Leistungsbeurteilung im öffentlichen Dienst: zur Validität des analytischen Beurteilungssystems LBB-SYS

Fietze, Simon; Holtmann, Doris; Matiaske, Wenzel

Erstveröffentlichung / Primary Publication
Arbeitspapier / working paper

Werk statt
für Organisations- und
Personalforschung e.V.

Leistungsbeurteilung im öffentlichen Dienst

Zur Validität des analytischen Beurteilungssystems LBB-SYS

Simon Fietze, Doris Holtmann und Wenzel Matiaske

Berichte der Werkstatt für Organisations- und Personalforschung e.V., ISSN 1615-8261

Die Autoren:

Dr. Simon Fietze ist wissenschaftlicher Mitarbeiter an der Professur für Betriebswirtschaftlehre, insb. Leadership and Labour Relations an der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg und Mitglied der Werkstatt für Organisations- und Personalforschung e.V. Berlin.
E-Mail: simon.fietze@werkstatt-opf.de

Dr. Doris Holtmann ist wissenschaftliche Mitarbeiterin an der Professur für Betriebswirtschaftlehre, insb. Leadership and Labour Relations an der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg und Mitglied der Werkstatt für Organisations- und Personalforschung e.V. Berlin.
E-Mail: doris.holtmann@werkstatt-opf.de

Prof. Dr. Wenzel Matiaske ist Professor für Betriebswirtschaftlehre, insb. Leadership and Labour Relations an der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg, Forschungsprofessor am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) und Mitglied der Werkstatt für Organisations- und Personalforschung e.V. Berlin.
E-Mail: wenzel.matiaske@werkstatt-opf.de

Berichte der Werkstatt für Organisations- und Personalforschung e.V.

Bericht Nr. 24, Berlin 2012

ISSN 1615-8261

**Kontakt zur Werkstatt für
Organisations- und Personalforschung e.V.:**

Dr. Karin Reichel
Offenbacher Str. 5
14197 Berlin

email: kontakt@werkstatt-opf.de
Internet: www.werkstatt-opf.de

Leistungsbeurteilung im öffentlichen Dienst

Zur Validität des analytischen Beurteilungssystems LBB-SYS

Simon Fietze Doris Holtmann Wenzel Matiaske

1 Vorbemerkung

Mit der Vereinbarung des Tarifvertrages für den öffentlichen Dienst (TVöD) gehören Leistungsentgeltsysteme zum Standardinstrumentarium des Personalmanagements öffentlicher Organisationen (Holtmann 2008). Die Empirie zur Praxis des §18 TVöD zeigt allerdings eine große Varianz in der Anwendung und damit auch in der Bandbreite der eingesetzten Instrumente zur Bestimmung der Leistungsentgelte: Neben einfachsten Beurteilungsbögen finden vor allem Zielvereinbarungssysteme aber auch komplexere analytische Verfahren Anwendung (Matiaske et al. 2007; 2012). Insbesondere für letztere Verfahrensgruppe sollte die Validitätsprüfung – da sich beispielsweise Merkmale der Messsituation, der Sprachgebrauch oder implizite Normen verändern – eine regelmäßige Aufgabe sein (Kersting/Hornke 2003).

Der vorliegende Bericht stellt mit dem LBB-SYS ein analytisches Leistungsbeurteilungsinstrument mit partizipativen Elementen in den Mittelpunkt, welches wir im kommenden Abschnitt näher vorstellen. Der dritte Abschnitt dieses Beitrages erläutert zentrale Befunde zur wiederholten Validitätsprüfung des Instrumentes. Dabei orientieren wir uns, wie auch in der Vorgängerstudie (Holtmann et al. 2001), an den Kriterien der klassischen Teststheorie. Da das LBB-SYS nicht nur in der Praxis der Leistungsbeurteilung, sondern auch in der Forschung zur Beobachtung von individuellen und kollektiven Effekten von Leistungsentgelten eingesetzt wird, können wir in einem weiteren Abschnitt Befunde zur wahrgenommenen Verteilungs-, Verfahrens- und Rückkopplungsgerechtigkeit berichten. Wir schließen mit einer Zusammenfassung sowie einigen Hinweisen zu möglichen Verbesserungen im Aufbau und beim Einsatz des Instrumentes.

2 Kurz vorgestellt: Das LBB-SYS

Das LBB-SYS – seit 2005 ein Produkt der LBB-SYS GmbH (www.lbb-sys.de) – wurde bereits Mitte der 1990er zur Leistungsbeurteilung im kommunalen Bereich entwickelt. Seither wurde das Instrument im Detail laufend überarbeitet und systematisch verbessert. Dies betrifft sowohl die Auswahl und Formulierung von Beurteilungskriterien als auch die Handhabung und insbesondere die informationstechnische Implementierung des Instrumentes, die Dateneingabe und die Aufberei-

tung der rückgekoppelten Ergebnisse. Der grundlegende Aufbau des Systems blieb allerdings weitgehend unverändert.

Das LBB-SYS ist ein analytisches Leistungsbeurteilungsinstrument, das in der Typologie der Dienst- und Betriebsvereinbarungen zum § 18 TVöD von Trittel et al. (2010, S. 40) dem „Selektion“ überschriebenen Basistypus zuzuordnen ist. Zielsetzung des LBB-SYS ist die Messung des Leistungsverhaltens und eine entsprechende Differenzierung von Leistungsentgelten. Angestrebt wird allerdings kein „exklusives“ Leistungsentgeltsystem, sondern es werden in Betriebs- oder Dienstvereinbarungen „organisationsspezifisch konkretisierende Regelungen“ über die Zuordnung von Punktwerten zu Leistungsentgelten getroffen. Insofern enthält das LBB-SYS im Verständis von Trittel et al. auch Elemente des Typus „Partizipation“. Die Beteiligung beschränkt sich dabei nicht auf die kollektive Ebene der Mitbestimmung. Die Mitarbeiter sind insbesondere auch im Rückkopplungsprozess beteiligt.

