

# Wissenschafts-Praxis-Forum

## Konzeption und Prüfung eines multimodalen Systems der Leistungsbeurteilung: Individualmodul

Heinz Schuler, Benedikt Hell, Peter Muck, Karlheinz Becker und Andreas Diemand

**Zusammenfassung.** Der Beitrag bildet den Auftakt zu einer dreiteiligen Praxisreihe, in der die Entwicklung eines multimodalen Leistungsbeurteilungssystems für die Sparkassen-Finanzgruppe geschildert wird, das aus drei Teilverfahren besteht (Individual-, Gruppen- und Auszubildendenmodul). In diesem ersten Teil der Reihe werden die theoretische Konzeption, die empirische Überprüfung und die praktische Einführung des Individualmoduls dargelegt, das zur mehrperspektivischen Beurteilung von Einzelpersonen dient.

Besonderes Augenmerk wurde in der Konzeptionsphase auf die Auswahl der Beurteilungsverfahren gelegt. Die drei Verfahren Verhaltensverankerte Einstufungsskala, Verhaltensbeobachtungsskala und Verhaltensrangprofil erwiesen sich als am geeignetsten für die gegebene Problemstellung und wurden der Konstruktion des Beurteilungssystems zugrunde gelegt. Die Skalenkonstruktion erfolgte in einem mehrstufigen partizipativen Prozess. Auf Basis der Daten der Probeanwendung (N=1518) konnten die Verfahren empirisch überprüft und revidiert werden. Die Analysen belegen die Brauchbarkeit aller drei eingesetzten Skalenformate und zeigen deren Äquivalenz. Neben substanziellen Übereinstimmungen der verschiedenen Beurteilungsrichtungen liefern die unterschiedlichen Perspektiven zusätzliche relevante Information, was den Nutzen ihrer Einbeziehung unterstreicht.

Schlüsselwörter: Leistungsbeurteilung, 360°-Feedback, Verhaltensverankerte Einstufungsskala, Verhaltensbeobachtungsskala, Verhaltensrangprofil, Multimodalität

Conception and test of a multimodal system of performance assessment: Individual module

This is the first contribution of three in a series of reports on the development of a multimodal system of performance assessment for the Sparkassen Financial Group. The system consists of three modules (individual, group and apprentice module). In this first part of the series, the theoretical conception, the empirical examination and the practical implementation of the individual module is described, which suits the purpose of a multi-perspective evaluation of individuals.

In the conception phase, special attention was paid to the selection of the scale format. The three formats Behavior Anchored Rating Scale, Behavioral Observation Scale and Behavior Ranking Profile proved as most suitable for the given situation and were the basis for the construction of the assessment system. The scale construction took place in a multi-level participative process. Participants (N=1518) helped to empirically examine and revise the measures. The analyses prove the capability of all three scale formats and show their equivalence. Apart from substantial agreement, the different assessment perspectives provide additional relevant information which underline the utility of their inclusion.

Key words: performance appraisal, 360°-Feedback, behaviorally anchored rating scale, behavior observation scale, behavior ranking profile, multimodality

### Vorbemerkung

Systematische Leistungsbeurteilung gehört zu den wirksamsten personalpsychologischen Methoden und führt – insbesondere in Verbindung mit Zielsetzung

und Feedback – nicht nur zu erheblicher Leistungssteigerung (Guzzo, Jette & Katzell, 1985), sondern gehört auch zu den wichtigsten Instrumenten der Entwicklung und Förderung von Mitarbeitern (Schuler & Marcus, in Druck).

In der Sparkassen-Finanzgruppe werden seit längerem Anstrengungen unternommen, die Verfahren der Auswahl und Förderung von Mitarbeitern und Führungskräften zu einem ganzheitlichen System weiterzuentwickeln, der so genannten *eignungsdi-*

---

Für ihre wertvolle Mitarbeit in verschiedenen Projektphasen danken wir Frau Dipl.-Psych. Barbara Hirth, Herrn Dr. Stefan Höft, Frau Dipl.-Oec. Virag Jóny, Frau Dipl.-Psych. Iris Kallo und Frau Dipl.-Psych. Tina Laubach.

*gnostischen Kette* (Backhaus & Wagner, 2003). Innerhalb dieses Systems sind auf einheitlicher anforderungsanalytischer Basis die Personalauswahl, -beurteilung, -entwicklung und Potenzialdiagnose zu einer kohärenten Sequenz integriert. Der zuletzt entwickelte Baustein dieses Systems war ein multimodales Verfahren der Leistungsbeurteilung, das in drei Teilverfahren gegliedert ist – das *Individualmodul* zur Einschätzung von Einzelpersonen, das *Gruppenmodul* zur Beurteilung von Teamprozessen und Gruppenleistungen sowie ein eigenständiges Verfahren zur förderungsorientierten Beurteilung von *Auszubildenden*.

In einer dreiteiligen Reihe werden diese drei Teilverfahren des Leistungsbeurteilungssystems der Sparkassen-Finanzgruppe – intern als MLB-Multimodal® bezeichnet – geschildert. Der vorliegende Beitrag stellt das Individualmodul dar, in folgenden Ausgaben dieser Zeitschrift werden das Gruppenmodul und das Auszubildendenmodul vorgestellt.

## Überblick über den Gesamtprozess

Das Individualmodul des Leistungsbeurteilungssystems wurde in einem partizipativ angelegten Prozess erstellt. Er umfasste folgende Arbeitsschritte, die zunächst nur cursorisch aufgelistet werden sollen:

- Zielformulierung und Festlegung der Zielprioritäten
- Bestandsaufnahme und Klärung der Rahmenbedingungen
- Festlegung der Zielgruppen
- Klärung arbeitsrechtlicher Fragen und Partizipationsfragen
- Literaturlauswertung
- Sichtung repräsentativer Leistungsbeurteilungssysteme (inkl. der Anforderung der Systeme)
- Auswertung bestehender Anforderungsanalysen / Expertenbefragungen zur Absicherung der Aktualität und Repräsentativität
- Aufbau eines Versuchsplans
- Festlegung der Auswertung
- Konstruktion der Skalen
- Vorerprobung und psychometrische Überprüfung
- Entscheidung zwischen mehreren Verfahrensalternativen
- Festlegung eines Gewichtungsalgorithmus
- Probeanwendung in repräsentativen Verbandsgebieten
- Psychometrische Überprüfung, Verständlichkeits- und Akzeptanzprüfung
- Abschließende Abstimmung
- Konzeption der Beurteilertrainings
- Vorbereitung der Einführung

Im Folgenden wird auf die Kernaspekte des Gesamtprozesses eingegangen, insbesondere auf die aus der im Unternehmen bereits etablierten Potenzialanalyse entlehnten Anforderungsdimensionen, die drei verwendeten Skalen der Probeanwendung sowie ihre Konstruktion und empirische Überprüfung.

