

Systeme zur Leistungsbeurteilung

Wenn berufliche Leistungsbeurteilungen zu positiv ausfallen, wird meistens den Führungskräften die Schuld daran gegeben. Der Vorwurf lautet dann, sie könnten nicht richtig beurteilen, seien zu milde oder scheuten davor zurück, schlechte Beurteilungen auszusprechen. In der Regel liegt es aber nicht an den Vorgesetzten, sondern an den eingesetzten Methoden. Die Autoren wollen mit ihrem Beitrag den Blick für die Herausforderungen einer soliden Leistungsbeurteilung schärfen und Lösungsmöglichkeiten aufzeigen.

Andreas Frintrup,
Dipl.-Ökonom, ist
Geschäftsführer der
S&F Personalpsychologie
Managementberatung
GmbH und Vorstand
der HR Diagnostics AG
in Stuttgart.



Prof. Dr. Heinz Schuler
hat den Lehrstuhl für
Psychologie an der
Universität Hohenheim
inne. Er forscht zur
Berufseignungsdiag-
nostik sowie zur Beur-
teilung und Förderung
beruflicher Leistung.



Die Forschung zur Leistungsbeurteilung hat in den vergangenen Jahrzehnten umfangreiche Fortschritte gemacht (vgl. Schuler 2004). Neben der Entwicklung optimierter Instrumente wurde zum Beispiel erkannt, dass in der betrieblichen Praxis drei Ebenen der Beurteilung und Rückmeldung unterschieden werden sollten.

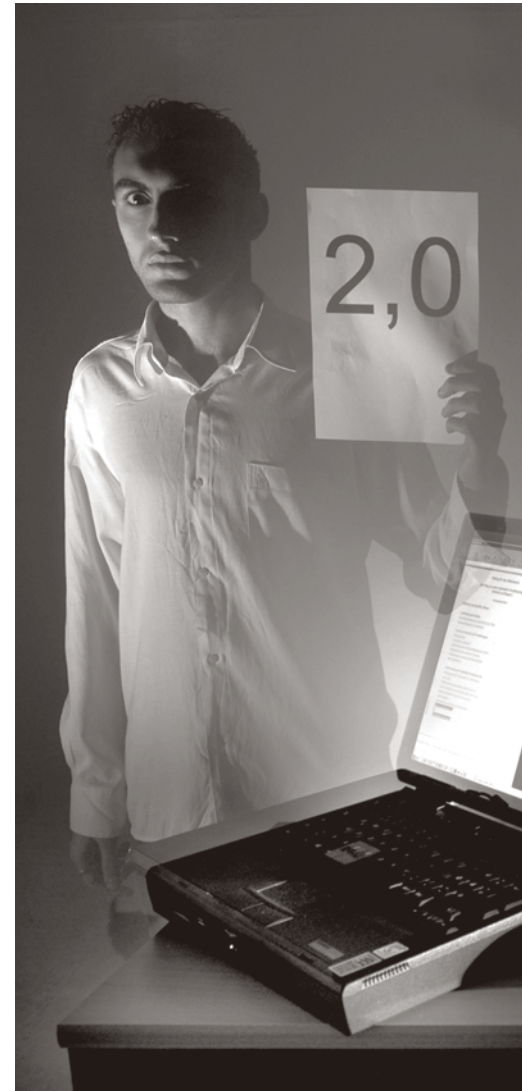
Day-to-Day-Feedback bezeichnet eine zeitlich unmittelbare positive und negative Leistungs- und Verhaltensrückmeldung an Mitarbeiter. Umso zeitnäher und konkreter eine Rückmeldung erfolgt, desto stärker sind die verhaltens- und leistungssteuernden Wirkungen und die Akzeptanz. Day-to-Day-Feedback kann daher auch als unmittelbar wirksame Maßnahme der Personalentwicklung verstanden werden.

Regelbeurteilungen sind regelmäßig, zum Beispiel jährlich wiederkehrende Beurteilungen der Leistung und des Verhaltens in der jeweils zurückliegenden Periode. Ihre Ziele sind Gehalts- und Bonusfindung sowie die individuelle Ableitung von Maßnahmen der langfristigen Personalentwicklung.

Potenzialbeurteilungen nutzen Prozesse und Methoden für eine zukunftsorientierte Einschätzung weitergehenden Potenzials für komplexere, anforderungsreichere Tätigkeiten.

