Gruppen herausbilden: „Fächer mit milder (musische Fächer und Religion), mit mittlerer (Nebenfächer) und mit strenger Beurteilung (Hauptfächer)“. Die Möglichkeiten der Differenzierung sind außerdem bei milder Beurteilung stark reduziert. Hinzukommen inter- versus intraindividuelle, d. h. lehrerabhängige, Einfluß- bzw. Fehlervariablen (siehe oben). Damit wäre erneut die Notwendigkeit *testunterstützter Notengebung* demonstriert.

Die Möglichkeiten zur objektiven Leistungsbeurteilung sind freilich trotz des wachsenden Testangebots begrenzt, und diese Grenzen können mitunter in der Spezifität des Gegenstandes selbst liegen. Als Paradebeispiel hierfür sei die Aufsatzbeurteilung genannt. Andererseits wird m. E. zu Recht darauf hingewiesen, daß man — aus den verschiedensten Gründen — auf die Übung des Aufsatzes (und seine Bewertung) nicht verzichten sollte. Der Bedeutung des Gegenstandes angemessen beschäftigen sich die letzten zwei

Beiträge dieses Bandes ausführlicher mit der Zensierung von Schüleraufsätzen.

1.3.3.3. *Probleme der Aufsatzbeurteilung*

In drei aufeinander aufbauenden Untersuchungsreihen versuchten NICKEL und WIECZERKOWSKI, „in einer quasi-experimentellen Situation den Einfluß verschiedener vermutlich bedeutsamer Variablen auf die Bewertung von Aufsätzen zu erforschen“. Die methodisch sehr sorgfältig angelegte Studie, in der eine Reihe unabhängiger Beurteilergruppen (Schüler, Studenten der Erziehungswissenschaften und Psychologie sowie Referendare und praxiserfahrene Lehrer) Schüleraufsätze von Viertkläßkern unter weitgehend standardisierten Bedingungen beurteilte, erbrachte — bei systematischer, isolierender Variation einzelner Variablen — folgende Hauptergebnisse (siehe die vollständige Darstellung in Kap. 5.3.).

Keinen Einfluß auf die Aufsatzzensur hat demnach das Geschlecht, und zwar weder des Schülers bzw. Aufsatzschreibers noch des Lehrers bzw. Aufsatzbeurteilers. Ebenso wenig scheinen vorausgehende oder nachfolgende Arbeiten gleicher oder ähnlicher Art die Ausbildung individueller Bezugssysteme auf Seiten des Beurteilers zu beeinflussen.

Demgegenüber konnten die Autoren folgende Einflußgrößen ermitteln: Vor- bzw. Zusatzinformationen über den Schüler, insbesondere über dessen sonstige Schulleistungen; Praxiserfahrung der Beurteiler, sowohl allgemein in bezug auf das Unterrichten als auch speziell im Hinblick auf Aufsatzbewertungen; verschiedene Sprachkriterien des Aufsatzschreibers (Aufsatzlänge, Sprachrichtigkeit vs. Fehlerhaftigkeit, Differenziertheit des Ausdrucksstils, Einfallsreichtum und Originalität, Handlungsganzheit u. ä.). „Eine Analyse des Einflusses verschiedener Sprachkriterien bei annähernder Konstanzhaltung des Faktors ‚Länge‘ und ‚Sprachrichtigkeit‘ ergab, daß die drei Variablen ‚Originalität der Einfälle‘, ‚Differenziertheit des sprachlichen Ausdrucks‘ und ‚Flüssigkeit bzw. Abgeschlossenheit des Handlungsablaufs‘ allein für die Gesamtzensur bereits einen Schätzungseffekt von 52 % besitzen. Wenn also die Richtigkeit und die Länge der Darstellung als Grundlage für eine Beurteilungsdifferenzierung entfallen, läßt sich die Aufsatzzensur in beträchtlichem Umfang bereits allein aufgrund dieser drei Sprachkriterien voraussagen.“

Diese Befunde gewähren einen guten Einblick in den Bedingungskomplex, unter dem subjektive Urteile über Schülerleistungen, aufgezeigt am Beispiel der Aufsatzbeurteilung, zustandekommen. Es ist denkbar, daß diesen Ergebnissen ein größerer Allgemeinheitsgrad zukommt, zumindest was die Beurteilung sprachlicher Leistungen betrifft. Darüberhinaus liefert die Arbeit wichtige Kriterien für eine einheitlichere Beurteilungspraxis — ein Thema, dem der nachfolgende Beitrag gewidmet ist.

Das Ausmaß mangelnder Übereinstimmung in der Benotung von Schüleraufsätzen ist tatsächlich erschreckend, wenn man die von WENDELER (s. Kap. 5.4.) zitierten zahlreichen Literaturquellen als repräsentativ für die heutige Praxissituation ansehen darf (genauer: muß). Nach Meinung des Autors ist das Problem der Aufsatzbeurteilung „keineswegs ein Spezialproblem des Sprachunterrichts, sondern das Problem dieser Prüfungsform als solcher“. Dabei kann man sich den Gesamtvorgang zweiphasig vorstellen. „Der erste Schritt ist die *Leistungsermittlung*, bei der man die Gesamtleistung oder bestimmte Teilleistungen nach einem Bewertungsschlüssel als ‚richtig‘, ‚falsch‘, ‚gut‘ oder ‚schlecht‘ beurteilt, meist mit Punktwerten versieht, und bei der man daraufhin einen Gesamtpunktwert bestimmt.

Der zweite Schritt ist die *Zensierung*, d. h. die Zuordnung einer Zensur zu der errechneten Punktzahl.“ Die Crux der Aufsatzbewertung liegt in der ersten Phase, nämlich in den *Kriterien* der Leistungserfassung und deren *objektiven* Bestimmung. Abweichende Urteile über ein und denselben Aufsatz sind demnach vor allem Folgen unterschiedlicher Kriterienmaßstäbe.

Im Bemühen, einheitliche (verbindliche) Kriterien bzw. Kriteriensysteme für die Aufsatzbeurteilung zu entwickeln, waren besonders einige angelsächsische Versuche erfolgreich. So erstellten DIEDERICH, FRENCH und CARLTON auf induktivem Wege ein Kategoriensystem für die Aufsatzbewertung, das folgende 5 (faktorenanalytierte) *Urteilsdimensionen* enthält: *Ideen*, *innere Form* (Gliederung), *Lebendigkeit* (Originalität), *Sprachrichtigkeit* und *Wortwahl* (Flüssigkeit). Darüber sowie über eine Reihe weiterer Ansätze berichtet WENDELER ausführlich in seinem Beitrag. Die Übereinstimmung zahlreicher dieser Befunde mit den in Kap. 5.3. referierten ist unverkennbar und erhärtet ihre praktische Relevanz.

Im zweiten Teil seines Beitrags geht dann der Autor auf Fragen und Probleme der Objektivitätssteigerung im Kontext Aufsatzbeurteilung ein. Hierzu werden wiederum mehrere Ansätze zur Erfassung der *Urteilsobjektivität* referiert sowie zahlreiche praktische Maßnahmen zu ihrer Verbesserung erläutert. Die mit entsprechenden Beispielen versehenen Ausführungen sollten deutlich machen, „daß tatsächlich ein befriedigender Objektivitätsgrad erreichbar ist“.

