

Performance Ratings in der Kritik

Analyse von Vor- und Nachteilen skalenbasierter
Methoden der Mitarbeiterbeurteilung

Seminararbeit

Vorgelegt am 30. Juni 2018 von:

Anja Bohl

Begleitet durch: Monika Rohrer

Im Rahmen des CAS Performance & Compensation Management 2018 an der



Inhaltsverzeichnis

Abbildungsverzeichnis	II
1 Einleitung	1
1.1 Ausgangslage	1
1.2 Fragestellung und Zielsetzungen	1
1.3 Methodik und Aufbau der Arbeit.....	2
2 Theoretische Grundlagen zum Performance Rating.....	3
2.1 Einbettung ins Performance Management.....	3
2.2 Begriffsbestimmung.....	5
2.3 Gestaltungsformen	6
2.4 Entwicklung	7
3 Vorteile von Performance Ratings.....	9
3.1 Zusammenfassung der Leistungsbeurteilung und Transparenz.....	9
3.2 High & Bottom Performer Management	10
3.3 Vergütungsentscheide	11
4 Kritik an Performance Ratings.....	12
4.1 Beschränkte Aussagekraft.....	12
4.2 Psychologische und motivationale Wirkung	12
4.3 Subjektivität	14
4.4 Beurteilungsfehler.....	14
4.5 Geschönte Ratings	16
4.6 Forced Ranking und Forced Distribution	17
5 Fazit	21
Literaturverzeichnis	22
Anhang.....	24
Selbstständigkeitserklärung	25

Abbildungsverzeichnis

Abbildung 1: Einbettung von Performance Ratings ins Performance Management.....	3
Abbildung 2: Übersicht der verschiedenen Skalenarten.....	6
Abbildung 3: Zusammenstellung von Beurteilungsfehlern.....	15
Abbildung 4: Verteilung der Leistungsbeurteilungen	16
Abbildung 5: Beispiel einer alphabetischen Rating-Skala	24
Abbildung 6: Beispiel einer Rating-Skala mit Leistungsbeschreibungen.....	24

1 Einleitung

1.1 Ausgangslage

Stellen Sie sich vor, das Jahr neige sich dem Ende zu. Sie erinnern sich daran, dass morgen ein Termin mit ihrem Vorgesetzten ansteht. Zweck des Gesprächs wird es sein, dass sie gemeinsam Rückschau halten auf das vergangene Jahr, sich über ihre Stärken und Defizite austauschen und sich insbesondere auch über ihre Weiterentwicklung unterhalten. In diesem Zuge wird Ihnen auch mitgeteilt, mit welchem **Rating** sie das vergangene Jahr abgeschnitten haben. – Eine Situation, die den meisten Mitarbeitenden sehr vertraut vorkommen dürfte.

Ratings gehören heute zum **Standardinstrumentarium** im Performance Management. Die grosse Mehrheit der Unternehmen greift heute auf Ratings zurück, wenn es darum geht, die Leistung von Mitarbeitenden zu beurteilen und begrifflich zu machen. Dies geht soweit, dass Performance Ratings manchmal sogar als ultimatives Resultat des Performance Management Prozesses angesehen werden. Trotz ihrer florierenden Verbreitung stehen Performance Ratings in jüngster Zeit allerdings zunehmend in der **Kritik**. Forderungen, sie aus dem Performance Management zu verbannen, werden lauter (Armstrong, 2017, S. 119).

Hierzulande ist die Zürcher Kantonalbank (ZKB) das wohl bekannteste Beispiel für ein Unternehmen, dass in jüngster Zeit der zunehmenden Kritik an Performance Ratings gefolgt ist. Im Zuge einer umfassenden Anpassung ihres Performance Management Systems schaffte sie 2016 (unter anderem) die Benotung ihrer Mitarbeitenden ab (Imwinkelried, 2017). Die ZKB ist dabei in guter Gesellschaft anderer namhafter Unternehmen wie beispielsweise Accenture, Deloitte oder Microsoft. Sie alle haben klassische Performance Ratings aus ihren Personalbeurteilungsprozessen verbannt und teilweise durch **neue Ansätze** ersetzt (Armstrong, 2017, S. 127 ff.).

1.2 Fragestellung und Zielsetzungen

Vor diesem Hintergrund stellt sich die Frage, ob diese Entwicklung Zeichen einer grossen Wende im Performance Management ist. Stellen die genannten Beispiele gesonderte Einzelfälle dar oder ist davon auszugehen, dass man künftig vermehrt von der etablierten Praxis der Performance Ratings wegkommen wird? Um eine Antwort auf

diese Frage zu finden, setzt sich die vorliegende Arbeit damit auseinander, was dafür spricht Ratings auch in Zukunft als Instrument des Performance Managements einzusetzen. Gleichzeitig wird erarbeitet, was die Gründe sind, die genannte Unternehmen zum Schritt bewogen haben, künftig auf die Verwendung von Ratings zu verzichten. Das **Ziel** der Arbeit ist es somit, ein Verständnis für die Vor- und Nachteile von Performance Ratings zu entwickeln, um aus den gewonnen Erkenntnissen Aussagen dazu treffen zu können, ob Performance Ratings als integraler Bestandteil von Performance Management Systemen ausgedient haben.

1.3 Methodik und Aufbau der Arbeit

Zur Erreichung der beschriebenen Zielsetzung, bedient sich die vorliegende Arbeit einer Analyse der zum Thema vorhandenen **Literatur**. Wissenschaftliche und theoriegeleitete Beiträge sollen dabei ebenso aufgearbeitet werden, wie Beiträge, die den Blick in die aktuelle Unternehmenspraxis richten. Eine sowohl theorie- wie auch praxisorientierte Betrachtung des Themas soll somit Erkenntnisse für die Beantwortung der Leitfragen dieser Arbeit liefern.

Ausgangspunkt der Arbeit stellt die Aufarbeitung der theoretischen Grundlagen zum Thema Performance Ratings dar (Kapitel 2). In einem nächsten Schritt wird untersucht, welche Gründe für den Einsatz von Performance Ratings sprechen und somit zum Florieren der Mitarbeiterbenotungen beigetragen haben (Kapitel 3). Daraufhin folgt schliesslich die Würdigung der Stimmen, die sich kritisch gegenüber Ratings äussern. In diesem Zuge werden Nachteile und Probleme, die Performance Ratings mit sich bringen beleuchtet (Kapitel 4). Im abschliessenden Kapitel wird schliesslich ein Fazit über die daraus gewonnenen Erkenntnisse gezogen (Kapitel 5).

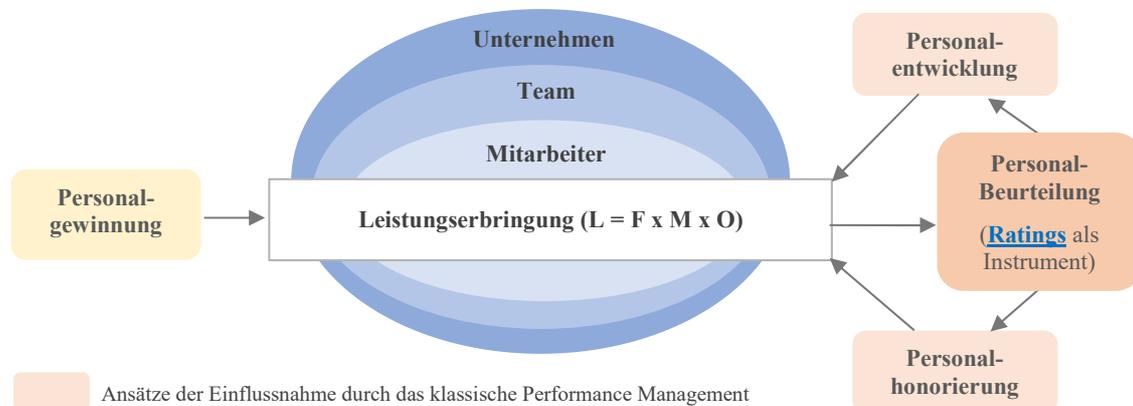
2 Theoretische Grundlagen zum Performance Rating

Dieses Kapitel widmet sich der theoretischen Erfassung und Einordnung des Themas der vorliegenden Arbeit. Dazu erfolgt zunächst eine thematische Einbettung in den Kontext des Performance Managements (Kapitel 2.1). Danach folgt die Definition und Erfassung des Begriffs des Performance Ratings (Kapitel 2.2.). Zum Ende des Kapitels wird schliesslich auf die geschichtliche Entwicklung von Ratings im Performance Management eingegangen (Kapitel 2.3).

