



**Heinz Schuler
(Hrsg.)**

Wirtschaftspsychologie

Beurteilung und Förderung beruflicher Leistung

**2., vollständig überarbeitete
und erweiterte Auflage**



Hogrefe

Beurteilung und Förderung beruflicher Leistung

Schriftenreihe
Wirtschaftspsychologie
herausgegeben von
Prof. Dr. Heinz Schuler

**Beurteilung und Förderung
beruflicher Leistung**

herausgegeben von
Prof. Dr. Heinz Schuler



**Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle**

Beurteilung und Förderung beruflicher Leistung

herausgegeben von
Heinz Schuler

2., vollständig überarbeitete und erweiterte Auflage



**Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle**

Prof. Dr. Heinz Schuler, geb. 1945 in Wien. Studium der Psychologie und Philosophie in München, Promotion 1973 und Habilitation 1978 in Augsburg. Nach Auslandsaufenthalten 1979 Professor und Institutsvorstand in Erlangen, seit 1982 Inhaber des Lehrstuhls für Psychologie der Universität Hohenheim, daneben Wissenschaftlicher Leiter der S & F Personalpsychologie Managementberatung in Stuttgart. Gründungsherausgeber der Zeitschrift für Personalpsychologie. Arbeitsschwerpunkte: Organisations- und Personalpsychologie, insbesondere Berufseignungsdiagnostik und Leistungsforschung.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar

Die erste Auflage des Buches ist 1991 in der Reihe „Beiträge zur Organisationspsychologie“ im Verlag für Angewandte Psychologie, Stuttgart, erschienen.

© by Hogrefe-Verlag, Göttingen · Bern · Toronto · Seattle 2004
Rohnsweg 25, D-37085 Göttingen

<http://www.hogrefe.de>

Aktuelle Informationen · Weitere Titel zum Thema · Ergänzende Materialien



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Satz: Grafik-Design Fischer, Weimar
Druck: AZ Druck und Datentechnik GmbH, Kempten
Printed in Germany
Auf säurefreiem Papier gedruckt

ISBN 3-8017-1604-X

Vorwort zur 2. Auflage

Seit dem Erscheinen der ersten Auflage dieses Bandes ist ein gutes Jahrzehnt vergangen, in dem das Personalwesen erhebliche Fortschritte gemacht hat. Speziell die personalpsychologischen Methoden der Berufseignungsdiagnostik und Leistungsbemessung zur Auswahl von Mitarbeitern sowie zur Beurteilung und Förderung ihrer beruflichen Leistung konnten sich zunehmend durchsetzen und bestimmen heute das Bild aufgeklärter Personalarbeit. Noch steht für die Leistungsbeurteilung keine vergleichbare Norm zur Verfügung, wie es die DIN 33430 für die Personalauswahl darstellt. Doch erkennbar ist auch auf diesem Gebiet das Bemühen um hochwertige Verfahren und um professionelle wie bedachtsame Einführung neuer Beurteilungsverfahren. Von einem Beurteilungssystem kann dann gesprochen werden, wenn verschiedene Verfahrenskomponenten in koordinierter Weise verbunden sind und abgestimmt werden mit kontrollierten Vorgehensweisen der Zielsetzung und des Feedbacks. Ein System dieser Art gehört zu den verlässlichsten Maßnahmen nachhaltiger Leistungsverbesserung und ist als Grundlage der Personalentwicklung sowie als Mittel der Qualitätskontrolle für die meisten Organisationen unentbehrlich. Freilich bedarf auch die Ausarbeitung und Einführung eines so wichtigen und durchaus heiklen Instruments seinerseits der Qualitätskontrolle, wofür heute elaborierte, aussagekräftige Methoden zur Verfügung stehen.

Die im vorliegenden Band versammelten Beiträge zeigen die Vielfalt der heutigen Möglichkeiten zur Beurteilung und Förderung beruflicher Leistung auf. Die besten Beurteilungsverfahren werden ebenso dargestellt wie ihre Grundlagen und Funktionen, desgleichen das Vorgehen bei Beurteilungsgesprächen, Möglichkeiten des Trainings der Leistungsmotivation und ein System des Leistungsmanagements von Gruppen. Eine Serie von vier Beiträgen berichtet über die Entwicklung, Prüfung und Einführung eines multimodalen Beurteilungssystems für Mitarbeiter, Führungskräfte, Auszubildende und Arbeitsgruppen in einem großen Unternehmen.

Verschiedenes konnte aus der ersten Auflage dieses Bandes nach Überarbeitung übernommen werden. Um den Fortschritten des Fachgebiets gerecht zu werden, wurde aber der größte Teil der Beiträge neu verfasst. Die Autoren, denen ich sehr herzlich für die Zusammenarbeit danke, sind entweder als Wissenschaftler oder als Praktiker mit Leistungsbeurteilung und Personalentwicklung befasst, in mehreren Fällen in beiden Funktionen. Sie alle haben sich bemüht, den Lesern anregende, sowohl methodisch anspruchsvolle als auch praktisch gut umsetzbare Darstellungen zu bieten.

Auf Grund der weitgehenden Neugestaltung hat der Band weit mehr Zeitaufwand erfordert als erwartet und steht deshalb auch später als zunächst geplant zur Verfügung. Für die Beteiligten hat sich die Unterschätzung des Arbeitsaufwands als wirksame Methode der Selbstmotivation damit ein weiteres Mal bestätigt. Heute ist Sonntag, der 7. September 2003. Von meinem Stehpult aus sehe ich die Spaziergänger im Hohenheimer Schlosspark und freue mich mit ihnen über den sonnigen Spätsommertag. Noch mehr allerdings freue ich mich über die Unterstützung durch Herrn Christian Steidel, der im Nebenzimmer ebenfalls dem Wetter trotzt und mit redaktionellen Arbeiten für dieses Buch beschäftigt ist. Er hat es in allen Entwicklungsschritten begleitet und so viel Unter-

stützung dafür geleistet (neben der Arbeit an zwei Enzyklopädiebänden), dass er als studentischer Mitarbeiter fast schon überqualifiziert ist. Ich danke ihm sehr herzlich für sein Engagement und werde für die nächste Beurteilung seiner Leistung die Einstufungsskala verlängern. Ebenfalls danke ich Frau Gisela Bieling, deren sorgfältige Arbeit beim abschließenden Korrekturlesen und bei der Anfertigung der Register eine große Hilfe war.

Allen Lesern wünsche ich eine interessante, zur Anwendung anregende Lektüre.

Stuttgart-Hohenheim, im September 2003

Heinz Schuler

Inhalt

1 Leistungsbeurteilung – Gegenstand, Funktionen und Formen	1
<i>Heinz Schuler</i>	
1.1 Gegenstand der Leistungsbeurteilung	2
1.2 Funktionen der Leistungsbeurteilung	3
1.3 Formen der Leistungsbeurteilung	5
1.3.1 Leistungskriterien	5
1.3.2 Leistungsmaße	6
1.3.3 Beurteilungsverfahren	9
1.4 Konzeption eines Beurteilungssystems	19
Literatur	22
2 Drei Ebenen der Leistungsbeurteilung: Day-to-day-Feedback, Regelbeurteilung und Potenzialanalyse	25
<i>Heinz Schuler</i>	
2.1 Day-to-day-Feedback, Regelbeurteilung und Potenzialanalyse – Einsatzbereiche, Merkmale und Funktionen	25
2.1.1 Day-to-day-Feedback	26
2.1.2 Regelbeurteilung	27
2.1.3 Potenzialanalyse	28
2.2 Beziehungen zwischen Day-to-day-Feedback, Regelbeurteilung und Potenzialanalyse	29
Literatur	31
3 Der Prozess der Urteilsbildung und die Qualität der Beurteilungen	33
<i>Heinz Schuler</i>	
3.1 Soziale Urteilsbildung	34
3.2 Urteilstendenzen	37
3.2.1 Mittelwertstendenzen	37
3.2.2 Streuungstendenzen	40
3.2.3 Korrelationstendenzen	41
3.3 Akkuratheit	42
3.4 Testtheoretische Qualitätsindikatoren	43
3.4.1 Objektivität	43
3.4.2 Reliabilität	45
3.4.3 Validität	48

3.5	Verwendungskriterien	51
3.5.1	Akzeptabilität	51
3.5.2	Praktikabilität und Nutzen	54
	Literatur	57
4	360°-Beurteilung	61
	<i>Martin Scherm</i>	
4.1	Einleitung	61
4.2	Modell der multiperspektivischen Beurteilung	64
4.3	Die Güte der 360°-Beurteilung	66
4.3.1	Reliabilität der Beurteilung	66
4.3.2	Validität der Beurteilung	70
4.4	Verbesserung durch Feedback? Befunde zur Effektivität von Beurteilungsprozessen	73
4.5	Auf dem Weg zu Exzellenz: Hinweise für die erfolgreiche Gestaltung von 360°-Beurteilungen	76
	Literatur	79
5	Selbstbeurteilung	83
	<i>Klaus Moser</i>	
5.1	Funktionen der Selbstbeurteilung beruflicher Leistung	83
5.2	Gütekriterien von Selbstbeurteilungen	85
5.2.1	Urteilstendenzen	85
5.2.2	Psychometrische Kriterien	87
5.2.3	Akkuratheit	88
5.3	Moderatoren der Güte von Selbstbeurteilungen	89
5.3.1	Personfaktoren	90
5.3.2	Organisations- und tätigkeitsbezogene Faktoren	91
5.3.3	Durchführungsbedingungen	92
5.4	Konsequenzen	94
5.5	Erklärungsansätze	95
	Literatur	97
6	Leistungsmessung und strategische Ausrichtung mithilfe der Balanced Scorecard	101
	<i>Daniel Sodenkamp</i>	
6.1	Einleitung	101
6.2	Die Balanced Scorecard (BSC) als System der Leistungsmessung	102
6.2.1	Mehrdimensionalität	102
6.2.2	Frühindikatoren	103
6.2.3	Einführung der Balanced Scorecard	104