Beförderungsentscheidungen zum Beispiel sollten nicht allein auf Basis der Beurteilung vergangener Leistungen erfolgen. Eine Potenzialaussage für eine Position mit höherer Verantwortung wäre auf dieser Informationsbasis nur bedingt möglich. Dafür bedarf es gesonderter Prozesse der Potenzialbeurteilung und -analyse, die die erforderlichen Kompetenzen der Zielposition zum Gegenstand haben. Freilich darf hierbei die bisherige berufliche Bewährung nicht unbeachtet bleiben.

Ebenso falsch wäre es, die alltäglichen Erfahrungen im Umgang mit Mitarbeitern nicht unmittelbar für die Führungsaufgabe



zu nutzen. Vorgesetzte, die Verhaltenseindrücke über lange Zeit sammeln und erst zu einem wesentlich späteren Zeitpunkt geballt zurückmelden, ernten dafür regelmäßig Reaktanz und eine geringe Umsetzungsbereitschaft. Ein in der Praxis geübtes Day-to-Day-Feedback dagegen erleichtert die Führungsarbeit.

Wenn zu gute Beurteilungen zum Problem werden



Schwierig und unbeliebt

bei vielen Vorgesetzten sind Leistungsbeurteilungen. Die Ursache für auffällig viele gute Bewertungen ist nach Ansicht der Autoren aber keineswegs immer zu große Milde.

gute Gesprächsführung auffangen. Daher sollte die Regelbeurteilung nicht mit Zielen überfrachtet werden.

Die Funktion, die die Regelbeurteilung am besten erfüllen kann, ist die Ex-post-Evaluation individueller Leistungsbeiträge und die Herstellung eines interpersonellen Vergleichs für die genannten Entscheidungszwecke. Sie ist nur nachrangig als Instrument der Personalentwicklung geeignet, weil das Gespräch über Entwicklungsbedarfe und das erfolgreiche Umsetzen von Maßnahmen zur Personalentwicklung unabdingbar auf die Offenheit und Mitwirkung der betroffenen Person angewiesen sind. Die Entwicklung einer funktionalen und zugleich akzeptablen Lösung für die Beurteilung beruflicher Leistung bedarf daher zunächst der Definition damit verbundener Ziele und Wirkungsweisen.

Systeme der Leistungsbeurteilung erfüllen vielfältige organisationale Funktionen: Sie kommunizieren berufliche Anforderungen an alle Stelleninhaber und wirken so verhaltenssteuernd, sie ermöglichen eine leistungsgerechte und faire Vergütungsstruktur, die Evaluation von Personalauswahl, Personalentwicklung sowie die zielgerichtete Weiterentwicklung der personellen Ressourcen eines Unternehmens. Sie machen das aktuelle Leistungsvermögen der Mitarbeiter auf aggregierter Ebene sichtbar und erlauben

Ziele und Funktionen der Regelbeurteilung

Auf Basis der Regelbeurteilung werden häufig wichtige personelle Entscheidungen getroffen – Gehaltsfindung und Eingruppierung, Bonusvergabe, Übernahme oder Trennung von Mitarbeitern. Naturgemäß beeinflussen die aus einer Bewertung und

die daraus zu erwartenden Folgen die Offenheit in einer solchen Gesprächssituation. Wenn Mitarbeiter fürchten müssen, dass das Eingestehen persönlicher Schwächen und Entwicklungsbedarfe mit Nachteilen einhergeht, wird ihre Bereitschaft zu einem offenen Feedbackgespräch enge Grenzen haben. Das lässt sich nur bedingt durch

einen personen- und bereichsübergreifenden Vergleich. Leistungsbeurteilungen sind aber auch eine fundierte Basis für individuelle Personalentscheidungen. Aus unternehmerischer Perspektive sind besonders die Verhaltens- und Leistungssteuerung von großer Bedeutung.