## Verwendete Anforderungsdimensionen

Der Auftraggeber strebt eine integrierte Konzeption seiner Personalförderinstrumente an. Um die Parallelität der Anforderungsdimensionen (also der Beurteilungskriterien) der Potenzialanalyse und des Leistungsbeurteilungssystems zu gewährleisten, wurden diese weitestgehend von dem in den Instituten der Sparkassen-Finanzgruppe bereits existierenden Potenzialanalyseverfahren übernommen. Die Anforderungsdimensionen lauten: „(1) Kundenorientierung und verkäuferische Fähigkeiten“, „(2) Kooperation und Teamfähigkeit“, „(3) Planung und Organisation“, „(4) Fachkompetenz“, „(5) Qualitätsorientierung / Selbstkontrolle / Arbeitssorgfalt“, „(6) Soziale Belastbarkeit“, „(7) Initiative und Erfolgsorientierung“. Weitere drei Dimensionen kommen bei der Einschätzung von Führungskräften zum Einsatz: „(8) Mitarbeiterorientierung“, „(9) Leistungsförderung“ und „(10) Steuerung und Koordination“ zur Einschätzung von Führungskräften. Damit basieren Leistungsbeurteilung und Potenzialanalysen auf den Ergebnissen der gleichen Anforderungsanalysen.

## Verwendete Beurteilungsverfahren

Insgesamt wurden 15 Beurteilungsverfahren in Erwägung gezogen, deren Brauchbarkeit durch eine Experteneinschätzung überprüft wurde. (Als Experten fungierten die zwölf Mitglieder des projektsteuernden Lenkungsausschusses, darunter neun leitende, mit Fragen der Personalentwicklung betraute Mitarbeiter der Sparkassen-Finanzgruppe und drei auf diesem Gebiet tätige Wissenschaftler.) Drei Beurteilungsverfahren wurden als besonders geeignet eingeschätzt, die Zielsetzungen und Bedürfnisse der Sparkassen zu erfüllen: die Verhaltensverankerte Einstufungsskala, die Verhaltensbeobachtungsskala und das Verhaltensrangprofil. Diese wurden für die Probeanwendung ausgewählt.

## Verhaltensverankerte Einstufungsskalen

Verhaltensverankerte Einstufungsskalen (VVE; Behaviorally Anchored Rating Scales [BARS], Smith &

Kendall, 1963) sind grafische Einstufungsskalen, bei denen Skalenstufen durch Kurzbeschreibungen typischen tatsächlichen Arbeitsverhaltens definiert werden (vgl. Kasten 1). Sie werden nach einem komplexen mehrstufigen Schema konstruiert: Zunächst definiert eine Gruppe von Beurteilern relevante Leistungsdimensionen und konkretisiert diese in Umschreibungen für jeweils gute, mittlere und schwache Leistung, die später als Skalenverankerung dienen. Eine zweite Beurteilerstichprobe formuliert dann konkrete Verhaltensbeispiele für jede Dimension, die anschließend von einer dritten unabhängigen Gruppe wiederum den Dimensionen zugeordnet werden. Verhaltensbeispiele, die diesen Prozess überstanden haben, werden von einer weiteren Stichprobe den einzelnen Skalenstufen zugeordnet. Übrig bleiben Verhaltensbeispiele, die sich, bei möglichst geringer Streuung der Einschätzungen, über die gesamte Skalenlänge verteilen. Ziel der VVE-Methode ist eine verhaltensnahe Beurteilung und die Hinlenkung der Beurteiler auf ein gemeinsames, arbeitsanalytisch fundiertes Beobachtungsschema. Die Beurteilung erfolgt pro Dimension auf einer neunstufigen Skala.

Zusammengefasst arbeitet die Verhaltensverankerte Einstufungsskala mit einer Gesamteinschätzung zu jeder der sieben (für Mitarbeiter) bzw. zehn (für Führungskräfte) Anforderungsdimensionen, deren Skalenbereiche mit illustrierenden typischen Verhaltensbeispielen verankert sind. Die VVE verbindet also die Orientierung an konkreten Verhaltensweisen mit einer dimensionsbezogenen, gestalthaften Integration dimensionaler Einschätzungen.

## Verhaltensbeobachtungsskalen

Gleichfalls auf einer Anforderungsanalyse basierend und teststatistisch abgesichert ist die Konstruktion von Verhaltensbeobachtungsskalen (VBS; Behavior Observation Scales [BOS], Latham & Wexley, 1977). Das Ergebnis entspricht formal einer großen Anzahl von einzelnen Likert-Items, bei denen Häufigkeitseinschätzungen ausschließlich beobachtbarer Verhaltensweisen jeweils auf einer Skala von 1 (fast nie) bis 5 (fast immer) abgegeben werden sollen (vgl. Kasten 2), die anschließend zu den hier verwendeten sieben bzw. zehn Dimensionen zusammengefasst werden. Man erhofft sich von diesen Häufigkeitsschätzungen im Vergleich zu anderen Skalenformaten verhaltensnähere, weniger subjektiv-bewertende Urteile und damit auch eine geeignete Grundlage gezielter Verhaltenssteuerung. Jeweils acht Items bilden eine Dimension ab. Im Gegensatz zu dem integrierten, dimensionsbezogenen Urteil der VVE besteht das Prinzip der VBS in der gesonderten Beurteilung einzelner Verhaltensweisen. Dieser Skalentyp ist deshalb besonders gut zur Verhaltenssteuerung geeignet, da Informationen über einzelne Verhaltensweisen, die separat beurteilt werden, vorliegen.

## Verhaltensrangprofil

Eine andere Intention wird mit dem Verhaltensrangprofil verfolgt (VRP; Brandstätter & Schuler, 1974; Schuler, 1991). Hier werden nicht Personen, sondern

<u>Einschätzung</u> (bitte ankreuzen)	<u>Verhaltensbeispiele</u>
+----+   9  ---+ +----+	* bietet seine Hilfe an, wenn ein Kollege unter Zeitdruck einen Arbeitsvorgang durchführen muss
+----+   sehr gute   8  ---  bis hervorragende +----+   Leistung	* schlägt alternative Verhaltensmöglichkeiten vor, wenn er Kritik an Kollegen / Mitarbeitern übt
+----+     7  ---+ +----+	* entwickelt gemeinsame Strategien und Vorgehensweisen mit seinen Kollegen / Mitarbeitern, um die Arbeit des Teams voranzubringen
...	* vertritt seine Position überzeugend, wenn seine Meinung ohne triftige Argumente kritisiert wird
	...

Kasten 1. Beispielitem Verhaltensverankerte Einstufungsskala (Dimension „Kooperation und Teamfähigkeit“; Ausschnitt)

1.	- bietet seine Hilfe an, wenn ein Kollege unter Zeitdruck einen Arbeitsvorgang durchführen muss	fast nie	①	②	③	④	⑤	fast immer
2.	- schlägt alternative Verhaltensmöglichkeiten vor, wenn er Kritik an Kollegen / Mitarbeitern übt	fast nie	①	②	③	④	⑤	fast immer
3.	- entwickelt gemeinsame Strategien und Vorgehensweisen mit seinen Kollegen / Mitarbeitern, um die Arbeit des Teams voranzubringen	fast nie	①	②	③	④	⑤	fast immer
4.	- vertritt seine Position überzeugend, wenn seine Meinung ohne triftige Argumente kritisiert wird	fast nie	①	②	③	④	⑤	fast immer
...	...							