6.3 Die BSC als Strategie-System	105
6.4 Leistungsmessung und strategische Ausrichtung mithilfe der BSC – ein Beispiel	106
6.5 Die BSC im deutschen Sprachraum	110
6.6 Anwendererfahrungen	111
6.7 Fazit aus psychologischer Sicht	113
Literatur	113
7 Strategische Elemente in Leistungsbeurteilungen	115
<i>Klaus Moser</i>	
7.1 Strategische Ziele	115
7.2 Determinanten strategischer Verzerrungen	117
7.3 Diskussion	120
Literatur	123
8 Multimodale Leistungsbeurteilungssysteme für Mitarbeiter, Führungskräfte und Auszubildende in der Kreditwirtschaft – Ziele und Inhalte	125
<i>Karlheinz Becker, Andreas Diemand und Heinz Schuler</i>	
8.1 Leistungsoptimierung als Ziel der Projektentwicklung	125
8.2 Komponenten des Multimodalen Leistungsbeurteilungssystems MLB	127
8.3 Verbesserte Kollektivleistung durch Gruppenbeurteilung	129
8.4 Wissenschaftliche Fundierung und praktische Erprobung von MLB	130
8.5 Spezifika von MLB Junior	131
8.6 Erste Erfahrungen bei der Einführung von MLB	132
Literatur	132
9 Entwicklung eines multimodalen Systems zur Beurteilung von Individualeistungen	133
<i>Heinz Schuler, Peter M. Muck, Benedikt Hell, Stefan Höft, Karlheinz Becker und Andreas Diemand</i>	
9.1 Einführung	133
9.2 Verwendete Anforderungsdimensionen	134
9.3 Verwendete Beurteilungsverfahren	135
9.3.1 Verhaltensverankerte Einstufungsskalen	135
9.3.2 Verhaltensbeobachtungsskalen	136
9.3.3 Verhaltensrangprofil	136
9.3.4 Vergleich der Merkmale und Funktionen der drei Beurteilungs- verfahren	139

9.4 Entwicklungsprozess	141
9.4.1 Auswahlprozess für die Beurteilungsmodule	141
9.4.2 Entwicklung der Beurteilungsverfahren: Item- und Skalen- konstruktion	142
9.5 Probeanwendung	144
9.5.1 Beurteilungsrichtungen	144
9.5.2 Ergebnisse	145
9.6 Zusammenfassung und Fazit	156
Literatur	157
10 Entwicklung eines multimodalen Systems zur Beurteilung von Gruppenleistungen	159
<i>Peter M. Muck, Heinz Schuler, Karlheinz Becker und Andreas Diemand</i>	
10.1 Zielsetzung und theoretischer Hintergrund	159
10.2 Konzeption des Gruppenmoduls	161
10.2.1 Grundlegendes Modell	161
10.2.2 Charakterisierung des Skalentyps	166
10.3 Empirische Prüfung	168
10.3.1 Zielsetzungen der empirischen Prüfung	168
10.3.2 Methode	168
10.3.3 Ergebnisse der empirischen Prüfung – Analysen auf der Ebene des Individuums	170
10.3.4 Ergebnisse der empirischen Prüfung – Analysen auf der Ebene der Gruppe	177
10.4 Abschließende Ausgestaltung des Gruppenmoduls	181
10.5 Zusammenfassende Diskussion und Fazit	183
Literatur	184
11 Entwicklung eines multimodalen Systems zur Leistungsbeurteilung von Auszubildenden	187
<i>Yvonne Klingner, Heinz Schuler, Andreas Diemand und Karlheinz Becker</i>	
11.1 Grundlagen	187
11.1.1 Hintergrund und Zielsetzung	187
11.1.2 Überblick über den Gesamtprozess	188
11.2 Anforderungsdimensionen	189
11.3 Beurteilungsverfahren	192
11.3.1 Verhaltensbeobachtungsskala	193
11.3.2 Verhaltensrangprofil	193
11.4 Beurteilungsmodule	196
11.5 Skalenkonstruktion	197
11.5.1 Itemkonstruktion	197
11.5.2 Vorstudie	197
11.5.3 Hauptstudie	198

11.6	Das Beurteilungsgespräch	208
11.7	Basiszeugnis für die Abschließende Beurteilung	210
11.8	Beurteilertraining	211
11.9	Fazit	211
	Literatur	212
12	Die Wirkung von Zielsetzungen auf die Leistung	215
	<i>Uwe Kleinbeck</i>	
12.1	Die Bedeutung von Zielen für Leistungen	215
12.1.1	Die Befundlage: Hohe spezifische Ziele fördern die Leistung ..	216
12.2	Merkmale von leistungswirksamen Zielen	217
12.3	Moderatoren der Zielsetzungswirkung	218
12.4	Wirkmechanismen von Zielen	223
12.5	Die Bedeutung von Erfolg und Misserfolg für die langfristige Zielverfolgung	224
12.6	Gruppenziele und Gruppenleistung	227
12.7	Die Anwendung von Erkenntnissen der Zielsetzungsforschung in Organisationen	232
	Literatur	233
13	Förderung von Gruppenleistungen mit dem Partizipativen Produktivitätsmanagement (PPM)	239
	<i>Klaus-Helmut Schmidt</i>	
13.1	Einleitung und Problemstellung	239
13.2	Theoretischer Hintergrund von PPM	241
13.3	Merkmale des PPM-Ansatzes	242
13.4	Entwicklungsschritte bei der Anwendung von PPM	244
13.5	Wirkungen von PPM	247
13.6	Neuere Entwicklungen und weitere Nutzungsmöglichkeiten von PPM	249
	Literatur	251
14	Beurteilungsgespräch, Zielsetzung und Feedback	255
	<i>Peter M. Muck und Heinz Schuler</i>	
14.1	Funktionen des Beurteilungsgesprächs	255
14.2	Prinzipien der Gesprächsführung	257
14.2.1	Gesprächsform und -stil	257
14.2.2	Gesprächsregeln	260
14.3	Wirkung von Feedback	264
14.3.1	Wirkung von Anerkennung und Kritik	265
14.3.2	Reaktionen beim Empfänger	268

14.4	Zielsetzung	272
14.5	Beispielhafter Ablauf eines Beurteilungsgesprächs	273
14.5.1	Voraussetzungen für ein gelungenes Gespräch	273
14.5.2	Struktur eines Beurteilungsgesprächs	276
14.6	Sonderfälle von Beurteilungsgesprächen	280
14.6.1	Feedback an Auszubildende	280
14.6.2	Feedback an Vorgesetzte	282
14.6.3	Feedback bei Gruppenbeurteilungen	283
	Literatur	286
15	Das Beurteilungs- und Fördergespräch als Instrument der Personalentwicklung bei der SEB AG	291
	<i>Erich Barthel und Marcella Kessel</i>	
15.1	Einleitung	291
15.2	Ziele des Beurteilungs- und Fördergesprächs im Rahmen der Personalentwicklung	292
15.3	Das Beurteilungs- und Fördergespräch	292
15.3.1	Entwicklung des Systems	292
15.3.2	Aufbau des Beurteilungs- und Fördergesprächs	293
15.3.3	Das Fördergespräch	296
15.4	Verknüpfung mit anderen Instrumenten der Personalentwicklung	299
15.4.1	Das Personal-Portfolio	299
15.4.2	Das Assessment Center (AC)	301
15.4.3	Das Führungskräfte-Feedback	301
15.5	Erfahrungen und Ausblick	302
16	Der Funktionskreis „Leistungsförderung“ – eine Skizze	305
	<i>Heinz Schuler</i>	
16.1	Zum Defizit der Führungsforschung	305
16.2	Das Modell	306
16.3	Komponenten des Funktionskreises	308
16.3.1	Leistungsmotivation	308
16.3.2	Attribution	309
16.3.3	Leistungsbeurteilung und Feedback	311
16.3.4	Leistungszufriedenheit	313
16.3.5	Selbstbild und Selbstvertrauen	313
16.3.6	Erfolgserwartung	314
16.3.7	Zielsetzung und Commitment	315
16.3.8	Anstrengung	317
16.3.9	Fähigkeiten	318
16.4	Leistungsförderung oder Leistungsmanipulation?	319
	Literatur	320