1.4. Ausblick

Die Schulleistungsdiagnostik ist Bestandteil einer umfassenderen *pädagogischen* Diagnostik und steht als solche im Dienste unterrichtlicher und erzieherischer Ziele. Sie kann und will nie Selbstzweck sein. Andererseits stellt sie in mannigfacher Weise die notwendigen Informationen für unumgängliche pädagogische Entscheidungen zur Verfügung, die ohne diese Hilfen

nicht oder nur unzulänglich zu fällen wären. Diese Bewertung diagnostischer Funktionen und Möglichkeiten bedeutet keine Technokratisierung des Unterrichts, wie vielfach behauptet wird, sondern auf rationaler Basis begründete Lehr-/Lernprozesse, wofür wir uns allerdings aussprechen, nicht zuletzt unter dem Gesichtspunkt pädagogischer und bildungspolitischer Forderungen wie der der Verwirklichung von Chancengleichheit im Bildungsgang.

Demgegenüber gilt es, auf die Grenzen diagnostischer Möglichkeiten hinzuweisen. Diagnostische Hilfen in Form von Testergebnissen oder anderen Urteilen (z. B. Aussagen via Schätzskaalen) können dem Lehrer die pädagogische Entscheidung selbst nie abnehmen. Diese Einschränkung gilt gleichermaßen im Hinblick auf die praktischen Funktionen und die theoretischen Voraussetzungen pädagogischer Leistungsbeurteilung, die sich ja keineswegs problemlos darstellt. Doch welcher Forschungs- oder Praxisbereich könnte dies heute schon für sich beanspruchen?

Daraus den Schluß zu ziehen, man könne auf die gebotenen — begrenzten — Möglichkeiten diagnostischer Hilfen in der Schule verzichten, wäre m. E. allerdings verhängnisvoll. Ich glaube nicht, daß die Praxis (hier: tagtäglich abgegebener Schülerbeurteilungen) besser sein kann als der jeweilige Erkenntnisstand wissenschaftlicher Forschung hierzu. Beide sind vielmehr in einem gegenseitigen Bedingungsgefüge zu sehen; paradigmatisch verweise ich in diesem Zusammenhang auf die Diskussion um die sog. Selektions-versus Klassifikationsproblematik im Rahmen schulischer Differenzierungsmodelle (vgl. HOPF 1973 sowie bereits HELLER 1970). Hier — wie so oft — sind Theorie und Praxis aufeinander angewiesen, will man sich nicht in praxisferne Modelle oder kurzschlüssige Praktiken verlieren. Dieser Überzeugung entspricht auch die Grundintention dieses Werkes.

Wir möchten unseren kurzen Überblick nicht beschließen, ohne auf zwei Desideratä hingewiesen zu haben. Zum einen wünschte man sich endlich eine umfassende *Theorie der pädagogischen Diagnostik*, die es bislang nur ansatzweise gibt (z. B. GUTHKE 1972). Zum andern dürften die Ausführungen in diesem Band die Notwendigkeit vor Augen führen, *persönlichkeits-* und *sozialpsychologische* Fragestellungen in einer solchen Theorie — stärker als in den vorliegenden Ansätzen — zu berücksichtigen (vgl. Ulich & MERTENS 1973). Das Forschungsfeld und die praktische Bewährung schulischer Leistungsbeurteilung harren weiterer Initiativen.

1.5. Literaturverzeichnis

- Fischer, G. H. (Hrsg.): Psychologische Testtheorie. Huber, Bern 1968.
Fricke, R.: Über Meßmodelle in der Schulleistungsdiagnostik. Schwann, Düsseldorf 1972.
Furck, C. L.: Das pädagogische Problem der Leistung in der Schule. Beltz, Weinheim 1961, 1964 (2. Aufl.).

- Gaude, P. u. Teschner, W. P.*: Objektivierete Leistungsmessung in der Schule. Diesterweg, Frankfurt/M. usw. 1970.
- Guthke, J.*: Zur Diagnostik der intellektuellen Lernfähigkeit. VEB Deutscher Verlag der Wissenschaften, Berlin 1972.
- Heller, K.*: Aktivierung der Bildungsreserven. Huber und Klett, Bern und Stuttgart 1970.
- Heller, K.*: Intelligenzmessung. Neckarverlag, Villingen 1973.
- Heller, K.; Rosemann, B. und Gaedike, A.-K.*: Planung und Auswertung empirischer Untersuchungen. Klett, Stuttgart 1974.
- Hentig, H. von*: Systemzwang und Selbstbestimmung. Klett, Stuttgart 1968, 1970 (3. Aufl.).
- Hofer, M. und Weinert, F. E.* (Hrsg.): Pädagogische Psychologie. Grundlagen-texte 2 zum Funk-Kolleg Päd. Psych. Fischer Taschenbuch, Frankfurt/M. 1973.
- Hopf, D.*: Möglichkeiten und Grenzen der Anwendung von Tests. In: *Hofer, M. und Weinert, F. E.* (Hrsg.): Pädagogische Psychologie. Frankfurt/M. 1973.
- Ingenkamp, K.* (Hrsg.): Die Fragwürdigkeit der Zensurengebung. Beltz, Weinheim usw. 1971, 1972 (3. Aufl.).
- Ingenkamp, K.* (Hrsg.): Tests in der Schulpraxis. Beltz, Weinheim usw. 1971, 1973 (3. Aufl.).
- Michel, L.*: Allgemeine Grundlagen psychometrischer Tests. In: Hb. d. Psychol., Bd. 6: Psychol. Diagnostik (Hrsg. *Heiß, R.*). Hogrefe Göttingen 1964.
- Neander, J.*: Objektivierete Lernerfolgsmessung in der Gesamtschule — Fortschritt für wen? Die Deutsche Schule, 65 (1973), 35—47.
- Rosenshine, B.*: Die Beobachtung des Unterrichts in der Klasse. In: *Hofer, M. und Weinert, F. E.* (Hrsg.): Pädagogische Psychologie. Frankfurt/M. 1973.
- Stene, J.*: Einführung in RASCHS Theorie der psychologischen Messung. In: *Fischer, G.* (Hrsg.): Psychologische Testtheorie. Bern 1968.
- Symposium des Europarats in Berlin vom 11.—19. 11. 1971. Unveröffentl. Bericht der Arbeitsgruppe 2 (Tutor: Dr. *Weiß, R.*, Protokollant: Dipl.-Psych. *Büscher, P.*).
- Ulich, D. und Mertens, W.*: Urteile über Schüler. Beltz, Weinheim usw. 1973.
- Wendeler, J.*: Standardarbeiten — Verfahren zur Objektivierung der Notengebung. Beltz, Weinheim usw. 1969, 1973 (5. Aufl.).
- Zielinski, W.*: Die Beurteilung von Schülerleistungen. In: Funkkolleg „Pädagogische Psychologie“ (Studienbegleitbrief 12). Beltz, Weinheim usw. 1973.

6. Autorenverzeichnis

Peter Büscher, Jahrgang 1936, studierte Psychologie und Erziehungswissenschaft an den Universitäten Heidelberg und Mannheim (Diplom-Prüfung 1970). Der Autor ist zur Zeit als Sonderschullehrer an der Gehörlosenschule Heidelberg tätig und war von 1971—73 Leiter der Bildungsberatungsstelle Mannheim. Außerdem ist er Lehrbeauftragter für pädagogische Diagnostik an der Universität Mannheim (Lehrstuhl Prof. Dr. L. Michel).

