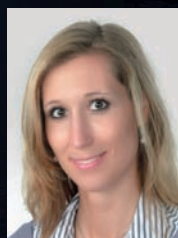

Entwicklung und Evaluation von Leistungsbeurteilungssystemen Mit vertretbarem Aufwand zu aussagekräftigen

Sehr feine, auf den individuellen Arbeitsplatz bezogene Leistungsbeurteilungsskalen und eine offene Informationspolitik zahlen sich für Organisationen aus. Dies ergab eine empirische Evaluationsstudie zu einem neuen Leistungsbeurteilungssystem. Die Autoren stellen den Hintergrund der Untersuchung vor, erklären die Vorgehensweise und erläutern abschließend die Ergebnisse.



Prof. Dr. phil. habil. Uwe Peter Kanning ist Professor für Wirtschaftspsychologie an der Hochschule Osnabrück.



Jennifer Rustige, Diplom-Psychologin, ist Mitarbeiterin des französischen Energiekonzerns GDF SUEZ in Bangkok.



Dr. Jan Henning Möller, Diplom-Psychologe, ist Projektmitarbeiter an der Fakultät für Wirtschafts- und Sozialwissenschaften der Hochschule Osnabrück.



Nikolay Kolev, Diplom-Psychologe, ist Projektmitarbeiter an der Fakultät für Wirtschafts- und Sozialwissenschaften der Hochschule Osnabrück.

Skalen der Leistungsbeurteilung



Die Leistungsbeurteilung ist die Grundlage für die richtige Platzierung von Mitarbeitern im Unternehmen, die Planung gezielter Personalentwicklungsmaßnahmen, die Formulierung von Leistungszielen sowie die Verteilung von Gehaltszuschlägen. Viele Leistungsbeurteilungssysteme leiden jedoch darunter, dass die Leistungskriterien nicht spezifisch für den jeweiligen Arbeitsplatz abgeleitet wurden und die Vergabe der Punktwerte den Führungskräften extrem große Interpretationsspielräume lässt. In den Augen der Mitarbeiter erscheinen Leistungsbeurteilungen daher nicht selten als ein Akt der Willkür. Der Vorteil derartiger Systeme liegt darin, dass sie kostengünstig entwickelt werden können. Aus der Perspektive der Personalagnostik stellen sie jedoch kein aussagekräftiges Instrument zur Leistungsmessung dar.

Allerdings ist der fehlende Arbeitsplatzbezug nur ein Problem von vielen. Ein Praxisbeispiel, das so oder so ähnlich in vielen Unternehmen anzutreffen sein dürfte, verdeutlicht dies (vgl. Abb. 1). Die Mängel eines solchen Verfahrens treten offen zutage.

Mängel von Leistungsbeurteilungssystemen

1. Die Leistungsbereiche sind inhaltlich weder definiert noch gegeneinander deutlich abgegrenzt. Je nach Führungskraft ordnet diese eine bestimmte Leistung einmal der Arbeitsweise und ein anderes Mal der Kundenorientierung zu. Für die Personalplatzierung und für die Personalentwicklung sind die Ergebnisse der Bewertung kaum aussagekräftig. Gleiche Punktwerte bedeuten bei verschiedenen Führungskräften jeweils etwas anderes.
2. Die Punktwerte sind inhaltlich nicht definiert. Eine bestimmte Leistung führt daher, abhängig vom individuellen Bezugssystem der Führungskraft, zu unterschiedlichen Bewertungen.
3. Die Bewertungsskala hat sehr viele Stufen. Der Unterschied zwischen zum Beispiel sieben und acht Punkten dürfte inhaltlich schwer zu bestimmen und gegenüber dem Mitarbeiter kaum trennscharf zu kommunizieren sein. Letztlich gaukelt die feine Abstufung einen Differenzie-

rungsgrad vor, der mit den Mitteln der unsystematischen Beobachtung im Arbeitsalltag kaum zu realisieren sein dürfte.