17 Leistungsmotivation als globales Konstrukt – ein Förderversuch einzelner Teilaspekte	323
<i>Corinna Wall und Heinz Schuler</i>	
17.1 Einleitung	323
17.2 Das Konstrukt Leistungsmotivation	323
17.3 Drei Teilaspekte aus theoretischer und praktischer Sicht	325
17.3.1 Schwierigkeitspräferenz	326
17.3.2 Kompensatorische Anstrengung	328
17.3.3 Leistungsstolz	331
17.4 Evaluation des Trainingsprogramms	333
17.4.1 Ergebnisse	334
17.4 Diskussion	340
Literatur	343
18 Eigenleistung in Beruf und Gesellschaft	347
<i>Hans Lenk</i>	
18.1 Einführung	347
18.2 Eigenhandeln und Eigenleistung als Elemente menschlichen Lebens ..	347
18.3 Zum Begriff der Eigenleistung	348
18.4 Vergleichs- und Bewertungsprinzipien der Eigenleistung	350
18.5 Zur Bedeutung der Eigenleistung im täglichen Leben	351
18.6 Abschließende Thesen zur Eigenleistung	354
Literatur	357
19 Overcoming Halo and Leniency – A New Method of Merit Rating	359
<i>Hermann Brandstätter und Heinz Schuler</i>	
19.1 The problem	359
19.2 Description of the method	360
19.3 Working efficiency	361
19.4 Index for rating quality	363
References	363
Personenregister	365
Sachregister	373

1 Leistungsbeurteilung – Gegenstand, Funktionen und Formen

Heinz Schuler

Leistungsbeurteilung gehört, zumal wenn sie gekoppelt ist mit Zielsetzung und Feedback, zu den wirksamsten personalpsychologischen Maßnahmen. Der leistungsfördernde Effekt von Beurteilungen kann mit etwa einer halben Standardabweichung des durchschnittlichen Leistungswerts am betreffenden Arbeitsplatz geschätzt werden (Guzzo, Jette & Katzell, 1985), das bedeutet eine durchschnittliche Anhebung der individuellen Leistung von Prozentrang 50 auf etwa Prozentrang 70.

Unter den vielfältigen Aufgaben einer Führungskraft gehört die Leistungsbeurteilung gleichwohl meist zu den weniger geschätzten. Dies gilt vor allem im Falle formeller Beurteilungsverfahren, festgelegter Termine und geforderter Beurteilungsgespräche. Nicht immer wird eine förmliche Beurteilung praktiziert, und nicht immer ist sie erforderlich. Doch auch ohne Formulare und Regeln ist die Einschätzung von Leistung und Leistungsfähigkeit eine unabdingbare Grundlage der Führung und Zusammenarbeit im Berufsleben. Wie sollte ein Vorgesetzter die Entscheidung über eine Beförderung treffen, wie ein Kollege sich mit der Bitte um Rat an den kompetentesten Kollegen wenden können, wenn sie nicht in der Lage wären, Fähigkeiten und Verhaltensbereitschaften bei anderen einzuschätzen.

Der Kern der Kritik im Unternehmen gilt deshalb auch häufig nicht dem Prinzip des Beurteilens als solchem, sondern dem Zwang, Formulare zu verwenden, sich an Beurteilungsdimensionen zu orientieren, die nicht dem eigenen Urteilsprozess entsprechen, und den Mitarbeitern Beurteilungswerte zu erklären, die sich an der Vorgabe der Normalverteilung zu orientieren haben. Beurteiler vertrauen gewöhnlich ihren eigenen globalen Einschätzungen mehr als den elaborierten Urteilen, die mithilfe der Beurteilungformulare zu Stande kommen.

Probleme mit der Leistungsbeurteilung sind deshalb zu einem Großteil Management-Probleme und nicht allein Probleme der Durchführungs- und Skalierungstechnik, als die sie in der psychologischen Forschung über lange Jahre behandelt wurden. Zu wenig wurde berücksichtigt, wie die Beurteilung zum Umgangsstil in einer Organisation passt, welche Vorgeschichte dort die Einstellungen zur Leistungsbestimmung prägt, in welchem Kontext die Beurteilung steht und welche Konsequenzen von den Ergebnissen abhängen, inwieweit die Methoden allen Beteiligten verständlich und für sie akzeptabel sind. Ein Großteil des Aufwands, den die Einführung eines Beurteilungsverfahrens mit sich bringt, ist deshalb für den Prozess der Anpassung oder besser „Einpassung“ des Systems in die Organisation zu leisten. In manchen Fällen wird die Entscheidung vernünftigerweise dahingehend ausfallen, dass man auf ein ausgefeiltes System ganz verzichtet. Wie die weiteren Ausführungen und die anderen Beiträge in diesem Band zeigen, lautet aber die Entscheidung gewöhnlich nicht „hochkompliziertes System oder völliger Verzicht

auf Beurteilung“, sondern „Welche der vielen möglichen Verfahrensweisen erfüllt unter den gegebenen Bedingungen am besten den spezifischen angestrebten Zweck und dient in ihrer Gesamtwirkung bestmöglich den Zielen der Organisation und den Interessen der Beteiligten“.

1.1 Gegenstand der Leistungsbeurteilung

Leistungsbeurteilung – und noch stärker ihre Kritik – wird meist in dreierlei Hinsicht zu eingeschränkt gesehen: zum Einen beschränkt auf den Gegenstand „sachliche Arbeitsleistung“ im engen technisch verstandenen Sinn von Produktivität; zum Zweiten als quasi hoheitlicher Akt, exekutiert von einem Vorgesetzten an einem Untergebenen; das dritte Missverständnis besteht in der Auffassung, das natürliche Objekt der Leistungsbeurteilung sei der Mensch, genauer gesagt die Einzelperson. In dem bescheidenen Maße, wie Publikationen überhaupt Wirkungen auf das Handeln in Organisationen haben, trugen alle Autoren, die sich zum Thema Leistungsbeurteilung („Mitarbeiterbeurteilung“!) geäußert haben – der Verfasser dieses Beitrags nicht ausgenommen – zur Verfestigung dieser Beschränkungen bei; nicht weniger allerdings auch all jene, die für die Beurteilung von Leistungen in der Organisationspraxis verantwortlich sind.

Zu Punkt Eins: Arbeitsleistung in einem Betrieb ist tatsächlich häufig sachlichen Charakters und bedeutet „Umsatz abgeschlossener Leasingverträge“, „Fehlerquote beim Einbau von Kühlschranksmotoren“ oder „Kenntnis des europäischen Patentrechts“. Nicht seltener allerdings ist Leistung eine Verhaltens- oder Ergebnisqualität, die sich auf das Agieren in einem Netzwerk sozialer Beziehungen oder auf die Wirkungsweise in einem Feld symbolischer Repräsentationen bezieht. Dann ist Leistung beispielsweise „Ermunterung desinteressierter Schüler zu eigenständigen Unterrichtsbeiträgen“, „Förderung der Kooperation zwischen rivalisierenden Abteilungen“ oder „Beeinflussung von Verbrauchern zu Gunsten umweltschonender Verpackungen“.

Zu Punkt Zwei: Verbreitetste Vorstellung von einer Leistungsbeurteilung ist die formelle Einschätzung eines Mitarbeiters durch seinen Vorgesetzten. Eine Vielfalt von Urteilen über leistungsbezogenes Verhalten wird dabei übersehen – beispielsweise die Eindrücke, die sich Kollegen oder Konkurrenten voneinander bilden, die Einschätzung des eigenen Verhaltens, die jemandem durch die Beobachtung seiner Wirkungen auf andere ermöglicht wird, die Reaktion des Publikums auf Schauspieler oder Politiker, die Beurteilung der Kreditwürdigkeit eines Antragstellers durch seine Bank. Auch dort, wo „traditionellerweise“ die Leistung von Mitarbeitern durch ihre Vorgesetzten beurteilt wird, kommt darin nicht notwendigerweise die hierarchische Struktur von Organisationen zum Ausdruck; sie kann auch Teil einer offenen, konsensuell vereinbarten gegenseitigen Verhaltenssteuerung sein.

Zu Punkt Drei: Leistungen werden nicht nur von Einzelpersonen erbracht oder diesen zugeschrieben, sondern auch von der Mannschaft eines Viererbobs, der Werbeabteilung eines Konsumgüterherstellers, von Arbeitgebern und Arbeitnehmern im Ausfechten von Tarifvereinbarungen, von einem Orchester bei der Aufführung einer Symphonie, von einem Team freiwilliger Helfer beim Verteilen von Hilfsgütern an Erdbebenopfer. In all diesen Aktivitäten stecken Leistungen von Einzelpersonen, die in vielen Fällen auch als solche beurteilbar sind. Oft ist aber Gegenstand der Bewertung die Leistung einer Gruppe,

einer Organisation, sogar eines Wirtschafts- oder Gesellschaftssystems. Leistungsziele sind prinzipiell auf jeder Ebene der Kollektivität formulierbar, ihr Erreichen ist in den meisten Fällen zumindest in Teilen oder Aspekten auch erfassbar. Nicht immer allerdings sind Zuschreibungen von Verantwortlichkeit auf beliebigen Ebenen möglich, sinnvoll oder erwünscht.

