

# Vergütung nach „gemessener“ Zielerreichung: Zufallsergebnisse untergraben die Motivation

Johannes Reich  
Wolfram Püchert

„Management by Objectives“ ist eines der nicht mehr ganz so neuen populären Managementschlagwörter: Mitarbeiter werden über Ziele geführt. Verbindet man das mit Entgeltbestandteilen, nennt man es „Pay-for-Performance“. Wer seine Ziele erreicht, bekommt mehr, wer sie nicht erreicht, bekommt weniger Geld. Die Ratgeberliteratur zu diesem Thema fordert, dass die Ziele „s.m.a.r.t.“ sind: specific, measurable, attainable, realistic and timely bzw. spezifisch, messbar, erreichbar, realistisch und zeitlich konkret. Eine einprägsame Formel! Aber sind derartige Ziele mit Entgeltbindung wirklich zielführend?

Eine einfache Betrachtung, die die Unwägbarkeiten dieser Welt in Form der sich natürlicherweise ergebenden Parameterstreuung einbezieht, weckt erhebliche Zweifel an dieser Konzeption. Objektiv messbare Zielerreichungen an Entgelt zu binden heißt, einen Rechtsanspruch auf Vergütung bei Erreichung bestimmter Parameterwerte zu vereinbaren. Die natürliche Streuung dieser Parameterwerte macht die Vereinbarungen zu einem Lotteriespiel für alle Beteiligten und demotiviert die Beschäftigten à la longue gerade wegen der hohen Anreizwirkung von Geld.

### EIN PRAXISBEISPIEL

In der Wartungsabteilung eines Softwareunternehmens lösen die Mitarbeiter Kundenprobleme (sogenannte Meldungen), ca. 100 im Jahr. Anschließend werden sie von den Kunden auf einer Skala von 1 für überragend bis zu 10 für indiskutabel bewertet. Es gibt die Zielgröße „Anzahl der Meldungen“, aber auch die Qualität der Problemlösungen soll gemessen werden. Da bieten sich die Kundenbewertungen an, denn wer könnte die Qualität besser beurteilen als die Kunden! Das Management stellt aufgrund der

bisherigen Daten fest, dass die Bewertungsmittelwerte der Kollegen in der Vergangenheit ungefähr bei 4,5 lagen. Man setzt den für 100 % Bonus zu erreichenden mittleren Qualitätswert auf durchaus anspruchsvolle 4,4–4,5 fest und jedes Zehntel weniger ergibt 10 % Bonus mehr und umgekehrt.

Ein Jahr später sichtet das Management die Zahlen zweier seiner Mitarbeiter: Einer hat 3,98, der anderer 4,81 erreicht. Ein *scheinbar* klarer Fall: Mitarbeiter 1 gehört zu den Guten und erhält 150 % Bonus, während Mitarbeiter 2 dagegen abfällt und nur 60 % Zielerreichung attestiert bekommt. Damit liegt er leider unter der 70%-Schwelle und erhält gar nichts, heißt es doch im Bonusrahmenwerk: „Damit ein Mindestgrad an Leistung erreicht wird, wird bei 70 % der Gesamtzielerreichung ein Schwellenwert festgelegt. Das bedeutet, dass kein Bonus ausgezahlt wird, wenn die Gesamtzielerreichung unter 70 % liegt.“

Dabei sei an dieser Stelle angemerkt, dass Unternehmen dazu übergehen, nicht nur einmalige variable Gehaltsbestandteile an solche „Leistungsmessungen“ zu binden, sondern auch Gehaltserhöhungen, die einen viel längerfristigen Effekt haben. Bei einem Jahreseinkommen von 70.000 € mag eine fehlerhafte Bonifizierung von 1.000 € zumindest finanziell verschmerzbar sein. Eine 1-%-Differenz bei einer Gehaltserhöhung macht jedoch im Verlauf von 30 Jahren über 20.000 € aus – ohne Zinseffekte!

### METHODENKRITIK

Wo liegen nun die Probleme? Sind solche Ziele vielleicht doch nicht s.m.a.r.t.? Zunächst ist die verwendete Bewertung eine Ordinalskala, die eine Reihung festlegt, aber keine Aussage macht, ob die Bedeutung aller Intervalle identisch ist, ob also 6 genauso viel besser ist als 3 im Vergleich etwa zu 5 und

2. Damit sind mathematische Operationen wie etwa Mittelwertbildung auf Basis dieser Zahlen eigentlich nicht sinnvoll, sondern nur der Vergleich von Verteilungen. Aber auch bei Schulnoten wird dies häufig außer acht gelassen und so wollen wir über diesen Einwand „pragmatisch“ hinwegsehen.