4. Die Punktwerte für einen einzelnen Mitarbeiter werden im direkten Vergleich zur Leistung der Kollegen vergeben. Ein und dieselbe Leistung wird daher in verschiedenen Abteilungen unterschiedlich bewertet. Bei einem internen Arbeitsplatzwechsel verändert sich die Leistung des Mitarbeiters nur dem Schein nach, ohne dass dies eine Entsprechung in der Realität hätte.
5. Über alle Leistungskriterien hinweg wird ein Mittelwert berechnet. Dies führt dazu, dass Informationen über das individuelle Leistungsprofil verloren gehen. Ein Mitarbeiter, der in allen Bereichen fünf Punkte erzielt, scheint die gleiche Leistung zu erbringen wie ein Mitarbeiter, der extrem schwache Leistungen in manchen Bereichen durch extrem starke Leistungen in anderen rechnerisch kompensiert. In der Arbeitsrealität gelingt eine solche Kompensation nur selten tatsächlich.
6. Jede Leistungsdimension wird durch eine einzige Bewertung erfasst. So schlagen Messfehler (die immer auftreten) direkt auf die Beurteilung durch.

Aus diagnostischer Sicht stellt ein solches System ein sehr unbefriedigendes Messinstrument dar (vgl. Kanning 2004). Das Ergebnis der Leistungsmessung hängt extrem von Faktoren ab, die nichts mit der eigentlichen Arbeit der Beurteilten zu tun haben (mangelnde Objektivität). Messfehler werden nicht – wie in der Diagnostik üblich – durch Mehrfachbewertungen desselben Merkmals abgefedert (mangelnde Reliabilität). Die Definition dessen, was Leistung ist, liegt letztlich im Ermessensspielraum der Führungskraft (unklare Validität).

Ein neues System

In Kooperation mit zwei Organisationen aus dem öffentlichen Sektor (Kreisverwaltung und Handwerkskammer) wurde jeweils ein neues Leistungsbeurteilungssystem entwickelt, das den skizzierten Probleme-

tig ist dabei die spezifische Definition. Wenn beispielsweise in zwei Clustern „Führung“ als Leistungsdimension identifiziert wird, unterscheiden sie sich in aller Regel inhaltlich voneinander, da erfolgreiche Führung in unterschiedlichen Kontexten – zum Beispiel in Abhängigkeit vom Qualifikationsniveau der Mitarbeiter – unterschiedliche Verhaltensweisen erfordert.

Verhaltensverankerte Beurteilungsskalen Im nächsten Schritt werden die Ergebnisse des Interviews in verhaltensverankerte Beurteilungsskalen umgesetzt, die unter anderem auch in Assessment-Centern oder strukturierten Einstellungsinterviews Verwendung finden (vgl. Schuler 2004). Abbildung 3 zeigt eine solche Skala, die die Führungskräfte später im Zuge der Leistungsbeurteilung ausfüllen müssen.

Es lässt sich erkennen, dass im Gegensatz zu den in Abbildung 1 dargestellten Skalen nicht nur die Leistungsdimension, sondern auch die Punktwerte der fünfstufigen Skala inhaltlich definiert werden. Ein Vergleich mit anderen Mitarbeitern entfällt (Förderung der Objektivität und Validität). Darüber hinaus ergibt sich der Gesamtpunktwert für eine Leistungsdimension (im Beispielfall drei Punkte) aus drei unabhängigen Einzelbewertungen verschiedener Subdimensionen. Dies wirkt sich positiv auf die Messgenauigkeit (Reliabilität) der Beurteilung aus. Je nach Skala und Cluster schwankt die Anzahl der Subdimensionen zwischen drei und sechs.

Führungskräfteschulung Wenn für alle Cluster fünf bis acht Leistungsdimensionen mit verhaltensverankerten Beurteilungsskalen in der Projektgruppe abgestimmt und mehrfach überarbeitet sind, folgt eine eintägige Führungskräfte-schulung. Die Führungskräfte lernen dabei grundlegende diagnostische Prinzipien