Kasten 2. Beispielitem Verhaltensbeobachtungsskala (Dimension „Kooperation und Teamfähigkeit“; Ausschnitt)

*Merkmale einer einzelnen Person* in eine Rangreihe gebracht und auf Übereinstimmung mit der – arbeitsanalytisch ermittelten – Rangfolge von Arbeitsplatzanforderungen verglichen. Ziel ist dementsprechend nicht der interpersonale Vergleich, sondern die Überprüfung, inwieweit Position und Stelleninhaber zueinander passen. Angaben über das Leistungsniveau können bei dieser Verfahrensweise allerdings nicht gewonnen werden. Grundprinzip des VRP ist der ipsative Vergleich, d.h. die Anforderungsdimensionen werden in eine Rangreihe bezüglich ihrer Bedeutsamkeit für den Arbeitsplatz gebracht. In gleicher Weise werden die Anforderungsdimensionen gemäß ihrem Zutreffen auf die zu beurteilende Person in eine Rangreihe gebracht, sodass ein unmittelbarer Vergleich zwischen Arbeitsplatz und Person (also auch zwischen Soll und Ist) möglich ist. Die Übereinstimmung von Arbeitsplatzrangreihe und Personrangreihe (Rangkorrelationskoeffizient) gibt Aufschluss über das „Zusammenpassen“ von Arbeitsplatz und Person. Nahezu über drei Jahrzehnte war das Verhaltensrangprofil das einzige Beurteilungsverfahren dieser Funktionsweise. Erst vor kurzem wurde von Staufenbiel und Kleinmann (2002) ein vergleichbares Verfahren vorgestellt.

### Vergleich der Merkmale und Funktionen der drei Beurteilungsverfahren

Zusammenfassend enthalten die drei Skalierungsverfahren weitgehend die gleichen Beurteilungsaussagen. Dabei handelt es sich um Beschreibungen konkreter

Verhaltensweisen – z. B. „bietet seine Hilfe an, wenn ein Kollege unter Zeitdruck einen Arbeitsvorgang durchführen muss“ –, die ebenso wie die übergeordneten Dimensionen – z. B. Kooperation und Teamfähigkeit – aus Anforderungsanalysen abgeleitet worden sind und von Experten (Arbeitsplatzinhaber, deren Vorgesetzte sowie leitende Mitarbeiter der Personalentwicklung) als anforderungsrelevant eingestuft wurden. Alle drei Verfahren weisen deshalb engen Bezug zu den in der Sparkassen-Finanzgruppe bereits etablierten Potenzialanalyseverfahren auf. Während in diesen – dem Prognosecharakter der Verfahren entsprechend – eine deutliche *Eigenschaftsorientierung* im Vordergrund steht, sind die hier geprüften Beurteilungsverfahren explizit verhaltensorientiert formuliert, was sowohl im Dienste ihrer Steuerungsfunktion (Verhaltenslernen), ihrer Transformierbarkeit in Zielvereinbarungen als auch der Kommunizierbarkeit ihrer Ergebnisse steht. Gleichzeitig ist dadurch sichergestellt, dass Führungskräfte ihr Augenmerk auf explizit erfolgsrelevantes *Verhalten* ihrer Mitarbeiter richten.

Vergleicht man die drei Beurteilungsverfahren miteinander, so bietet jedes Verfahren bestimmte Vorteile. Die Verhaltensverankerte Einstufungsskala (VVE) und die Verhaltensbeobachtungsskala (VBS) sind beides Einstufungsskalen; in dieser Charakteristik sind sie sich ähnlich, nicht aber hinsichtlich des Bewertungsgegenstands: Bei der Verhaltensbeobachtungsskala werden einzelne Verhaltensaussagen bewertet, bei der Verhaltensverankerten Einstufungsskala wird eine dimensionale Interpretation zugrunde

gelegt, vom Beurteiler also eine Integration von Einzelbeobachtungen gefordert. Als Ergänzung dazu bietet sich das Verhaltensrangprofil (VRP) als Rangordnungsverfahren an, das die Relationen zwischen Person und Arbeitsplatz im Sinne einer relativen Erfüllung der Anforderungen in den Vordergrund rückt. Daraus resultiert, dass für die Verhaltensverankerte Einstufungsskala vor allem ihre Flexibilität spricht (nicht jedes einzelne angesprochene Verhalten muss zwangsläufig beobachtet worden sein), für die Verhaltensbeobachtungsskala der engere Bezug zum Day-to-day-Feedback und damit die Möglichkeit der Verhaltenssteuerung und für das Verhaltensrangprofil der unmittelbare Vergleich zu den Tätigkeitsanforderungen, wodurch Maßnahmen der Arbeitsplatzzuordnung und Personalentwicklung begünstigt werden.

## Entwicklungsprozess

Die gesamte Verfahrensentwicklung und -einführung stand im Rahmen eines umfangreicheren Diskussionsprozesses unter den Beteiligten und Betroffenen (Mitarbeiter, Führungskräfte, Personalentwicklung, Entscheidungsinstanzen im Unternehmen). Alle Komponenten der zur Zeit in Forschung und Praxis diskutierten Beurteilungsquellen (z. B. Conway, Lombardo & Sanders, 2001) bzw. des 360°-Feedback (z. B. Scherm & Sarges, 2002) wurden in Erwägung gezogen, zudem die Beurteilung von Gruppenleistungen.

## Auswahlprozess für die Beurteilungsmodule

Mittels einer Expertenbefragung unter Personalreferenten der Organisation wurde festgestellt, welche Beurteilungsmodule tatsächlich zu entwickeln waren. Dabei wurde ermittelt, welche Beurteilungsrichtungen (also welche Beurteiler-Beurteilten-Kombinationen) am besten für den Einsatz in den Instituten der Sparkassen-Finanzgruppe geeignet sein könnten. Anhand dreier Kriterien wurden verschiedene Module der Beurteilung durch die Experten eingeschätzt:

- *Eignung zur Leistungsbeurteilung* meint die Einschätzung: „Welche Erfolgskriterien kämen Ihrer Meinung nach zur Leistungsbeurteilung der Mitarbeiter/innen in Frage?“
- *Erfassbarkeit von Leistung* meint die Einschätzung: „Welche der angeführten Erfolgskriterien lassen sich erfassen?“
- *Akzeptanz des jeweiligen Beurteilungsmoduls* meint die Einschätzung: „Inwieweit würden die verschiedenen Leistungsmerkmale von den Mitarbeiter/innen (und dem Personalrat) akzeptiert werden?“

Folgende Beurteilungsmodule wurden durch die Experten eingeschätzt:

- Beurteilung Mitarbeiter durch Führungskraft
- Beurteilung Führungskraft durch Mitarbeiter
- Beurteilung durch Kollegen
- Selbstbeurteilung Mitarbeiter
- Beurteilung durch Kunden
- Beurteilung durch interne Kunden
- Beurteilung von Gruppenleistungen durch eigene und Fremdgruppen

Außerdem wurden „Objektive Leistungsmaße“ (z. B. Verkaufszahlen) als mögliches Beurteilungskriterium anhand der oben genannten Kriterien eingeschätzt. Schließlich hatten die Befragten die Möglichkeit, geeignete Beispiele objektiver Leistungsmaße zu nennen.