Tatsächlich wirken sich auf den Erfolg neben der Anstrengung der Kollegen viele Zufälle aus. Das eine Problem ist schwieriger als das andere, der eine Kunde ungeduldiger als der andere, mal ist ein hilfreicher Kollege greifbar, mal nicht, usw. Diese Zufälle führen dazu, dass bei angenommener gleicher Anstrengung und Fähigkeit der Mitarbeiter die Bewertungen der Arbeit durch die Kunden nicht immer dieselbe sein, sondern ganz im Gegenteil erheblich streuen wird.

Tatsächlich stammen die beiden Werte von 4,81 und 3,98 aus einer Modellrechnung. In dieser wurde von einer statistisch zufälligen Verteilung der Bewertungen ausgegangen, bei der positive Bewertungen häufiger als schlechte vorkommen, was die Praxis recht zutreffend wiedergibt.<sup>1</sup>

Die in *Abbildung 1* gezeigten Daten stammen also aus einer Modellrechnung,

<sup>1</sup> Die Verteilung heißt Gammaverteilung (z. B. <http://de.wikipedia.org/wiki/Gammaverteilung>), die mit den Werten  $p=2$ ,  $b=2$  parametrisiert wurde – der genaue Charakter ist jedoch unerheblich, da die Aussagen dieses Artikels für jede nicht entartete Verteilung gelten.

Johannes Reich ist Betriebsrat in der Softwareindustrie. Arbeitsschwerpunkte: Entgeltsysteme, Arbeit und Gesundheit, Weiterbildung.  
e-mail: Johannes.Reich@sophoscape.de  
Wolfram Püchert war Betriebsrat in der Softwareindustrie. Arbeitsschwerpunkte: Entgeltsysteme, Arbeitszeit, Reorganisationen, Betriebsänderungen.  
e-mail: weepee@gmx.de

die davon ausgeht, dass Mitarbeiter 1 und Mitarbeiter 2 ihre Arbeit wie eineiige Zwillinge völlig identisch ausführen und eine Streuung allein durch zufällige Einflüsse, die nicht durch die Mitarbeiter kontrolliert werden können, entsteht. Es werden die Häufigkeiten der Bewertungen von Mitarbeiter 1 und 2 gezeigt, die sich aus zwei verschiedenen Stichproben zu je 100 Werten ergaben, zusammen mit der theoretischen Verteilung der Grundgesamtheit.

Würde man unendlich viele Werte ziehen, käme genau die vorgegebene theoretische Verteilung heraus. Die Verschiedenheit der beiden Häufigkeitsverteilungen ist also allein der Endlichkeit der Stichproben geschuldet, die mit 100 Werten nicht gerade klein ist. Offensichtlich führt diese Endlichkeit der Stichprobe zu einer Streuung der Mittelwerte der Stichproben. Die durchschnittliche Abweichung dieser Mittelwerte der Stichproben vom tatsächlichen Mittelwert der Grundgesamtheit wird auch als Standardfehler bezeichnet.

Betrachtet man die Verteilung der Mittelwerte von 32 aufeinanderfolgend gezogenen Stichproben, was 32 identischen Mitarbeitern entspricht, die vom Kunden je ein Jahr lang für 100 Fehlermeldungen bewertet wurden, so zeigt sich das in *Abbildung 2* veranschaulichte Bild: Die Mittelwerte verteilen sich über einen Bereich von ca. 3,9–5,1, allein aufgrund der Tatsache, dass jede einzelne 100er-Serie der Bewertungen eines Mitarbeiters nur eine beschränkte Stichprobe darstellt.