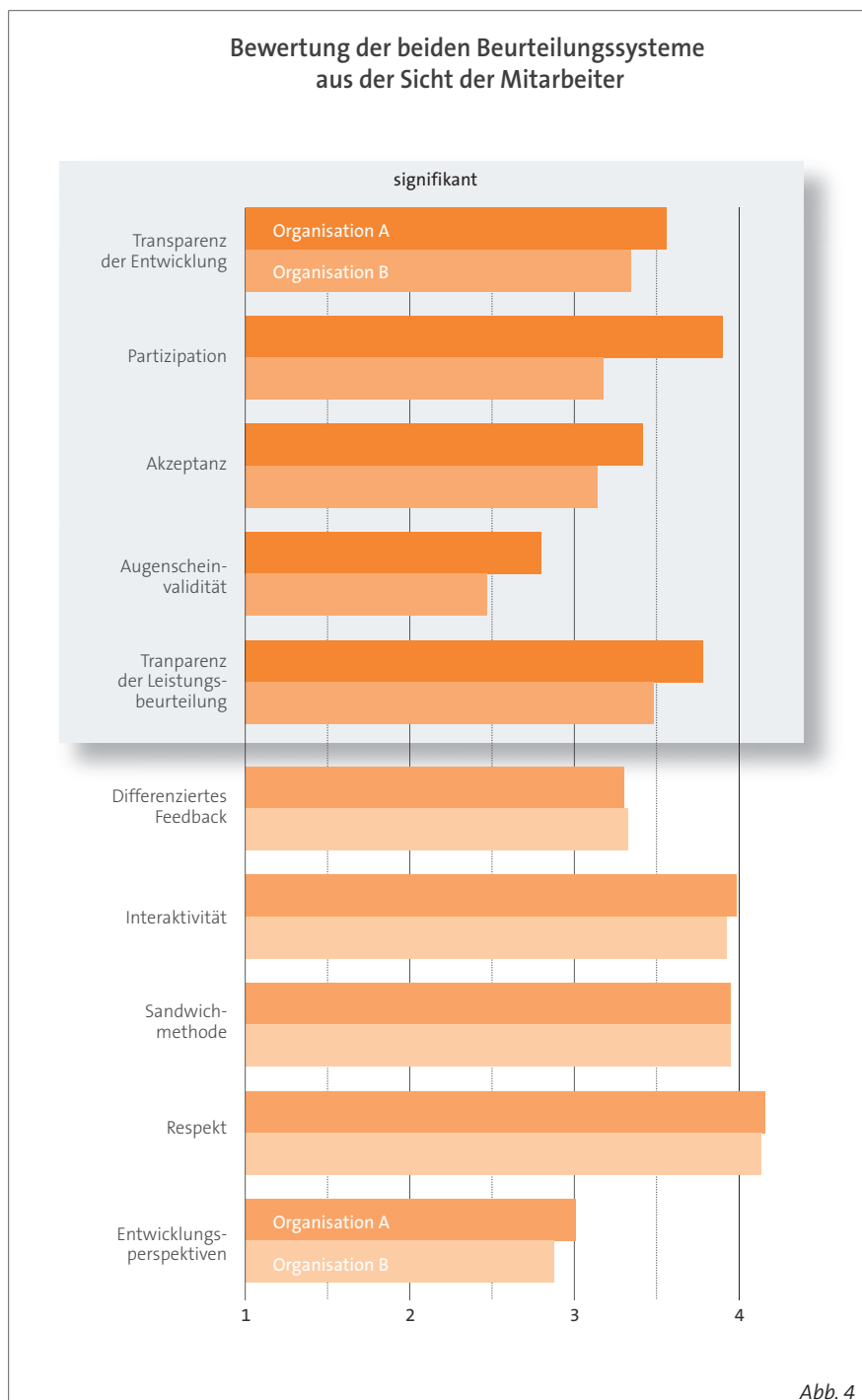


Abb. 4

sowie den Einsatz der neuen Beurteilungsskalen kennen. In Rollenspielen üben sie das Verhalten im Feedbackgespräch. Dieses dient später dazu, den Mitarbeitern das Ergebnis der Beurteilung mitzuteilen, sie zu einer Leistungssteigerung zu motivieren und gegebenenfalls

über weitere gemeinsame Schritte (z. B. Personalentwicklungsmaßnahmen) nachzudenken.

In Organisation A wurden der gesamte Entwicklungsprozess sowie die Ergebnisse der Evaluation in mehreren Informationsveranstaltungen für alle Mitarbeiter

vollkommen transparent dargestellt. Zudem konnten alle Mitarbeiter einige Monate vor der eigentlichen Leistungsbeurteilung sämtliche Beurteilungsskalen im Intranet abrufen. Organisation B entschied sich gegen eine umfassende Aufklärung der Mitarbeiter. Hier wurden die Mitarbeiter erstmals im Feedbackgespräch mit den neuen Skalen konfrontiert. Über den Entwicklungsprozess erfuhren sie eher per ‚Flurfunk‘.