Arbeitsschwerpunkt: Psychologische Diagnostik im Bildungswesen.

Wichtigste Veröffentlichungen: Probleme der Differenzierung. Hörgeschädigtenpädagogik, 25 (1971), 1—12; Mannheimer Intelligenztest (zus. mit W. Conrad u. a.). Weinheim (Beltz) 1971; Operationalisierte Lernziele und pädagogische Taxonomien. Hörgeschädigtenpädagogik, 27 (1973), 31—42.

Anschrift: 6806 Viernheim, Eichenstraße 14.

Walter Fingerhut, Jahrgang 1946, studierte Psychologie an der Universität Marburg (Diplom-Prüfung 1971). Der Autor ist zur Zeit Wissenschaftlicher Angestellter im Fachbereich Psychologie der Universität Marburg.

Arbeitsschwerpunkt: Pädagogische Psychologie.

Veröffentlichungen: Der Beitrag biographischer Daten von Schülern und Lehrern zur Vorhersage von Schulnoten (zus. mit H.-P. Langfeldt). Vortrag auf dem 28. Kongr. Dt. Ges. Psychol. 1972 in Saarbrücken, Göttingen (Hogrefe), Kongreßbericht im Druck; Erfahrungen mit dem Allgemeinen Schulleistungstest AST 4 (zus. mit H.-P. Langfeldt). Psychol. in Erz. u. Unterr., 20 (1973), 249—257; Entwicklung eines Lehrer-Einstellungs-Fragebogens LEF-3 (zus. mit H.-P. Langfeldt). Z. f. Entwicklungspsychol. u. Päd. Psychol., 6 (1974), im Druck.

Anschrift: 3551 Cappel, Goethestraße 40.

Anne-Katrin Gaedike, Jahrgang 1949, studierte Psychologie an den Universitäten Braunschweig und Hamburg (Diplom-Prüfung 1972) sowie Pädagogik an der Pädagogischen Hochschule Rheinland, Abteilung Bonn. Die Autorin ist zur Zeit Wissenschaftliche Assistentin am Psychologischen Seminar der Abteilung Bonn der Päd. Hochschule Rheinland.

Arbeitsschwerpunkte: Experimentelle und Angewandte Psychologie, Psychologische Diagnostik und Methodenlehre.

Veröffentlichung: Planung und Auswertung empirischer Untersuchungen — Einführung für Pädagogen, Psychologen und Soziologen (zus. mit K. Heller und B. Rosemann). Stuttgart (Klett) 1974.

Anschrift: 53 Bonn, Rittershausstraße 23.

Kurt Heller, Jahrgang 1931, studierte Psychologie und Erziehungswissenschaft an den Universitäten Freiburg/Br. und Heidelberg (Diplom-Prüfung 1964, Promotion 1968). Der Autor ist ordentl. Professor für Psychologie an der Pädagogischen Hochschule Rheinland, Abteilung Bonn, und Lehrbeauftragter an der Universität Heidelberg im Fachbereich Erziehungswissenschaft.

Arbeitsschwerpunkte: Begabungs- und Bildungsforschung, Psychologische Diagnostik und Beratung im Bildungswesen, Sonderpädagogische Psychologie.

Wichtigste Veröffentlichungen: Modell eines Guidance-Systems für Abiturienten und Studenten (zus. mit E. Demel u. G. Schorre). In: Bildung in neuer Sicht, Bd. 20, Villingen (Neckarverlag) 1969; Aktivierung der Bildungsreserven. Bern u. Stuttgart (Huber/Klett) 1970; Heidelberger Hörprüf-Bild-Test HHBT für Schulanfänger (zus. mit A. Löwe). Villingen (Neckarverlag) 1972; Intelligenzmessung — Zur Theorie und Praxis der Begabungsdiagnostik in Schule und Son-

derpädagogik. Villingen (Neckarverlag) 1973; Wortschatztests für Schbehinderte WST(Sb) 4—9 (zus. mit B. Schirmer). Weinheim (Beltz) 1973; Wortschatztests für Blinde WST(Bl) 4—8 (zus. mit B. Köhn). Weinheim (Beltz) 1973; Planung und Auswertung empirischer Untersuchungen (zus. mit B. Rosemann u. A.-K. Gaedike). Stuttgart (Klett) 1974; Kognitiver Fähigkeits-Test KFT 4—13 (zus. mit A.-K. Gaedike u. H. Weinläder). Weinheim (Beltz) 1974. Mitherausgeber der „Schriftenreihe zur Bildung und Rehabilitation Sehgeschädigter“ (Schindele) und „Sehgeschädigte“ — Intern. Wiss. Arch. (Schindele).

Anschrift: 6903 Neckargemünd, In den Wingert 6.

Ralf Horn, Jahrgang 1942, studierte Psychologie an der Universität Freiburg/Schweiz und Frankfurt/Main (Diplom-Prüfung 1967). Der Autor ist wiss. Leiter der Beltz Test Gesellschaft in Weinheim/Bergstr. und Lehrbeauftragter an der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz, Abteilung Worms.

Arbeitsschwerpunkte: Testentwicklung, Forschungsmethoden.

Wichtigste Veröffentlichungen: Lernziele und Schülerleistung. Weinheim (Beltz) 1972, 1973 (3. Aufl.); Lernziel-Test Mathematik 5. Schuljahr (zus. mit B. Andelfinger). Weinheim (Beltz) 1973; Sozialschicht und Intelligenzleistung. Psychol. in Erz. u. Unterr., 21 (1974).

Anschrift: 684 Lampertheim, Daimlerstraße 49.

Hans-Peter Langfeldt, Jahrgang 1943, studierte Psychologie und Erziehungswissenschaft an den Universitäten Tübingen und Marburg (Diplom-Prüfung 1971). Der Autor ist zur Zeit Wissenschaftlicher Assistent an der Pädagogischen Hochschule Heidelberg, Fachbereich Sonderpädagogik.

Arbeitsschwerpunkte: Pädagogische Psychologie und Diagnostik.

Veröffentlichungen: Der Beitrag biographischer Daten von Schülern und Lehrern zur Vorhersage von Schulnoten (zus. mit W. Fingerhut). Vortrag auf dem 28. Kongreß Dt. Ges. Psychol. 1972 in Saarbrücken, Göttingen (Hogrefe), Kongreßbericht im Druck; Erfahrungen mit dem Allgemeinen Schulleistungstest AST 4 (zus. mit W. Fingerhut). Psychol. in Erz. u. Unterr., 20 (1973), 249—257; Entwicklung eines Lehrer-Einstellungs-Fragebogens LEF-3 (zus. mit W. Fingerhut). Z. f. Entwicklungspsychol. u. Päd. Psychol., 6 (1974); Entscheidungstheoretische Aspekte der Umschulungsdiagnostik in die Sonderschule für Lernbehinderte. Z. f. Heilpäd., 25 (1974).

Anschrift: 6903 Neckargemünd, Waldstraße 14.

Erich Langhorst, Jahrgang 1925, studierte nach der 1. und 2. Lehrerprüfung Psychologie an der Universität Bonn (Diplom-Prüfung 1962, Promotion 1966). Der Autor ist Akademischer Oberrat für Psychologie an der Pädagogischen Hochschule Rheinland, Abteilung Bonn.