2.1 Einbettung ins Performance Management

Performance Ratings sind ein etabliertes Instrument des Personalmanagements. Sie werden heute von vielen Unternehmen im Rahmen des Performance Managements bzw. der Personalbeurteilung eingesetzt. Die genannten Begriffe werden daher im Folgenden in ihrer Bedeutung erfasst und zueinander in Beziehung gesetzt. Abbildung 1 fasst die Inhalte dieses Kapitels grafisch zusammen und stellt die Begriffe und Konzepte, die nachfolgend erarbeitet werden, zueinander in Beziehung.

Abbildung 1: Einbettung von Performance Ratings ins Performance Management



(Eigene Darstellung in Anlehnung an die Modelle von: Fombrun, Tichy & Devanna, 1984, S. 41; Meyer-Ferreira, 2015, S. 163 und Hilb, 2008, S. 19)

Performance Management gilt als ein zentraler Prozess zur Steuerung von Unternehmen (Donkor & Slobodjanjuk, 2014). Meyer-Ferreira (2015, S. 148) erklärt, dass es im **Performance Management** um die «Sicherstellung von wettbewerbsfähigen Unternehmensleistungen» geht. Ein Unternehmen müsse in der Lage sein, seine Fähigkeiten – die mehr als die Summe der Fähigkeiten einzelner Mitarbeiter sind – in Leistung umzusetzen. Armstrong (2017) definiert Performance Management als: «[...] a systematic

and continuous process for improving organizational performance by developing the performance of individuals and teams» (S. 31). Durch Performance Management bezwecken Unternehmen also Leistung systematisch zu steuern, um dadurch auf Individual-, Team- und Unternehmensebene bessere Resultate erzielen zu können.

Aus diesen Ausführungen wird ersichtlich, dass im Performance Management zwischen verschiedenen **Ebenen der Leistungserbringung bzw. -steuerung** unterschieden werden kann. Performance Management im weiteren Sinne kann als übergeordnete Verantwortung auf Ebene des Gesamtunternehmens und damit als Aufgabe der Unternehmensführung verstanden werden. Im engeren Sinne bezieht sich Performance Management indes auf die Steuerung von Verhalten und Ergebnissen auf Team- sowie individueller Ebene. Im Verständnis eines modernen Personalmanagements wird Performance Management im engeren Sinne als eine der zentralen Aufgaben und Leistungsbereiche der HR-Funktion angesehen (Meyer-Ferreira, 2015, S. 241 ff.).

Die Leistung des einzelnen Mitarbeiters definiert Meyer-Ferreira (2015, S. 162) als Produkt seiner Fähigkeiten (F) und Motivation (M) sowie der organisationalen Rahmenbedingungen (O), welche die Leistungserbringung beeinflussen. In einer Gleichung ausgedrückt stellt er folgende Formel auf: $L = F \times M \times O$. Die Aufgabe des Performance Managements sieht er darin, die organisationalen (O) und individuellen Voraussetzungen (F & M) so zu beeinflussen und zu gestalten, dass die Mitarbeitenden in der Lage sind, ihre bestmögliche Leistung zu erbringen. Als strategische **Handlungsfelder** des Performance Managements und damit Ansätze zur Einflussnahme auf die Mitarbeiterleistung sieht Meyer-Ferreira (S. 163) die folgenden drei Bereiche¹:

- **Personalbeurteilung:** Bezieht sich in erster Linie auf die Variable der Mitarbeiterleistung (L). Zweck ist es, das Leistungsvermögen der Mitarbeitenden zu eruieren, ihr Potential einzuschätzen und im Dialog mit dem Mitarbeitenden auf die Entwicklung seiner Leistung hinzuarbeiten (Studer, 2009, S. 8). Die Erfassung von Leistung und Potential geschieht dabei anhand festgelegter Kriterien und systematischer Beurteilungsprozesse (Seibt, Hormel & Kröger, 2017, S. 252).

¹ Zusammen mit der Personalgewinnung werden diese drei Handlungsfelder als zentralen Funktionen des Personalmanagements angesehen (Hilb, 2008, S. 14 ff.). Der Personalgewinnung scheint im Rahmen des Performance Managements allerdings eine untergeordnete Rolle zuzukommen, da die Mitarbeiterleistung logischerweise erst erfolgen kann, nachdem das Personal angestellt wurde.

- **Personalentwicklung:** Bezieht sich in erster Linie auf die Variable Mitarbeiterfähigkeiten (F). Zweck ist es, die Mitarbeitenden zu fördern und ihre Fähigkeiten und Kompetenzen im Einklang mit den Bedürfnissen des Unternehmens und des Mitarbeitenden weiterzuentwickeln, damit zukünftige Erfolgspotentiale gesichert werden können (Hilb, 2008, S. 129).
- **Personalhonorierung:** Bezieht sich in erster Linie auf die Variable Mitarbeitermotivation (M). Ziel ist es, durch die Lohn- und Kompensationsgestaltung gezielte Anreize zu setzen und positiv auf die Motivation der Mitarbeitenden einzuwirken.

Die genannten Handlungsfelder stehen dabei nicht isoliert nebeneinander. Vielmehr weisen sie enge Wirkungszusammenhänge und Interdependenzen auf (Meyer-Ferreira, 2015, S. 162 f.; vgl. auch Hilb, 2008, S. 20). Um die Leistung der Mitarbeitenden in den verschiedenen Feldern zu beeinflussen, kann eine Vielzahl von methodischen Ansätzen und Instrumenten zum Einsatz kommen. **Instrumente** des Performance Managements sind bspw. die Definition, Messung und Kontrolle von Zielen, Standards und Kompetenzen (Armstrong, 2017, S. 5). Wie eingangs dieses Kapitels erwähnt, stellen auch Performance Ratings ein Instrument dar, welches von vielen Unternehmen im Rahmen der Personalbeurteilung eingesetzt wird. Im nachfolgenden Kapitel wird das Instrument, das im Zentrum dieser Arbeit steht, begrifflich erfasst.

2.2 Begriffsbestimmung

Das Wort «Performance» stammt aus dem Englischen und kann mit «Leistung» übersetzt werden. Das Wort «Rating» stammt vom englischen Verb «to rate» ab, was so viel bedeutet wie «bewerten», «beurteilen» oder «einschätzen». Daraus wird ersichtlich, dass der kombinierte **Begriff «Performance Rating»** im Kontext des Performance Managements für die Bewertung von Mitarbeiterleistungen steht. Armstrong (2017, S. 119 f.) erklärt, dass Performance Ratings dazu verwendet werden, die Gesamtleistung des Mitarbeitenden zusammengefasst anhand einer Bewertungsskala auszudrücken. Ein Beurteiler (in der Regel die Führungskraft) wählt dabei den Punkt (bzw. die Ausprägung) auf der Skala aus, bei dem sich die Beschreibung am besten mit der wahrgenommenen Leistung des Mitarbeitenden deckt. Zwischen den verschiedenen Ausprägungen der Bewertungsskala besteht dabei jeweils eine bestimmte Reihung bzw. Rangfolge.