Der Begriff der Leistung sollte also nicht auf die Verhaltensergebnisse einzelner Personen beschränkt werden, der der Leistungsbeurteilung nicht auf die Feststellung dieser Ergebnisse durch eine hierarchisch übergeordnete Instanz. Leistung ist hier vielmehr zu verstehen als Beiträge einer Organisationseinheit zu den Zielen der Gesamtorganisation, oder – noch allgemeiner – eines Subsystems zu einem übergeordneten System. Als Methoden der Leistungsbeurteilung kommen demgemäß nicht nur die zur Mitarbeiter einschätzung üblicherweise benutzten Skalen in Betracht, sondern praktisch alle Verfahren der Evaluation empirisch erfassbarer Sachverhalte – von der intuitiv-ganzheitlichen Eindrucksbildung bis zum ausgefeiltesten formalen Bewertungsalgorithmus.

Wenn in diesem Beitrag wie im ganzen Band in erster Linie doch von individueller Beurteilung die Rede ist, so entspricht dies der Häufigkeit der Verbreitung und den vorliegenden Untersuchungsergebnissen. Zumindest in einigen der nachfolgenden Kapitel und gelegentlich an Beispielen wird illustriert, dass auch überindividuelle Einheiten und andere Beurteilungsrichtungen als die „von oben nach unten“ gemeint sind.

1.2 Funktionen der Leistungsbeurteilung

Einer der ersten und grundlegenden Schritte bei der Erwägung, ob ein Beurteilungsverfahren eingeführt werden soll, und gegebenenfalls welche Form in Frage kommt, ist die Festlegung der Ziele, die damit erreicht werden sollen. Eine Hilfestellung bei der Zielformulierung kann es sein, Ziele der Organisation und solche der Personen zu unterscheiden (Ilgen, 1993). Zu den Zielen der Organisation gehört es vor allem, die Aufgaben des „Human Resource Management“ zu erfüllen, das heißt Anforderungen zu kommunizieren und Kontingenz zwischen dem Arbeitsverhalten und diesen Anforderungen herzustellen, also Anreize und Belohnungen leistungsgerecht zu gestalten. Daneben gehört zu den Zielen eines Beurteilungsverfahrens für die Organisation, Grundlagen für Entscheidungen auf Systemebene bereitzustellen, das sind beispielsweise Kriterien für die Überprüfung von Personalauswahlverfahren und von Trainingsmaßnahmen. Die Zwecke, die das Beurteilungsverfahren für die Mitarbeiter der Organisation haben kann, liegen vor allem in der Information über Anforderungen und Bewertungen, in Hinweisen, Unterstützung und Motivation zu zielgerechtem Verhalten sowie in leistungsfördernden Belohnungen. Im Abschnitt „Wirkungen“ des Kapitels „Beurteilungsgespräch“ wird erörtert, inwieweit die letztgenannten Ziele durch Leistungsbeurteilung erfüllt werden.

Die in Tabelle 1 wiedergegebene Gliederung von Funktionen der Leistungsbeurteilung verzichtet auf die explizite Trennung von Zielen der Organisation und der Person und ist stattdessen stärker an den praktischen Verwendungsmöglichkeiten orientiert.

Leistungsbeurteilungen können also eine Vielfalt von Zwecken erfüllen. Das Problem ist, sie können dies nicht alles gleichzeitig. Bei Methoden im Rechnungswesen oder im Marketing sieht man es als selbstverständlich an, dass sie zweckentsprechend eingesetzt werden müssen, nicht aber wenn es um Leistungsbeurteilung geht. Ein erheblicher Teil

Tabelle 1: Funktionen der Leistungsbeurteilung

1. Leistungsverbesserung durch Verhaltenssteuerung (Feedback für die Beurteilten)
2. Planung, Auswahl und Gestaltung von Maßnahmen der Personalentwicklung
 - individuell: Maßnahmen zur Erhöhung der individuellen Bewährungswahrscheinlichkeit und Einsatzbreite (u. a. Aufgabenerweiterung, Verhaltenstraining, Fort- und Weiterbildung)
 - kollektiv: Personalentwicklungsplanung
3. Personelle Entscheidungen auf individuellem und kollektivem Niveau
 - individuell: Platzierung, Beförderung, Versetzung, Übernahme, Kündigung
 - kollektiv: Personalplanung
4. Gestaltung von Arbeitsbedingungen (Arbeitsplatz und Arbeitsumgebung), Ausgangspunkt von Organisationsdiagnose und Organisationsentwicklung
5. Gehalts- und Lohnbestimmung
6. Individuelle Beratung und Förderung von Mitarbeitern
7. Verbesserung der Führungskompetenz der Vorgesetzten
8. Evaluation von Selektionskonzepten, personellen Entscheidungen, Maßnahmen der Personalentwicklung, Programmen der Organisationsentwicklung, Anreiz- und Verstärkungssystemen
9. Artikulation von Anforderungen an Arbeitstätigkeit und soziales Verhalten (Leistungsbeurteilung als Instrument der betrieblichen Sozialisation, der Motivierung und Verhaltenssteuerung)
10. Hervorhebung der Bedeutung leistungsorientierter Personalentwicklung in der Organisation

der Unzufriedenheit über Beurteilungssysteme dürfte auf das Missverständnis zurückgehen, mit ein und demselben Verfahren gleichzeitig alle Ziele erreichen zu können. Schon seit langem (z. B. von Meyer, Kay & French, 1965) wird darauf hingewiesen, dass Entscheidungen seitens der Organisation einerseits und Lernen und Coaching andererseits inkompatible Zielgrößen sind.

Als besonders schwer vereinbar hat es sich erwiesen, Gehaltsentscheidungen und Maßnahmen der Personalentwicklung auf der Basis der gleichen Beurteilungsdaten zu veranlassen. Personalentwicklung, also die Gesamtheit der Maßnahmen in Leistungsorganisationen zur Förderung der berufsbezogenen Qualifikation der Beschäftigten, ist in hohem Maße auf die Eigenaktivität des Mitarbeiters beim Feststellen von Defiziten und Förderungsbedarf, danach auch von Trainingswirkungen angewiesen. Vermutlich ist die Personalentwicklung sogar das Anwendungsfeld für die Selbstbeurteilung schlechthin. Wenn

aber gleichzeitig und auf Grund der gleichen Beurteilungsdaten über die Gehaltszulage oder die Beförderung entschieden werden soll, so behindert das ein offenes Ansprechen der Defizite, die ja eine der Grundlagen für die Auswahl von Förderungsmaßnahmen sein können. Zu Förderungszwecken wurden deshalb Methoden wie das Verhaltensrangprofil vorgeschlagen (s. Abschnitt „Methoden“), die auf einen Vergleich zwischen Mitarbeitern explizit verzichten und stattdessen die Anforderungen am gegebenen Arbeitsplatz mit den relativen Stärken und Schwächen einzelner Personen vergleichen, um daraus Förderungserfordernisse abzuleiten. Verzerrungen von Urteilsdaten resultieren freilich nicht nur von Seiten des Beurteilten, sondern auch des Beurteilers. Erwartet beispielsweise ein Mitarbeiter, der das Verhalten seines Vorgesetzten beurteilt, dass seine Einschätzung zur Beförderung dieses Vorgesetzten beitragen wird, so kann sein Urteil eventuell stärker an den erwünschten Konsequenzen orientiert sein als an der zunächst intendierten Feedbackfunktion. Oder wenn die Leistung eines Teamleiters an den Erfolgen der ganzen Gruppe gemessen wird, so dürfte es in seinem Interesse sein, Indikatoren zu sammeln, die den Teamerfolg eher über- als unterschätzen. Ein Großteil der häufig beklagten Urteilstendenzen, insbesondere der Beschönigungstendenz, dürfte durch die eigene Betroffenheit des Urteilenden von seinem Urteil bedingt sein.

Ein anderer Antagonismus ergibt sich zwischen dem individuellen Feedback und der qualitativen Personalplanung. Während ersterem mit einem informellen, anlassbezogenen kurzen Gespräch am besten gedient ist, bedarf letztere vergleichbarer und aggregations-tauglicher Daten. Noch höher ist der methodische Anspruch an Leistungsdaten, der zur Evaluation von Personalauswahlverfahren zu stellen ist. In nicht wenigen Fällen sind die Auswahlverfahren von besserer Qualität als die Kriteriendaten, die als Maßstab ihrer Brauchbarkeit verwendet werden.

1.3 Formen der Leistungsbeurteilung

1.3.1 Leistungskriterien

Der Beitrag einer Person oder Gruppe zu den Zielen der Organisation kann auf vielfältige Weise eingeschätzt werden – zum Beispiel als globale Bewertung, die in eine Rangreihe der verglichenen Personen oder Gruppen mündet, oder aber differenziert nach sehr kleinen Einheiten, etwa des nonverbalen Verhaltens gegenüber Kunden. In keinem Fall ergibt sich die Beurteilungseinheit von selbst, immer steht eine „Theorie“ des Beurteilers dahinter, welche Beobachtungseinheiten als Indikatoren des Erfolgs aufgefasst werden können (vgl. Landy & Farr, 1983).