Die Ergebnisse der Expertenbefragung sind im Folgenden aufgeführt:

- Das Modul *Beurteilung Mitarbeiter durch Führungskraft* wurde als unstrittig festgelegt. „Mitarbeiter“ werden Personen ohne Führungsfunktion genannt.
- Die Entwicklung eines Moduls *Beurteilung Führungskraft durch Mitarbeiter* wurde bejaht, allerdings mit der Maßgabe, besser von Feedback als von Beurteilung zu sprechen. Betont wurde, dass der Einsatz dieses Moduls in besonderem Maße eine Kommunikations- und Vertrauenskultur im Unternehmen voraussetze, aber auch fördere. Eine Selbstbeurteilung der Führungskraft wurde zusätzlich als sinnvoll und notwendig bezeichnet.
- Das Modul *Beurteilung durch Kollegen* sollte aufgrund der Gefahr der Ablehnung durch die Praxis nicht weiter verfolgt werden (vgl. hierzu auch Druskat & Wolff, 1999).
- Das Modul *Selbstbeurteilung des Mitarbeiters* wurde im Prozess der Leistungsbeurteilung als wichtig und notwendig gesehen, sollte jedoch nicht den Stellenwert anderer Module (wie z. B. „Beurteilung Mitarbeiter durch Vorgesetzte“) erreichen.
- Das Modul *Beurteilung durch Kunden* sollte trotz größerer Bedenken erprobt werden.
- Beim Modul *Beurteilung durch interne Kunden* wurden starke Analogien zur „Beurteilung durch Kunden“ gesehen. Als interne Kunden werden andere Abteilungen der Organisation verstanden; durch diesen Begriff soll der Dienstleistungscharakter der Arbeit auch innerhalb der Organisation betont werden. Die Idee wurde bei der Gruppenfremdbeurteilung aufgegriffen und weiterverfolgt (vgl. Muck, Schuler, Becker & Diemand, in Druck).
- Bei dem Modul *Beurteilung von Gruppenleistungen* gab es erhebliche Bedenken. Als interessant wurde die Möglichkeit aufgenommen, dass Grup-

pen sich selbst beurteilen. Diese Möglichkeit sollte im Konstruktionsprozess aufgegriffen und auf Effizienz überprüft werden. Gleichzeitig erschien es hier (wie schon bei der Beurteilung durch interne Kunden angedeutet) möglich, die Zusammenarbeit mit einer anderen Gruppe beurteilen zu lassen (so genannte Gruppenfremdbeurteilung; Einzelheiten in Muck et al., in Druck).

## Entwicklung der Beurteilungsverfahren: Item- und Skalenkonstruktion:

Die Itemkonstruktion wurde angesichts der Tatsache, dass bei der Erprobung des Multimodalen Leistungsbeurteilungssystems neben dem Verhaltensrangprofil sowie den Verhaltensbeobachtungsskalen auch die Verhaltensverankerte Einstufungsskala zum Einsatz kommen würde, eng an das theoretisch geforderte Vorgehen für Verhaltensverankerte Einstufungsskalen angelehnt. Da die Leistungsdimensionen durch die Dimensionen des in der Sparkassen-Finanzgruppe bereits etablierten Potenzialanalyseverfahrens im Wesentlichen vorgegeben waren, konnten die Arbeitsschritte a) und b) übersprungen werden. Die weiteren Schritte sind dagegen spezifisch für die dezidierte Verhaltensorientierung des Beurteilungsverfahrens (im Gegensatz zur Eigenschaftsorientierung der Potenzialanalysen):

- a) Definition von Leistungsdimensionen
- b) Überprüfung der Leistungsdimensionen
- c) Entwicklung von Qualitätsdefinitionen (Leistungsniveau gut/mittel/schlecht)
- d) Entwicklung von Verhaltensbeispielen zu den verschiedenen Leistungsdimensionen für verschiedene Leistungsniveaus (nach der Critical Incident Technique)
- e) Editierung der Verhaltensbeispiele
- f) Reallokation (Rückzuordnung) der Verhaltensbeispiele
- g) Einstufung verbliebener Verhaltensbeispiele bzgl. Leistungsniveau
- h) Ermittlung der Effektivitätsmittelwerte pro Verhaltensbeispiel und Auswahl verbliebener Verhaltensbeispiele mit unterschiedlichen Mittelwerten für jede Dimension
- i) Ziel: Abdecken gesamter Leistungsniveau-Spannbreite; Skalenabstufungen in möglichst gleich großen Intervallen

Um diesen Entwicklungsprozess innerhalb einer angemessenen Zeit und mit einem möglichst geringen Distributionsaufwand unter Einbezug vorhandener Experten – dabei handelte es sich in erster Linie um Personalreferenten der Organisation – in verschiedenen Instituten der Sparkassen-Finanzgruppe durch-

führen zu können, wurde die Itemkonstruktion auf folgende Arbeitsschritte komprimiert:

Expertenrunde 1: Generierung der Verhaltensaussagen

Expertenrunde 2: Einschätzung der Verhaltensaussagen bezüglich ihrer Eignung zur Leistungsbeurteilung

Expertenrunde 3: Reallokation verbleibender geeigneter Verhaltensaussagen zu den Beurteilungsdimensionen

Expertenrunde 4: Skalenverankerung verbleibender geeigneter Verhaltensaussagen

Nach der Itemkonstruktion (Expertenrunde 1 und 2) folgte demnach die Skalenkonstruktion für die Verhaltensverankerten Einstufungsskalen (Expertenrunde 3 und 4), die – aus Gründen der inhaltlichen Parallelität der Skalen – als Grundlage für die Konstruktion der Verhaltensbeobachtungsskalen sowie des Verhaltensrangprofils diene. Die folgenden Arbeitsschritte konkretisieren das skizzierte Vorgehen:

*Generierung Verhaltensaussagen:* Auf Basis der 10 Beurteilungsdimensionen wurden insgesamt 386 positiv und 889 negativ formulierte Verhaltensaussagen im Sinne effektiven Verhaltens am Arbeitsplatz generiert.

*Prüfung und Umformulierung der Verhaltensaussagen durch unabhängige Beurteiler:* Die formulierten Verhaltensaussagen wurden in einem ersten Durchgang gesichtet, aussortiert und teilweise gekürzt.