2.3 Gestaltungsformen

Die Möglichkeiten zur Gestaltung von Personalbeurteilungen und Rating Systemen sind vielfältig (vgl. bspw. Seibt et al., 2017). Die unterschiedlichen Erscheinungsformen können anhand verschiedener Gestaltungsdimensionen unterschieden und beschrieben werden. Nachstehend wird auf einige zentrale Gestaltungsmöglichkeiten eingegangen.

In **zeitlicher Hinsicht** können Personalbeurteilungen anlassbezogen (z.B. Probezeitende) oder periodisch erfolgen, wie Seibt et al. (2017, S. 253) erklärt. Periodische Beurteilungen werden entweder fortlaufend während des Jahres (z.B. monatlich) oder bezogen auf einen bestimmten Stichtag (z.B. Jahresende; zwecks Festlegung variabler Vergütungsbestandteile) vorgenommen. Kommen Performance Ratings bei der Personalbeurteilung zum Einsatz, werden diese in der Regel **jährlich** bis zu einem bestimmten Stichtag an Mitarbeitende vergeben. Eine laufende Vergabe und Anpassung scheint wenig sinnvoll, wenn man davon ausgeht, dass sich das Leistungsverhalten von Mitarbeitenden über kürzere Perioden nicht plötzlich grundlegend verändern kann.

In der Praxis finden verschiedene **Arten von Skalen** Anwendung (vgl. Studer, 2009, S. 18 f.; Armstrong, 2017, S. 121; Seibt et al., 2017, S. 8). Sie können zum Beispiel auf einem numerischen (z.B. 1, 2, 3) oder einem alphabetischen (z.B. A, B, C) Bewertungsraster basieren. Es kann aber auch sein, dass Unternehmen eine rein begriffliche Um-/Beschreibung der Ratingklassen verwenden (z.B. hervorragend, sehr gut, gut, befriedigend, unbefriedigend). Seibt et al. (2017, S. 8) hat die in der Praxis am häufigsten verwendeten Skalenart in folgender Darstellung zusammengestellt:

Abbildung 2: Übersicht der verschiedenen Skalenarten

Skala	Beschreibung
Skalenwertbeschreibungen	Vorgabe der textlichen Beschreibungen (z. B. „Bleibt ruhig und gelassen in Stresssituationen“)
Nominalskala	An Stelle von textlichen Beschreibungen werden einzelne Begriffe verwendet
Numerische Skala	Die Antworten werden bestimmten numerischen Skalen (z. B. von 1 – gar nicht bis 5 – sehr gut) zugeordnet
Grafische Skala	Grafische Darstellung der Ergebnisse, z. B. Smileys: 

(Quelle: Seibt et al., 2017, S. 8)

Weiterer Gestaltungsspielraum besteht in Bezug auf die **Anzahl der Bewertungsstufen**. Eine Frage, der sich die Unternehmen dabei stellen müssen ist, ob sie eine gerade oder eine ungerade Anzahl an Rating-Stufen anwenden wollen. Studer (2009) erläutert, dass beide Varianten Befürworter haben: *«Den Angehörigen der Fraktion «gerade Anzahl Stufen» ist ganz wichtig, dass es keine mittlere Stufe gibt. Nur so würde man sich klar für «über oder unter der Mitte» entscheiden müssen. Ansonsten könnte man sich feigerweise für die wenig aussagende Mitte – also den Durchschnitt entscheiden. Die Angehörigen der Fraktion «ungerade Anzahl Stufen» argumentieren hingegen, dass wir alle doch ziemlich durchschnittlich sind. [...] Die ganz Genialen und die ganz Schwachen sind anzahlmässig ziemlich wenige Personen»* (S. 14). Einer Studie von e-reward (2014, zit. in Armstrong, 2017, S. 120) zufolge, greifen Unternehmen am häufigsten auf eine Unterteilung in fünf verschiedene Beurteilungsklassen zurück (61% der Befragten). Einige exemplarische Beispiele für mögliche Bewertungsskalen sind im Anhang angefügt.

2.4 Entwicklung

Studer (2009, S. 6) erklärt, dass sich die Menschen bereits in der Steinzeit als Jäger und Sammler darauf verlassen mussten, dass sie Freund und Feind voneinander unterscheiden können. Ihr Überleben war davon abhängig die Stärke bzw. Schwäche ihres Gegners in kürzester Zeit richtig einschätzen zu können. Erste Artefakte von strukturierten Beurteilungen sollen in China gefunden worden sein und auf das dritte Jahrhundert zurückgehen (Armstrong, 2017, S. 5; Studer, 2009, S. 6). Verbreitung fanden strukturierte Beurteilungen dann anfangs des 20. Jahrhunderts aufgrund des amerikanischen Militärs. Von den USA zu uns kommend, haben sich systematische Personalbeurteilungen dann nach dem 2. Weltkrieg auch hierzulande sowie in der Wirtschaftspraxis durchgesetzt (Armstrong, 2017, S. 6; Studer, S. 6). Ein Instrument und Schlagwort, das in den 1960er und 70er aufkam, war das «Management by Objectives» (MbO), das auf den Begründer Peter Drucker (1955) zurückgeht.

Beurteilungen anhand von Notensystemen sind aus dem schulischen Kontext bekannt. Im Rahmen von Personalbeurteilungen wurden Bewertungsskalen erstmals zu Beginn des 20. Jahrhunderts im US Militär angewendet und florieren seither auf beiden Seiten des Atlantiks. Eine Erscheinung, die seit Beginn des 21. Jahrhunderts in Unternehmung-

en beobachtet werden kann, sind strikte Verteilvorgaben bei der Vergabe der Ratings. Diese Praxis ist unter den Begriffen «Forced Ranking» bzw. «Forced Distribution» bekannt (siehe dazu Kapitel 4.6). Heute bauen viele Performance Management Systeme auf der Vergabe von Ratings und nicht selten auch auf strengen Verteilvorgaben auf (Armstrong, 2017, S. 6).

Da die Verwendung von Mitarbeiterratings allerdings in letzter Zeit immer lauter werdender Kritik ausgesetzt ist, sind verschiedene prominente Unternehmen – hierzu-lande bspw. die Züricher Kantonalbank (ZKB) – jüngst davon weggekommen, ihre Mitarbeiter anhand starrer Bewertungssysteme zu beurteilen (Imwinkelried, 2017). Die Frage, ob es heute noch angebracht und richtig ist, Ratings bei der Personalbeurteilung heranzuziehen, wird allerdings **kontrovers diskutiert**. Während Unternehmen in anderen umstrittenen Aspekten des Performance Managements bereitwilliger von langjährigen Praktiken abkommen (z.B. Ersetzen des Jahresendgesprächs mit regelmässigen, unterjährigen Mitarbeitergesprächen), scheint der Abschaffung von Ratings mit mehr Zurückhaltung begegnet zu werden (Armstrong, 2017, S. 120). Dennoch zeichnet sich in der Praxis die Tendenz ab, dass Unternehmen vermehrt von Performance Ratings wegkommen (z.B. Adobe, Accenture, Deloitte, Microsoft) (Armstrong, 2017, S. 66-73 / 127-128).