Manchmal widersprechen sich auch diese Beobachtungen. Beispielsweise mag bei einem Mitarbeiter im Pharma-Außendienst positiv gewertet werden, wenn er täglich acht Ärztebesuche macht und nicht manchmal sechs und andere Male zehn, da man die Erfahrung gemacht hat, dass konstanter Einsatz auf Dauer wirksamer ist. Nun könnte es sein, dass dies zwar für die Mehrzahl der Mitarbeiter gilt, nicht aber für alle. Im Einzelfall könnte jemand eine wirksamere Selbststeuerung dadurch zu Stande bringen, dass er seine Leistungsschwankungen ausnützt und trotz überdurchschnittlicher Schwankungsbreite gute Verkaufserfolge erzielt. In diesem Fall erfüllt er die Anforderung „Verkaufsleistung“, nicht aber die Anforderung „Regelmäßigkeit“. In einer Gesamtbewertung mag

ersteres höher gewichtet werden; wenn aber befürchtet wird, dass die Verhaltensvariabilität Modellcharakter auch für Kollegen hat, deren Leistung sie weniger zuträglich ist, wird man anders denken.

Ein Vorgesetzter, der die Bedingungen des Zustandebringens von Leistung bei jedem seiner Mitarbeiter und für die Arbeitsgruppe als ganze kennt, wird sich hüten, allzu einfache Richtlinien rigide anzuwenden. Stattdessen wird seine Beurteilung immer auch seine Erfahrungen und seine Einschätzung der Zusammenhänge sowie der mutmaßlichen Urteilswirkungen enthalten.

Zumeist werden Beurteilungen anhand mehrerer Einzelaspekte oder -kriterien vorgenommen. Die Herkunft dieser Aspekte ist nicht beliebig, sondern sollte anforderungsanalytisch fundiert sein. Eine der Hauptaufgaben der psychologischen Arbeitsanalyse ist es, Anforderungen an den Stelleninhaber sichtbar zu machen. Es ist deshalb nahe liegend, wenn nicht zwingend, die Beurteilung von Leistungen an eben diesen Anforderungen zu orientieren. Als besonders zweckmäßiges Verfahren der Anforderungsanalyse hat sich die Methode der „Kritischen Ereignisse“ oder „Schlüsselereignisse“ nach Flanagan erwiesen (Flanagan, 1954), mittels deren Verhaltensereignisse von Stelleninhabern gesammelt werden, die sich als erfolgsrelevant gezeigt haben. Der wichtige Grundgedanke ist also, dass nicht beliebige Eigenschaften oder Verhaltensweisen beurteilt werden, sondern nur solche, die erkennbar zu Erfolg oder Misserfolg beitragen. Aber auch verschiedene standardisierte Verfahren der Arbeitsanalyse kommen hierfür in Betracht (Gael, 1988). Ein Beispiel für eine gründliche arbeitsanalytische Fundierung eines Beurteilungsverfahrens wird bei Schuler, Funke, Moser und Donat (1995) geschildert. In diesem Fall wurde ein eigenes Analyseverfahren entwickelt, um hiervon sowohl die Leistungsbeurteilung als auch ein Auswahlverfahren für Ingenieure abzuleiten.

1.3.2 Leistungsmaße

Die Einzelaspekte oder -kriterien können in Form verschiedener Leistungsmaße erhoben werden. Etwa kann dasjenige *Verhalten* beurteilt werden, das man als erfolgsrelevant einschätzt, es können *Fähigkeiten* der Stelleninhaber Gegenstand der Beurteilung sein oder es kann bemessen werden, inwieweit die durch die Tätigkeitsbeschreibung vorgegebenen *Einzelaktivitäten* tatsächlich ausgeführt, oder die vereinbarten *Ziele* erreicht wurden. Eine systematische Auflistung der in Frage kommenden Leistungsmaße mit jeweils einem Beispiel aus der Tätigkeit eines Mitarbeiters im Pharma-Außendienst wird in Tabelle 2 gegeben.

Es bieten sich also viele Möglichkeiten, Komponenten der Leistung oder Indikatoren für Leistung zu beschreiben. Allesamt allerdings sind sie nur Surrogate für das, was uns eigentlich interessiert, für das Konstrukt „Gesamtleistung“. Diese Gesamtleistung, das „letztliche Kriterium“ (Smith, 1976) im Sinne des Beitrags einer Person oder Gruppe zu den Zielen der Organisation, ist schlechthin als solche nicht erfahrbare. Wir können uns ihr nur nähern, indem wir Komponenten oder Indikatoren kombinieren und dadurch möglichst viele Facetten der Gesamtleistung berücksichtigen.

Aber selbst mit einer Mehrzahl einzelner Leistungsindikatoren kann eine sinnvolle Leistungsbeurteilung nur dann vorgenommen werden, wenn sie tatsächlich in Bezug zum Konstrukt Gesamtleistung stehen. Dies kann erstens durch arbeitsanalytische Fundierung des Beurteilungsverfahrens sichergestellt werden, zweitens durch konstruktanalytische

Tabelle 2: Verschiedene Leistungsmaße

1. Tätigkeiten
Beispiel: Niedergelassene Ärzte zu Informationsveranstaltungen einladen
2. Kenntnisse
Beispiel: Mit der einschlägigen pharmakologischen Literatur vertraut sein
3. Ziele
Beispiel: Das Image des Unternehmens in der Öffentlichkeit verbessern
4. Ergebnisse
Beispiel: Den Marktanteil eines eingeführten Produkts erhöhen
5. Verhalten
Beispiel: Die Wirkungsweise neuer Präparate prägnant erklären
6. Fähigkeiten/Eigenschaften
Beispiel: Kontaktfähigkeit

Prüfung der Zusammenhänge zwischen den verschiedenen Komponenten. Dass Gesamtleistung erfasst wird, ist allein durch die Plausibilität der Leistungsmaße keineswegs sichergestellt: Die Validität (vereinfacht gesagt, die Richtigkeit) auch gut ausgearbeiteter Verfahren der Leistungsbeurteilung liegt im Durchschnitt nicht höher als die Validität guter einzelner Methoden der Personalauswahl, ihre Objektivität (gemessen als Übereinstimmung verschiedener Beurteiler) liegt unter der Objektivität der meisten Auswahlverfahren (s. Kapitel 3 in diesem Band).

Welches der Beurteilungsmaße das jeweils richtige ist, hängt vor allem von der Zielsetzung und von der Registrierbarkeit ab. Zum Beispiel sind Tätigkeiten zumeist konkret beschreibbar, und es ist mit hoher Übereinstimmung zwischen Beurteilern beobachtbar, ob jemand diese Tätigkeit ausführt oder nicht. (Bei der Frage allerdings, wie gut er sie ausführt, ist die Übereinstimmung bereits geringer.) Vergleiche zwischen Mitarbeitern lassen sich damit jedoch nur an gleichen Arbeitsplätzen ziehen, auch ist der Bezug zur Gesamtleistung umso fraglicher, je kleiner das Element ist.

Gewissermaßen der Gegenpol auf der Abstraktionsdimension sind globale Eigenschafts- oder Ergebniskonstrukte wie „Leistungsbereitschaft“, „Verlässlichkeit“ oder „Arbeitsqualität“. Sie kommen dem Gesamtkriterium schon recht nahe. Ihr Einsatzbereich sind v. a. Beurteilungen, die als Grundlage von Entscheidungen mit Prognosecharakter dienen, also z. B. vor Beförderungen. Hier fordert auch die meist nur teilweise Bekanntheit der künftigen Anforderungen die Einstufung des „Potenzials“ für Weiterentwicklung und künftige Leistung. Der Vorteil eigenschaftsbezogener Beurteilungen liegt in ihrem hohen Allgemeinheitsgrad und der dadurch möglichen Vergleichbarkeit über verschiedene Arbeitsplätze hinweg. Nachteile sind die geringe begriffliche Eindeutigkeit und die Anfälligkeit für Urteilstendenzen (v. a. Halo- oder Überstrahlungseffekt und Beschönigungstendenz, vgl. Kapitel 3). Diese Fehleranfälligkeit kann eher in standardisierten

Beurteilungssituationen mit mehreren Beurteilern (z. B. im Assessment Center) ausgeglichen werden als in einer Einschätzung durch einen einzelnen Beurteiler.

Wo Ergebnisse nicht allgemein („Arbeitsqualität“), sondern konkret beurteilt werden („Anzahl der Kundenreklamationen“), ist die Beurteilung einfacher und oft objektiver und sowohl auf individuellem Niveau als auch auf der Ebene aller anderen Einheiten möglich. Die Erfassbarkeit der tatsächlich relevanten Leistungsindikatoren ist allerdings vorausgesetzt („Pünktlichkeit“ mag objektiv erfassbar sein, aber von Nutzen als Kriterium ist sie nur, wenn sie tatsächlich einen Erfolgsbeitrag leistet.). Ergebnisbezogene Beurteilungen haben vor allem den Vorteil, die offensichtlichste Repräsentation des Konstrukts „Gesamtleistung“ darzustellen, auch wenn sie immer nur Ausschnitte hiervon zeigen. Sie können je nach Bedarf kalibriert werden (z. B. „Gewinn der Geschäftsstelle“ als große Einheit, „Fehlzeiten der Mitarbeiter“ als kleinere), sie lassen den Beurteilten Verhaltensspielraum, das Ziel mit selbstgewählten Mitteln zu erreichen, und sie fördern ihr Kosten-Nutzen-Denken.