Gleichwohl ist das Durchführen von Leistungsbeurteilungen eine schwierige und unbeliebte Führungsaufgabe. Es gilt dabei, das Verhalten von Mitarbeitern zu bewerten und das Ergebnis der Bewertung mit ihnen zu besprechen. Täte man das nicht, fiel die verhaltens- und leistungssteuernde Funktion der Beurteilung gänzlich weg. Mitarbeiter stehen dem Prozess in der Regel offener gegenüber als die beurteilenden Vorgesetzten, diese scheuen häufig Rückmeldungsgespräche. In der Praxis ist daher oft ein Trend zu milden und wohlwollenden Urteilen zu beobachten, was dazu führt, dass bei Verwendung von Beurteilungsskalen der Mittelwert der Beurteilungen in der Regel deutlich oberhalb der Skalenmitte liegt, also weit mehr als die statistisch zu erwartenden 50 Prozent der Beurteilten besser als der organisationale Durchschnitt bewertet werden. Dieser Effekt ist beständiger Gegenstand unternehmensinterner Diskussionen über die richtigen Beurteilungsgrundsätze und Methoden, denn wenn Beurteilungen im Durchschnitt zu gut ausfallen, resultieren daraus erhebliche Probleme für die Praxis des HR-Managements.

Durch eine zu geringe Differenzierung zwischen den Beurteilten versagen die Funktionen der Leistungsbeurteilung. Im Extremfall (z. B. bei sehr kurzen Beurteilungsskalen) werden große Anteile der Beurteilten gleich beurteilt. Das widerspricht der empirischen Evidenz, dass es selbst in hohen Leistungsbereichen noch erhebliche Leistungsvarianzen zwischen Mitarbeitern gibt, die sich mit zunehmender Größe der Stichprobe immer näher einer Normalverteilung annähert, und erschwert somit die differenzierte und leistungsgerechte Verteilung von Bonuszahlungen oder Funktionen der Gehaltsfindung und -differenzierung.

Weshalb zu gute Beurteilungen?

Zu gute Bewertungen können unterschiedliche Ursachen haben – sie können einerseits in der Person des Urteilenden liegen, aber auch rein methodisch bedingt sein. Beide Ursachen lassen sich durch den Einsatz eines probaten Beurteilungsinstruments und eines begleitenden Trainings gut beheben. Dazu muss jedoch genau verstanden werden, welche Umstände zu undifferenzierten und zu positiven Beurteilungen beitragen.

Neben offensichtlichen Effekten gibt es eher verdeckt und unbewusst ablaufende Prozesse bei der Beurteilung. Für die meisten Beurteilungstendenzen (sowohl im Sinne stark positiver oder negativer Verzerrungen) dürfte die Abhängigkeit der Beurteilung von dem Beurteilenden die wichtigste Erklärungsvariable sein. Jede Beurteilung sagt auch etwas über den Urteilenden selbst aus, und daher wägen Urteilende durchaus taktisch ab, welchen Effekt eine zu negative Beurteilung auf ihr Image als Vorgesetzter haben oder inwieweit eine künftige Arbeitsbeziehung durch eine zwar realistische, aber negative Bewertung getrübt werden könnte.

Noch ein anderes Kalkül ist weitverbreitet: Werden wirklich gute Mitarbeiter sehr positiv bewertet, empfehlen sie sich dadurch für höhere Aufgaben und den internen Arbeitsmarkt. Es liegt nicht ohne Weiteres im Interesse des jeweiligen Vorgesetzten, dass seine besten Mitarbeiter fluktuieren (auf gleichem Wege lassen sich allerdings auch ungeliebte Mitarbeiter wegloben).

Korrekte Beurteilungen

In wenigen Anwendungsfeldern der Personalpsychologie besteht eine größere Differenz zwischen dem, was wissenschaftlich sinnvoll ist, und dem, was in der Praxis geschieht, als auf dem Gebiet der Leistungsbeurteilung. Es gibt drei zentrale Anforderungen, die an ein glei-

chermaßen funktionales wie praktikables Beurteilungssystem zu stellen sind: 1. Anforderungsbezug und organisationale Passung, 2. methodische Tauglichkeit, 3. eine partizipative und transparente Einführung in die Organisation.

Generische Beurteilungssysteme lassen sich selten so anpassen, dass sie methodischen, akzeptanz- und qualitätsmäßigen Anforderungen einer spezifischen Organisation entsprechen. Sind sie erst einmal eingeführt, sind partielle Verbesserungen nur schwer möglich und führen nur in sehr seltenen Fällen zu einem tauglichen Gesamtsystem. Auch Akzeptanz lässt sich im Regelfall nur durch unmittelbare Partizipation der künftigen Beteiligten und Betroffenen bei der Entwicklung des Verfahrens sicherstellen.