Bewertungsgruppe (Dimension)	Beispiel für ein Beurteilungskriterium (Item)
1. Arbeitseinsatz	Leistungsbereitschaft des Mitarbeiters
2. Leistungsergebnis	Umsetzung von Arbeitsrichtlinien/-vorgaben
3. Zusammenarbeit	Fähigkeit mit anderen zusammenzuarbeiten
4. Arbeitssystematik	Sorgfalt bei der Arbeitsausführung
5. Informationsverhalten	Bereitschaft sich fachlich in eine Diskussion einzubringen
6. Einsetzbarkeit	Einsetzbarkeit bei wechselnden Aufgaben

Tab. 1: Bewertungsgruppen und exemplarische Beurteilungskriterien

Auf Basis eines Beurteilungsbogens, der sich in sechs Bewertungsgruppen oder Dimensionen gliedert, werden zentrale Aspekte des Arbeitsverhaltens beurteilt. Nach der Revision des Instrumentes auf Grund der ersten Validierungsstudie 2001 wird jede Bewertungsgruppe durch sechs Kriterien konkretisiert. Ausgedrückt in der Sprache der Sozialforschung: Das Instrument umfasst sechs Dimensionen, die jeweils durch sechs Items operationalisiert werden. Die Dimensionen und jeweils eine typisierende Formulierung sind in Tabelle 1 zusammengestellt. Das Beurteilungssystem lässt sich modular um weitere Gruppen für spezifische Funktionen oder Organisationen erweitern. Führungskräfte werden beispielsweise anhand drei zusätzlicher Bewertungsgruppen – Verantwortung und Pflichterfüllung sowie auf Mitarbeiter und Organisation bezogenes Führungsverhalten – beurteilt. Wir konzentieren uns im Folgenden ausschließlich auf den Basiskatalog der in Tabelle 1 aufgeführten Dimensionen.

Die Bewertungsskalen zu jedem Kriterium oder Item umfassen jeweils acht Stufen: Neben der sogenannten Normalleistung können in fünf Stufen jeweils zehn Leistungspunkte vergeben werden. Ferner sind zwei Stufen vorgesehen, um zu notieren, dass die Leistungen bezüglich eines Kriteriums nicht den erwarteten An-

forderungen an den Mitarbeiter entsprechen. Inwieweit negative Beurteilungen zu einem Abzug von Punkten führen, bleibt der jeweiligen Organisation zur Regelung im Rahmen der Mitbestimmung überlassen.

Die Beurteilungen (Erstbeurteilung) finden i.d.R. jährlich und durch jeweils zwei Beurteilende – die direkte Führungskraft sowie eine weitere Person auf Leitungsebene – statt. Liegen die Ergebnisse beider Beurteilenden im Rahmen einer zuvor festgelegten Bandbreite, bildet das gemittelte Ergebnis die Grundlage für die Festlegung des Leistungsentgeltes. Überschreitet die Differenz in den Urteilen dagegen eine zuvor vereinbarte Grenze erfolgt eine Moderation zwischen den Urteilenden (Zweitbeurteilung). Im Fall einer erfolglosen Vermittlung wird ein dritter Urteilender hinzugezogen (Drittbeurteilung).

Die Rückkopplung der Ergebnisse erfolgt durch die Führungskraft im Rahmen eines Mitarbeitergesprächs. Im Konfliktfall kann der Mitarbeiter den Betriebs- oder Personalrat hinzuziehen. Das regelmäßige Gespräch institutionalisiert das Feedback zwischen Führungskraft und Mitarbeiter auf Grundlage der Leistungsbewertung und kann als Zielvereinbarungsgespräch ausgestaltet werden.

3 Verlässlichkeit und Gültigkeit des LBB-SYS

Die im Folgenden vorgestellten Auswertungen basieren auf vier jährlichen Regelbeurteilungen (2007 – 2010) aus drei kommunalen Organisationen. Dabei handelt es sich um eine Stadtverwaltung (rund 18.000 Einwohner), ein Landratsamt (rund 130.000 Einwohner) und das Spitalstift einer Kommune. Wie Tabelle 2 zeigt liegen insgesamt 3.554 Bewertungen von 1.778 Beschäftigten vor. Dabei überwiegen mit rund 62 % die weiblichen Beschäftigten. Diese Population unterscheidet sich auf Grund anderer Tätigkeitsbereiche deutlich von derjenigen der Vorgängerstudie (Holtmann et al. 2001), welche Leistungsbeurteilungen in Baubetriebshöfen und damit weit überwiegend männlich besetzten Arbeitsplätzen zur Grundlage hatte.

Jahr	Frauen		Männer		Bewertungen insg.	
	abs.	in %	abs.	in %	abs.	in %
2007	240	60,9	154	39,1	788	22,2
2008	256	61,1	163	38,9	838	23,6
2009	286	61,8	177	38,2	925	26,0
2010	314	62,5	188	37,5	1003	28,2
Insgesamt	1096	61,6	682	38,4	3554	

Tab. 2: Datenbasis (nur Erstbeurteilungen)

Auf dieser Datenbasis untersuchen wir zunächst die interne Konsistenz der Beobachtungsgruppen des LBB-SYS und deren faktorielle Struktur. Ferner analysieren wir die Inter-Rater- oder Übereinstimmungsverlässlichkeit zwischen den Beur-

teilern. Schließlich nutzen wir den Längsschnittcharakter der vorliegenden Datenbasis und analysieren die Verlässlichkeit des Instrumentes bei wiederholtem Messen.

3.1 Zeitpunktbezogene Verlässlichkeit und Gültigkeit

Die 36 Items der Skalen des LBB-SYS sind – abgesehen vom flachen Verlauf am linken Rand – näherungsweise normalverteilt und zeigen in ihren Häufigkeitsprofilen große Übereinstimmung. Das gemittelte Profil der Antworthäufigkeiten in Abbildung 1 ($\circ$) weist lediglich vereinzelt Bewertungen in den Kategorien 1 ($\bar{x} = 0,25$) und 2 ($\bar{x} = 9,25$), also unterhalb der als Normalleistung beschriebenen Kategorie 3, auf. D.h., nur bei einem Viertel der Items wurde überhaupt die Kategorie 1 verwendet und in die Kategorie 2 fallen im Durchschnitt weniger als 10 Urteile. Der Modalwert ist gleich der Bewertungstufe 6. Der Mittelwert für diese Kategorie beträgt $\bar{x} = 1115,28$ Beurteilungen.