Arbeitsschwerpunkte: Entwicklung der kognitiven Funktionen und allgemeinen Leistungsfähigkeit, Psychologie des Lese- und Rechtschreibeunterrichts, Lese-Rechtschreib-Schwäche.

Wichtigste Veröffentlichungen: Märchenbilder im Urteil von Kindern der Vorkriegszeit und Gegenwart — Ein Beitrag zur Psychologie des Bilderlebens der Sechs- bis Vierzehnjährigen. Bonn (Bouvier) 1967; Brennpunkte der pädagogischen Psychologie (Hrsg. zus. mit H. Nickel). Bern u. Stuttgart (Huber/Klett) 1973.

Anschrift: 53 Bonn, Luisenstraße 2.

Horst Nickel, Jahrgang 1929, studierte Psychologie und Erziehungswissenschaft an den Universitäten Marburg (Diplom-Prüfung 1961) und Erlangen-Nürnberg (Promotion 1965). Der Autor ist ordentl. Professor für Psychologie (Entwicklungs-

und Erziehungspsychologie) an der Universität Düsseldorf und Lehrbeauftragter an der Päd. Hochschule Rheinland, Abteilung Bonn.

Arbeitsschwerpunkte: Kognitive Entwicklung und Entwicklungsförderung, Erzieherverhalten, Unterrichtspsychologie.

Wichtigste Veröffentlichungen: Die visuelle Wahrnehmung im Kindergarten- und Einschulungsalter. Bern und Stuttgart (Huber/Klett) 1967; Entwicklungspsychologie des Kindes- und Jugendalters. Bern (Huber), Band 1: 1973 (2. Aufl.), Band 2: 1974; Brennpunkte der pädagogischen Psychologie (Hrsg. zus. mit E. Langhorst). Bern und Stuttgart (Huber/Klett) 1973; Beiträge zur Psychologie des Lehrerverhaltens. München (Reinhardt) 1974; Angstfragebogen für Schüler AFS (zus. mit W. Wiczercowski u. a.). Braunschweig (Westermann) 1974. Mitherausgeber der Zeitschrift „Psychologie in Erziehung und Unterricht“ (Reinhardt).

Anschrift: 5309 Meckenheim, Berliner Straße 25.

Bernhard Rosemann, Jahrgang 1940, studierte Psychologie und Soziologie an den Universitäten Heidelberg (Diplom-Prüfung 1969) und Mainz (Promotion 1972). Der Autor ist Wissenschaftlicher Assistent am Psychologischen Seminar der Abteilung Bonn der Päd. Hochschule Rheinland.

Arbeitsschwerpunkte: Vorgesetzte und Mitarbeiter — Rollenerwartungen und interpersonales Verhalten. Diss. Mainz 1972; Führungsverhalten und Rollenerwartung in informellen Gruppen. Psychol. u. Praxis, 17 (1973), H. 4; Planung und Auswertung empirischer Untersuchungen — Einführung für Pädagogen, Psychologen und Soziologen (zus. mit K. Heller u. A.-K. Gaedike). Stuttgart (Klett) 1974.

Anschrift: 53 Bonn, Ubierweg 3.

Jürgen Wendeler, Jahrgang 1937, studierte Psychologie an den Universitäten Berlin (FU) und Frankfurt/Main (Diplom-Prüfung 1962) sowie an der Techn. Hochschule Darmstadt (Promotion 1972). Der Autor ist seit 1965 Wissenschaftlicher Angestellter am Deutschen Institut für Internationale Pädagogische Forschung (Abteilung Pädagogische Psychologie) in Frankfurt/Main.

Arbeitsschwerpunkte: Konstruktion und Anwendung von psychologischen Tests und Schulleistungstests.

Wichtigste Veröffentlichungen: Elementare Lernprinzipien. Weinheim (Beltz) 1967; Standardarbeiten — Verfahren zur Objektivierung der Notengebung. Weinheim (Beltz) 1969, 1973 (5. Aufl.); Intelligenztests in Schulen. Weinheim (Beltz) 1970, 1971 (2. Aufl.); Stilanalyse als Testkonstruktion. Weinheim (Beltz) 1973; Frankfurter Denkaufgaben für 3. bis 6. Klassen. Weinheim (Beltz) 1973; Satzlehre 6+ (Stufentest). Weinheim (Beltz) 1973; Wortlehre 6+ (Stufentest). Weinheim (Beltz) 1974.

Anschrift: 6079 Sprendlingen, August-Bebel-Straße 22.

Wilhelm Wiczercowski, Jahrgang 1927, studierte Psychologie an den Universitäten Hamburg und Åbo/Finnland (Lizentiat 1960, Promotion 1963). Der Autor ist ordentl. Professor für Psychologie an der Universität Hamburg.

Arbeitsschwerpunkte: Probleme der Zweisprachigkeit und der Lehrer-Schüler-Interaktion, Lernpsychologie.

Wichtigste Veröffentlichungen: Bilinguismus im frühen Schulalter. Diss. Helsinki 1963; Frühe Zweisprachigkeit. München (Reinhardt) 1965; Erwerb einer zweiten Sprache im Grundschulalter. Hannover (Schroedel) 1972; Lernpsychologische Grundlagen des programmierten Unterrichts. In: H. Nickel u. E. Langhorst (Hrsg.), Brennpunkte der pädagogischen Psychologie, Bern u. Stuttgart (Huber/Klett) 1973.

Anschrift: 2 Hamburg 1 (Norderstedt), Moorkamp 15 f.

7. Personenregister

- Ahrens, G. 271, 301
Allport, G. W. 65
Amelang, M. 59, 79, 125, 128, 264
Anastasi, A. 96
Argyle, M. 254
Arnold, W. 55, 63
As, D. 77
Aschersleben, K. 120, 259
Atkinson, J. W. 64, 66
- Baldwin, T. S. 161
Bales, R. F. 239
Baker, C. T. 62
Bandura, A. 73
Bäumler, G. 66
Bastine, R. 77
Baur, Rita 83 f.
Beech, H. R. 73
Beer, F. 48 ff., 81
Bellak, L. 273 f., 294
Berlyne, D. E. 65
Bernstein, B. 75, 85
Biglmaier, F. 159
Bittmann, F. 236
Bleidick, M. 242
Block, J. H. 151, 156
Bloom, B. S. 171, 179 f., 184
Blöschl, Lilian 49
Bobertag, O. 39
Bormuth, J. R. 154
Bottenberg, E. H. 54, 58
Brandstätter, H. 66
Breitenbach, W. 66
Brickenkamp, R. 77
Brophy, J. E. 262
Buggle, F. 67
Burger, R. 48, 52, 262
Burt, C. 47, 55
Büscher, P. 137 ff., 161
Butterfield, E. C. 62
- Carlton, S. T. 314
Carmical, L. 58
Carter, R. S. 258
Cattell, R. B. 44, 66 f., 69
Claus, G. 96
Clostermann, J. 48
Cogan, M. L. 44
- Coleman, J. C. 58
Cowen, E. L. 66
Cox, F. N. 66, 145 f., 152, 155
Cranach, M. 235, 246
Cronbach, L. 53, 96, 106, 134, 152, 245 f.
Cunningham, Shirley 266
- D'Andrade, R. 64
Dann, H. D. 256
Denig, F. 15, 39
Diederich, P. B. 271, 305 f., 314
Dohse, W. 261
Donat, H. 238
- Ebel, R. L. 155, 160, 189, 191, 198
Ebner, H. 96
Elias, R. M. 59, 83
Endrigkeit, F. 118
Engels, F. 64
Entwistle, N. J. 266
Erlemeier, N. 250, 272, 305
Ewert, O. 48
Eysenck, H. J. 70, 73
Eysenck, Sybil 71
- Falk, R. 271
Feather, H. 64
Feldhusen, J. F. 59, 66, 83
Fenner, H. J. 66, 68, 77, 248
Ferdinand, W. 256, 258, 265, 267
Festinger, L. 65
Fingerhut, W. 15, 38 ff., 46, 79, 117 f., 125, 132, 253 ff., 264
Finlayson, D. S. 320
Fippinger, F. 38, 42, 47, 49, 125, 130, 258, 263 f.
Fischer, G. H. 26, 97, 129 f., 131
Flitner, A. 259, 261
Foppa, K. 73
Franke, H. 66 f.
Frankel, E. A. 58
French, J. W. 77, 314
Frenz, H.-G. 235, 246
Fricke, R. 25, 131, 142, 144, 148 f., 153, 155
Fröhlich, W. 46
Funke, G. H. 15, 40, 256, 265
Furck, C. L. 14