Besonders schwierig wird es, wenn nachträgliche Korrekturen erforderlich sind, um zum Beispiel offenkundig zu gute Beurteilungsergebnisse zu korrigieren. Zunächst muss die konkrete Ursache der Urteilsverschiebungen zum positiven Ende der eingesetzten Bewertungsskala gefunden werden. Liegt es an dem verwendeten Skalenniveau oder den Verankerungen der Skalen, sind große Anstrengungen zur Optimierung erforderlich.

Wenn zum Beispiel eine untaugliche drei- oder vierstufige Beurteilungsskala ohne konkrete Verhaltensverankerung der Beurteilungsstufen vorliegt, stellt der Wechsel hin zu einer ungeraden Skalierung (das entspricht eher der Normalverteilung von Verhaltensausrägungen als geradzahlige Skalenstufen) mit fünf oder sieben Stufen eine deutliche Verbesserung der Skalenlänge dar. Außerdem sind dadurch bessere Reliabilitäts- und sogar Akzeptanzwerte zu erwarten. Wenn jedoch nicht gleichzeitig auch die Verhaltensverankerung angepasst wird, ist dieser Fortschritt so marginaler Art, dass die damit verbundenen Mühen wenig Fortschritt bringen werden.

Eine gute Beobachtungsskala soll den Beurteiler unterstützen, akkurate, mög-

lichst unverzerrte Beurteilungen ihrer Mitarbeiter (oder auch über sich selbst oder den eigenen Vorgesetzten) vorzunehmen. Effekte zu negativer oder zu positiver Beurteilungen im Sinne einer systematischen Verschiebung der abgegebenen Urteile zu dem einen oder anderen Ende der Skala gilt es zu vermeiden.

Sachgerechte Skalierung

Welche Voraussetzungen sind an eine geeignete Skalierung zu stellen? Eine sachgerechte Skalierung sollte

- Milde und Strenge vermeiden, insbesondere durch klare Verhaltens- oder Häufigkeitsverankerung der zu beurteilenden Kriterien.
- verhaltensorientiert sein, also konkretes vergangenes Verhalten der Person statt hypothetisches Verhalten erfassen. (Tut jemand etwas, oder meint der Urteilende nur, jemand könnte es tun? Beispiel einer schlechten Verankerung: „Kann komplexe Sachverhalte verständlich erklären“ – besser wäre die Formulierung: „Erklärt komplexe Sachverhalte verständlich“. „Ist nicht gewissenhaft“ ist ebenfalls eine schlechtere Beurteilungsverankerung als die konkretere Aussage „kommt häufig zu spät zu Terminen“. Die Einschätzung, nicht gewissenhaft zu sein, ‚stempelt‘ einen Mitarbeiter ab, statt ihm konkret zu zeigen, was besser zu machen wäre.)
- eine Skalenlänge zwischen fünf und neun Urteilsstufen umfassen. Diese Skalenlängen haben sich als messtechnisch überlegen erwiesen und weisen die beste Messgenauigkeit auf (Schuler 2004).
- eine ungerade Anzahl an Stufen mit einer klaren Mitte haben – also zum Beispiel fünf, sieben oder neun Stufen statt vier, sechs oder acht. Auch das erhöht die Reliabilität der Beurteilung und steigert maßgeblich die Akzeptanz, weil mittlere Leistungsträger nicht einem der dichotomen Extreme „schlechte Leistung“ oder „gute Leistung“ zugeordnet werden müssen. Ein Extremfall schlechter Skalierung ist die mittlerweile in der Praxis kaum noch anzutreffende vierstufige Gliederung. Hier wird qua Skala die Mitarbeiterschaft in die zwei Hälften „gut“ und „schlecht“ eingeteilt. Das ist nicht nur methodisch ein Problem, sondern vor allem ein Garant für mangelnde Akzeptanz bei den Mitarbeitern (falls sich die Beurteiler nicht auf die Verwendung der positiven Skalenhälfte beschränken).
- im höheren Leistungsbereich besser differenzieren als im mittleren. Wenn eine gute Personalauswahl erfolgt, sind die Mitarbeiter im Durchschnitt nicht mittelmäßig, sondern besser (verglichen mit der Bewerbergruppe). Deshalb muss auch die Leistungsbeurteilung daran angepasst sein und primär im höheren Leistungsbereich Unterschiede zwischen Mitarbeitern aufzeigen.