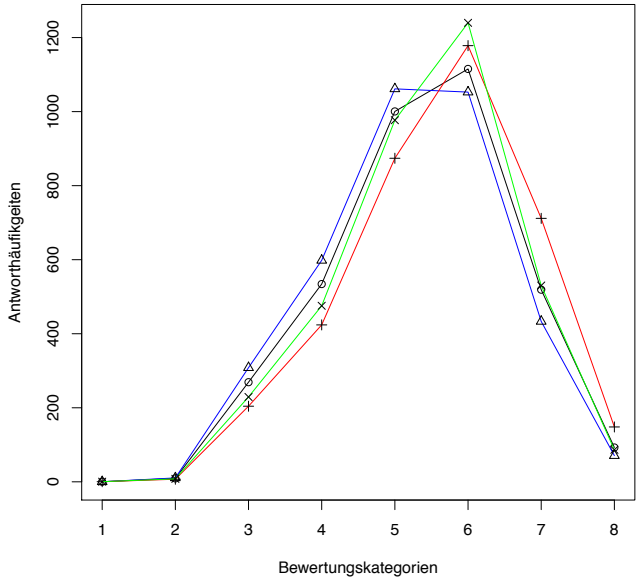

Abb. 1: Häufigkeitsprofile der Items

Abbildung 1 visualisiert auch die mittleren Häufigkeitsprofile nach Zugehörigkeit zu den Clustern einer Konfiguration auf Basis einer multidimensionalen Skalierung bzw. hierarchischen Clusteranalyse (Anhang A). Auffallend ist die hohe Ähnlichkeit der Profilverläufe. Cluster 3 ($\times$) und Cluster 2 ($+$) differieren geringfügig in der Höhe der Mittelwerte. Darüber hinaus ist die Verteilung der Items des Cluster 2 leicht nach rechts verschoben; es handelt sich also um weniger schwierige Items i.S. der Testtheorie. Cluster 1 ($\triangle$) enthält dagegen Items, die im Mittel geringfügig weniger gute Bewertungen mit sich bringen und daher entsprechend als

schwierigere Items gelten können. Der Modalwert dieser Gruppe, die den größten Teil der Items umfasst, entspricht der Ausprägung 5. Die Unterschiede zwischen den Clustern sind allerdings insgesamt gering und es überwiegt der Eindruck hoher Ähnlichkeit bezüglich der Items.

Die Übereinstimmung der Häufigkeitsprofile spiegelt sich in ähnlichen Lage- und Streuungskennwerten der Beurteilungskriterien (Anhang B, Tabelle 10) wieder. Ferner liegen die korrigierten Trennschärfen – die Korrelation des Beurteilungskriteriums mit der Summe der Beurteilungskriterien einer Gruppe ohne das jeweils betrachtete Item – durchweg im akzeptablen Bereich ($0,697 \leq r_{it} \leq 0,863$) eng beieinander.

Dimension	t_1	t_2	t_3	t_4	Gesamt
BG 1	0,95	0,94	0,96	0,95	0,95
BG 2	0,90	0,90	0,92	0,91	0,91
BG 3	0,92	0,91	0,93	0,92	0,93
BG 4	0,92	0,91	0,93	0,92	0,93
BG 5	0,93	0,93	0,94	0,93	0,93
BG 6	0,93	0,93	0,94	0,94	0,94

Tab. 3: Interne Konsistenzen der Bewertungsgruppen (Cronbachs α)

Ein wichtiges Kriterium zur Beurteilung der Verlässlichkeit ist Cronbachs α, ein Maß der internen Konsistenz einer Skala. Hohe Werte des zwischen 0 und 1 variierenden Koeffizienten drücken aus, dass die Beobachtungskriterien eine gemeinsame Skala bilden. Die internen Konsistenzen der Dimensionen oder Subskalen (Tabelle 3) fallen sowohl für die aggregierten Daten über alle Jahre als auch in differenzierter Analyse bezogen auf die Messzeitpunkte mit $\alpha \geq 0,90$ sehr hoch aus. Die auf die vier Zeitpunkte bezogenen Analysen lassen kein systematisches Muster oder Trend erkennen.

Ein weiterer Aspekt der Validitätsprüfung ist die Analyse der internen Struktur des Instrumentariums. Ein analytisches Instrument, das sich wie das LBB-SYS aus mehreren Bewertungsgruppen oder Subskalen zusammensetzt, sollte auch empirisch eine entsprechende Struktur erkennen lassen. Zum Zweck dieser Analyse betrachten wir zunächst die Korrelationen der Subskalen, skizzieren im Anschluss die Ergebnisse einer explorativen Faktorenanalyse und diskutieren im dritten Schritt kurz die Ergebnisse einer konfirmatorischen Faktorenanalyse.

Die Korrelationen der summierten Bewertungsgruppen fallen mit $r \geq 0,77$ recht hoch aus. Insbesondere die Gruppen 2 (Leistungsergebnis) und 4 (Arbeitssystematik) korrelieren stark. Berücksichtigt man bei Berechnung der Korrelation, dass die Variablen keine perfekte Reliabilität aufweisen und korrigiert den Rohwert für diese Minderung, steigt die Korrelation für den Zusammenhang dieser beiden Gruppen auf $r = 1,0$. D.h., beide Bewertungsgruppen messen (nahezu) Gleiches.

Diese Problematik zeigt sich auch bei Faktorenanalysen des Datensatzes. Tabelle 11 (Anhang C) zeigt, dass sich rund 74 % der Varianz mittels sechs Faktoren

Skala	BG 1	BG 2	BG 3	BG 4	BG 5	BG 6
BG 1	–	0,99	0,82	0,96	0,95	0,91
BG 2	0,92	–	0,85	1.00	0,95	0,91
BG 3	0,77	0,78	–	0,84	0,84	0,83
BG 4	0,90	0,92	0,78	–	0,97	0,94
BG 5	0,89	0,88	0,78	0,90	–	0,96
BG 6	0,86	0,84	0,77	0,88	0,89	–

Rohwerte in der unteren, minderungskorrigierte Werte in der oberen Dreiecksmatrix

Tab. 4: Interkorrelationen der LBB-SYS Skalen

aufklären lassen. Allerdings trägt der sechste Faktor λ_6 mit 3 % kaum mehr zur Varianzaufklärung bei und bündelt im Unterschied zu den anderen Faktoren keine einzelne Bewertungsgruppe. Vielmehr laden die hoch korrelierten Bewertungsgruppen 2 und 4 zumindest mit einer Reihe von Items gemeinsam auf dem fünften Faktor λ_5. Die übrigen Bewertungsgruppen sind mehr oder weniger deutlich, d.h. nicht immer mit allen Items bzw. mit einem Faktor korreliert.