3 Vorteile von Performance Ratings

Um die Leitfragen der vorliegenden Arbeit zu beantworten, wird in diesem Kapitel beleuchtet, welche Gründe das Florieren von Performance Ratings begründet haben. Nachfolgend werden daher die zentralsten Argumente aufgezeigt, die für die Anwendung von Ratings bei der Personalbeurteilung sprechen. Dies sind insbesondere die komprimierte Zusammenfassung der Leistungsbeurteilung und die Transparenz schaffende Wirkung der Ratings (Kapitel 3.1), das Management von über- und unterdurchschnittlich performenden Mitarbeitenden (Kapitel 3.2) sowie das Verwenden von Ratings für das Treffen von Vergütungsentscheidungen (Kapitel 3.3).

3.1 Zusammenfassung der Leistungsbeurteilung und Transparenz

Ratings bringen die Einschätzung des Beurteilenden in stark komprimierter Form zum Ausdruck und dienen daher als eine Art **Zusammenfassung der Leistungsbeurteilung** (Armstrong, 2017, S. 123). Der Anwendung von Rating Methoden liegt dabei die Annahme zu Grunde, dass es möglich und wünschenswert ist, die Leistung (oder auch das Potential) von Mitarbeitenden akkurat und konsistent anhand einer Skala zu messen, kategorisieren und auszudrücken (Armstrong, S. 121).

Murphy and Cleveland (1995, zit. in Armstrong, 2017, S. 121) erklären, dass Ratings aus der **Perspektive des Unternehmens** dazu dienen, die «persönliche» Leistungseinschätzung des Beurteilenden transparent und zu einer Art «öffentlichen» Statements zu machen. Darüber hinaus ermögliche es, eine einfache Dokumentation der Beurteilung. Für Beurteiler sind Skalen meist einfach anzuwenden und zu verstehen (Müller, 2017, S. 4). Firmen die Ratings einsetzen, berufen sich im Weiteren darauf, dass Ratings den Vergleich über grössere Gruppen hinweg ermöglichen und Transparenz über die Leistung der Mitarbeitenden schaffen (Maass, 2013).

Ratings werden in aller Regel als Grundlage für das Mitarbeitergespräch herangezogen (Maass, 2013). Bei diesem Anlass wird es üblicherweise gegenüber dem Mitarbeitenden kommuniziert. Für den Mitarbeitenden wird dadurch ersichtlich, wie seine Leistungen auf der Bewertungsskala eingeordnet werden. Aus der **Perspektive der Mitarbeitenden** befriedigen Ratings somit in gewissem Masse ihr grundlegendes und natürliches Bedürfnis, zu erfahren wo sie stehen (Armstrong, 2017, S. 123; vgl. auch Donkor &

Slobodjanjuk, 2014). Erlauben einen klaren Vergleich zwischen Selbst- und Fremdbild, was gemäss Seibt et al. (2017, S. 253) einen zentralen Nutzen von Personalbeurteilung für den Mitarbeitenden darstellt.

3.2 High & Bottom Performer Management

Eine grundlegende Annahme des Performance Managements ist, dass Mitarbeitende sich entwickeln können. Dass mit entsprechenden Anstrengungen und Massnahmen bessere Leistungen und Ergebnisse erreicht werden können (Studer, 2009, S. 24). Um die richtigen Personalentwicklungsmassnahmen einleiten zu können, gilt es ausfindig zu machen, wo bzw. bei wem welcher Bedarf besteht. Dazu sind Leistung, Verhalten und Entwicklungspotential der Mitarbeitenden zu beurteilen. Eine Funktion, die durch die Personalbeurteilung sichergestellt werden soll (Seibt et al., 2017, S. 251).

Besonders die Identifikation von potentialträchtigen Leistungsträgern sowie leistungsschwachen Mitarbeitenden gilt in diesem Zusammenhang als eine zentrale Aufgabe des Performance Managements. Im englischen Sprachgebrauch werden für diese Mitarbeiterkategorien auch die Begriffe «High Performer» oder «High Potentials» bzw. «Bottom Performer» oder «Underperformer» verwendet. In Bezug auf die erst Mitarbeiterkategorie, die überdurchschnittlich leistenden Mitarbeitenden, liegt es dabei im Interesse von Unternehmen, diese zu (be-)fördern und in entsprechenden Schlüsselrollen zu entwickeln. Bezüglich der Underperformer gilt es dagegen geeignete Entwicklungsmassnahmen zu ergreifen, um die Leistung der Mitarbeitenden auf das gewünschte Niveau anzuheben. Denn das Nichterreichen von Zielen kann unter Umständen ernsthafte Probleme für das Unternehmen mit sich bringen (Studer, 2009, S. 24). Falls eine Besserung nicht in nützlicher Zeit erreichbar scheint, gilt es alternativ daran zu denken, sich ganz vom Mitarbeitenden zu trennen (Studer, S. 25).

Unternehmen die mit Mitarbeiterratings arbeiten, berufen sich darauf, dass sie die **systematische Identifikation** von über- bzw. unterdurchschnittlich performenden Mitarbeitenden ermöglichen (Armstrong, 2017, S. 124). Durch die Zuteilung eines Ratings wird für das Unternehmen ersichtlich, welcher Stufe der Bewertungsskala der Mitarbeitende zugeordnet bzw. wie gut oder schlecht seine Leistung eingeschätzt wird. Ratings bilden somit eine Möglichkeit, um nach oben oder unten auffallende Ausreisser

zu identifizieren und darauf aufbauend entsprechende Massnahmen einzuleiten. Ratings werden daher als praktikables Instrument angesehen, um High bzw. Bottom Performer auszumachen und diese wichtigen Mitarbeiterpopulationen entsprechend zu managen.

3.3 Vergütungsentscheide

Ein wichtiger Anwendungsbereich von Personalbeurteilungen ist das Treffen von Vergütungsentscheiden (Seibt et al., 2017, S. 253). Leistungsbeurteilungen werden also gerne herangezogen, um zu bestimmen wie Mitarbeiter entschädigt werden sollen (insbesondere bezüglich Salärenwicklung sowie variable Vergütungen). In diesem Zusammenhang wird auch von «performance pay» (vgl. bspw. Gratz, 2009) - zu Deutsch «leistungsbezogener Vergütung» gesprochen. Besonders Rating basierte Mitarbeiterbeurteilung, werden in aller Regel auch für das Treffen von Entlohnungsentscheiden genutzt (Armstrong, 2017, S. 124). Befürworter sehen in Performance Ratings aufgrund ihres leicht quantifizierbaren Charakters als ein praktikables Mittel, um Leistungsbeurteilungen unmittelbar in Zuteilungen vorhandener Geldmittel zu übersetzen. Dies dürfte einer der Gründe dafür sein, weshalb sich Unternehmen schwertun von Performance Ratings wegzukommen.

Donkor & Slobodjanjuk (2014) stellen fest, dass der eigentliche Hauptzweck des Performance Managements (die Beurteilung und Weiterentwicklung von Mitarbeiterleistungen) heute immer mehr durch die vergütungsbezogene Nutzung von Performance Management Systemen verdrängt wird. Gleichzeitig sagen sie voraus, dass künftig eine verstärkte Entkopplung von Leistungsbeurteilung und Vergütungsentscheid stattfinden wird. Armstrong (2017, S. 124; 135-136) stellt sich diesbezüglich auf den Standpunkt, dass eine leistungsbezogene Vergütung auch ohne Performance Ratings möglich sei. Dazu sei es nötig die Entlohnung zeitlich wie auch prozessual von der Mitarbeiterbeurteilung zu entkoppeln. Vergütungsentscheide seien basierend auf dem Wert des Mitarbeitenden für das Unternehmen (z.B. hoch, mittel, tief, keinen) zu fällen. Diesen einzuschätzen und zu benennen obliege der Führungskraft, so Armstrong. Inwiefern sich eine solche Einschätzung tatsächlich von einer Mitarbeiterbeurteilung mittels eines Ratings im klassischen Sinne unterscheidet, scheint allerdings fraglich.