In der Praxis allerdings sind Ergebnisbeurteilungen zumeist schwieriger, als es den Anschein hat. Inwieweit das Ergebnis auf die einzelne Person oder Gruppe zurückgeht und nicht anderen Einflussbedingungen zuzuschreiben ist, lässt sich oft nicht zufriedenstellend klären. Zudem läuft man Gefahr, Zähl- und Messbares überzubewerten auf Kosten des wirklich Wichtigen, aber weniger eindeutig Feststellbaren. Dadurch besteht die Gefahr der dysfunktionalen Verhaltenssteuerung (z. B. Verkaufen um jeden Preis statt langfristig orientierter Pflege der Kundenkontakte). Vergleiche zwischen verschiedenen Arbeitsplätzen sind nur sehr eingeschränkt möglich. Schließlich eignet sich die ergebnisbezogene Beurteilung für Zielsetzungsgespräche eigentlich nur dann, wenn die „Übersetzung“ der Leistungsziele in zielführendes Verhalten gelingt, denn nur dann sind Verbesserungen zu erwarten und können Förderungsmaßnahmen an die Beurteilung angeschlossen werden.

Damit ergibt sich die Einschätzung des *Verhaltens* als Kernbereich der Leistungsbeurteilung. Voraussetzung ist allerdings die tatsächliche Beobachtbarkeit oder zumindest Erschließbarkeit des Arbeitsverhaltens sowie die Kenntnis von Verhaltens-Erfolgs-Zusammenhängen. Die Vorteile verhaltensbezogener Beurteilung liegen vor allem im direkten Anforderungsbezug und in der relativen Zugänglichkeit für Selbst- und Fremdbeeinflussung. Damit bietet sie sich als Basis von Feedback, Beratung, Unterstützung und weiteren Maßnahmen der Personalentwicklung an.

Doch auch die potenziellen Nachteile der Verhaltensbeurteilung dürfen nicht übersehen werden. Hierbei ist zu nennen, dass sie häufig nur sprachlich kaschierte Eigenschaftsbeurteilungen darstellen („arbeitet zuverlässig“). Dies ist nicht nur auf Nachlässigkeit bei der Konstruktion des Beurteilungsverfahrens zurückzuführen, sondern entspricht durchaus unseren Urteilsgewohnheiten: Beurteiler bilden sich nicht einen Gesamteindruck von einer Person, indem sie akkurate Verhaltensbeobachtungen sammeln und daraus das Urteil synthetisieren, sondern, gerade umgekehrt, indem sie von einem Globaleindruck ausgehen und sich bei Verhaltenseinstufungen fragen „Wäre es für diese Person typisch, sich so oder so zu verhalten?“ (Feldman, 1998, vgl. Kapitel 3). Dementsprechend sind Verhaltensbeurteilungen im Prinzip ebenso anfällig für Beurteilungstendenzen wie Eigenschaftsbeurteilungen; sie können allerdings durch exakte Skalendefinition und die Beschränkung auf kleine Beobachtungseinheiten verbessert werden – wodurch allerdings andererseits der Bezug zum Gesamtkriterium und die Beobachtbarkeit leiden. Vielleicht sollte noch ergänzt werden, dass Ratschläge bezüglich des angemessenen Verhaltens zwar

bei gewissen Tätigkeiten und speziell in Lernphasen immer ihren Platz haben, in anderen Fällen – vor allem bei erfahrenen Personen, die komplexe Tätigkeiten ausüben – aber auch als „erzieherisch“ und anmaßend empfunden werden können.

Es gibt also kein Leistungsmaß, von dem sich sagen ließe, dass es in allen Fällen ideal sei. Will man ein verlässliches Bild von der Leistung oder Leistungsfähigkeit einer Person oder Gruppe erzielen, wird man meist nicht umhinkommen, auf verschiedenen Ebenen mehrere Kriterien zu sammeln und zu kombinieren (Schuler & Schmitt, 1987). In mehreren Kapiteln dieses Bandes werden Vorschläge und Beispiele vorgebracht, wie verschiedene „Modalitäten“ des Beurteilens zu einem Beurteilungssystem integriert werden können, das auch einer Vielfalt von Zielen gerecht wird. Eine weitergehende Erörterung der Vor- und Nachteile eigenschafts-, verhaltens- und ergebnisbezogener Beurteilungen findet sich bei Schuler und Marcus (2004).

1.3.3 Beurteilungsverfahren

Die Beurteilung beruflicher Leistung ist nicht zwingend an die Verwendung formaler Skalierungsverfahren gebunden. In vielen, vor allem kleineren Organisationen begnügt man sich mit informeller Einschätzung der Qualifikation der Mitarbeiter, die einen gewissen expliziten Charakter oft erst zu dem Zeitpunkt bekommt, zu dem Personalentscheidungen anstehen. In manchen Fällen wird auch bewusst auf formelle Beurteilungen verzichtet, weil man deren psychometrische oder sozialpsychologische Probleme fürchtet. Zumindest die psychometrischen Probleme werden durch intuitive Bildung und Verwertung des Eindrucks natürlich nicht vermieden, indem man sie sich nicht vor Augen führt und dadurch auch nicht kontrollieren kann.

Was Probleme des Miteinander-Umgehens betrifft, kann allerdings die unbedachte, ungeschickte oder gar rücksichtslose Verwendung formeller Verfahren tatsächlich Schäden anrichten, die den Nutzen der Beurteilung überwiegen. Es ist wichtig, sich als Entscheidungsträger und Beurteiler vor Augen zu führen, dass der Einsatz auch des technisch ausgefeiltesten Verfahrens die Verwender nicht der Verantwortung zu faires, „sozial validen“ Vorgehensweisen enthebt. Bei Dickinson (1993) und im Kapitel Beurteilungsgespräche werden diese Probleme eingehender erörtert.

Um höheren Ansprüchen an die Reliabilität und die Validität der Beurteilungen zu genügen, bedient man sich zumeist gebundener Urteilstechniken in Form von Skalierungsmethoden. Diese Skalierungsmethoden sind von den im vorigen Abschnitt besprochenen Beurteilungsebenen weitgehend unabhängig; im Einzelfall legt die Methode gleichwohl nahe, sich z. B. eher auf Verhaltensweisen, auf Ergebnisse oder auch auf Eigenschaften als Einheiten zu beziehen. Über die Verbreitung formalisierter Beurteilungsverfahren informieren Liebel und Oechsler (1987).

Objektive Leistungsmaße

Objektiven Daten als Indikatoren von Leistungsergebnissen werden verschiedene Vorzüge gegenüber anderen Beurteilungsformen zugesprochen, darunter die, geeignete Repräsentationen der Leistungsziele darzustellen, von den Unzulänglichkeiten subjektiver Einschätzung nicht beeinträchtigt zu sein und Affinität zur betriebswirtschaftlichen Nutzen-Kosten-Kalkulation aufzuweisen. Allerdings sind auch ihre Nachteile nicht zu über-

sehen, wobei vor allem Attributionsprobleme (Inwieweit ist ein Mitarbeiter tatsächlich für das Leistungsergebnis verantwortlich?), Reliabilitätsdefizite (Würde die Leistungsermittlung zu einem anderen Zeitpunkt genauso ausfallen?) und häufig unbefriedigende Relevanz (Aussagekraft) ins Auge fallen. Dessen ungeachtet kommen objektive Daten, wo erhebbar, zumindest im Sinne erstrebenswerter Multimodalität als ergänzende Leistungsinformation in Betracht.

Viswesvaran (2002) klassifiziert im Anschluss an Schmidt objektive Leistungsmaße in *Produktivitätsdaten* als direkte Maße und *Personaldaten* als indirekte Maße. Produktivitätsdaten können quantitativ (etwa Verkaufszahlen oder Anzahl produzierter Einheiten) oder qualitativ (z. B. Anzahl und Bedeutsamkeit von Fehlern) ausgeprägt sein. In manchen Fällen enthalten Produktivitätsdaten sowohl Quantitäts- als auch Qualitätsinformation, etwa als Anzahl akzeptierter Patente oder Veröffentlichungen in angesehenen (also ihrerseits nach Qualität auswählenden) Fachzeitschriften. Personaldaten sind ihrerseits keine direkten Produktivitätsmaße, sondern lassen lediglich Schlüsse auf Produktivität zu (z. B. Fehlzeiten oder Unfälle) oder sind Konsequenzen von Leistungseinschätzungen (etwa Beförderungen oder Gehalt).