Modell	Passungskennwerte				
	df	χ^2	GFI	$AGFI$	$RMSEA$
1-F.-Modell	594	17118,0	0,707	0,672	0,088
6-F.-Modell	579	9439,7	0,845	0,821	0,065

Tab. 5: Passungskennwerte der konfirmatorischen Modelle

Zur genaueren Klärung der Faktorenstruktur führten wir verschiedene konfirmatorische Faktorenanalysen durch. Zur Modellschätzung fanden Produkt-Moment-Korrelationen und eine Maximum-Likelihood Schätzung Anwendung. Hier sollen nicht alle Modelle detailliert, sondern nur die zentralen Befunde vorgestellt werden. Genauer haben wir ein Drei- und Fünf-Faktoren-Modell untersucht. Sowohl ein Drei-Faktor-Modell, analog zur oben berichteten Clusterlösung auf Basis der Häufigkeitsprofile, als auch ein Fünf-Faktoren-Modell, das die hoch korrelierenden Bewertungsgruppen 2 und 4 zusammenfasst, müssen aus Sicht der konfirmatorischen Faktorenanalyse als fehlspezifiziert gelten. Vielmehr verweisen die Modifikationsindices dieser Modelle auf veränderte Zuordnungen einzelner Items zu Faktoren. Letzteres gilt auch für das der Konstruktion entsprechende Sechs-Faktoren- und das alternativ getestete General-Faktor-Modell. Die Passungskennwerte in Tabelle 5 zeigen jedoch, dass das Sechs-Faktoren-Modell sowohl für die Anpassungswerte – der *goodness of fit (GFI)* sowie der um die Freiheitsgrade korrigierte *adjusted goodness of fit (AGFI)* sollten gegen 1,0 streben – als auch das Maß der Residuen – der *root mean square error of approximation (RMSEA)* sollte gegen 0 streben – den Daten gut angepasst ist. Der Test für den *GFI* gegen das Null-Modell fällt ebenso signifikant aus wie auch der Test auf verbesserte An-

passung zwischen dem einfacheren General-Faktor- und dem komplexeren Modell und Berücksichtigung des Zuwachses an Komplexität.

3.2 Inter-Rater-Verlässlichkeit

Ein zusätzliches Kriterium zur Einschätzung der Verlässlichkeit einer Messung ist die Inter-Rater-Reliabilität, d.h. der Grad der Übereinstimmung der Ergebnisse, die verschiedene Beurteiler mit einem Instrument erzielen. Das LBB-SYS fördert die Übereinstimmung mittels Beurteilerschulung und fordert diese durch den verbindlichen Grad an Übereinstimmung in den Urteilen. Zur Berechnung der Inter-Rater-Reliabilität sind daher nur die Daten der unmoderierten ersten Beurteilungsrunde verwendbar, um die Inter-Rater-Verlässlichkeit nicht zu überschätzen. Zu erwarten sind allerdings auch für diese Daten deutlich positive Korrelationen.

	t_1	t_2	t_3	t_4	Gesamt
Mittlere IRR	0,61	0,64	0,61	0,67	0,63
Minimum	0,47	0,54	0,50	0,55	0,52
Maximum	0,67	0,71	0,72	0,72	0,71

Tab. 6: Kennzahlen zur Inter-Rater-Reliabilität (IRR)

Tabelle 6 stellt die diesbezüglichen Befunde zusammen. Die Korrelationen auf Ebene der Einzelkriterien korrelieren im Mittel zwischen 0,61 und 0,67, d.h. die gemeinsame Varianz liegt zwischen 37 und 45 %. Die minimale und die maximale Korrelation zeigen, dass die Streuung um die mittlere Korrelation gering ist und keine Ausreißer zu verzeichnen sind.

3.3 Retest-Verlässlichkeit

Die Retest-Reliabilität, d.h. der Grad der Übereinstimmung von Messergebnissen bei denselben Personen zu mehreren Zeitpunkten, ist der klassische Indikator für die Verlässlichkeit eines Messinstrumentes. Dieser ist im Fall von Leistungsbeurteilungen allerdings nur begrenzt aussagekräftig, weil die Beurteilung und das jährliche Feedbackgespräch zwischen Vorgesetzten und Mitarbeitern eine Verhaltensänderung intendieren. Dennoch wollen wir diesen Indikator der Systematik halber und auch, um das Datenmaterial auszuschöpfen, kurz vorstellen.

	t_1	t_2	t_3
t_2	0,70		
t_3	0,64	0,72	
t_4	0,56	0,70	0,78

Tab. 7: Kennzahlen zur Retest-Reliabilität (IRT)

Tabelle 7 zeigt die Korrelationen zur Retest-Reliabilität im Längsschnitt für die hier analysierbaren Beurteilungswellen. Erwartbar ist sowohl eine positive Korrelation der Urteile über die Zeit als auch eine Abschwächung dieser Korrelation mit zunehmender Zeitdauer. Beide Hypothesen bestätigen die Korrelationskoeffizienten auf Basis der summierten Urteile in Tabelle 7. Die Kenntnis der Beurteilung im Vorjahr senkt den Vorhersagefehler der Beurteilung für das Folgejahr im Durchschnitt um mehr als 50 % ($\bar{x}(r)^2 = 0,73^2 = 0,54$). Bemerkenswert ist, dass die Konsistenz der Urteile bzw. des Verhaltens mit Kontinuität der Beurteilungen zunimmt. Möglicherweise handelt es sich um einen wechselseitigen – Führungskräfte oder Beurteiler bzw. Mitarbeiter oder Beurteilte betreffenden – Lerneffekt.