4 Kritik an Performance Ratings

Nachdem das letzte Kapitel den Argumenten für die Verwendung von Performance Ratings gewidmet war, wechselt der Fokus in diesem Kapitel auf die vorherrschende Kritik an Performance Ratings. Im Folgenden wird somit erarbeitet und analysiert, inwiefern es Gründe gibt, die dafür sprechen bei der Personalbeurteilung künftig von Performance Ratings abzusehen. Auf die beschränkte Aussagekraft von Ratings (Kapitel 4.1) wird dabei ebenso eingegangen wie auf ihre motivationale Wirkung (Kapitel 4.2). Im Anschluss daran bilden die Subjektivität von Ratings (Kapitel 4.3), Beurteilungsfehler (Kapitel 4.4), geschönte Ratings und die vielfach kritisierte Praxis von «Forced Rankings / Distribution» (Kapitel 4.5) Gegenstand dieses Kapitels.

4.1 Beschränkte Aussagekraft

Wie in Kapitel 3.1 erklärt, handelt es sich beim Performance Rating um eine stark zusammengefasst Leistungseinschätzung. Eine beschränkte Aussagekraft ist einer der Punkte, der entsprechend an Performance Ratings bemängelt wird. Armstrong (2017) stellt dazu fest: *«to sum up the total performance of a person with a single rating is a gross over-simplification of what may be a complex set of factors influencing that performance»* (S. 124). Weiter führt er dazu aus: *«Even when performance is rated against defined standards the ratings do not convey what the employee did or did not do in sufficient detail»* (S. 122). Armstrong macht damit den berechtigten Punkt geltend, dass es die **starke Abstraktion** der Mitarbeiterbeurteilung verunmöglicht, daraus konkrete und handlungsanleitende Informationen zu gewinnen. Ein Mitarbeiter kann also von seinem Rating nicht ableiten, was er zu unternehmen hat, um sich bzw. seine Leistung weiterzuentwickeln. Dem Performance Rating alleine ist deshalb eine geringe verhaltensbeeinflussende Wirkung zuzuschreiben. Ein über das Rating hinausgehendes oder sogar allenfalls ersetzendes Feedback scheint unabdingbar, um auf das Ziel hinzuwirken, die Leistung von Mitarbeitenden zu verbessern und weiterzuentwickeln.

4.2 Psychologische und motivationale Wirkung

Im Zentrum des Performance Managements steht das Ziel, die Leistung von Mitarbeitenden weiterzuentwickeln und zu verbessern. Kritische Stimmen werfen Ratings allerdings auch eine kontraproduktive Wirkung vor, was mit negativen Effekten der

Noten auf die Motivation der Mitarbeitenden begründet wird. Armstrong (2017) meint dazu: «*To label people as 'average' or 'below average', or whatever equivalent terms are used, is both demeaning and demotivating*» (S. 6). Besonders starren Leistungsbewertungssystemen in Kombination mit Forced Rankings oder Distribution (siehe Kapitel 4.6) wird angelastet, dass sie für die **Mitarbeitermotivation** nicht förderlich sind (Maass, 2013). Häufig führen schlechte oder nicht mit der Selbstwahrnehmung der Mitarbeitenden übereinstimmende Ratings sogar dazu, dass die Mitarbeitenden kündigen (Maass, 2013).

Hofherr (2015) setzt sich in diesem Zusammenhang mit einer Untersuchung des NeuroLeadership Institute² auseinander. In der Studie wurden 52 Unternehmen erforscht, die ein klassisches Performance Management System durch ein fortschrittliches, unkonventionelles System ersetzt haben. Anstatt von numerischen Bewertungen wurden die Führungskräfte dazu angehalten, die Leistung ihrer Mitarbeitenden zu beschreiben nachdem sie sich in Form von regelmässigen Interaktionen mit diesen auseinandergesetzt hatten. David Rock (zit. in Hofherr, 2015), Direktor des NeuroLeadership Institute, äussert sich zu den Ergebnissen der Studie folgendermassen: «*employees from all different performance levels said they were happier without numerical ratings*».

Im Weiteren liefert Rock (zit. In Hofherr, 2015) folgende Erklärung für die negativen, psychologischen Effekte von Ratings: «*We human beings are very focused on social interactions. We are very obsessed with our social status. Calling people, a number activates a deep sense of danger in people. It's just how we are built.*» Statt den Fokus zukunftsgerichtet auf Defizite und Entwicklungspotentiale der Mitarbeitenden zu legen, ersorgen sich Führungskräfte und Mitarbeitende daher oft mehr darum, welches Ratings sie verteilen sollen bzw. erhalten werden. Rating Systeme lösen bei Führungskräften wie auch Mitarbeitenden oft sogar eine Art **Unbehagen** aus (Imwinkelried, 2017). Besonders, wenn sie in heikle Mitarbeitergesprächen thematisiert werden müssen. In solch unangenehmen Situationen neigen Beurteiler deshalb dazu Ratings zu schönen, wie eine Studie von Saffie-Robertson und Brutus (2014) belegt. Auf diesen und weitere Beurteilungsfehler wird im Rahmen der nachfolgenden Kapitel näher eingegangen.

² Das NeuroLeadership Institute ist eine globale Forschungsorganisation, die sich insbesondere mit Neuroscience und Leadership Themen auseinandersetzt (www.neuroleadership.com).

4.3 Subjektivität

Zumindest in der Theorie unterliegen Personalbeurteilungen grundsätzlich gewissen methodischen Anforderungen und Güteansprüchen. Bezogen auf die allgemeinen Gütekriterien (Objektivität, Reliabilität und Validität) besteht bei der Beurteilung von Mitarbeitenden insbesondere die Erwartung, dass diese aufgrund objektiver Kriterien zu erfolgen hat, also frei von subjektive Einflüssen und Interpretationen der Beurteilenden (Seibt et al., 2017, S. 266-267). Echte Objektivität ist bei der Personalbeurteilung allerdings nicht möglich (Müller, 2017, S. 7; Seibt et al., 2017, S. 266-267).

Da Personalbeurteilungen von Personen (meist Führungskräften) durchgeführt werden, unterliegen sie bis zu einem gewissen Teil unvermeidbar der Subjektivität (Müller, 2017, S. 7). Was unter einer guten bzw. schlechten Performance verstanden wird, ist zu weiten Teilen abhängig von «Performance Verständnis» und der Leistungseinschätzung des Beurteilenden. Daran ändert auch nichts, wenn Ratings bei der Beurteilung verwendet werden. Auch die Vergabe von (scheinbar objektiven) Noten basiert grundsätzlich auf der **subjektiven Einschätzung** des Bewertenden. Konsistenz über verschiedene Manager hinweg scheint daher schwer erreichbar (Armstrong, 2017, S. 124). Wenn überhaupt, dann am ehesten durch Kalibrierung und Abstimmung der Bewertenden, was einen hohen Aufwand mit sich bringt.

Wirft man Performance Ratings also vor, sie würden der Subjektivität des Beurteilenden unterliegen, so scheint diese weit verbreitete Kritik berechtigt zu sein. Festzuhalten bleibt aber gleichzeitig, dass selbiges auch an anderen Methoden der Personalbeurteilung bemängelt wird (Seibt et al., 2017, S. 266-267). In dieser Hinsicht scheinen Performance Ratings somit kein grundsätzlich schlechteres Instrument als andere Personalbeurteilungsmethoden zu sein.