Nicht immer sind klassische Einstufungs- und Bewertungsskalen aus den vorgenannten Aspekten heraus am besten für ein System der Leistungsbeurteilung geeignet.

Quotensysteme – erzwungene Verteilung

Eine in der Praxis häufig gewählte Antwort hierauf ist, eine bestimmte Ergebnisverteilung zu erzwingen. Hierfür werden klare Quoten vorgegeben, wie viele Mitarbeiter in welche distinkte Bewertungskategorie zugeordnet werden dürfen. So werden dann zum Beispiel nur zehn Prozent der besten, 20 Prozent der zweitbesten, 40 Prozent der mittleren, 20 Prozent der zweitschlechtesten und zehn Prozent der schlechtesten Kategorie zugeordnet. Der Ansatz klingt bestechend, weil sich dadurch auch faktisch Verschiebungen zu Extremwerten und insbesondere zu positive Bewertungen vermeiden lassen. Im Ergebnis entsteht zumindest optisch eine Verteilung ähnlich einer Normalverteilung.

Der Preis dafür indes ist hoch: Einerseits entspricht die so erzwungene Verteilung fast nie den realen Gegebenheiten und zeichnet daher nur ein theoretisches Verteilungsideal, nicht aber die Realität nach. Statistisch gesehen ist diese erzwungene Verteilung jedoch keine Messung. Im sozialen Sinne wird hierdurch viel Unmut seitens der Mitarbeiter erzeugt, denn einerseits schürt dieses Nullsummenspiel (wenn ein Mitarbeiter eine Kategorie besser beurteilt wird, muss dafür ein anderer schlechter eingruppiert werden) eher den Einsatz von Ellenbogen als dass es teambezogenes Engagement und gegenseitige Unterstützung fördert. Andererseits entsteht ein reales Problem für den Vorgesetzten, wenn die Leistung seiner Mitarbeiter dieser künstlichen Vorgabe gar nicht entspricht. Er ist dann genötigt, seine Urteilskompetenz hintenanzustellen und Bewertungen abzugeben, die nicht zum wirklichen Bild passen. Dabei ist nicht nur vorstellbar, sondern sogar häufig anzutreffen, dass alle Mitarbeiter nur mittlere Leistungen erbringen oder in einer Abteilung sehr viele Leistungsträger und nur wenige unterdurchschnittliche Kollegen versammelt sind. Dann gerät der Vorgesetzte in Erklärungsnot, wenn

Verhaltensverankerte Einstufungsskala

Bitte bewerten Sie anhand der Verhaltensbeispiele für „INNOVATION“ die Leistungen Ihres Mitarbeiters in dieser Dimension.

Einschätzung (bitte ankreuzen)	Verhaltensbeispiele
überdurchschnittliche Leistung	9 • bringt selbst kreative Vorschläge und Beispiele aus verschiedenen Bereichen vor
	8 • überträgt neue Inhalte aus Literatur oder Vorträgen auf eigene Aufgaben
	7 • findet sich sehr schnell in neuen Fachgebieten zurecht
durchschnittliche Leistung	6 • erkennt Vorteile und Nutzen von neuen Entwicklungen für seine Aufgaben
	5 • kann neue Inhalte vermitteln und anwenden
	4 • ist Entwicklungen und Neuerungen gegenüber aufgeschlossen
unterdurchschnittliche Leistung	3 • informiert sich laufend über neue Aspekte seiner Aufgaben
	2 • ist an Neuerungen interessiert, spricht aber nur nach Aufforderung über diese Aspekte
	1 • verwendet herkömmliche Lösungswege ohne Bereitschaft zur Neuerung

Quelle: Schuler et al. 1995

Abb. 1

seine Bewertungen zwar der Vorgabe, aber nicht dem offen wahrnehmbaren Bild des Teams entsprechen – akzeptanzförderlich ist das nicht.

Alternative Skalierungen

Das Ziel für den Einsatz einer so erzwungenen Verteilung ist vor allem, zu gute Verteilungen zu vermeiden. Das kann durch andere Skalierungsformen häufig besser und sozialverträglicher erreicht werden, ohne den Pfad einer soliden Messung zu verlassen.