4 Akzeptanz der Leistungsbewertung

Ein personalwirtschaftliches Instrument soll nicht nur solide Messungen im Sinne sozialwissenschaftlicher Gültigkeit liefern, sondern auch Praktikabilität aufweisen. Mit Blick auf die Anwendbarkeit gilt daher die Akzeptanz im personalwirtschaftlichen Kontext als wichtiges Nebengüte-Kriterium zur Beurteilung von Messinstrumenten.

Im Rahmen unserer Begleitforschung zur leistungsorientierten Besoldung im „New Public Management“ konnten wir das hier vorgestellte Instrumentarium mehrfach als Referenzmodell verwenden. In verschiedenen Projekten war es möglich, das LBB-SYS in Befragungsprogramme einzubetten, die verschiedenen Standardvariablen von Mitarbeiterbefragungen wie die Arbeitszufriedenheit, den wahrgenommenen Handlungsspielraum der Tätigkeit oder die subjektive Gerechtigkeit erfassten. Im Folgenden beziehen wir uns auf ein Subsample des vorgestellten Datensatzes von 182 Mitarbeitern (65,9 % Frauen) des Landratamtes, in welchem wir 2010 eine zusätzliche Mitarbeiterbefragung durchführen konnten. Das Design der Studie ermöglicht die Kombination von Leistungsbeurteilungs- und Befragungsdaten der Mitarbeiter (Matiaske/Weller 2008).

Insbesondere die wahrgenommene Gerechtigkeit, hier in Anlehnung an die Differenzierung von Verteilungs-, formaler Verfahrens- und Interaktionsgerechtigkeit (Colquitt 2001) operationalisiert, gilt in der Personalforschung als Standard zur Beurteilung der Akzeptanz von Entgeltsystemen. Die Verteilungsgerechtigkeit umfasst die durch das Verfahren ermittelten materiellen Ergebnisse. Die Gerechtigkeit des Verfahrens ist für die Legitimität und damit für die Akzeptanz von Verteilungsergebnissen von Bedeutung. Dies gilt auch für die Interaktionsgerechtigkeit, womit die Rückkopplung des Vorgesetzten angesprochen ist.

Die hier verwendete Skala zur Messung der Gerechtigkeit gliedert sich in drei Subskalen mit jeweils fünf Items. Deren Verlässlichkeit ist als zufriedenstellend bis gut einzustufen. Cronbachs α beträgt 0.92 für die Verteilungs-, 0,90 für die formale Verfahrens- und 0,76 für die Interaktions-Gerechtigkeit. Der Handlungsspielraum wird hier über eine Kurzskala in Anlehnung an Müller-Böling (1978) gemessen ($\alpha = 0,87$). Zur Operationalisierung der Arbeitszufriedenheit verwenden wir eine

Frage, die dem Sozio-oekonomischen Panel entnommen ist (Matiaske/Mellewigt 2001).

Das LBB-SYS sollte mit den Subskalen der Gerechtigkeit nicht negativ korrelieren, um in der untersuchten Gruppe als akzeptiert zu gelten. Darüber hinaus sollte die Leistungsbewertung die Arbeitszufriedenheit nicht negativ beeinflussen. Mit dem Handlungsspielraum, der organisatorischen Voraussetzung zur Entfaltung von (intrinsischer) Leistungsmotivation, erwarten wir dagegen – sowohl mit den Ergebnissen der Leistungsbewertung als auch mit der Arbeitszufriedenheit – positive Korrelationen. Tabelle 8 gibt eine Übersicht der Befunde.

	LBB	Gerecht.	V-Gerecht.	F-Gerecht.	I-Gerecht.	AZ
Gerecht.	-0.02					
V-Gerecht.	-0.12	0.85				
F-Gerecht.	0.04	0.88	0.60			
I-Gerecht.	0.06	0.79	0.49	0.59		
AZ	0.08	0.08	0.14	0.07	-0.05	
Handlung	0.30	0.10	0.00	0.25	-0.02	0.17

Tab. 8: Korrelate der Leistungsbeurteilung

Insgesamt entsprechen die Korrelationen zur Akzeptanz der Leistungsbeurteilungen den Erwartungen. Die Ergebnisse der Leistungsbeurteilung (LBB) korrelieren zwar schwach negativ allerdings nicht signifikant mit der Verteilungsgerechtigkeit (V-Gerecht. $r = -0,12; p = 0,085$). Für die formale Gerechtigkeit des Verfahrens und die Interaktionsgerechtigkeit sind dagegen keine nennenswerten Zusammenhänge zu verzeichnen. Auch die Arbeitszufriedenheit bleibt von der Leistungsbeurteilung unbeeinflusst. Ein mittlerer signifikanter Zusammenhang besteht dagegen erwartungsgemäß mit dem Handlungsspielraum als Voraussetzung der individuellen Leistungsentfaltung.

5 Resümee

Die hier diskutierten Befunde zur Güte des Leistungsbeurteilungs- und -bewertungsystems LBB-SYS weisen dieses Instrumentarium aus Perspektive der klassischen Testtheorie als solides Messinstrument aus. Die Kennwerte zur Verlässlichkeit im Quer- und Längsschnitt sowie zur Inter-Rater-Reliabilität sind gut. Zufriedenstellend sind auch die Hinweise zur Dimensionalität des Instrumentes. Allerdings bilden die Subskalen kein orthogonales System, sondern die Dimensionen sind erwartbar stark interkorreliert. Darüber hinaus korrelieren auch einzelne Beurteilungskriterien stark mit Items anderer Beurteilungsgruppen. Gegebenenfalls wäre die Revision einzelner Items angezeigt, um eine bessere Trennung der Dimensionen zu erzielen. Da die Subskalen jedoch insgesamt zu einem Summenscore verrechnet werden, ist statistische Unabhängigkeit der Dimensionen oder Bewertungsgruppen nicht erstrebenswert.