*Einschätzen der Verhaltensaussagen:* Durch eine weitere Experteneinschätzung wurde die Anzahl der Verhaltensaussagen noch einmal reduziert.

Insgesamt 528 geeignete, davon 188 positive und 340 negative Verhaltensaussagen konnten auf die beschriebene Weise gewonnen werden, die als Grundlage zur Skalenkonstruktion dienten. Die auf Grund der Einschätzung erforderlich gewordenen Änderungen wurden durchgeführt.

*Reallokation der Verhaltensaussagen zu den Beurteilungsdimensionen:* Die Reallokation geeigneter Verhaltensaussagen zu den Beurteilungsdimensionen und die Skalenverankerung der 528 verbliebenen geeigneten Verhaltensaussagen sind die beiden abschließenden Schritte der Konstruktion der Beurteilungsverfahren. Diese wurden durch sechs Diplom-Psychologen durchgeführt.

*Skalenverankerung verbleibender Verhaltensaussagen:* Die vor der Reallokation als geeignet eingeschätzten 528 Verhaltensaussagen (und nicht nur die nach der Reallokation haltbaren 370) wurden durch die teilnehmenden Experten auf einer neunstufigen Skala (von *negativ im Sinne der Berufsausübung* bis *positiv im Sinne der Berufsausübung*) hinsichtlich ihrer Bedeutung für die Berufsausübung bewertet.

(Dass nicht nur die 370 nach der Reallokation resultierenden Items bewertet wurden, lag daran, dass aus Ökonomiegründen die Expertenrunden 3 und 4 gleichzeitig durchgeführt wurden.) Die korrekt zugeordneten und nicht für die Skalen unmittelbar verwendeten Beurteilungsaussagen wurden in ergänzende Listen aufgenommen, um den Beurteilern die Bedeutung der Dimensionen zu veranschaulichen.

## Probeanwendung

### Beurteilungsrichtungen

In der Probeanwendung wurden nach dem nunmehr abgeschlossenen Itemauswahlprozess folgende Module der Individualbeurteilung getestet (die Namen der Module wurden mehrfach abgeändert, es werden deshalb zur Vereinheitlichung die Namen der Endversionen nach der Probeanwendung benutzt, auch wenn diese zum Teil von denen der Probeanwendung abweichen):

- *Führungskraft / Vorgesetzter beurteilt Mitarbeiter* (d.h. Person ohne Führungsfunktion; auf sieben Anforderungsdimensionen)
- *Vorgesetzter* (d.h. übergeordnete Führungskraft) *beurteilt Führungskraft* (d.h. Mitarbeiter mit Führungsfunktion; anstatt von sieben werden hier zehn (sieben + drei) Dimensionen beurteilt)
- *Mitarbeiter gibt Feedback an Führungskraft*<sup>1</sup> (nur die drei Führungsdimensionen sind von den Mitarbeitern einzuschätzen)
- *Selbsteinschätzung Mitarbeiter* (auf sieben Anforderungsdimensionen)
- *Selbsteinschätzung Führungskraft / Vorgesetzter* (auf zehn Anforderungsdimensionen)
- *Kundenbeurteilung* (Operationalisierung der Anforderungsdimension „Kundenorientierung und verkäuferische Fähigkeiten“ über sechs geschlossene und eine offene Frage)

Abbildung 1 illustriert die verschiedenen Perspektiven.

### Ergebnisse

Im Zuge der Probeanwendung konnten in großem Umfang empirische Daten gewonnen werden, anhand deren die Brauchbarkeit der entwickelten Verfahren überprüft wurde. Die Auswertung erfolgte bezüglich verschiedener psychometrischer Kennwerte, Skalenvergleiche, Urteilertendenzen, Akzeptabilität und Praktikabilität. Im Folgenden werden die wichtigsten Ergebnisse referiert.

<sup>1</sup> Hiermit sind auch das Feedback der Mitarbeiter an Vorgesetzte und das Feedback der Führungskräfte an Vorgesetzte gemeint.

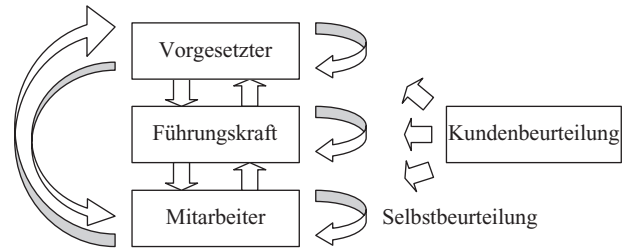


Abbildung 1. Beurteilungsschema der Probeanwendung

### Beschreibung der Stichprobe

Insgesamt wurden 1518 Beurteilungen durchgeführt, wovon der größte Anteil auf Beurteilungen zurückgeht, in denen die Mitarbeiter entweder die (Selbst-) Beurteiler oder die Beurteilten waren (vgl. Tabelle 1).

An der Erprobung beteiligten sich sechs Institute. Die Teilnahme an der Erhebung war freiwillig.

### Vergleich Selbstbeurteilungen und Fremdbeurteilungen

Mittelwertsvergleiche zwischen Selbst- und Fremdbeurteilung lassen sich sowohl für die Führungskräfteebene (vgl. Abbildung 2 und Abbildung 3) als auch für die Mitarbeiterenebene (vgl. Abbildung 4) mit allen drei verwendeten Beurteilungsinstrumenten (VVE, VBS, VRP) durchführen; bei den Abbildungen handelt es sich um eine Auswahl aus den möglichen Vergleichen. Zusammenhangsanalysen sind aufgrund der unterschiedlichen Konzeption des VRP nur mit der VVE und der VBS je separat auf der Führungskräfteebene (vgl. Tabelle 2) und der Mitarbeiterenebene (vgl. Tabelle 3) möglich.

Die Zusammenhänge zwischen Selbst- und Fremdbeurteilungen liegen im Bereich dessen, was nach den vorliegenden Metaanalysen beim Einsatz qualifizierter Beurteilungsinstrumente zu erwarten ist (Conway & Huffcutt, 1997; Moser, 1999). Die Selbstbeurteilungen fallen etwas differenzierter und etwas positiver aus als die Fremdbeurteilungen. Angesichts der Skalenlänge leisten die Beurteilungen allerdings auch noch im oberen Bereich eine gute Differenzierung. Leichte positive Mittelwertstendenzen sollten nicht verhindert werden, da sie aufgrund der positiven Vorselektion der Mitarbeiter durchaus rational sind und überdies den Feedbackprozess erleichtern.