Frühere Ansätze bestanden in der Bildung einer Rangreihe aller zu beurteilenden Mitarbeiter hinsichtlich aller bewertungsrelevanten Merkmale. Angenommen, es gäbe fünf Bewertungskriterien A, B, C, D, E und zehn Mitarbeiter. Dann würde für jedes Merkmal A bis E eine Liste aller zehn Mitarbeiter entstehen, wobei jeweils der Erstgenannte der Leistungsstärkste und der auf Platz zehn gelistete Mitarbeiter der Leistungsschwächste wäre. Über alle fünf Merkmale hinweg bekäme dann jede Person einen durchschnittlichen Gesamtrangplatz zugewiesen. Bei dieser Methode entsteht das Problem, wie zwei gleich starke Mitarbeiter einzuordnen sind. Die gleichen Rangplatzabstände entsprechen hier nicht immer den tatsächlichen Leistungsabständen. Noch mehr gegen den Einsatz spricht allerdings, dass hier die unmittelbar vergleichende Bewertung offensichtlich wird und auch den Mitarbeitern transparent gemacht werden müsste. Dieser Ansatz mag daher als rein interne Arbeitshilfe für einen Vorgesetzten ein hilfreiches Instrument sein, sich gewahr zu werden, wer die wirklichen Champions im eigenen Team sind – als offizielles Leistungsbeurteilungssystem ist dieser Ansatz nicht geeignet.

Die Forschung hat jedoch neuere Skalenformate hervorgebracht, die allen vorgenannten Kriterien genügen, zu nahezu idealtypischen Normalverteilungen in der Bewertung führen und gleichzeitig

Sequenzielle Prozentrangskala

Über gute Fachkenntnisse verfügt:

Wer umfangreiches Wissen auf seinem Gebiet besitzt, praktische Erfahrung hat, seine Kenntnisse erfolgreich anwenden kann.

	Erfüllt, was seine Fachkenntnisse betrifft, nicht die Anforderungen, die an ihn gerichtet sind.
	Seine Fachkenntnisse liegen über den Mindestanforderungen (trifft auf etwa 90 Prozent der Mitarbeiter an vergleichbaren Arbeitsplätzen zu).
	Hat mindestens ebenso gute Fachkenntnisse wie etwa 55 Prozent der anderen Mitarbeiter an vergleichbaren Arbeitsplätzen.
	Hat mindestens ebenso gute Fachkenntnisse wie etwa 25 Prozent der anderen Mitarbeiter an vergleichbaren Arbeitsplätzen.
	Hat mindestens ebenso gute Fachkenntnisse wie nur zehn Prozent der anderen Mitarbeiter an vergleichbaren Arbeitsplätzen.
	Hat mindestens ebenso gute Fachkenntnisse wie nur fünf Prozent der anderen Mitarbeiter an vergleichbaren Arbeitsplätzen.

Quelle: Schuler 2004

Abb. 2

effizient und akzeptabel in der Handhabung sind. Hierzu zählen zum Beispiel die verhaltensverankerte Einstufungsskala und die sequenzielle Prozentrangskala.

Durch die konkrete Beschreibung von Verhaltensweisen und die Wahl einer in allen Leistungsbereichen differenzierenden Skalenlänge trägt die verhaltensverankerte Einstufungsskala dazu bei, individuelles Verhalten objektiv und akkurat zu bewerten und interpersonale Differenzen nicht zu verwischen (vgl. Abb. 1).

Die sequenzielle Prozentrangskala (vgl. Abb. 2) ist asymmetrisch abgestuft, um Beschönigungstendenzen entgegenzuwirken. Empirisch gelingt dies sehr gut, die durch sie erzielbaren Verteilungen folgen nahezu perfekt einer Normalverteilung. Sie ist deshalb auch hervorragend geeignet, um innerhalb höherer Leistungsbereiche gute Differenzierungen zu erbringen. Die sequenzielle Vorgehensweise verlangt vom Beurteiler gründlichere Reflexion und ergibt deshalb reliablere (verlässlichere) Beurteilungswerte. Aufgrund der Skalierungsform lassen sich mit dieser Skala nicht nur Verhaltensbeurteilungen, sondern (bei entsprechender Definition und Schulung der Verwender) auch eigenschaftsbezogene Beurteilungen erfassen.