Die externen Korrelate der Leistungsbeurteilungsergebnisse indizieren darüber hinaus externe Kriteriumsvalidität. Die Korrelation mit Daten aus einer Mitarbeiterbefragung zeigen, dass der Handlungsspielraum der Tätigkeit – wie erwartbar – Voraussetzung zur Entfaltung von Leistungsmotivation auf individueller Ebene ist. Arbeitszufriedenheit und wahrgenommene Gerechtigkeit bleiben dagegen weitgehend unberührt von den Ergebnissen der Leistungsbeurteilung. Prinzipiell stößt das Instrumentarium in der Anwendungspraxis also nicht auf Ablehnung oder Widerstände. Allerdings ist für eine Komponente der Gerechtigkeitswahrnehmung – der Verteilungsgerechtigkeit – eine schwach negative, wenn auch nicht-signifikante Korrelation zu verzeichnen. Dies verweist auf grundlegende Probleme von Leistungsentgelten im öffentlichen Dienst, deren Lösung eine grundlegende Leitvorstellung des vorgestellten Leistungsbewertungs und -beurteilungssystems ist.

Literatur

Colquitt, J. A. (2001): On the dimensionality of organizational justice: A construct validation of a measure. In: Journal of Applied Psychology, vol. 86, pp. 386–400.

Holtmann, D./Matiaske, W./Möllenhoff, D./Weller, I. (2001): Leistungsbeurteilung im öffentlichen Dienst: Zur Validierung des Leistungsbeurteilungs- und -bewertungssystems LBB-SYS. Arbeitspapier, Nr. 4, Berlin, Werkstatt für Organisations- und Personalforschung.

Holtmann, D. (2008): Funktionen und Folgen von Leistungsbeurteilungen: Eine Studie zur Einführung eines personalwirtschaftlichen Standardinstrumentariums in öffentlichen Verwaltungen. München, Mering.

Kersting, M./Hornke, L. F. (2003): Qualitätssicherung und -optimierung in der Diagnostik: Die DIN 33430 und notwendige Begleit- und Folgeinitiativen. In: Psychologische Rundschau, 54. Jg., S. 175–178.

Matiaske, W./Holtmann, D./Fietze, S. (2012): Zum Stand der Leistungsvergütung im öffentlichen Dienst: Ergebnisse einer Kommunalbefragung. Arbeitspapier, Berlin, Werkstatt für Organisations- und Personalforschung.

Matiaske, W./Holtmann, D./Weller, I. (2007): Leistungsvergütung im öffentlichen Dienst: Erwartungen und erste Erfahrungen – Ergebnisse einer Kommunalbefragung. In: Matiaske, W./Holtmann, D. (Hg.): Leistungsvergütung im öffentlichen Dienst. München, Mering. S. 79–86.

Matiaske, W./Mellewigt, T. (2001): Arbeitszufriedenheit: Quo vadis? Eine Längsschnittsuntersuchung zu Determinanten und Dynamik von Arbeitszufriedenheit. In: Die Betriebswirtschaft, 61. Jg., S. 7–24.

Matiaske, W./Weller, I. (2008): Leistungsorientierte Vergütung im öffentlichen Sektor: Ein Test der Motivationsverdrängungsthese. In: Zeitschrift für Betriebswirtschaft, 78. Jg., S. 35–60.

Müller-Böling, D. (1978): Arbeitszufriedenheit bei automatisierter Datenverarbeitung. München, Wien.

Trittel, N./Schmidt, W./Müller, A./Meyer, T. (2010): Leistungsentgelt in den Kommunen: Typologie und Analyse von Dienst- und Betriebsvereinbarungen. Berlin.

A Deskription der Items des LBB-SYS

Bewertungs-gruppe	Item	Ausprägung								Cluster
		1	2	3	4	5	6	7	8	
		absolute Häufigkeiten								Nr.
BG 1	1	0	1	179	410	771	1305	705	181	2
	2	0	4	196	409	891	1189	674	189	2
	3	1	9	232	497	976	1248	502	87	3
	4	0	9	256	531	1008	1115	491	142	1
	5	1	9	389	613	980	1017	447	96	1
	6	0	9	209	412	775	1175	788	184	2
BG 2	1	0	4	258	646	1086	1079	431	48	1
	2	0	13	281	524	985	1134	530	85	1
	3	0	19	311	566	1021	1065	470	100	1
	4	0	7	213	457	962	1258	557	98	3
	5	1	7	183	467	971	1113	689	121	2
	6	0	6	307	429	1085	1129	535	61	1
BG 3	1	1	9	213	487	928	1252	584	78	3
	2	0	11	276	611	1071	1083	435	61	1
	3	0	12	376	692	1200	1013	250	9	1
	4	0	9	192	446	1036	1231	547	91	3
	5	1	7	161	359	824	1260	769	171	2
	6	0	26	426	660	1105	1009	302	24	1
BG 4	1	0	7	188	410	904	1193	758	92	2
	2	0	9	273	458	928	1144	659	81	2
	3	0	8	277	444	1059	1276	436	52	3
	4	1	11	241	453	920	1108	654	164	2
	5	0	5	289	595	991	1065	536	71	1
	6	0	9	323	658	1150	962	396	54	1
BG 5	1	0	7	308	672	1022	1060	385	64	1
	2	1	15	298	652	1055	1041	401	55	1
	3	1	10	240	561	1087	1072	450	97	1
	4	0	8	208	437	882	1117	709	151	2
	5	0	4	281	582	1106	1104	393	48	1
	6	0	14	330	568	939	1102	472	93	1
BG 6	1	0	6	248	522	900	1174	555	113	3
	2	0	6	279	542	1058	1104	445	84	1
	3	1	11	335	616	1106	1015	367	67	1
	4	0	8	300	611	1123	1017	401	58	1
	5	0	12	315	582	1065	973	495	76	1
	6	0	12	297	652	1048	948	468	93	1