4.4 Beurteilungsfehler

Wie eben erklärt, ist Objektivität bei der Personalbeurteilung kaum erreichbar. Zur unvermeidbaren Subjektivität treten zudem unbewusste Beurteilungsfehler hinzu, die den Beurteilern in der Regel unbewusst unterlaufen (Müller, 2017, S. 7). In der Literatur werden diese Beurteilungsfehler auf kognitive Verzerrungen der Beurteilenden zurückgeführt. Wissenschaftlich ist dieses Thema gut erforscht und die entsprechenden Phänomene empirisch belegt. In Abbildung 3 findet sich eine Zusammenstellung typ-

ischster Beurteilungsfehler. Die Übersicht basiert auf entsprechender Literatur, die sich mit dem Thema auseinandersetzt (vgl. dazu Müller, 2017, S. 3-7; Armstrong, 2017, S. 124-124; Studer, 2009, S. 35-36; Blickle, 2014, S. 284-285).

Abbildung 3: Zusammenstellung von Beurteilungsfehlern

Beurteilungsfehler	Beschreibung
Halo-Effekt	Tendenz des Beurteilers, von einem bekannten Aspekt der Leistungsbeurteilung Schlüsse über einen unbekanntem Leistungsaspekt zu ziehen. Bekannte Merkmale überstrahlen also die Beurteilung unbekannter Merkmale.
Hang zur Mitte	Tendenz, Mitarbeitende mit einem Rating nahe der Mitte der Skala zu bewerten, selbst wenn ihre Leistung gegen oben oder unten davon abweicht. (Gegenteil vom Effekt «Hang zu Extremen»)
Hang zu Extremen	Tendenz, Mitarbeitende aufgrund von Milde extrem positiv (Mildetendenz) oder aufgrund von Strenge extrem negativ (Strengetendenz) zu bewerten. (Gegenteil vom Effekt «Hang zur Mitte»).
Kontrast-Effekt	Tendenz des Beurteilers, Mitarbeitende mit anderen zu vergleichen statt sie an den gesetzten Standards zu messen.
Erster-Eindruck-Effekt	Tendenz des Beurteilers, Leistungseinschätzungen stark von ersten positiven oder negativen Eindrücken färben zu lassen.
Gleichheits-Effekt	Tendenz des Beurteilers, Mitarbeitende besser zu bewerten, wenn sie ihm ähnlich sind.
Zuordnungsfehler	Tendenz, schlechte Leistungen dem Mitarbeiter selbst anzulasten, während Erfolge eher externen Faktoren zugeschrieben werden.
Rezenz-Effekt	Tendenz, verhältnismässig unwichtige Ereignisse, die sich erst vor kurzem zugetragen haben, stärker in die Beurteilung einfließen zu lassen, als wichtige Ereignisse und Leistungen, die aber bereits weiter zurück liegen.
Stereotypisierung	Tendenz aufgrund von Vorurteilen generalisierende Schlüsse über die zu Beurteilenden zu ziehen, ohne Unterschieden zwischen Individuen angemessen Rechnung zu tragen.

(Quelle: eigene Darstellung; basierend auf den Ausführungen von Müller, 2017, S. 3-7; Armstrong, 2017, S. 124-124; Studer, 2009, S. 35-36 sowie Blickle, 2014, S. 284-285)

Seibt et al. (2017, S. 268) zufolge, fallen Beurteilungsfehler ganz besonders ins Gewicht, wenn die Beurteilung zu einem Gesamturteil über die Leistung des Mitarbeitenden verdichtet wird – also bspw. in Form eines Ratings. Eine Studie von Scullen & Mount (2000) konnte belegen, dass Ratings zu einem grösseren Teil auf Verzerrungen des Beurteilers als auf der Leistung des Beurteilten basieren. Kritikern zufolge sagen Ratings somit scheinbar mehr über den Beurteiler als den eigentlich Beurteilten aus (Samanta & Blumenschein, 2018). Es wird allerdings auch die Meinung vertreten, dass

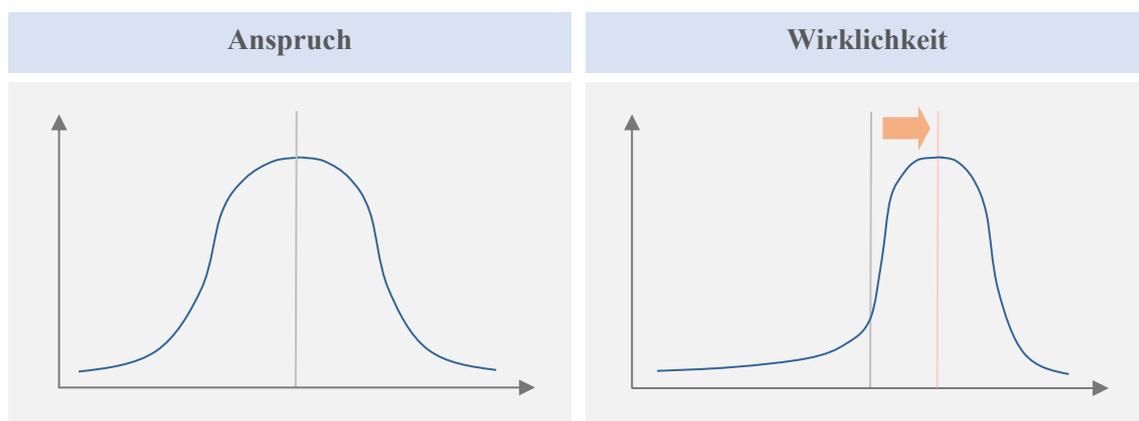
den aufgezeigten Beurteilungsfehlern zumindest in gewissem Masse entgegengewirkt werden kann - bspw. durch Sensibilisierung, Führungskräfte trainings, Einsatz verschiedener Beurteiler oder technische Korrekturmaßnahmen (Müller, 2017, S. 7-8; Studer, 2009, S. 36).

4.5 Geschönte Ratings

In der Theorie erwarten Unternehmen, dass die Leistung ihrer Mitarbeitenden einer **Normalverteilung** folgt (Schnelli, 2013; Studer, 2009, S. 14-15). Schnelli (2013) erklärt: *«Werden die individuellen Leistungen der Mitarbeitenden eines Teams visuell dargestellt, dürfte man eine klassische, ausgeglichene Glockenkurve erwarten. Das heisst, der höchste Punkt der Kurve liegt etwa in der Mitte der Erwartungen. Die meisten Mitarbeitenden liefern also gute Arbeit, haben aber auch Schwachpunkte oder Verbesserungspotential».*

In der Realität neigen Führungskräfte allerdings dazu ihre Mitarbeitenden tendenziell besser zu bewerten und Ratings zu schönen (vgl. Beurteilungsfehler «Tendenz zur Milde»). Die Kurve der Leistungsbeurteilungen ist daher wie in Abbildung 4 dargestellt meist nach rechts verschoben. Das bedeutet, dass die Leistung von Mitarbeitenden mehrheitlich als überdurchschnittlich beurteilt wird. Diese **Rechtsverschiebung** der Leistungsbeurteilungskurve bei der Vergabe von Ratings gilt als empirisch belegtes Phänomen (Schnelli, 2013).

Abbildung 4: Verteilung der Leistungsbeurteilungen



(Quelle: Referat Carsten Schlichting zit. in Schnelli, 2013)

Gemäss einer Studie des Unternehmens Hostettler, Kramarsch & Partner (hkp group)³ ist das Phänomen auf verschiedene Ursachen zurückzuführen. Wie Cartsten Schlichting, Partner bei der hkp group von Schnellli (2013) zitiert wird, ist ein **Grund** für die oft geschönten Mitarbeiterbeurteilungen darin zu sehen, dass es Führungskräften an der Bereitschaft und auch an der Fähigkeit zur Differenzierung mangle. Im Weiteren seien die Bewertungen über verschiedene Führungskräfte hinweg meist nicht aufeinander abgestimmt. Laut Norbert Thom (zit. in Imwinkelried, 2017), emeritierter Berner Betriebswirt, sind zu gute Mitarbeiterbewertungen das Resultat eines psychologischen Deals zwischen Vorgesetzten und Mitarbeitenden. Gute Ratings gegen Mitarbeiterloyalität sei der «Handel» zwischen den Parteien, so Thom. Auch die Tatsache, dass Führungskräfte in der Regel Konflikte scheuen oder negative Auswirkungen auf ihre eigene Person fürchten, führt zur Vergabe von geschönten Ratings (Müller, 2017, S. 5).