Entwicklung von Systemen der Leistungsbeurteilung

Diese Auflistung von Qualitätsanforderungen an die verwendeten Skalen und Beispielen guter Skalierung soll nicht davon ablenken, dass viele weitere Voraussetzungen gegeben sein müssen, damit ein taugliches Gesamtsystem entsteht. Gerade aufgrund seiner Tragweite (wichtige Entscheidungen für alle Mitarbeiter hängen davon ab) zählt ein Leistungsbeurteilungssystem zu den sensibelsten Instrumenten der Personalarbeit und hat großes Potenzial, betrieblichen Nutzen zu stiften, Energien und Verhalten in eine gewünschte Richtung zu lenken und ein zentrales Element im Gesamtsystem des Performance Management zu wer-

den. Es hat aber auch das Potenzial, ein Betriebsklima nachhaltig zu zerstören und eine egomanische Unternehmenskultur zu begründen. Ein leichtfertiger Umgang mit der Entwicklung und Einführung einer Leistungsbeurteilung ist also keinesfalls zu empfehlen.

Zu den Konstruktionsschritten einer soliden Leistungsbeurteilung zählen folgende Punkte (verkürzt):

1. Bestandsaufnahme und Bewertung des bisherigen Systems,
2. Zielfestlegung,
3. Sicherstellung der sozialen Akzeptanz eines Veränderungsprozesses, auch bei dem Mitbestimmungspartner,
4. Definition von Zielgruppen, Beurteilern und Urteilsquellen,
5. Durchführung einer Anforderungsanalyse zur Ermittlung relevanter Verhaltens- und Leistungsanforderungen; Nutzung von Ergebnissen aus anderen Prozessen (z. B. einer Anforderungsanalyse für die Personalauswahl),
6. empirische Bestimmung der Beurteilungskriterien,
7. Wahl der Skalierungsweise (Skalenart, -niveau und -länge),
8. Skalenkonstruktion, statistische Erprobung und Überarbeitung,
9. Probeanwendung, Optimierung (ggf. IT-Umsetzung),
10. Einführung und Beurteilertraining.

Wer sich an diese Prozessfolge hält und auf jeder einzelnen Stufe psychometrische Qualitätskriterien zum Anspruch erhebt, wird ein Beurteilungssystem entwickeln, das gleichermaßen solide misst, differenzierende Werte liefert, zur Leistungsverbesserung und Mitarbeiterförderung führt, den betrieblichen Frieden nicht gefährdet und praktikabel anzuwenden ist.

Summary

When Performance Is Overrated

Conducting a performance evaluation is a difficult and unpopular management task. The objective of this process is to assess the conduct of employees and to discuss the results of this assessment with them. In practice, a tendency to overrate employee performance can be frequently observed. When assessment rating scales are used, the average value of the assessments then lies significantly above the middle of the scale. The validity of the performance evaluation is compromised due to this minimal differentiation between the employees evaluated. The authors list three essential requirements for a functional and practicable evaluation system: 1. focus on job responsibilities and organizational fit, 2. methodological suitability, 3. a participatory and transparent introduction. The article describes in detail the prerequisites for an appropriate evaluation scale. The classic classification and evaluation scales are not always the best to use. Quota systems that dictate a certain distribution of results do not measure performance in the statistical sense. The authors are convinced that the objective of avoiding overly favorable performance evaluations can be better achieved by using different kinds of scales. As examples thereof, the authors elaborate in particular on the behavior-anchored classification and sequential percentage ranking scales.

Literatur

- Dickinson, T. L. (1993): Attitudes about performance appraisal, in: H. Schuler / J. L. Farr / M. Smith (Eds.): *Personnel selection and assessment: Individual and organizational perspectives*, Hillsdale, 141–161
- Farr, J. L. (1991): Leistungsfeedback und Arbeitsverhalten, in: H. Schuler (Hg.): *Beurteilung und Förderung beruflicher Leistung*, Göttingen
- Marcus, B. / Schuler, H. (2006): Leistungsbeurteilung, in: H. Schuler: *Lehrbuch der Personalpsychologie*, Göttingen, 433–469
- Schuler, H. (Hg.) (2004): *Beurteilung und Förderung beruflicher Leistung*, 2. Aufl., Göttingen
- Schuler, H. / Marcus, B. (2004): Leistungsbeurteilung, in: H. Schuler (Hg.): *Enzyklopädie der Psychologie. Organisationspsychologie I – Grundlagen und Personalpsychologie (D/III/3)*, Göttingen