Tab. 9: Häufigkeitsprofile der Items

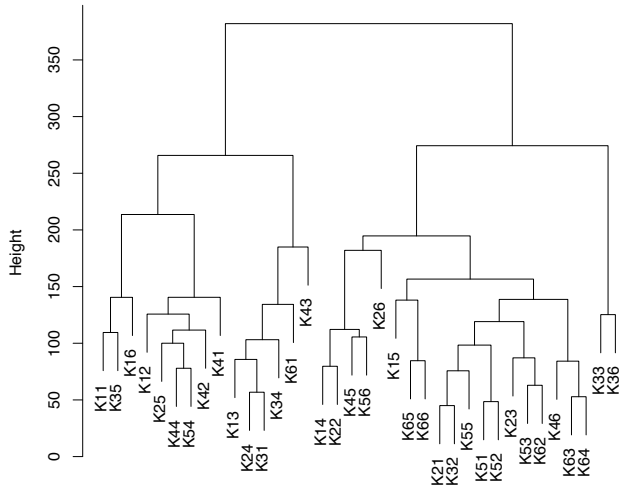

Abb. 2: Hierarchische Clusteranalyse (AGNES) der Häufigkeitsprofile

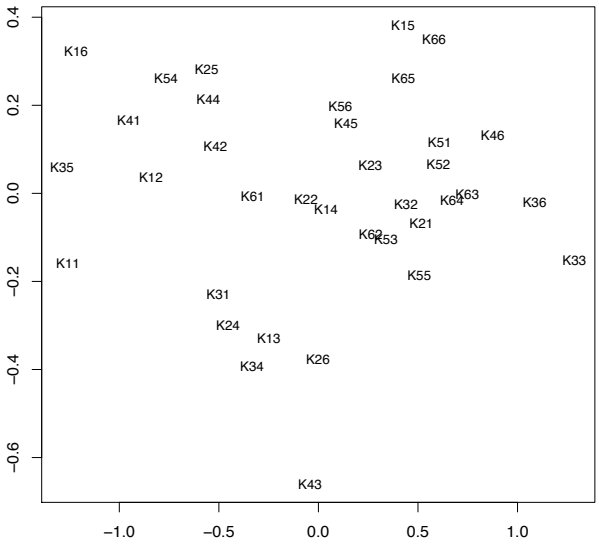

Abb. 3: Multidimensionale Skalierung (SMACOF N-MDS) der Häufigkeitsprofile

B Statistiken der Subskalen des LBB-SYS

Bewertungs-gruppe	Item	$\bar{x}$	s	α ohne Item i	korrigiertes r_{it}
BG 1	1	4.700	1.204	0.942	0.855
($\alpha = .95$)	2	4.645	1.226	0.941	0.858
	3	4.428	1.176	0.941	0.863
	4	4.409	1.233	0.941	0.863
	5	4.219	1.274	0.945	0.825
	6	4.689	1.259	0.944	0.835
BG 2	1	4.256	1.152	0.920	0.811
($\alpha = .91$)	2	4.373	1.221	0.915	0.851
	3	4.298	1.246	0.919	0.822
	4	4.496	1.171	0.916	0.841
	5	4.562	1.200	0.921	0.802
	6	4.372	1.191	0.934	0.697
BG 3	1	4.481	1.175	0.924	0.838
($\alpha = .93$)	2	4.265	1.180	0.929	0.797
	3	4.017	1.113	0.925	0.832
	4	4.490	1.148	0.927	0.817
	5	4.733	1.192	0.928	0.805
	6	4.027	1.190	0.928	0.806
BG 4	1	4.613	1.180	0.909	0.805
($\alpha = .93$)	2	4.471	1.228	0.912	0.783
	3	4.361	1.144	0.919	0.725
	4	4.545	1.264	0.908	0.809
	5	4.327	1.218	0.912	0.782
	6	4.165	1.183	0.910	0.798
BG 5	1	4.203	1.192	0.923	0.773
($\alpha = .93$)	2	4.202	1.191	0.920	0.799
	3	4.338	1.189	0.917	0.822
	4	4.601	1.238	0.917	0.820
	5	4.250	1.147	0.917	0.823
BG 6	6	4.300	1.253	0.924	0.770
($\alpha = .94$)	1	4.451	1.223	0.927	0.800
	2	4.321	1.191	0.920	0.861
	3	4.178	1.198	0.924	0.825
	4	4.215	1.176	0.923	0.833
	5	4.268	1.232	0.926	0.807
	6	4.250	1.237	0.933	0.755

Tab. 10: Statistiken der Subskalen

C Ergebnisse der Faktorenanalyse (MinRes)

Bewertungs-		Hauptkomponenten					
gruppe	Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
BG 1	1	0.02	0.10	0.13	0.71	0.03	-0.01
	2	0.17	0.08	-0.01	0.71	0.02	0.03
	3	0.07	0.14	0.09	0.45	0.18	0.15
	4	0.08	0.06	0.24	0.35	0.21	0.16
	5	0.06	0.02	0.45	0.34	0.13	-0.01
	6	0.02	0.05	0.45	0.18	0.18	0.24
BG 2	1	0.23	0.09	0.04	0.24	0.40	-0.03
	2	0.19	0.11	0.04	0.28	0.26	0.22
	3	0.35	0.08	-0.05	0.29	0.19	0.19
	4	0.09	0.09	0.19	0.24	0.32	0.15
	5	0.08	0.04	0.34	0.16	0.26	0.19
	6	-0.02	0.19	-0.02	0.07	0.66	-0.03
BG 3	1	-0.06	0.89	0.00	0.09	-0.05	0.01
	2	-0.06	0.83	-0.02	0.00	0.14	-0.10
	3	0.09	0.78	0.02	-0.04	0.07	-0.04
	4	0.15	0.66	0.00	0.08	-0.03	0.15
	5	0.04	0.73	0.01	0.11	-0.03	0.03
	6	0.13	0.71	0.10	-0.10	0.02	0.10
BG 4	1	0.04	0.13	0.20	0.16	0.39	0.17
	2	0.35	0.00	0.05	-0.03	0.53	0.05
	3	0.01	0.11	0.14	0.06	0.54	0.08
	4	0.08	0.06	0.55	0.13	0.11	0.13
	5	0.49	0.00	0.23	0.05	0.14	0.03
	6	0.58	0.13	0.12	-0.01	0.14	-0.01
BG 5	1	0.11	0.29	0.41	0.06	0.14	-0.14
	2	0.23	0.10	0.57	0.07	0.03	-0.15
	3	0.24	0.11	0.45	0.12	0.09	-0.07
	4	0.18	0.12	0.52	0.07	-0.03	0.22
	5	0.35	0.16	0.34	0.01	0.04	0.13
	6	0.39	0.20	0.04	0.13	0.11	0.11
BG 6	1	0.69	-0.03	0.07	0.08	0.01	0.10
	2	0.73	0.08	0.07	0.08	-0.07	0.05
	3	0.77	0.06	-0.04	0.04	0.06	0.01
	4	0.68	0.11	0.06	0.01	0.03	0.06
	5	0.61	0.04	0.06	0.11	0.10	0.02
	6	0.57	0.07	0.12	0.16	0.03	-0.22
Eigenwerte		7.05	5.81	4.55	4.29	3.88	1.12
Varianz in %		0.20	0.16	0.13	0.12	0.11	0.03
kumuliert		0.20	0.36	0.48	0.60	0.71	0.74