Eine Studie von Sackett, Zedeck und Fogli (1988) machte auf den geringen Zusammenhang zwischen objektiven und subjektiven Leistungsmaßen sowie zwischen verschiedenen objektiven Leistungsmaßen untereinander aufmerksam. Das Interesse dieser Studie richtete sich primär auf die Übereinstimmung zwischen maximaler und typischer Leistung von Supermarktkassiererinnen (Abfertigung eines Warenkorbs nach Ankündigung bzw. ohne Ankündigung einer Bewertung). Selbst nach Reliabilitätskorrektur betrug die Korrelation in der Gruppe der Neueingestellten nur $r = .16$, bei den seit längerem Beschäftigten $.36$. Der Zusammenhang zwischen Geschwindigkeit und Genauigkeit lag zwischen $r = .02$ und $.19$. Überdies war auch die Übereinstimmung der Vorgesetztenurteile mit den objektiven Leistungsmaßen gering und betrug für die verschiedenen Kombinationen von typisch-maximal und Geschwindigkeit-Genauigkeit zwischen $.01$ und $.36$. Damit weichen die Ergebnisse von Sackett et al. (1988) nicht wesentlich vom gewichteten Durchschnittswert ab, den Heneman (1986) mit $r = .27$ errechnete, und belegen ein weiteres Mal das Erfordernis multimodaler Messungen, soweit das Ziel nicht in der Messung spezifischer Leistungsfacetten, sondern in der „Gesamtleistung“ besteht.

Freie Eindruckschilderung

Die geringste methodische Gebundenheit liegt bei der sprachlich freien Schilderung beurteilungsrelevanter Aspekte vor. Freie Eindruckschilderungen beziehen sich vor allem auf Eigenschafts- und Verhaltenskriterien; insbesondere Führungskräfte werden häufig auf diese Weise beurteilt. Die wichtigste Funktion freier Formulierungen dürfte in Einsatzbereichen liegen, wo es nicht entscheidend auf die Vergleichbarkeit der Beurteilungsergebnisse ankommt – vor allem im Rahmen individueller Beratung und Förderung von Mitarbeitern. Soweit die freie Eindruckschilderung auch als „Potenzialeinschätzung“ verstanden wird, also als Äußerung über die mutmaßliche Qualifikation für andere Positionen oder die Entwicklungsfähigkeit, fällt die geringe Reliabilität dieser Form der Beurteilung ins Gewicht. Um ein Mindestmaß an Beurteilerübereinstimmung zu gewährleisten, können Urteilsaspekte oder Auswahllisten als Formulierungshilfen angeboten werden. Vielfach, z. B. in weiten Bereichen des öffentlichen Dienstes, wird Beurteilern

zur Ergänzung der vorzunehmenden Skalierungen die Möglichkeit freier Ergänzungen gegeben. Zuweilen ist die freie Eindrucksschilderung auch gekoppelt mit Gesprächs- und Zielsetzungsverfahren.

Ein Vorteil der freien Eindrucksschilderung wird darin gesehen, den Konstruktionsaufwand von Skalierungsverfahren zu vermeiden. Eine weitere, von den Verwendern als Vorteil betrachtete Eigenheit ist die Möglichkeit, jedem Mitarbeiter durch individuelle Charakterisierung und in Betonung der für den Beurteiler bedeutsamen Urteilsaspekte gerecht werden zu können, sich nicht einem Schema vorgegebener Aspekte und Abstufungen beugen zu müssen. Ihr wichtigster Nachteil ist ihre geringe Reliabilität, u. a. ihre Abhängigkeit von der Stimmungslage des Beurteilers und der gerade erinnerten Verhaltensereignisse und Eindrücke. Auch vom Zeitaufwand und Ausdrucksvermögen des Beurteilers ist die freie Eindrucksschilderung abhängig, was die Objektivität (Übereinstimmung verschiedener Beurteiler) beeinträchtigt (vgl. Kapitel „Qualität“). Durch Training und Absprache zwischen den Beurteilern können diese Probleme gemindert werden.

Einstufungsverfahren

Die weitaus meisten der in Wirtschafts- und Verwaltungsorganisationen verwendeten systematischen Beurteilungsverfahren können der Kategorie der Einstufungsverfahren zugerechnet werden (Cascio, 1987). Methodisches Prinzip ist die Zuordnung – „Einstufung“, engl. „rating“ – von Verhaltensbeobachtungen, Ergebnis- oder Merkmalseinschätzungen zu einer mehrstufigen Skala. Die Skalenstufen sind gewöhnlich durch Zahlenwerte, Adjektive, Adverbien oder Verhaltensbeschreibungen verankert. Die Skalenlänge beträgt üblicherweise zwischen 5 und 9 Stufen, was als Reliabilitätsoptimum angesehen wird (Bernardin & Beatty, 1984). Um alle wichtigen Facetten des Leistungsbereichs abzudecken, werden etwa 5 bis 20 Skalen dieser Art verwendet. Einstufungsverfahren können, wie auch die meisten anderen der im Folgenden beschriebenen Methoden, prinzipiell für alle Tätigkeitsbereiche und Positionsebenen eingesetzt werden.

Ein Beispiel für eine einfache Skalenform, die sogenannte *Graphische Einstufungsskala* (Bezeichnung, wie auch einige weitere Skalenbezeichnungen, nach Brandstätter, 1970), wird in Abbildung 1 gegeben.

Die verbale Verankerung der Skalenpunkte trägt zur Verringerung von Urteilstendenzen bei, wie sie bei Einstufungsskalen besonders stark auftreten (z. B. unterschiedliche Mittelwertstendenzen, also verschiedene „Strenge“ oder „Milde“ beim Beurteiler; zur Definition und Illustration von Urteilstendenzen vgl. Kapitel 3).

Kontaktfähigkeit						
1	2	3	4	5	6	7
ist im Umgang mit anderen befangen und ängstlich		ist unkompliziert und wird von anderen akzeptiert			findet selbst zu schwierigen Menschen leicht Kontakt	

Abbildung 1: Beispiel für eine Graphische Einstufungsskala

Welche Bedeutung hat die Verhaltensdimension „INNOVATION“ (z. B. Neuerungen erkennen, aufgreifen und umsetzen) für die Leistung Ihres Mitarbeiters					
0	1	2	3	4	5
keine	geringe		mittlere		hohe

Bitte bewerten Sie anhand der Verhaltensbeispiele für „INNOVATION“ die Leistungen Ihres Mitarbeiters in dieser Dimension.														
<p>Einschätzung (bitte ankreuzen)</p>	<p>Verhaltensbeispiele</p>													
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center; border: 1px solid black; width: 40px; height: 30px; margin-bottom: 5px;">9</td> <td rowspan="3" style="border: none; padding-left: 10px; vertical-align: middle;"> überdurchschnittliche Leistung </td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">8</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">7</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">6</td> <td rowspan="3" style="border: none; padding-left: 10px; vertical-align: middle;"> durchschnittliche Leistung </td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">5</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">4</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">3</td> <td rowspan="4" style="border: none; padding-left: 10px; vertical-align: middle;"> unterdurchschnittliche Leistung </td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">2</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px; margin-bottom: 5px;">1</td> </tr> <tr> <td style="text-align: center; border: 1px solid black; height: 30px;">1</td> </tr> </table>	9	überdurchschnittliche Leistung	8	7	6	durchschnittliche Leistung	5	4	3	unterdurchschnittliche Leistung	2	1	1	<ul style="list-style-type: none"> – bringt selbst kreative Vorschläge und Beispiele aus verschiedenen Bereichen vor – überträgt neue Inhalte aus Literatur oder Vorträgen auf eigene Aufgaben – findet sich sehr schnell in neuen Fachgebieten zurecht – erkennt Vorteile und Nutzen von neuen Entwicklungen für seine Aufgaben <ul style="list-style-type: none"> – kann neue Inhalte vermitteln und anwenden – ist Entwicklungen und Neuerungen gegenüber aufgeschlossen – informiert sich laufend über neue Aspekte seiner Aufgaben – ist an Neuerungen interessiert, spricht aber nur nach Aufforderung über diese Aspekte <ul style="list-style-type: none"> – verwendet herkömmliche Lösungswege ohne Bereitschaft zur Neuerung – beteiligt sich nicht an Fachgesprächen über Neuentwicklungen/Innovationen – steht neuen Ideen reserviert gegenüber – gibt bei ungewöhnlichen Lösungsansätzen schnell auf
9	überdurchschnittliche Leistung													
8														
7														
6	durchschnittliche Leistung													
5														
4														
3	unterdurchschnittliche Leistung													
2														
1														
1														

Abbildung 2: Beispiel für eine Verhaltensverankerte Einstufungsskala (aus Schuler, Funke, Moser & Donat, 1995, S. 64)

Eine Form der Einstufungsskala, mit der man Urteilstendenzen zu vermindern sucht, ist die *Verhaltensverankerte Einstufungsskala*. Sie geht auf Smith und Kendall (1963) zurück. Bei diesem Verfahren werden zu jedem Urteilsaspekt positive, wertneutrale und negative Verhaltensbeispiele zusammengestellt und auf relative Eindeutigkeit des Skalenwerts vorgeprüft. Diese Verhaltensbeispiele, die gewöhnlich aus Anforderungsanalysen stammen, werden als Markierungen verschiedener Skalenpositionen vorgegeben und dienen als beispielhafte Verankerungen für das einzustufende Verhalten. Ein Beispiel wird in Abbildung 2 dargestellt, es bezieht sich auf Mitarbeiter im Bereich Forschung und Entwicklung (aus Schuler et al., 1995, S. 64).