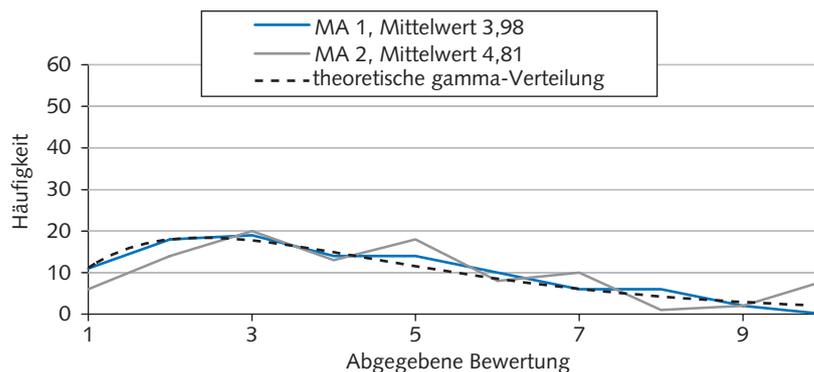
Je kleiner die Stichproben, desto mehr streuen ihre Mittelwerte. Ist die Stichprobengröße im Extremfall nur noch 1, so ist die Streuung dieser Werte gerade die ursprüngliche Streuung. Im Falle der Beurteilungen würde die Streuung damit den gesamten Bereich von 1–10 umfassen mit einer gewissen Häufung guter Noten. Je größer die Stichprobe, desto stärker nähert sich die Verteilung der Mittelwerte einer Gauss'schen Normalverteilung an.

Die vom Manager festgestellten Unterschiede lassen sich im vorgestellten Fall also nicht auf die einzelnen Mitarbeiter zurückführen und stellen damit auch keine Leistungsunterschiede dar. Sie sind verfahrensbedingte Artefakte.

## METHODISCHE KONSEQUENZEN

Was folgt aus dieser Kritik für Entgeltsysteme mit Zielkennziffern? Von einer ernstzunehmenden Leistungsperspektive

**Abb. 1: Verteilung der Mitarbeiterbewertungen**



Quelle: Darstellung der Autoren.

WSI MITTEILUNGEN

aus dürften Unterschiede zwischen Mitarbeitern erst gemacht werden, wenn man nachweisen kann, dass ihre Leistung sich auch tatsächlich in irgendeiner Hinsicht „signifikant“ unterscheidet.

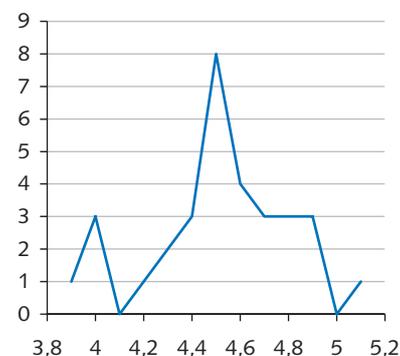
Um überhaupt sinnvoll „leistungsbezogene“ Ergebnisunterschiede zu identifizieren, ist es daher notwendig, die Streuung zu kennen, die sich allein durch die endlichen Stichproben ergibt. Diese hängt sowohl von der statistischen Verteilung der Werte der Grundgesamtheit wie auch von der Stichprobengröße ab. In der Statistik ist die Definition von „Signifikanz“ in gewisser Weise willkürlich und entscheidet über die Irrtumswahrscheinlichkeiten. Gewöhnlich werden 5% Irrtumswahrscheinlichkeit noch als akzeptabel angesehen. Im Einzelfall, je nach den möglichen Konsequenzen einer Fehlentscheidung, kann dieser Wert auch deutlich kleiner gewählt werden.

Arbeitet man mit mitarbeiterbezogenen Leistungsparametern, müssten für einen sinnvollen Vergleich also immer die mitarbeiterbezogenen Verteilungen miteinander verglichen werden<sup>2</sup> und keinesfalls nur die Mittelwerte. Folglich lassen sich aus dem Vergleich von Einzelwerten kaum sinnvolle praktische Konsequenzen ableiten. Die unmittelbare Bindung von messbaren Ergebnissen an Entgelt ist jedoch eine solche praktische Konsequenz, noch dazu eine mit erheblichen weitergehenden Folgen.

## PRAKTISCHE KONSEQUENZEN

Welche praktische Relevanz hat die methodische Kritik? Man könnte einwenden, dass diese formalen Betrachtungen nur Glasperlenspiele sind. Das Gegenteil ist der Fall!

**Abb. 2: Verteilung der Mittelwerte von 32 Mitarbeitern**



Quelle: Darstellung der Autoren.

WSI MITTEILUNGEN

Zunächst ist es gerade erwünscht, dass sich externe und eigene Einflüsse in der gleichen Größenordnung bewegen. Kann man seine Zielerreichung gar nicht beeinflussen, so gibt es selbstverständlich keinen Zusammenhang mit der eigenen Leistung, was nicht günstig für die Steuerung mit Zielen ist. Kann man sie vollständig selbst beeinflussen, gibt es also gar keine relevanten äußeren Einflüsse, spielt offensichtlich die Interaktion mit der Umwelt und insbesondere mit dem Unternehmen keine Rolle mehr, was aus Unternehmenssicht ebenfalls nicht erstrebenswert ist.