Evaluationsstudie als Messhilfe

Ein Leistungsbeurteilungssystem kann sein Potenzial zur Motivierung der Mitarbeiter nur dann optimal entfalten, wenn die Mitarbeiter dieses Verfahren als ein aussagekräftiges Instrument zur Bewertung ihrer Leistung akzeptieren. Im Rahmen einer Evaluationsstudie sollte daher untersucht werden, wie die Mitarbeiter das neue Leistungsbeurteilungssystem erleben. Dabei interessierte unter anderem der Vergleich zwischen den beiden Organisationen. Hierzu wurde ein Fragebogen entwickelt, dessen Struktur die nebenstehende Tabelle zeigt.

Die eine Hälfte der Skalen im Fragebogen zielt auf das fertige Leistungsbeurteilungssystem sowie dessen Entwicklung. Die andere Hälfte untersucht den Ablauf der Feedbackgespräche. Die Items des Fragebogens mussten jeweils mithilfe einer fünfstufigen Zustimmungsskala (1 = „stimme überhaupt nicht zu“ bis 5 = „stimme völlig zu“) bearbeitet werden. Eine Überprüfung der Messgenauigkeit der eingesetzten Fragebogenskalen ergab durchgängig sehr zufriedenstellende Reliabilitätswerte (Cronbachs Alpha > .70). An der Befragung beteiligten sich 237 Mitarbeiter der Organisation A (52 % weiblich, 48 % männlich) und 111 Mitarbeiter der Organisation B (55 % weiblich, 45 % männlich).

Fragebogen zur Wahrnehmung der Leistungsbeurteilungssysteme aus Sicht der Mitarbeiter			
Konzept	Beispiel-Item	Anzahl der Items	Messgenauigkeit
Partizipation der Mitarbeiter an der Entwicklung des Systems	„Der Personalrat war Mitglied der Planungsgruppe. Diese Einbindung in die Entwicklung des neuen Beurteilungssystems ist hinreichend.“	3	(.83)
Transparenz der Entwicklung	„Ich hatte den Eindruck, dass während der Entwicklung des Beurteilungssystems mit offenen Karten gespielt wurde.“	4	(.81)
Transparenz des Beurteilungssystems	„Ich wusste, was mich bei der Beurteilung erwartet.“	3	(.81)
Akzeptanz	„Ich finde die Einteilung der einzelnen Arbeitsplätze in die Cluster hinreichend differenziert.“	4	(.85)
Augenscheinvalidität	„Jemand, der in dieser Leistungsbeurteilung gut abschneidet, bringt im Berufsalltag auch tatsächlich gute Leistung.“	3	(.77)
Differenziertes Feedback	„Meine Führungskraft hat ihre Beurteilung gut begründet.“	4	(.90)
Interaktivität im Feedbackgespräch	„Ich hatte im Gespräch Gelegenheit, zur Beurteilung Stellung zu nehmen.“	4	(.86)
Sandwichmethode	„Meine Führungskraft hat mir zunächst positive Rückmeldung gegeben, bevor sie mir kritische Punkte nannte. Meine Führungskraft hat die Rückmeldung mit einem positiven Aspekt meiner Leistung abgeschlossen.“	4	(.71)
Respektvoller Umgang	„Während des Beurteilungsgesprächs hat meine Führungskraft mich mit Respekt behandelt.“	9	(.95)
Entwicklungsperspektive	„Ich habe das Gefühl, dass das Beurteilungsgespräch mir dabei hilft, mich beruflich weiterzuentwickeln.“	5	(.91)

Ergebnisse

Alles in allem wiesen die Ergebnisse auf eine aus Sicht der Mitarbeiter gelungene Neukonstruktion der Leistungsbeurteilungssysteme sowie insgesamt zufriedenstellend abgelaufene Feedbackgespräche

hin. Bei nahezu allen zehn Kriterien lagen die Werte im mittleren bis oberen Bereich (vgl. Abb. 4). Eine Ausnahme bildeten die Kriterien „Augenscheinvalidität“ sowie „Entwicklungsperspektive“. Hier zeichnete sich ein Nachbesserungsbedarf ab.