- Gaedike, Anne-Katrin 46 ff.
 Gagné, R. M. 171
 Gärtner-Harnach, Viola 66, 69
 Garvin, A. D. 139 f.
 Gaude, P. 14, 29, 158, 191, 198, 219
 Gebauer, T. 52, 262
 Gerlicher, K. 67
 Getzels, J. W. 58 f., 83
 Giesen, H. 264
 Glaser, R. 138, 161
 Gleser, G. 53, 134
 Glidewell, J. L. 83
 Goodwin, W. 73
 Göller, A. 256
 Good, T. L. 262
 Gosling, G. W. H. 306, 317, 319
 Gottschaldt, K. 236
 Graham, G. T. 152, 155
 Grau, U. 271 f., 298, 300, 302, 310
 Graumann, C. F. 235 f.
 Gronlund, N. E. 184, 186, 188, 191, 194, 197 ff.
 Guilford, J. P. 47, 52, 171, 248 f.
 Gulliksen, H. 96, 158
 Guttman, L. 152, 154

 Haase, M. 222
 Hadley, T. S. 258
 Hahn, H. 314
 Hambleton, R. K. 151
 Harris, C. W. 151
 Hartog, P. 311 f., 319
 Haseloff, O. W. 83, 96
 Hasemann, K. 235, 246, 248 f., 253
 Hastings, G. R. 174
 Heckhausen, H. 60, 63 ff., 72 ff., 75
 Heller, K. 13 ff., 21, 23 f., 30, 35, 50, 63, 96, 220, 235, 264
 Hentig, H. v. 13 f.
 Herbig, M. 155
 Hess, R. D. 74
 Hielscher, H. 85
 Hill, K. T. 66 f.
 Hitpass, J. 52, 262, 264
 Hively, W. 153
 Höger, D. 39 f., 48, 256, 259, 263
 Höhn, Elfriede 249, 253
 Hofer, M. 248, 253
 Hoffmann, M. 96, 250
 Hofstätter, P. R. 297
 Hopf, D. 35

 Hopp, A. D. 39, 55, 254 f.
 Hoppe, F. 61
 Horn, R. 169 ff., 173 ff., 177, 179, 222 ff.
 Horn, W. 42
 Horst, P. 96
 Hudelson, E. 318

 Ingenkamp, K.-H. 31, 38, 131, 134, 158, 161, 169, 192, 222, 257 f., 261, 271, 310, 312, 321
 Iseler, A. 317
 Israel, J. 77

 Jackson, W. 59, 83, 153
 Jacobson, Leonore 262, 305
 Janssen, J. R. 52, 262
 Jantzen, W. 41
 Jörg, H. 271 f., 298

 Kagan, J. 62, 75
 Kahl, J. A. 84
 Kalveram, K. T. 38
 Karnick, R. 271
 Keil, W. 118
 Kemmler, Lilly 41, 50, 55, 57 ff., 79, 81 ff., 248, 265
 Kießling, A. 309
 Kiwitz, H. 256, 267
 Klauer, K. 137, 155, 169
 Klausmeier, H. J. 66, 73
 Kleiter, E. 254
 Kniel, A. 84
 Knoche, W. 256
 Kohl, G. 264
 Kornmann, R. 53, 116, 118
 Koschat, H. 71, 81
 Kötter, L. 271 f., 298, 300, 302, 310
 Krapp, A. 53, 116
 Kropp, R. P. 172 f.
 Kruse, Lenelis 64
 Kühn, R. 79, 125, 128
 Kutalek, L. 49 f., 81

 Landsheere, de, G. 271
 Langfeldt, H. P. 15, 38 ff., 79, 96 ff., 117 f., 125, 132, 253 ff., 261, 264
 Langheinrich, D. 57
 Langhorst, E. 79, 233 ff.
 Latscha, F. 79
 Laurien, H.-R. 271

- Lehmann, E. 310, 312
 Lienert, G. A. 39, 53, 55, 57, 96, 104 f.,
 107, 111 f., 116, 129, 134, 158, 160,
 162, 210, 215, 219 f., 254 f., 277, 297,
 316
 Lietzmann, W. 309
 Lippit, R. 77
 Livingston, S. A. 141
 Löschenkohl, E. 47, 263
 Löser, G. 16, 44, 69, 263, 266
 Lückert, H.-R. 59, 63
- Madaus, H. 58
 Maehr, M. L. 262
 Mager, R. F. 169, 174
 Magnusson, D. 52, 96
 Mai, N. 134
 Mandl, H. 53, 66, 116
 McClelland, D. C. 60, 62, 66, 83
 McKeachie, W. J. 77
 Meehl, P. E. 106
 Meili, R. 48
 Merkens, H. 151
 Mertens, W. 35, 234, 246, 253
 Merz, F. 254
 Metzelder, L. 69
 Meyer, W. V. 61
 Michel, L. 122, 134
 Miller, G. W. 265
 Mischel, W. 73
 Mittenecker, E. 96
 Mitzlaff, S. 84
 Mollenhauer, K. 83, 85
 Möller, C. 170
 Moss, H. A. 62
 Müller-Fohrbrodt, G. 256
 Mussen, P. H. 304
- Neander, J. 21
 Nelson, V. L. 62
 Newell, R. 58
 Nickel, H. 66, 68, 77, 79, 236, 266,
 271 ff., 304
 Niesbet, J. D. 320
 Novick, H. R. 151
 Nunnally, J. C. 189, 191, 193, 198, 205,
 209
- Oevermann, U. 83 ff.
 Orlik, P. 255, 262
 Osburn, H. G. 154
- Page, E. 153
 Patterson, R. 153
 Peel, E. A. 53
 Petrat, G. 264 f.
 Pidgeon, D. A. 53
 Popham, W. J. 142, 146, 155
 Porter, R. B. 44, 67, 69
- Rachman, S. 70, 73
 Rank, T. 71, 81 f., 259
 Rasch, G. 25, 149
 Rasof, B. 58
 Ravitch, M. M. 151
 Rhodes, E. C. 311 f.
 Rogge, K.-E. 64
 Röhm, H. 250, 253
 Roloff, F.-A. 57
 Rosemann, B. 158 ff., 182 ff.
 Rosen, B. C. 64, 83
 Rosenfeld, G. 62, 64
 Rosenshine, B. 30
 Rosenstil, L. V. 66
 Rosenthal, R. 75, 262, 305
 Rösler, H.-D. 81
 Ross, H. 73
 Roth, H. 240
 Rubovitz, P. C. 262
 Ruch, G. M. 318
 Russel, J. T. 53
 Ryans, D. G. 76, 244
- Sader, M. 118, 266
 Sander, A. 118
 Sarason, S. B. 66 ff.
 Seitz, W. 16, 44, 67, 69, 263, 266
 Shavelson, R. J. 141
 Sherif, M. B. 77
 Shipman, V. 74
 Shoemaker, D. M. 154
 Siegel, S. 277
 Simons, H. 57, 263, 265
 Sixtl, F. 100, 254
 Skolnik, A. 62
 Skurnik, L. S. 171
 Sontag, L. W. 62
 Specht, Heike 266
 Spitzmüller, Ruth 83, 85
 Suermann, Irmgard 52
 Süllwold, F. 114 f., 261

