

**Rainer König**

## **Johann Carl Friedrich und die Norm - 8 Thesen zur Normalverteilung**

1. Bei Personal-Beurteilungen sollte man zwischen den Aufgaben und Arbeiten der Personalverwaltung einerseits und denen der beurteilenden Führungskräfte andererseits unterscheiden:
  - Die einen haben neben rein administrativen Tätigkeiten (die z.B. mit Einstellung, Entlohnung, Fehlzeiten etc. zu tun haben) vor allem die Qualifikation, Entwicklung und Einsatz(planung) des Personals aus der Perspektive der Gesamtorganisation zu leisten.
  - Die anderen sind primär für die Betreuung, Qualifikation, Entwicklung und den Einsatz der ihnen direkt unterstellten Mitarbeiter zuständig.
  - Man kann das auch so ausdrücken: Die Personalverwaltung agiert auf der Makroebene, die Führungskräfte auf der Mikroebene des Systems Verwaltung. Beide treffen sich auf der Mesoebene, aber dazu später.
2. Auf der Makroebene benötigt die Personalverwaltung objektiv richtige (d.h. neutral überprüfbare) Daten über die Qualifikation und Leistung des Gesamt-Personals. Personal-Beurteilungen stellen dabei eine wichtige Datenquelle dar. Durch sie erhält die Verwaltung auch eine Gesamtschau über das Leistungsniveau und damit indirekt auch über die Qualifikation des Personals.
3. Die Qualität dieser Daten hängt aber von der Qualität ihrer Erhebung, d.h. von der Qualität der einzelnen Beurteilung ab. Von daher liegt es nicht nur im Interesse der Personalverwaltung, die Gesamtheit der Beurteilungen überblicken zu können. Sie muss darüber hinaus auch einschätzen können, ob sie einigermaßen objektiv und gerecht zustande gekommen sind. Mit anderen Worten: Die Verwaltung muss die Qualität der Beurteilungen beurteilen können. Und das ebenfalls so objektiv und gerecht wie möglich. Hier gibt es allerdings ein strukturelles Problem: Das Prüfen der Beurteilung durch die Verwaltung findet auf der Makroebene, ihre Erzeugung durch die je subjektive Beurteilung der Führungskräfte aber auf der Mikroebene statt. Direkt ist der Personalverwaltung diese Ebene nicht zugänglich.
4. Sie braucht deshalb Makro-Maßstäbe, die auch für die Mikroebene gültig sind. Diese Messlatten müssen einerseits quasi in die Mikroebene hineinstoßen und andererseits von ihr, d.h. von den beurteilenden Führungskräften unabhängig sein. Als Lösung bieten sich Vergleiche mit anderen Kommunen/Verwaltungen oder – noch besser - der Vergleich mit anderen Ebenen/Adressaten an, die die gleiche Leistung derselben Gruppe von Beurteilten (in Ausnahmefällen sogar einer einzigen beurteilten Person) aus unterschiedlichen Perspektiven beurteilen. So etwas erlauben z.B. 360-Grad-Beurteilungen oder systematisch erhobene Kundenbefragungen und – Beschwerden. Auf diese Weise ist zumindest ein indirekter Rückschluss auf die Qualität der Beurteilung durch die Führungskraft möglich: weicht z.B. deren Beurteilung zum Verhaltensmerkmal „Kundenumgang“ signifikant von den Ergebnissen der Zufriedenheitsbefragung bei den Kunden des Mitarbeiters ab, liegt eine Differenz vor, die geklärt werden sollte.
5. Häufig wird in diesem Zusammenhang auch die Normalverteilung (im Folgenden: GNV) des großen deutschen Mathematikers Johann Carl Friedrich Gauß als Maßstab genannt. Ihr zufolge sind Verteilungen, die durch Überlagerung einer großen Zahl von unabhängigen Einflüssen entstehen, unter schwachen Voraussetzungen annähernd normalverteilt. Für biologische, wirtschaftliche oder naturwissenschaftliche Vorgänge soll sich derart a priori ein Mittelwert z.B. als Trend berechnen lassen. Für soziale Verteilungen bestätigt die Gauß´sche Kurve im Prinzip nur die Alltagsweisheit, dass das Normale am häufigsten vorkommt und im Durchschnitt die Ausnahme die Ausnahme und eben nicht die Regel ist.