Die bessere Selbsteinstufung verglichen mit der Fremdeinstufung ist auf Mitarbeiterenebene ausgeprägter als auf Führungskräfteebene. Auf Vorgesetztebene ist diesbezüglich keine Differenz mehr festzustellen (nur VVE). Führungskräfte werden von ihren

Tabelle 1. Anzahl der Beurteilungen

Beurteilungsrichtung	Häufigkeit	Prozent	Parallel eingesetzte Methoden
Selbstbeurteilung Führungskraft (FK)	87	5,7	VVE, VBS, VRP
Selbstbeurteilung Mitarbeiter (MA)	404	26,6	VVE, VBS, VRP
Selbstbeurteilung Vorgesetzter (V)	23	1,5	VVE, VBS, VRP
V beurteilt FK	90	5,9	VVE, VBS, VRP
MA beurteilt FK / V	338	22,3	VVE, VBS
FK / V beurteilt MA	453	29,8	VVE, VBS, VRP
FK beurteilt V	123	8,1	VVE, VBS
<b>Gesamt</b>	<b>1518</b>	<b>100,0</b>	

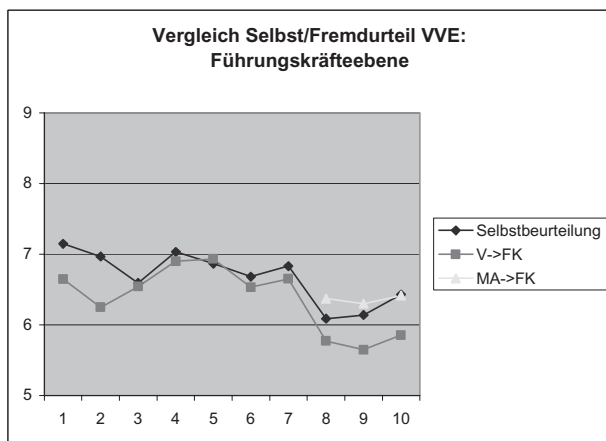


Abbildung 2. Mittelwertsvergleiche auf der Führungskräfteebene für das Beurteilungsverfahren VVE

Anmerkungen: Verankerung 1=„verbesserungsbedürftige Leistung“ bis 9=„hervorragende Leistung“, auf der Abszisse sind die Dimensionen in der gleichen Reihenfolge wie in Abschnitt 2 eingeführt abgetragen; V=Vorgesetzter, FK=Führungskraft, MA=Mitarbeiter.

Vorgesetzten strenger eingestuft als von ihren Mitarbeitern.

Die Übereinstimmung der Beurteilungen (vgl. Tabellen 2 u. 3) ist generell zwischen Mitarbeitern und Vorgesetzten sowie zwischen Führungskräften und Vorgesetzten höher als zwischen Mitarbeitern und Führungskräften (mit Ausnahme der Dimension „Qualitätsorientierung“). Die Fremdurteile (MA mit V → FK; vgl. die Anmerkungen zu Tabelle 2) korrelieren jeweils höher miteinander als die Fremdurteile mit der Selbsteinstufung.

Die Ähnlichkeiten der Profilverläufe der Selbst- und Fremdbeurteilungen sind im Falle der VVE geringer als bei der VBS. Die größte Ähnlichkeit findet sich beim Verhaltensrangprofil.

Bemerkenswert ist, dass trotz moderater Korrelationen der Profilverlauf von Selbst- und Fremdbeur-

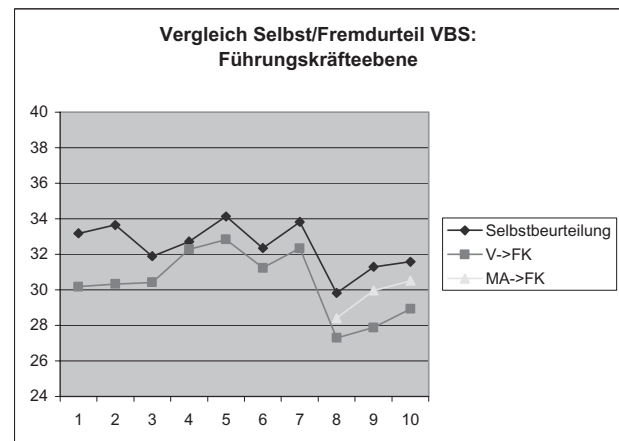


Abbildung 3. Mittelwertsvergleiche auf der Führungskräfteebene für das Beurteilungsverfahren VBS

Anmerkungen: Aggregationen von 8 Items; Verankerung 1=„fast nie“ bis 5=„fast immer“; auf der Abszisse sind die Dimensionen in der gleichen Reihenfolge wie in Abschnitt 2 eingeführt abgetragen; V=Vorgesetzter, FK=Führungskraft, MA=Mitarbeiter.

teilungen auf allen Beurteilerebenen und bei allen Beurteilungsverfahren sehr ähnlich ist. Dies spricht dafür, dass für die Differenzen eher kontextbezogene Gesichtspunkte ausschlaggebend sein dürften als Unterschiede in der Einschätzung der relativen anforderungsbezogenen Stärken und Schwächen.

### Beurteilerübereinstimmung

Der einzige Fall, in dem die Übereinstimmung zwischen den Beurteilern einer Ebene errechnet werden kann, ist die Beurteilung von Führungskräften durch ihre Mitarbeiter mittels VBS. Es ergibt sich eine mittlere Übereinstimmung von .41 (Cronbachs  $\alpha$ ). Bei größeren Gruppen erhöht sich die Beobachterübereinstimmung: alle Mitarbeitergruppen mit  $N \geq 14$  weisen einen Übereinstimmungswert von mindestens .60 auf.

Tabelle 2. Zusammenhangsanalysen auf der Führungskräfteebene

	FK <sub>Selbst</sub> mit V	FK <sub>Selbst</sub> mit MA	MA mit V → FK
<b>VVE:</b>			
Allg. Anforderungsdimensionen	.34 (54–60)	-	-
Führungsdimensionen	.26 (56–57)	.12 (79–80)	.35 (55)
<b>VBS:</b>			
Allg. Anforderungsdimensionen	.22 (54–60)	-	-
Führungsdimensionen	.20 (57)	.17 (79–80)	.33 (54–55)

*Anmerkungen:* Angegeben sind die über die Dimensionen gemittelten Produkt-Moment-Korrelationen zwischen den Urteilen, Angaben in Klammern= Anzahl der Urteilspaare, V=Vorgesetzter, FK<sub>Selbst</sub>=Selbstbeurteilung Führungskraft, MA=Mitarbeiter. MA mit V → FK= Korrelation zwischen Mitarbeiterbeurteilung und Vorgesetztenbeurteilung bezogen auf die gleiche Führungskraft.

Tabelle 3. Zusammenhangsanalysen auf der Mitarbeiterebene

	MA <sub>Selbst</sub> mit FK	MA <sub>Selbst</sub> mit V
<b>VVE:</b>	.25 (240–295)	.28 (59–64)
<b>VBS:</b>	.19 (230–290)	.22 (63–66)

*Anmerkungen:* gemäß Tabelle 2; auf Mitarbeiterebene wurden nur die sieben allgemeinen Anforderungsdimensionen verwendet.

Dies liegt deutlich über den in der Literatur berichteten Werten und lässt sich dadurch erklären, dass extreme Meinungen weniger ins Gewicht fallen.