Ein Vergleich der beiden Organisationen förderte ein interessantes Ergebnismuster zutage. In sämtlichen Kriterien, die sich auf die Entwicklung sowie die wahrgenommene Qualität des Leistungsbeurteilungssystems bezogen, schnitt die Organisation A signifikant besser ab als die Organisation B. Bei den verbleibenden fünf Kriterien, die sich auf die Gestaltung der Feedbackgespräche beziehen, fielen die Unterschiede nicht signifikant aus.

Offensichtlich hatten sich die Entwicklung sehr viel feinerer, auf den individuellen Arbeitsplatz bezogener Beurteilungsskalen (16 versus 5 Cluster) sowie die deutlich offenere Informationspolitik für die Organisation A ausgezahlt. Dies hat jedoch keinen spürbaren Einfluss auf die Gestaltung der Feedbackgespräche. Sie hängt letztlich vom Verhalten der Führungskraft in der konkreten Situation ab und wurde für beide Organisationen auf gleiche Weise in identischen Führungskrafttrainings geschult.

Fazit

Die Leistungsbeurteilung ist unter anderem eine zentrale Informationsquelle zur richtigen Platzierung der Mitarbeiter im Unternehmen sowie für den gezielten Einsatz von Personalentwicklungsmaßnahmen. Ähnlich bedeutsam ist sie ganz unmittelbar für das Erleben und Verhalten der Mitarbeiter. Nur wenn die Mitarbeiter ein gleichsam zutreffendes und differenziertes Feedback über ihre Arbeitsleistung erhalten, können sie ihr Verhalten zielbezogen ausrichten und gegebenenfalls optimieren. Diese Aufgaben kann die Leistungsbeurteilung allerdings nur dann

erfüllen, wenn sie auf den jeweiligen Arbeitsplatz abgestimmte, objektive, reliable und valide Aussagen über die Leistungen des Arbeitsplatzinhabers ermöglicht.

Einfache Leistungsbeurteilungsskalen (vgl. Abb. 1) bieten hierfür keine ausreichende Basis. Leistungskriterien wie „Arbeitsqualität“, „Teamfähigkeit“ oder „Führungskompetenz“ bleiben weitgehend Worthülsen, wenn jede Führungskraft sie vor dem Hintergrund der eigenen Erfahrung und Meinung mit Leben füllen und einen eigenen Bewertungsmaßstab entwickeln muss. Ein gleicher Punktwert, den zwei Führungskräfte vergeben, steht dann mitunter für etwas völlig Unterschiedliches. Die Zahlen sagen nichts Verbindliches aus und können daher

auch nicht sinnvoll im Personalmanagement genutzt werden.

Das vorgestellte Projekt zeigt, wie man mit einem vertretbaren Aufwand aussagekräftigere Leistungsbeurteilungsskalen entwickeln und implementieren kann. Zentrale Bausteine dabei sind die Einbindung der Mitarbeiter, die Entwicklung verhaltensverankerter Beurteilungsskalen auf dem Fundament empirischer Anforderungsanalysen sowie die Schulung der Führungskräfte.

Literatur

- Kanning, U. P. (2004): Standards der Personal diagnostik, Göttingen
 Kanning, U. P. / Pöttker, J. / Klinge, K. (2008): Personalauswahl: Leitfaden für die Praxis, Stuttgart
 Schuler, H. (Hg.). (2004): Beurteilung und Förderung beruflicher Leistung, Göttingen

Summary

Development and Assessment of Performance Evaluation Systems

Many performance evaluation systems suffer from the weaknesses that performance criteria have not been developed specifically enough for the individual job position and that managers have extreme leeway for interpretation in awarding point values. Additionally, these systems are afflicted by many other problems including, for example, inadequate objectivity and a lack of reliability. As a result, a new performance evaluation system was developed in cooperation with two public sector organizations to counteract these problems. The following steps were completed during the process: formation of a project group, clustering of job positions, preparation of job descriptions and derivation of performance dimensions, development of conduct-based assessment scales and, at the end of the process, management personnel training. Subsequently, an evaluation study was conducted using a questionnaire to determine how employees perceive the new performance evaluation system. The results of the study confirmed that the design of the new performance evaluation system had been successful and that the new system requires fewer revisions.