- Scheier, J. H. 66
 Schell, H. 66
 Schirmer, Brigitte 50
 Schlesinger, I. M. 154
 Schlüter, P. 66, 68, 77, 266
 Schmitz, G. F. 48, 263 f.
 Schnell, H. 49 f., 81
 Schröter, G. 271, 301, 310, 321
 Schulze, W. 259, 262, 264
 Schwarz, E. 263
- Steinhauser, F. 47
 Steinkamp, G. 79, 262
 Stelzl, W. 47, 131
 Stene, J. 25
 Stoddard, G. D. 318
 Stoltz, L. M. 83
 Stoker, H. W. 172 f.
- Tausch, R. 73, 76 f.
 Taylor, H. C. 53, 58
 Tent, L. 47, 49, 235, 244, 254, 259,
 264, 266
 Teschner, W. P. 14, 29, 158, 191, 198,
 219
 Tewes, U. 69, 71, 265 f.
 Thiersch, H. 77, 79 f.
 Thomae, H. 233, 235, 245 f.
 Thyen, H. 39
 Tismer, K.-G. 250, 272, 305
 Todt, E. 50, 263, 266
 Treffinger, D. J. 55, 83
 Tyler, R. W. 199
- Überla, K. 38
 Uhr, R. 265
 Ulich, D. 35, 234, 246, 253
 Ulshöfer, R. 271, 309, 313, 321
 Undeutsch, U. 52
- Vagt, G. 59, 83, 264
 Vargas, J. S. 145 f.
 Vontobel, L. 62, 65
- Walker, Helen 96
 Wang, M. C. 138, 162
 Wasna, Maria 61 f., 236
 Weber, A. 49, 271, 306, 314
 Wechsler, D. 115
 Weingardt, E. 256, 259
 Weinert, F. 68, 71, 93
 Weis, V. 15, 39, 164
 Weiss, R. 47, 49 f., 56, 58, 255, 258 f.,
 261, 263, 302
 Wendeler, J. 32, 79, 149, 164, 309 ff.
 Wendt, D. 297
 Wenzl, A. 55
 Wewetzer, K. H. 55, 59
 White, R. H. 77
 Wicznerkowski, W. 79, 271 ff., 310
 Willing, M. H. 318
 Winer, W. J. 287, 290
 Winterbottom, M. 61, 64
 Wiseman, S. 319
 Wolpe, J. 73
 Wood, R. 171
 Wright, B. D. 149
- Yates, A. 53
- Zielinski, W. 18, 32, 66, 68, 82, 260,
 265
 Zigler, E. 62
 Ziler, H. 48
 Zimbardo, P. G. 67
 Zimmermann, K. W. 40

8. Sachregister

- Alternativantwort 187 ff., 198
— freie 198
— eingeschränkte 198
Antworten 123 f., 194
— freie 123
— gebundene 123, 187 ff.
— nicht gebundene 196
Ähnlichkeitsfehler → Beurteilungsfehler
Angst 66 ff.
— Auswirkungen auf Leistung 66
Ängstlichkeit → Angst
— und Intelligenz 67
— und Schulleistung 67 f.
Äquivalenzkoeffizient 212
Äquivalenznorm 122
Anreizwert (von Aufgaben) 64 f.
Arbeitshaltungen 63
Aufgaben/analyse → Iteanalyse
— -bewertung 205
— -gültigkeit 144
— -homogenität 147
— -reihung 205
— -schwierigkeit 124, 130, 142 f., 149
— -trennschärfe 143 f.
Aufsatzbeurteilung 3 f., 271 ff., 309 ff.,
Aufsatzskala 318
Aufstiegsverhalten → Wertvorstellungen 82 ff.
Auswertungsobjektivität → Objektivität
- Beobachtung 233 ff.
— fraktionierte 236
— Gelegenheits- 235, 246
— in standardisierten Situationen 236
— kontrollierte 236
— systematische 235 f.
Beobachtungs/deskription 237
— -pläne 235
— -Skalen 240
Beschreibung 235, 237 ff.
— standardisierte 239
best answer type → Mehrfachwahlitem
Bewertungsprozeß 163
Beurteilung 233 ff., 241, 253 f.
— Kategorienschema zur 234
— Urteils- und Verhaltensstrategien zur 254
Beurteilungsfehler 246 ff., 253
- Chi-Quadrat-Analyse 146 f.
class-room-test → informelle Tests 158 f.
concurrent-validity → Validität,
kriteriumsbezogene
confidential limits → Vertrauensgrenze
content validity → Validität,
inhaltsbezogene
correct-answer-type → Mehrfachwahlitem
criterion-level → Lernkriterium
curriculare Validität → Validität
cut-off-point → Leistungsstandard
- Denken
— divergentes → Kreativität 59
— konvergentes 59
direktives Verhalten (und Schulleistung) 78 f.
Distraktoren 190
Durchführungsobjektivität → Objektivität
- Eichung → Normierung
Eichstichprobe 119
Empirisches Relativ → Relativ
Ergänzungsantworten 196 f.
error of leniency → Beurteilungsfehler
Erwartungseffekt → Beurteilungsfehler
Essay-test → Kurzaufsatzantwort
extended-response-question → Antwort,
freie
Extraversion → Schulleistung
- facet-design-method 154
Faktorenanalyse von Zeugnissen 38 ff.
Fehler, konstanter 103
— systematischer 103
Fehler/theorie → Testtheorie
— -streuung → Standardmeßfehler
— -wert 24, 103 f., 112, 129
fill-in-answer → Ergänzungsantwort
Furcht vor Mißerfolg 61 f., 66
- Generalisierungstheorie (Cronbach)
152 f.
generosity-error → Beurteilungsfehler
Grundmuster (kognitive Stile) 27, 171 f.