	λ_1	λ_2	λ_3	λ_4	λ_5
λ_2	0.76				
λ_3	0.77	0.58			
λ_4	0.75	0.71	0.68		
λ_5	0.63	0.68	0.59	0.64	
λ_6	0.36	0.24	0.33	0.41	0.33

Tab. 11: Oblimin-rotierte Ladungsmatrix und Interkorrelationen der Faktoren

Berichte der Werkstatt für Organisations- und Personalforschung e.V.:

01. **Weller, I./Steffen, E. 2000:** Ergebnisse einer Mitarbeiterbefragung bei der Lynx Consulting Group/Bielefeld. Berlin.

02. **Bendel, K. 2000:** Zufriedenheit von Nutzerinnen und Nutzern mit ambulanten Pflegedienstleistungen. Forschungsbericht. Berlin.

03. **Bendel, K./Matiaske, W./Schramm, F./Weller, I. 2000:** „Kundenzufriedenheit" bei ambulanten Pflegedienstleistern. Bestandsaufnahme und Vorschläge für ein stresstheoretisch fundiertes Messinstrument. Berlin.

04. **Holtmann, D./Matiaske, W./Möllenhoff, D./Weller, I. 2001:** Leistungsbeurteilung im öffentlichen Dienst. Zur Validierung des Leistungsbeurteilungs- und -bewertungssystems LBB-SYS. Berlin.

05. **Martin, A./Purwin, J. 2001:** Soziale Fähigkeiten in Arbeitsgruppen. Eine empirische Studie zur Ermittlung der Kooperationsfähigkeit. Berlin.

06. **Weller, I. 2001:** Fluktuationsneigung und Commitment. Eine empirische Betrachtung bei F&E-Mitarbeitern. Berlin.

07. **Matiaske, W./Holtmann, D./Weller, I. 2002:** Anforderungen an Spitzenführungskräfte. Retrospektive und Perspektive: Eine empirische Untersuchung. Berlin.

08. **Jütte, W./Matiaske, W. 2002:** Regionale Weiterbildungsnetzwerke. Eine exemplarische Analyse. Berlin.

09. **Holtmann, D./Matiaske, W./Weller, I. 2002:** Transparenz und Kommunikation als Erfolgsfaktoren von Leistungsbeurteilungen im öffentlichen Dienst. Vorstellung eines Forschungsprojektes. Berlin.

10. **Erbel, C. 2003:** Personalmanagement, Mitarbeiterverhalten und Kundenzufriedenheit im Dienstleistungskontakt. Eine empirische Analyse. Berlin.

11. **Weller, I./Matiaske, W. 2003:** Gütekriterien und faktorielle Struktur des IMC-Gitters zur Messung von Leistungs-, Macht- und Anschlussmotiven. Berlin.

12. **Schlese, M./Schramm, F. 2004:** Beschäftigungsbedingungen in der Gebäudereinigung – eine Analyse des Sozioökonomischen Panels. Berlin.

13. **Schramm, F./Zeitlhöfler, I. 2004:** Personalpolitik an Hochschulen. Eine Studie anhand der HWP – Hamburger Universität für Wirtschaft und Politik. Berlin.

14. **Bekmeier-Feuerhahn, S./Eichenlaub, A. 2004:** Ein Markenzeichen für die Universität: Wie kann die Identität der Universität in einem Bild verdichtet werden? Berlin.

15. **Schlese, M./Schramm, F. 2004:** Implikationen der Tarifverträge zur Leiharbeit für die Tarif- und Beschäftigungsbedingungen im Gebäudereiniger-Handwerk. Berlin.

16. **Weller, I./Matiaske, W. 2008:** Gütekriterien einer deutschsprachigen Version der Mini Markers zur Erfassung der „Big Five". Berlin.

17. **Wigger, A. 2008:** Managing organizational change: Application of the Biomatrix theory to the transformation of a non-profit organization, Berlin.

18. **Matiaske, W./Tobsch, V./Fietze, S. 2009:** Erfolgs- und Kapitalbeteiligung von Beschäftigten in Deutschland. Abschlussbericht einer repräsentativen Befragung, Berlin.

19. **Weller, I./Matiaske, W. 2009:** Leistungsorientierung und der Wechsel des Rahmens. Ein Erklärungs- und Messansatz für Extra-Rollenverhalten, Berlin.

20. **Fietze, S./Matiaske, W. 2009:** Podcast in der Lehre: Bericht über den Einsatz an der Helmut-Schmidt-Universität, Berlin.

Berichte der Werkstatt für Organisations- und Personalforschung e.V.:

21. **Fritz, M./Issa, N./Müller, G./Tuchtfeldt, S./Fietze, S./Kattenbach, R. 2011:** Der Arbeitskraftunternehmer. Erschöpfung und Arbeitszufriedenheit im JD-R Modell, Berlin.
22. **Olejniczak, M. 2011:** Arbeit im Kontext des SGB II - Personalwirtschaftliche Aspekte des Neuen Steuerungsmodells, Berlin.
23. **Olejniczak, M. 2011:** Hartz IV als Dienstleistung, Berlin.
24. **Fietze, S./Holtmann, D./Matiaske, W. 2012:** Leistungsbeurteilung im öffentlichen Dienst. Zur Validität des analytischen Beurteilungssystems LBB-SYS, Berlin.