Zudem stellt sich die Frage nach Anreizeffekten. Für Zieldefinitionen werden zu Recht gerade die Bereiche für inte-

<sup>2</sup> Eine Möglichkeit wäre der Wilcoxon-Mann-Whitney-Test, <http://de.wikipedia.org/wiki/Wilcoxon-Mann-Whitney-Test>.

ressant erachtet, bei denen externe und eigene Einflüsse in einem ausgewogenen Verhältnis stehen. Hier kann die Fähigkeit des Menschen, komplexe Situationen zu meistern, zur vollen Entfaltung kommen. Für solche komplexen Tätigkeiten kommt hinzu, dass sie sich in der Regel nicht hinreichend standardisieren lassen, sodass Stichproben häufig tatsächlich keine hundert, sondern nur wenige Elemente umfassen.

Was passiert, wenn trotzdem unbeirrt an der unmittelbaren Verknüpfung von scheinbar messbarer Leistung mit Entgelt festgehalten wird? Sind die Voraussetzungen für eine solche unmittelbare Verknüpfung nicht gegeben, ergibt sich zwangsläufig, dass nicht Leistung, sondern statistische Artefakte durch Entgelt „honoriert“ werden. Aufgrund der starken Anreizwirkung von Geld wird jedoch eine ausbleibende Anerkennung leicht als Bestrafung empfunden. Bekanntermaßen reagieren Menschen auf als ungerecht empfundene Bestrafung mit ausgesprochener Verstimmung, während eine als ungerechtfertigt empfundene Belohnung häufig keine bleibenden Spuren hinterlässt. Daraus ergibt sich, dass eine unmittelbare

re konsequente Verknüpfung von Entgelt und scheinbar messbarer Leistung unter diesen Bedingungen einen Mechanismus darstellt, der über die Zeit eigentlich jeden nachhaltig demotiviert.

#### FAZIT

Die Aversion der Beschäftigten und ihrer Interessenvertretung gegenüber Leistungsmessung mit unmittelbarer Entgeltwirkung ist also gerechtfertigt. Die Bewertung quantitativer Unterschiede ist weitaus komplizierter, als es vom Management in der Regel praktiziert wird. Eine Arbeitsleistung an Hand von „messbaren“ Merkmalen zu bewerten, ist sicher ein bequemes Vorgehen, enthebt es doch den Beurteilenden von der mühsamen Pflicht, sich inhaltlich mit der Materie auseinanderzusetzen. Der tatsächlich erbrachten Leistung wird es aber selten gerecht.

Da die vorgebrachten Argumente nur geringe statistische Spezialkenntnisse voraussetzen, wie sie etwa Betriebswirten vermittelt werden, deutet die Popularität der Verbindung des s.m.a.r.t.-Konzepts mit Entgeltbestandteilen eher auf eine hohe Folgebereitschaft gegenüber sugges-

tiven Konzepten – man könnte auch sagen „Hypeanfälligkeit“ – der für diese Systeme Verantwortlichen hin. Eigentlich wäre es die Aufgabe des Managements, sich hier gegenüber den Zeitgeistströmungen zu emanzipieren und seinen Aufgaben ernsthaft nachzugehen.

Was bedeutet das für die Interessenvertretungen der Beschäftigten? Entgeltbestandteile an einfachen Mittelwerten – oder noch schlimmer an Einzelwerten – festzumachen, ist, wie gesehen, nahezu töricht. Es führt dazu, dass Unternehmen wertvolle Zeit mit statistischen Belanglosigkeiten verschwenden und die Beschäftigten systematisch demotivieren. Leistungsentgelte sind in Deutschland jedoch nicht nur den Grundsätzen sondern auch der Höhe nach mitbestimmungspflichtig (BetrVG §87(1) 10,11). Die Interessenvertretungen sind damit de jure für diese Systeme mitverantwortlich. Vom Gesetzgeber gewollt, haben die Interessenvertretungen einen annähernd ebenso großen Einfluss auf die Gestaltung dieser Entgeltformen wie der Arbeitgeber. Es wird höchste Zeit, dass sie dieser Mitverantwortung gerecht werden und ihre Gestaltungsmacht wahrnehmen.