### Einschätzung der Bedeutsamkeiten

Bei der Probeanwendung stuften alle Beurteiler die Bedeutsamkeiten der Anforderungsdimensionen ein. Für die VVE und VBS wurde eine normative (vs. VRP=ipsativ) Bewertung vorgenommen; beim VRP wurden die Anforderungsdimensionen gemäß ihrer Bedeutsamkeit in eine Rangreihe gebracht. Die Bedeutsamkeit der Anforderungsdimensionen wird sowohl von Mitarbeitern als auch von Führungskräften im Rahmen der Selbstbeurteilung höher eingeschätzt als bei der Fremdbeurteilung.

In der Sparkassen-Finanzgruppe werden die drei Tätigkeitsfelder „Markt“, „Betrieb“ und „Stab“ unterschieden. Diese Tätigkeitsfelder sollten sich in ihrem Anforderungsprofil unterscheiden. Substanzielle Differenzen ergeben sich für VVE und VBS lediglich zwischen den Tätigkeitsfeldern „Markt“ und „Nicht-Markt“ (= Stab und Betrieb). Erwartungsgemäß wird der Dimension „Kundenorientierung“ im Marktbereich die größte, im Nicht-Marktbereich die geringste Bedeutsamkeit unter allen Dimensionen zugesprochen. Dementsprechend ist im Anschluss an die Probeanwendung eine Modifikation dieser Anforderungsdimension für die Gruppe der Nicht-Marktmit-

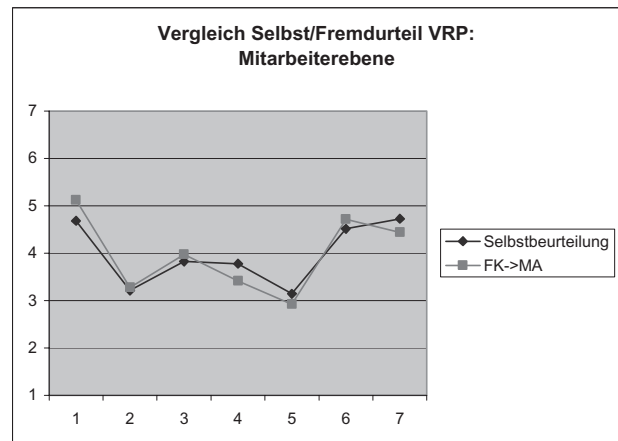


Abbildung 4. Mittelwertsvergleiche auf der Mitarbeiterebene für das Beurteilungsverfahren VRP

*Anmerkungen:* angegebene Werte=gemittelte Rangplätze, auf der Abszisse sind die Dimensionen in der gleichen Reihenfolge wie in Abschnitt 2 eingeführt abgetragen; FK=Führungskraft, MA=Mitarbeiter.

arbeiter vorgenommen worden (neue Dimension: Interne Kundenorientierung).

Anders verhält es sich mit den Bedeutsamkeitseinschätzungen, die beim VRP durch die Rangreihung der Stellenanforderungen zustande kommen. Hier unterscheiden sich die Dimensionen in ihrer Bedeutsamkeit im Vergleich zu den anderen Skalentypen deutlicher. (Das Verfahren ergibt seinem ipsativen Prinzip gemäß zwar immer die gleiche Rangplatzsumme, ermöglicht aber selbstverständlich unterschiedliche Profile.) Das VRP leistet also nicht nur bei der Beurteilung der Person eine deutlichere Merkmalsdifferenzierung, sondern auch beim Arbeitsplatz.

### Itemanalyse der VBS

Die Erfassung jeder Urteilsdimension mit jeweils acht Einzeleinstufungen bei der VBS erlaubt die Berechnung einer Itemanalyse. Die Berechnungen ergeben

mit einer Ausnahme für alle Dimensionen bei der Selbstbeurteilung wie bei der Fremdbeurteilung von Mitarbeitern sowie von Führungskräften und Vorgesetzten zufriedenstellende bis hohe interne Konsistenzen. Sie liegen für die Selbstbeurteilungen mit einer Ausnahme zwischen  $\alpha=.76$  und  $\alpha=.86$ . Die Konsistenzen der Fremdbeurteilungen liegen in einem Wertebereich von  $\alpha=.78$  bis  $\alpha=.92$ .

### Zusammenhangsanalysen der drei Beurteilungsskalen

Will man die drei eingesetzten Skalen untereinander vergleichen, so ist zunächst eine Transformation der Skalen durchzuführen. Während die VVE und die VBS normative Skalen sind, stellt das VRP eine Ipsativskala dar. Um die drei Skalen vergleichen zu können, ist es erforderlich, die normativen Werte der beiden erstgenannten Skalen zu ipsativieren, also in eine Rangreihe zu bringen. Es stellt sich die Frage, ob die nach der Transformation erhaltenen Rangreihen große Ähnlichkeiten oder markante Unterschiede aufweisen.

Abbildung 5 zeigt die Zusammenhänge zwischen den Rangreihen bei der Selbstbeurteilung je separat für die Ebene der Mitarbeiter und für die Ebene der Führungskräfte. Bei den Führungskräften ist für das Instrumentenpaar VVE-VBS eine signifikant höhere Korrelation ( $r=.58$ , bei dieser wie bei allen in diesem Abschnitt berichteten Korrelationen handelt es sich um Rangkorrelationen) zu beobachten als für die anderen beiden Instrumentenpaare (VVE-VRP:  $r=.48$ , VBS-VRP:  $r=.46$ ). Im Unterschied hierzu ist bei den Mitarbeitern keine so deutliche Tendenz zu sehen. Während die Korrelationskoeffizienten der Instrumentenpaare VVE-VBS und VVE-VRP ähnlich hoch sind ( $r=.59$  bzw.  $r=.58$ ), ist der Zusammenhang des Instrumentenpaares VBS-VRP signifikant niedriger ( $r=.49$ ). Allerdings weist auch auf der Ebene der Mitarbeiter die Korrelation zwischen VVE und VBS den höchsten Wert auf.

Die Ergebnisse deuten darauf hin, dass sich bei der Selbsturteilung die VVE und die VBS auch nach der Ipsativierung weniger voneinander unterscheiden als vom VRP. Für den Einsatz der Leistungsbeurteilungsskalen in der Sparkassen-Finanzgruppe stellt sich die Frage nach ihrer Eignung. Anhand des obigen Ergebnisses ist insofern zu empfehlen, eine der beiden normativen Skalen (VVE, VBS) auszuwählen. Das VRP scheint eine andere Perspektive darzustellen. Aus diesem Grund könnte es gut mit der VVE oder der VBS kombiniert werden, um zusätzliche Information über eine Person zu liefern – zumal das Verhaltensrangprofil durch den methodenimmanenten Vergleich zwischen Person und Arbeitsplatz ja in seinen Funktionen über Einstufungsskalen hinausgeht, nämlich Maßnah-

men der Personalentwicklung und Arbeitsplatzzuordnung sowie der Arbeitsplatzgestaltung nahe legt.