Der Vorteil Verhaltensverankerter Einstufungsskalen wird vor allem im Anforderungsbezug und in der deutlichen Verhaltensorientierung gesehen, die sie zu einer brauchbaren Grundlage für Personalentwicklungszwecke machen. Der Konstruktionsaufwand für ein solches Verfahren ist allerdings relativ hoch (Schilderung der Konstruktion z. B. bei Domsch & Gerpott, 1985).

Einfacher aufgebaut sind *Verhaltensbeobachtungsskalen*, die von Latham und Wexley (1977) vorgeschlagen wurden. Sie basieren auf dem sogenannten Likert-Format und umfassen fünf Skalenstufen, deren Extremausprägungen mit adverbialen Häufigkeitsbezeichnungen markiert sind. Grundsätzlich soll bei diesem Verfahren nur beobachtbares Verhalten eingestuft werden. Die relevanten Aussagen werden auf arbeitsanalytischem Weg gewonnen, zumeist mittels der Methode der Kritischen Ereignisse nach Flanagan (1954). Die einzelnen Skalen werden nach teststatistischen Prinzipien geprüft und meist nach Leistungsdimensionen geordnet (zur Skalenentwicklung siehe Latham, Fay und Saari, 1979). Je nach Zielsetzung werden die Einzelwerte nach Dimensionen summiert oder zu einem Gesamtwert zusammengefasst, der durch die große Auswahl von Einzel-skalen (etwa 50) relativ reliabel ist. Beispiele für Verhaltensbeobachtungsskalen sind in Abbildung 3 dargestellt (verändert nach Latham & Wexley, 1981, S. 210).

Kommt pünktlich zur Arbeit						
fast nie	1	2	3	4	5	fast immer
Sagt im Gespräch mit Gästen „bitte“ und „danke“						
fast nie	1	2	3	4	5	fast immer
Hält die Aschenbecher sauber						
fast nie	1	2	3	4	5	fast immer
Vermeidet Klatsch über das Privatleben der Kollegen						
fast nie	1	2	3	4	5	fast immer
Fragt die Gäste, ob sie mit allem zufrieden sind						
fast nie	1	2	3	4	5	fast immer

Abbildung 3: Verhaltensbeobachtungsskalen

Die Konstruktion von Verhaltensbeobachtungsskalen ist nicht ganz so aufwendig wie die von Verhaltensverankerten Einstufungsskalen, die Handhabung ist weniger erklärungsbedürftig. Die testtheoretische Datenqualität (Reliabilität und Validität) ist ungefähr die gleiche. Bezüglich der Akzeptabilität werden für Verhaltensbeobachtungsskalen positivere Werte berichtet als für Graphische Einstufungsskalen (Tziner, 1986).

Kennzeichnungs- und Auswahlverfahren

Anders als bei den meisten Einstufungsverfahren werden bei Kennzeichnungs- und Auswahlverfahren die Verhaltensaussagen nicht nach Beurteilungsdimensionen geordnet, sondern gemischt vorgegeben. Dadurch sollen Halo-(Überstrahlungs-)Effekte vermindert werden. Auch sind den Beurteilern gewöhnlich die vorgeprüften Skalenwerte der Einzelaussagen unbekannt, wodurch Mittelwertstendenzen verringert werden sollen.

Die *Gemischte Aussagenliste mit freier Wahl* (check list) verlangt vom Beurteiler die Kennzeichnung, ob die Einzelaussagen auf die beurteilte Person zutreffen oder nicht. Bei vorbestimmtem Skalenwert kann aus den Kennzeichnungen ein Gesamtwert ermittelt werden. Ein Beispiel wird in Abbildung 4 dargestellt (verändert nach Knauff, 1948, zit. aus Bernardin & Beatty, 1984, S. 67).

Die Möglichkeit, einen Index der Urteilsqualität für einzelne Beurteilungen zu bestimmen, bietet ein Verfahren dieser Art dann, wenn es wie die *Mixed Standard Scale* (Banz & Ghiselli, 1972) nach dem Prinzip einer hierarchisch eindeutig geordneten Skala (Guttman-Skala) aufgebaut ist. Abweichungen von der Aussagen-Hierarchie werden dann als Indikatoren für mangelhafte Urteilsqualität interpretiert.

	trifft zu	trifft nicht zu	
Die Schaufenster sind attraktiv gestaltet	<input type="checkbox"/>	<input type="checkbox"/>	(8,5)
Manchmal werden Brötchen verkauft, die auf den Boden gefallen sind	<input type="checkbox"/>	<input type="checkbox"/>	(1,4)
Lässt sich Arbeit von seinen Bäckern rückdelegieren	<input type="checkbox"/>	<input type="checkbox"/>	(3,1)
Ermutigt seine Mitarbeiter zur Initiative	<input type="checkbox"/>	<input type="checkbox"/>	(8,1)
Hat brauchbare Formulare entwickelt	<input type="checkbox"/>	<input type="checkbox"/>	(6,4)

Abbildung 4: Gemischte Aussagenliste mit freier Wahl (in Klammern stehen die dem Beurteiler unbekanntes Aussagengewichte)

Die *Gruppierte Aussagenliste mit Wahlzwang* (forced-choice) hat vor allem zum Ziel, Beschönigung der Urteile zu vermeiden. Der Beurteiler wählt zwischen zwei oder mehreren Aussagen, die gleich günstig erscheinen, sich aber für den Arbeitserfolg als unterschiedlich bedeutsam herausgestellt haben. Bei der Auswertung ergibt sich ein Qualifikationswert aus der Gewichtung der einzelnen Aussagen. Ein Beispiel gibt Tabelle 3 (nach Berkshire & Highland, zit. nach Bernardin & Beatty, 1984, S. 98).

Tabelle 3: Gruppierete Aussagenliste mit Wahlzwang (die Werte sind den Beurteilern nicht bekannt)

Aussage	Diskriminationsindex	Bevorzugungsindex
1. Hat Geduld mit langsam Lernenden	1,15	2,82
2. Zeigt beim Unterrichten Selbstvertrauen	0,54	2,75
3. Weckt Interesse und Aufmerksamkeit bei den Schülern	1,39	2,89
4. Teilt die Lernziele der nächsten Stunde im Voraus mit	0,79	2,85

Die Konstruktion eines Wahlzwangverfahrens, das von Wherry initiiert und von Sisson (1948) erstmalig geprüft wurde, erfolgt nach psychometrischen Prinzipien und ist relativ aufwendig. Die Einsatzmöglichkeit zur Beurteilung durch Vorgesetzte ist beschränkt auf Grund relativ geringer Akzeptanz bei Beurteilern wie Beurteilten und auf Grund der Verwendungslogik (der Beurteiler darf die Erfolgsrelevanz der beschriebenen Verhaltensweisen nicht kennen). Für Beurteilungen durch Außenstehende (Kunden, Elternsprecher) oder für Selbstbeurteilungen könnte sich das Verfahren dagegen besser eignen.

Mit der Zielsetzung, die Objektivität von Beurteilungen zu erhöhen, indem sie an tatsächliche Vorkommnisse geknüpft werden, wurde ein Beurteilungsverfahren entwickelt, das auf Flanagans Technik der Kritischen Ereignisse basiert (Flanagan, 1954). Hierbei werden Verhaltensweisen beobachtet und registriert, die erkennbar zu Erfolg oder Misserfolg im Arbeitsvollzug führen. Als Beurteilungsverfahren spielt diese Methode heute allerdings keine Rolle und wird deshalb nicht näher dargestellt, wohingegen Flanagans Prinzip der Kritischen Ereignisse als Verfahren der Anforderungsanalyse große Verbreitung gefunden hat, gerade auch bei der Ausarbeitung von Beurteilungsverfahren.

Rangordnungsverfahren

Hauptintention bei der Verwendung von Rangordnungsverfahren ist die Differenzierung zwischen Personen (oder Gruppen etc.), die bei Einstufungsverfahren nicht gewährleistet ist. Auch hat sich gezeigt, dass Rangordnungen relativ reliabel sind und Gesamtbewertungen deutlich zum Ausdruck bringen. In manchen Bereichen, z. B. im Verkauf, neuerdings auch zur Bewertung von Hochschulen, werden sie als Motivationsinstrument eingesetzt.

Rangreihen können als *direkte Rangreihe* oder auf dem Wege des *Paarvergleichs* gebildet werden. Die Bildung einer Rangreihe kann prinzipiell als Gesamtbewertung oder in Aufschlüsselung nach einzelnen Leistungsdimensionen erfolgen. Beispiele werden in Abbildung 5 dargestellt.

Bei „subjektiven“ Urteilen ist die Menge der reliabel einzustufenden Personen auf eine überschaubare Anzahl begrenzt, nicht dagegen bei Vorliegen „objektiver“ Leis-