- Grundtypen der Problemsituation (BLOOM) 179 f.
 Gültigkeit → Validität
 Gütekriterien 22 f., 101, 108 f.
- Halbierungs-Reliabilität 111
 Halo-Effekt → Beurteilungsfehler
 Hof-Effekt → Halo-Effekt
 Hoffnung auf Erfolg 61 f., 66
 Homogener Antwortvektor → Übereinstimmungskoeffizient
 Homogenität (von Testitems) 154 f.
 Homogenitätsprüfung 152
- Identifikation → Imitationslernen
 Identität → Nominalskala
 Imitationslernen 73
 Implizite Persönlichkeitstheorie → Beurteilungsfehler
 Index of agreement 153
 Informelle Tests 28 f., 96, 158 ff., 182 ff.
 Intelligenzfaktoren → Schulleistung
 Interpretationsobjektivität → Objektivität
 Interpretationsübung 199 f.
 interpretive exercises → Interpretationsübung
 Intervallskala 99, 102, 122, 130
 Item/analyse 123, 142, 146, 205 f.
 — -charakteristikkurve 150
 — -homogenität → Aufgabenhomogenität
 — -kennwerte 142
 — -konstruktion 186
 — -muster 154
 — -revision 209 ff.
 — -selektion 146, 209 ff.
 — -typen 186
- Kategoriensystem von BALES 239 f.
 Klassenbildung → Skalentransformation
 Koeffizient, LIVINGSTONEScher 152
 Kognitive Stile 73 f.
 Konsistenz/analyse 212
 — interne 153, 212
 — -reliabilität 111 f., 133
 Konstrukt 106 f.
 — -validität → Validität
 Korrelation 96
 Korrelationskoeffizient 105
 Kontrastfehler → Beurteilungsfehler
 Kreativität 58 f.
- Kriteriums/begriff 139 f., 160 f.
 — -bezogene Validität 106
 — -leistung 133
 Kurzantworten 196 ff.
 Kurzaufsatzantwort 198
- Lehrziele 170
 Lehrplangültigkeit → Validität, curriculare
 Lehrerverhalten → Schulleistung
 Lenkung → direktives Verhalten
 Leistungs/angst → Angst
 — -beurteilung 18 ff., 26 f., 30 ff., 79, 165 f., 184 f.
 — -bewertung 163 ff.
 — -ermittlung 312 f.
 — -feststellung 163 ff., 185
 — -messung 27, 167 ff.
 — -niveau 160
 — -normen 63
 — -standard 140, 160, 162
 — -verhalten und Wertvorstellungen 82 ff.
- Leistungsmotivation, äußere Anreizvariable der 64
 — Definition der 60
 — extrinsische (sachfremde) 72
 — Genese der 61, 74
 — Gütemaßstab der 60, 64
 — intrinsische 72
 — Schul- 63
 — TAT als Meßinstrument der 62
- Lern/effekt → Übungseffekt
 — -einheit 165
 — -kontrolltest 26, 165 f.
 — -kriterium 139
 — -leistungsstandard 139
 — -motivation 62
 — -steuerungstest 26, 165 f.
 — -zielanalyse 172
 — -zielanalyseschema 175 f.
- Lernziele: (BLOOM/HORN)
 — Anwendung 172 ff., 178
 — Evaluation 172 ff., 174
 — Synthese 172 ff.
 — Verstehen 172 ff.
 — Wissen 172 ff.
- Lernziele, Modelle der 164
 — operationalisierte 27 f., 169 ff.
 — Spezifikationstabelle der 184
 Logischer Fehler 243
 Lösungswahrscheinlichkeit 150

- matched-exercises → Ordnungs-
 antworten
 Maximum-likelihood-Verfahren 149
 Mehrfachwahl/alternativen 190
 — -antwort 123, 190 ff., 198 f.
 — -aufgabe 191
 — -item 190
 Messen 22, 97 ff.
 Meß/funktion 140
 — -instrumente, sequentiell-skalierte
 152
 — -theorie 102
 — -wert 97, 103, 151
 — -wiederholung 104
 Messung, kriteriumsbezogene
 137 ff., 146
 — normbezogene 138, 140, 151, 155
 — parallele 103
 Mildefehler → Beurteilungsfehler
 Modelllearning → Imitationslernen
 Modelltest 150
 Motivation → Leistungsmotivation
 multiple choice form 190 ff.
- Neuigkeitswert des Lernstoffes 64
 Nicht-direktives Verhalten → direktives
 Verhalten
 Nominalskala 99, 102
 Norm(en) 215
 — Äquivalent- 122
 — bei Schultests 121 f.
 — mittlere 119
 Normalverteilung 96
 Normierung (Standardisierung)
 119, 215 ff.
 Numerisches Relativ → Relativ
 Nullpunkt, absoluter 100
- Objektivität 22, 101, 107 f., 131
 — Auswertungs- 108, 117 f., 316, 320 f.
 — Durchführungs- 107, 111, 117 f.
 — Interpretations- 108, 118 f.
 — Urteils- 316 ff., 321
 — von Schultests 117 ff.
 Ogive 217 f.
 Ordinalskala 99, 102, 122, 130
 Ordnungsantworten 194
 Overachiever 57, 59, 68, 81
- Parallelität (Tests), inhaltliche 110
 — statistische 110
 Paralleltest-Reliabilität 110
- Personenmessung, itemfreie (RASCH-
 Modell) 149
 Persönlichkeitsdimensionen (und Schul-
 leistung) 65 ff.
 Populationsabhängigkeit → Test-
 theorie
 predictive validity → Validität,
 Vorhersage 65 ff., 242 ff.
 Prediktorvariable 162
 Projektionsfehler → Beurteilungsfehler
 Prototyp-Item 146 f.
 Prozentrangnormen 122, 216
 Prozentränge 216 ff.
 Prüfungsaufgaben, Erstellung von
 176 ff.
 Pygmalion-Effekt 262
- Rang/folge → Ordinalskala
 — -ordnung 130
 — -reihen-Umwandlung 220
 — -skala 22, 99 f.
 RASCH-Modell 131, 149
 Ratewahrscheinlichkeit (Aufgaben-
 antworten) 189, 193
 Ratingskala → Schätzsкала
 Reaktionstendenzen 189, 193
 Relativ, empirisches 97 f.
 — numerisches 98, 102
 Reliabilität 22, 101, 104, 109 ff., 133,
 212 ff.
 — Halbierungs- 111, 133
 — Konsistenz- 111, 133
 — Paralleltest- 110 f., 133
 — Retest- 109 f., 133
 Reliabilitäts/bestimmung (Schultests)
 109, 133
 — kriteriumsbezogene Messung 151
 — -koeffizient 129, 212 f.
 — -minderung 110
 — -schätzung, klassische 153
 Repräsentationsschluß 106
 Repräsentativität 131
 Resonanzoptimierung 75
 response-set → Reaktionstendenzen
 restricted response questions → Ant-
 wort, eingeschränkte
 Retest-Reliabilität → Reliabilität
 Rohpunktskala 119
- Schätzsкаlen 243 f.
 Schülerbögen 233