Schlichting (zit. in Schnellli, 2013) erklärt, dass unrealistische Mitarbeiterbeurteilungen zu verschiedenen **Problemen** führen: *«Die schwachen Mitarbeiter bekommen nicht das richtige Feedback, notwendige Kritik wird nicht adressiert. Bei den starken Mitarbeitern wirkt eine schwache Abstufung im Sinne von 'alle in unserem Team sind gut' demotivierend. Echte Top-Performer erhalten keine herausragende Bewertung und lernen so, dass sich zusätzliche Leistung nicht lohnt. Zudem kann es ein Unternehmen viel kosten, wenn zu viele leistungsabhängige Boni ausbezahlt werden müssen.»*

4.6 Forced Ranking und Forced Distribution

Wie im vorangegangenen Kapitel beschrieben sind Unternehmen die Ratings bei der Mitarbeiterbeurteilung einsetzen oft mit geschönten Mitarbeiterbeurteilungen (Rechtsverschiebung der Leistungsbeurteilungskurve) konfrontiert. Um dieser Problematik entgegenzuwirken, greifen Unternehmen oft auf technische Korrekturmassnahmen wie «Forced Ranking» oder «Forced Distribution» zurück. Instrumente, die anhaltender Kritik ausgesetzt sind (Armstrong, 2017; Maass, 2013). Nachfolgend werden diese beschrieben und darauf eingegangen, welchen Nutzen und welche Probleme mit den Methoden verbunden sind.

³ Die hkp group ist eine internationale Unternehmensberatung mit dem Fokus auf Performance Management, Talent Management und Vergütung (www.hkp.com/de).

Armstrong (2017, S. 82) erklärt und unterscheidet die beiden Methoden wie folgt:

- **Forced Ranking:** *«Forced ranking requires managers to place their staff in order from best to worst. The rank order is divided into predetermined categories expressed as percentiles that are defined as hierarchy of performance grades. Individuals are then placed in one of those grades according to their rank order.»*
- **Forced Distribution:** *«forced distribution [...] does not involve ranking but does require managers to conform to a laid down distribution of performance ratings.»*

Mit Forced Ranking Systemen werden Führungskräfte dazu gezwungen, ihre Mitarbeitenden nicht nur gegen vordefinierte Leistungsstandards, sondern auch mit anderen Mitarbeitenden zu vergleichen (Grote, 2005) - was auch entscheidende Kritik an der Methode begründet, wie später aufgezeigt wird. Dieser Aspekt wird bei der Forced Distribution stark entschärft. Beiden Methoden ist gemein, dass die Mitarbeitenden in verschiedene Mitarbeiterkategorien einzuteilen sind, wobei dafür bestimmte Verteilungsvorgaben vom Unternehmen vorgegeben werden.

Als Begründer von **Verteilungsvorgaben** bei der Vergabe von Mitarbeiterratings gilt Jack Welch, ehemaliger CEO des Unternehmens General Electric. Er führte die «20-70-10-Verteilung» bei GE ein, wonach alle Mitarbeiter des Unternehmens dieser Vorgabe entsprechend in drei Mitarbeiterkategorien einzuteilen waren. Die besten 20% der Mitarbeiter, welche sich durch ausserordentliche Leistungen auszeichneten, galten als Top-Performer und sollten entsprechend eingesetzt und gefördert werden. Die schwächsten 10% der Mitarbeiter hingegen, welche unterdurchschnittliche Leistungen erbrachten, sollten identifiziert werden, um entsprechend aus dem Unternehmen gemanagt werden zu können. Die verbleibenden 70% wurden der Mitarbeiterkategorie zugeteilt, welche die Erwartungen der Firma erfüllt und sich durch gute Leistungen auszeichnet. General Electric (2000) begründet diese Mitarbeiterpolitik wie folgt: *«A Company that bets its future on its people must remove that lower 10% and keep removing it every year – always raising the bar of performance and increasing the quality of its leadership. Not removing that bottom 10% early in their careers is not only a management failure, but false kindness as well [...]. Removing marginal performers [...] is doing the right thing for them; leaving them in place to settle into a career that will inevitably be terminated is not.»* (S. 6).

Der grundlegende **Nutzen** der beiden Methoden liegt darin, dass sich die Verteilung der Leistungsbeurteilungen der erwarteten Normalverteilung annähert und der in Kapitel 4.5 beschriebenen Rechtsverschiebung entgegengewirkt werden kann (Schlichting, zit. in Schnell, 2013). Damit einher geht eine stärkere Differenzierung zwischen den Leistungen der einzelnen Mitarbeiter (Grote, 2005; Schlichting, zit. in Schnell, 2013). Befürworter von solchen Systemen halten es deshalb für ein gutes Instrument, um Top wie auch Bottom Performer effektiv identifizieren zu können. (Armstrong, 2017, S. 134; Grote, 2005; vgl. auch Maas, 2013). Die Mitarbeiterkategorisierung soll dazu genutzt werden, um überdurchschnittlich genauso wie unterdurchschnittlich performende Mitarbeiter zu identifizieren und entsprechende Massnahmen daraus abzuleiten. (Armstrong, 2017, S. 119-120). Grote (2005) argumentiert des Weiteren, dass die Methoden eine echte Leistungskultur im Unternehmen fördere (Grote, 2005).

Dem gegenüber stellen sich Kritiker auf den Standpunkt, dass die Methoden auch mit wesentlichen Nachteilen verbunden sind. Eines der grundlegendsten **Probleme** wird darin gesehen, dass wenn die schwächsten 10% der Mitarbeiter aus der Unternehmung gemanagt werden, in etwa 10% der Mitarbeiter aus der «Mittelklasse» im Folgejahr zwangsläufig in die unterste Mitarbeiterkategorie herabgestuft werden (Armstrong, 2017, S. 134; Grote, 2005). In einem Artikel mit dem vielsagenden Titel «Dead Man's Curve» erklärt Meisler (2003, S. 47), dass die Anwendung des Prinzips der natürlichen Selektion nach Charles Darwin im Arbeitskontext, zu unzufriedenen Mitarbeitenden führe. Speziell, wenn deren Werthaltung auf traditionellen Werten wie Loyalität oder Sicherheit basiere. Kompetitiven Rating Systemen wird daher nach einigen Jahren bzw. Iterationen eine abnehmende, tendenziell destruktive Wirkung zugeschrieben (Grote, 2005; Meisler, 2003, S. 48). Ein weiterer entscheidender Kritikpunkt wird darin gesehen, dass Mitarbeiter zueinander in einen internen Wettstreit treten, statt im Sinne der Unternehmung zusammen zu arbeiten. Besonders aus Sicht von sehr leistungsstarken Mitarbeitenden können solche Systeme einen Anreiz dazu darstellen, dass sie versuchen unter allen Umständen mit leistungsschwachen Kollegen zusammen zu arbeiten, um keine schlechte Bewertung zu riskieren (Armstrong, 2017, S. 134; Maas, 2013). Ein nach innen statt nach aussen (auf Mitbewerber und Kunden) gerichteter Fokus sei die Folge von kompetitiven Beurteilungssystemen (Armstrong, 2017, S. 135). Und

schliesslich sind meist auch gewisse arbeitsrechtliche Herausforderungen sowie ein grosser administrativer Aufwand mit Forced Ranking / Distribution verbunden (vgl. Meisler, 2003, S. 45 / 48-49; Maass, 2013).