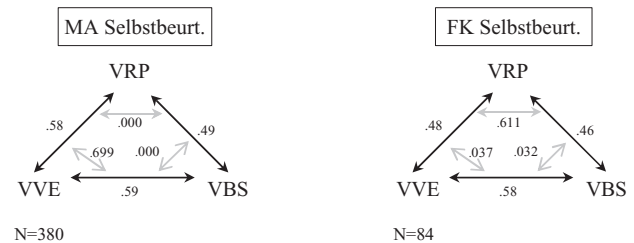


Abbildung 5. Interkorrelation von Selbstbeurteilungen mittels verschiedener Beurteilungsverfahren

Anmerkung: Die drei (äußeren) schwarzen Pfeile markieren Korrelationskoeffizienten, die drei (inneren) grauen Pfeile Signifikanzwerte.

Bei der Fremdbeurteilung (ohne Abbildung) zeigt sich ein ähnliches Bild: Die Korrelationen der Instrumentenpaare VVE-VBS sind sowohl für Mitarbeiter als auch für Führungskräfte signifikant höher als die der beiden anderen Instrumentenpaare (VVE-VRP, VBS-VRP). Insgesamt wird wie bei der Selbstbeurteilung deutlich, dass das VRP andere Beurteilungsinformationen beisteuert. Aus diesem Grund ist seine ergänzende Nutzung zu empfehlen.

### Kundenbeurteilung

Die Varianz der Kundenbeurteilungen unterlag extremen Varianzeinschränkungen dahingehend, dass nahezu ausnahmslos gute bis sehr gute Einstufungen durch die Kunden vorgenommen wurden. Damit ist erklärlich, dass die Übereinstimmung zwischen Kundenbeurteilungen und den übrigen Beurteilungen gering und inkonsistent wird. Dies hat zur Folge, dass die Kundenbeurteilung nur in Einzelfällen zur Urteilsdifferenzierung beiträgt. Da überdies nicht ausgeschlossen werden konnte, dass bei der Erfassung der Beurteilungen teilweise eine „Positivselektion“ der Kunden stattgefunden hat, wurde in der Organisation beschlossen, zunächst auf Kundenbeurteilungen zu verzichten.

### Zusammenfassung und Fazit

Die Ergebnisse der Probeanwendung belegen die Brauchbarkeit der drei Beurteilungsverfahren Verhaltensverankerte Einstufungsskala (VVE), Verhaltensbeobachtungsskala (VBS) und Verhaltensrangprofil (VRP). VVE und VBS können als äquivalent angesehen werden, während das VRP sich aufgrund teilweise anderer Funktionen als Ergänzung zu einer der beiden Einstufungsskalen anbietet.

Beurteilungen durch Führungskräfte (Vorgesetzte), durch Mitarbeiter und Selbstbeurteilungen tragen substantielle Qualifikationsinformation bei, sodass die weitere Einbeziehung aller drei Gruppen in den Beurteilungsprozess empfohlen werden kann. Für die Gruppe der Kunden gilt dies nur eingeschränkt. Hier hätte die Beurteilung stärker motivationale als informatorische Funktion.

Die Ableitung der Beurteilungsdimensionen aus den spezifischen Anforderungsanalysen findet darin ihre Bestätigung (Inhaltsvalidierung), dass alle Anforderungsdimensionen für alle Beurteilengruppen sowohl im Selbst- wie im Fremdurteil als bedeutsam eingeschätzt werden.

Die Anbindung an die gleiche Anforderungsgrundlage ermöglicht den unmittelbaren Bezug von Beurteilungsverfahren und Potenzialanalyseverfahren aufeinander. Die wesentlichen Unterschiede liegen in der Verhaltensorientierung der Beurteilungsverfahren gegenüber der Eigenschaftsorientierung der Potenzialanalyse sowie in der Standardisierung letzterer, während die Leistungsbeurteilung in den individuellen Tätigkeitskontext eingebettet ist. Die operationale Formulierung der einzelnen Beurteilungsaussagen begünstigt die Lenkung der Aufmerksamkeit der Beurteiler auf leistungsrelevantes Verhalten.

Die Beurteilungsverfahren haben die Erprobung unter der erschwerenden Bedingung bestanden, dass im Vorfeld keine tiefere Beurteilerschulung unternommen werden konnte. Inzwischen wurden zweitägige übungssintensive Beurteilungs- und Gesprächstrainings für Vorgesetzte und abgekürzte Beurteilungstrainings für beurteilende Mitarbeiter durchgeführt. Daneben können weitere Maßnahmen zur Verbesserung der Beurteilungsqualität getroffen werden, wie z.B. Feedback an die Beurteiler über Kennwerte ihrer Beurteilungen. Damit kann sichergestellt werden, dass die Multimodale Leistungsbeurteilung ein wichtiger Baustein zur Weiterentwicklung der Personalarbeit in der Sparkassen-Finanzgruppe wird.

## Literatur

- Backhaus, J. & Wagner, R. (2003). *Ausbilder-Taschenbuch 2003*. Stuttgart: Deutscher Sparkassen Verlag.
- Brandstätter, H. & Schuler, H. (1974). *Overcoming halo and leniency: A new method of merit rating*. Paper presented at the 18th International Congress of Applied Psychology, Montreal.

- Conway, J. & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 361–380.
- Conway, J. M., Lombardo, K. & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance, 14*, 267–303.
- Druskat, V. & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology, 84*, 58–74.
- Guzzo, R. A., Jette, R. D. & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology, 38*, 275–291.
- Latham, G. & Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. *Personnel Psychology, 30*, 255–268.
- Moser, K. (1999). Selbstbeurteilung beruflicher Leistung: Überblick und offene Fragen. *Psychologische Rundschau, 50*, 14–25.
- Muck, P. M., Schuler, H., Becker, K. & Diemand, A. (in Druck). Konzeption und Prüfung eines multimodalen Systems der Leistungsbeurteilung: Gruppenmodul. *Zeitschrift für Personalpsychologie*.
- Scherm, M. & Sarges, W. (2002). *360°-Feedback*. Göttingen: Hogrefe.
- Schuler, H. (1991). Leistungsbeurteilung – Funktionen, Formen und Wirkungen. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 11–40). Göttingen: Hogrefe.
- Schuler, H. & Marcus, B. (in Druck). Beurteilung beruflicher Leistung. In H. Schuler (Hrsg.), *Enzyklopädie der Psychologie (D/III/3)*. Göttingen: Hogrefe.
- Smith, P. & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149–155.
- Staufenbiel, T. & Kleinmann, M. (2002). PaiRS: Ein Skalierungsverfahren für die Eignungsdiagnostik. *Zeitschrift für Personalpsychologie, 1*, 27–34.

Prof. Dr. Heinz Schuler

---

Lehrstuhl für Psychologie (540F)  
 Universität Hohenheim  
 70593 Stuttgart  
 E-Mail: schuler@uni-hohenheim.de