- Schullaufbahnberatung 20 f.
- Schulleistung und
 - Ängstlichkeit 66 f.
 - Anpassung 49
 - Begabung 44
 - Extraversion 70 ff.
 - Familienkonstellation 82 ff.
 - Intelligenz 16, 40 ff., 47 ff.
 - Intelligenzfaktoren 15 f., 54 ff.
 - kognitive Faktoren 17 f., 47 ff., 54 ff.
 - kognitive Stile 74 f.
 - Kreativität 58 f., 83 f.
 - Lehrerverhalten 70 f., 80
 - Neurotizismus 67
 - nicht-kognitive Faktoren 18, 60 ff.
 - Persönlichkeit 44, 65 f., 242 f.
 - Selbstachtung 69 f.
 - soziokulturelles Milieu 80 f.
 - Schulnoten 42
 - Unterrichts Atmosphäre 14 f., 75 ff.
 - Werthaltungen 14 f., 63, 79 f., 82 f.
 - Schulleistung, Definition der 46
 - Indikator der 38
 - Konstrukt der 38 ff., 44
 - Prognose der 53
 - strukturelle Beschreibung der 38 ff.
- Schulleistungstests (Übersicht) 29 ff., 228
- Schulnoten
 - als Urteile 253
 - Funktion der 260 f.
 - und Intelligenz 37
 - Normalverteilung der 257
 - Objektivität der 257 f.
 - Reliabilität der 259 f.
 - Validität der 260 f.
 - Verteilungsformen der 255
- Schultests (Definition) 128 f.
 - Klassifikation von 26 f., 158 ff.
- Schwierigkeit → Aufgabenschwierigkeit
- Schwierigkeits/grad 65, 203, 205 f.
 - -index 142, 206
- Selektion, rationale (LIENERT) 212
- short-answer-response → Kurzantworten
- Skala 22, 98 f.
 - Beobachtungs- 240
 - graphische 244
 - numerische 244
 - Schätz- 243 f.
 - Rating- 243 f.
- Skalen/niveau 98
 - -transformation 100
- Skalogrammanalyse 152
- Stabilitätskoeffizient 212
- Stamm (des Items) 190
- Standard 138
 - absoluter 138
 - normativer 138
 - relativer 138
- Standard/-Äquivalent-Normen 120 f.
 - -meßfehler 112, 130
 - -norm 120 ff.
 - -normalverteilung 120
 - -schätzfehler 116
 - -schulnoten 120, 122
 - -werte 120, 219
- Stichprobe, repräsentative 119
- Summenprozentkurve → Ogive

- TAT → Thematischer Apperzeptions-
test
- Taxonomie (BLOOM) 172, 177 ff., 223 f.
- Test/analyse bei kriteriumsbezogenen
Tests 151
 - -angst 67
 - -batterien 125 f.
 - -eichung, personenfreie 149
 - -heft 202
 - -leitereinflüsse 118
 - -konstruktion 28, 103, 123, 183
 - -verhalten 105
- Testen 97 f.
- Testmodell, logistisches 169 f.
 - stochastisches 131
- Test/definition 22, 159
 - eindimensionaler 150
 - informeller 96, 159 f., 182 ff.
 - kriteriumsbezogener 138, 161 f.
 - lehrzielorientierter 137 f., 161
 - lernzielorientierter 137
 - normbezogener 161 f.
 - Schulleistungs- 96 ff.
 - standardisierter 29 f., 108, 137, 159 f.
- Testtheorie
 - Axiome der 102 f.
 - klassische 21 f., 96 ff., 102 f., 109 f.
 - Kritik der 24, 129 ff.
- Thematischer Apperzeptionstest (TAT)
61
- time-sampling → Beobachtung,
fraktionierte

- Transformation 119, 130, 219
 - Flächen- 120
 - lineare 120
- Trennschärfe 124, 142 f., 205 ff., 210
 - fehlende 143
 - -index 124, 143, 207 f.
 - klassische 145
 - lehrzielorientierte 145
 - negative 143
 - testkriterienbezogene 146 f.
- true-false-items → Alternativantwort
- T-Wert-Skala (Standardwerte) 121

- Übereinstimmungs/koeffizient 148 f., 153
 - -validität 115
- Übungseffekt (Testwiederholung) 110
- Underachiever 56, 58 f., 68, 81
- Untertests 125

- Validität 22 f., 101 f., 105 ff., 114 ff., 133, 215
 - curriculare 114, 134
 - inhaltliche 105 f., 144
 - Konstrukt- 106 f., 116, 153
 - kriteriumsbezogene 106, 115 ff., 130, 139
 - von Schultests 114 ff., 133
 - Übereinstimmungs- 106, 115
 - Vorhersage- 106, 115, 133
- Validitätsbestimmung bei kriterienbezogenen Meßverfahren 153
- Variabilität der Testwerte 140
- Varianz/analytischer Ansatz 154
 - beobachtete 129
 - Fehler- 105
 - Test- 105
 - wahre 105, 129
- Verhaltens/beobachtung 30 f., 233
 - -skalen (РОТН) 240
- Verhältnisskala 100, 102
- Verteilungsfunktion, cumulierte, logistische 150
- Vertrauensgrenzen 112 f.
- Vorhersagevalidität → Validität
- Vortest — Nachtest/Differenzen-Index (Dpp) 145
 - -Design 145
- Vierfelder/analyse 146 f.
 - -häufigkeit 148 f.

- Wahrer Wert 24, 103 f., 129

- Zensierung (Notengebung) 31 f., 253 ff., 312 f.
 - von Schulaufsätzen 271 ff., 309 ff.
- Zuordnungsantworten 194 f.
- Zuverlässigkeit → Reliabilität

Psychologie und Schulpädagogik

Heidelberger Autorengruppe: Steckbrief der Psychologie. Hrsg. von Klaus Eckart Rogge. 2., durchgesehene Auflage, 290 Seiten, DM 12,80. UTB 37

Hans-Jürgen Pfister (Hrsg.): Aspekte der Pädagogischen Psychologie. Psychologie im Studium der Lehrer. 236 Seiten, DM 14,—

Hans Maier / Hans-Jürgen Pfister: Die Grundlagen der Unterrichtstheorie und der Unterrichtspraxis. Ein Beitrag zur Phänomenologie des Unterrichts. 248 Seiten, DM 37,—

Wolfgang Fischer (Hrsg.): Schule und kritische Pädagogik. Fünf Studien zu einer pädagogischen Theorie der Schule. 176 Seiten, DM 19,—

Jörg Ruhloff: Ein Schulkonflikt wird durchgespielt. Beschreibung und Analyse. Mit einem Vorwort von Wolfgang Fischer. 112 Seiten, DM 12,—

Christa Berg: Die Okkupation der Schule. Eine Studie zur Aufhellung gegenwärtiger Schulprobleme an der Volksschule Preußens (1872—1900). 272 Seiten, DM 28,—

Bruno Krapf: Unterrichtsverlauf und programmierte Lernhilfen. 144 Seiten, DM 21,—

Georg E. Becker: Optimierung schulischer Gruppenprozesse durch situatives Lehrtraining. Mit Studienmaterialien und Trainingsunterlagen. 180 Seiten, DM 16,50 (Gruppenpädagogik — Gruppendynamik, Bd. 3)

David Warwick: Team Teaching. Grundlegung und Modelle. Aus dem Englischen übersetzt und bearbeitet von Rainer Winkel. 148 Seiten, DM 19,— (Gruppenpädagogik — Gruppendynamik, Bd. 5)

Ernst Meyer und Börje Forsberg (Hrsg.): Einführung in die Praxis der schulischen Gruppenarbeit. Materialien für Lehrer, Schüler und Eltern. 180 Seiten, DM 16,— (Gruppenpädagogik — Gruppendynamik, Bd. 8)

Quelle & Meyer Heidelberg