6. Gleichwohl sollten Personalverwaltungen bei der Qualitätsmessung ihrer Personal-Beurteilungen möglichst nicht auf die GNV zurückgreifen. Und das aus zumindest drei Gründen:
- Zum einen, weil die Gauß'sche Kurve ein rein statistischer Durchschnitts-Wert ist, den es im wirklichen Leben genau so wenig gibt wie z.B. das Klima<sup>1</sup> oder das Durchschnittseinkommen des Durchschnittsdeutschen. Solch abstrakte Größen ragen in die Mikroebene des konkreten Lebens überhaupt nicht hinein. Sie schweben vielmehr über ihm und stellen eine Art mathematischer Metawelt dar.
  - Zum zweiten ist immer zu klären, welche „Voraussetzungen“ genau vorliegen und ob diese wirklich „schwach“ sind. In sozialen Zusammenhängen ist das nur sehr bedingt bis gar nicht möglich. So ist bei Arbeitsleistungen eine andere statistische Verteilung wahrscheinlich repräsentativer, nämlich eine zweipflige Verteilung mit starkem Rechtsdrall. Demnach sind viele *tatsächlich* etwas besser als die durchschnittlich geforderte Arbeitsleistung, einige etwas schlechter. Einige wenige liegen signifikant über- bzw. unter den Durchschnittsbereich.<sup>2</sup>
  - Zum Dritten läuft man Gefahr, die abstrakte statistische Größe zur Norm für konkrete Beurteilungen zu erklären und diese damit zu verzerren. Bei Schulnoten soll das inzwischen tatsächlich schon stattfinden<sup>3</sup>.
7. Auf Seiten der beurteilenden Führungskräfte hat die GNV überhaupt nichts zu suchen.
- Zum einen, weil hier immer die je konkrete Person und ihre Leistung in dem betreffenden Zeitraum vor dem Hintergrund des jeweiligen Stellenprofils die Grundlage der Beurteilung sein muss. Wenn das trotzdem geschieht, wenn also die Benotung oder Beurteilung einer konkreten Leistung oder Person vor dem Hintergrund einer Normalverteilung geschieht, sind Messfehler vorprogrammiert. Ähnliches geschieht wohl auch in den Naturwissenschaften. So erfreute 1999 der us-amerikanische Klimaforscher Micheal E. Mann die IPCC-Gemeinde mit der Aussage, die Durchschnitts-Temperatur des Weltklimas verlaufe nach dem Bild eines Hockeyschlägers. Demnach haben sich die Durchschnitts-Temperaturen – wie vom Weltklimarat (IPCC) schon immer postuliert - nur und erst in der Neuzeit rapide nach oben entwickelt. Inzwischen weiß man, dass Mann Daten reihenweise manipulierte und unterschlug, um die Wirklichkeit der IPCC-Normkurve gemäß zurechtzubiegen.
  - Zum anderen, weil die Grundgesamtheit des zu führenden Teams mit ca. 10 Personen viel zu klein ist, um mit der GNV in Verbindung gebracht werden zu können.
8. Die Interaktion zwischen der Makroebene der Personalverwaltung und der Mikroebene der beurteilenden Führungskräfte sollte nach einem dreischrittigen Prüfverfahren durchgeführt werden:
- Schritt 1: Messen und Bewerten der Globaldaten durch die Verwaltung mit Hilfe der oben skizzierten Maßstäbe – also vor allem durch den Vergleich mit anderen Leistungsmessungen der gleichen Grundgesamtheit.
  - Schritt 2: Gespräche mit den Führungskräften. Diese Gespräche sollten nicht von Vorurteilungen über zu milde oder zu strenge Urteile geprägt sein. Einzig das Ziel, die Gründe für die Diskrepanzen zwischen beiden Messgrößen zu erkennen, sollte im Fokus der Klärungsgespräche stehen. Diese Gründe können auf Seiten der Beurteiler liegen. Sie können aber auch der Gesamtkultur geschuldet sein. Das ist immer dann der Fall, wenn sie implizit nur Bestnoten fordert, kritische Urteile sanktioniert oder als Fehlervermeidungskultur von Angst und Lüge geprägt ist.
  - Schritt 3: Gemeinsames Erarbeiten von Maßnahmen. Das können Seminare für Führungskräfte sein, um Messfehler wie z.B. Milde, Mitte, Halo, Nikolaus etc. zu vermeiden. Das können aber auch Maßnahmen z.B. zum Abbau einer Angstkultur sein. Hier muss dann ggf. auch die Personalverwaltung an sich arbeiten.

<sup>1</sup> Nach der klassischen Definition von J.v. Hann aus dem Jahre 1883 meint *Klima* "die Gesamtheit aller meteorologischen Erscheinungen, die den mittleren Zustand der Atmosphäre an irgendeiner Stelle der Erdoberfläche charakterisieren." Klima ist also das statistisch ermittelte Durchschnittswetter, berechnet über einen Zeitraum von 30 Jahren (so die seit den 30er Jahren des letzten Jahrhunderts gültige Konvention unter Klimaforschern).

<sup>2</sup> Auf diese Idee hat mich Uwe Haßelmann gebracht

<sup>3</sup> vgl. <http://www.sueddeutsche.de/karriere/kritik-an-guten-noten-nicht-zu-viele-einser-bitte-1.592366>.