Inspiziert vom Vorreiter General Electric wurden Forced Ranking und Forced Distribution Systeme in den letzten Jahren in zahlreichen weiteren namhaften Unternehmen (bspw. Microsoft, Yahoo) eingesetzt, wobei einige in der Zwischenzeit auch bereits wieder davon weggekommen sind (vgl. Armstrong, 2017, S. 66-73; Maass, 2013; Samanta & Blumenschein, 2018). Eichenwald (2012) schildert, dass der «Stacked Ranking» Prozess von den Managern von Microsoft als der destruktivste Prozess im Unternehmen bezeichnet wurde. Die kompetitive Kultur und die mangelnde Bereitschaft zur Zusammenarbeit, die dadurch im Unternehmen gesät wurden, werden gar als Ursache für die Misserfolge von Microsoft in der Dekade nach der Jahrtausendwende angesehen. Trotz ihrer weiten Verbreitung gelten Forced Ranking und Distribution Systeme heute also als umstritten.

5 Fazit

Performance Management Systeme, die auf einer systematischen Mitarbeiterbeurteilung mittels Ratings basieren, sind heute in den meisten Unternehmen Status Quo (Donkor & Slobodjanjuk, 2014). Der Fokus scheint dabei in der heutigen Anwendungspraxis häufig stark auf dem Instrument an sich sowie allfälligen Rating- und Verteilvorgaben zu liegen. Es fragt sich berechtigterweise, ob Zeit und Energie, welche Mitarbeitende, Führungskräfte, und HR-Fachleute in diesen Unternehmen für Ratings aufbringen, richtig investiert sind. Ob die skalenbasierten Mitarbeiterbeurteilungen zum eigentlichem Zweck des Performance Managements – Mitarbeiterentwicklung! – tatsächlich in genügendem Masse beitragen oder ob sie diesen im schlechtesten Fall sogar sabotieren.

In der vorliegenden Arbeit wurde aufgearbeitet, welche Vor- und Nachteile der Einsatz von Performance Ratings mit sich bringt. Es konnte eine Reihe von Argumenten aufgefunden gemacht werden, die *für* Ratings sprechen und wohl auch für deren Florieren in der Vergangenheit gesorgt haben. Gleichzeitig hat die Erarbeitung der Nachteile aufgezeigt, dass auch negative Aspekte mit der Anwendung von Ratings verbunden sind oder sein können. Während die Mitarbeiternoten teilweise auf klare Kritiker stossen (vgl. bspw. Armstrong, 2017), gibt es nach wie vor aber auch Vertreter des anderen Lagers – sogar für die heftig kritisierte Praxis der Forced Rankings / Distribution (vgl. bspw. Grote, 2005). Die Frage, ob Ratings künftig aus Performance Management Systemen zu verbannen sind, scheint ebenso brisant wie umstritten zu sein.

In der Praxis scheint sich angesichts dieser Diskussion allerdings die Tendenz abzuzeichnen, dass Unternehmen ihre starren und aufwändigen Rating Systeme vermehrt durch fortschrittliche und alternative Performance Management Modelle ablösen (Donkor & Slobodjanjuk, 2014). Die richtige Reaktion auf diese Entwicklung und die vorherrschende Debatte dürfte allerdings kaum darin liegen, Ratings überstürzt und per se aus dem Performance Management zu verbannen. Vielmehr scheint es angebracht, das Instrument anhängig vom jeweiligen Unternehmenskontext zu hinterfragen. Im Hinblick auf die zentrale Fragestellung dieser Arbeit kann festgehalten werden, dass die vorherrschende Kritik an Ratings durchaus berechtigt scheint. Deshalb davon auszugehen, dass Ratings als Instrument der Personalbeurteilung künftig ausgedient hätten, scheint allerdings unbegründet zu sein.

Literaturverzeichnis

- Armstrong, M. (2017). *Reinventing Performance Management: Building a culture of continuous improvement*. London/Now York: Kogan Page.
- Blickle, G. (2014). Leistungsbeurteilung. In: F. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 271-289). Berlin / Heidelberg: Springer.
- Donkor, C. & Slobodjanjuk, A. (2014). Trend: Trennung von Leistungsbeurteilung und Bonus. *HR Today*, 31. Oktober 2014. Abgerufen am 2. Juni 2018 unter: <https://www.hrtoday.ch/de/article/trend-trennung-von-leistungsbeurteilung-und-bonus>
- Drucker, P. (1955). *The Practice of Management*. London: Heinemann.
- Eichenwald, K. (2012). Microsoft's Lost Decade. *Vanity Fair*, 24. Juli 2012. Abgerufen am 17. Juni 2018 unter: <https://www.vanityfair.com/news/business/2012/08/microsoft-lost-mojo-steve-ballmer>
- Fombrun, C., Tichy, N. & M. Devanna. (1984). *Strategic Human Resource Management*. New York: Wiley.
- General Electric. (2000). *GE 2000 Annual Report* [Geschäftsbericht]. Abgerufen am 17. Juni 2018 unter: <https://www.ge.com/annual00/download/images/GEannual00.pdf>
- Gratz, D. (2009). *The peril and promise of performance pay: making education compensation work*. Plymouth: Rowman & Littlefield Education.
- Grote, D. (2005). Forced Ranking: Making Performance Management Work. *Harvard Business School Working Knowledge*, 14. November 2005. Abgerufen am 2. Juni 2018 unter: <https://hbswk.hbs.edu/archive/forced-ranking-making-performance-management-work#12>
- Hilb, M. (2008). *Integriertes Personal-Management: Ziele – Strategien – Instrumente* (18. Auflage). Köln: Luchterhand.
- Hofherr, J. (2015). What really happens when companies kill performance reviews. *The Boston Globe*, 17. November 2015. Abgerufen am 11. Juni 2018 unter: <https://www.boston.com/jobs/jobs-news/2015/11/17/what-really-happens-when-companies-kill-performance-reviews>

- Imwinkelried, D. (2017). Die ZKB setzt auf Vertrauen statt auf Noten. *Neue Zürcher Zeitung (NZZ)*, 9. November 2017). Abgerufen am 20. Mai 2018 unter: <https://www.nzz.ch/wirtschaft/die-zkb-setzt-auf-vertrauen-statt-auf-noten-ld.1327253>
- Maass, S. (2013). Forced Rankings: Wie Unternehmen die Underperformer aufspüren. *Die Welt*, 20. Dezember 2013. Abgerufen am 2. Juni 2018 unter: <https://www.welt.de/wirtschaft/karriere/article123143543/Wie-Unternehmen-die-Underperformer-aufspueren.html>
- Meisler, A. (2003). Dead Man's Curve. *Workforce Management*, Juni 2003, S. 44-49.
- Meyer-Ferreira, P. (2015). *Human Capital strategisch einsetzen: Neue Wege zum Unternehmenserfolg* (2. Auflage). Köln: Luchterhand.
- Müller, J. (2017). *Verzerrung bei Personalbeurteilungen durch Führungskräfte*. Diskussionspapier des Instituts für Organisationsökonomik. Münster: Westfälische Wilhelms-Universität.
- Saffie-Robertson, M. & Brutus, S. (2014). The impact of interdependence on performance evaluations: the mediating role of discomfort with performance appraisal. *The International Journal of Human Resource Management*, 23 (4), S. 525-48.
- Samanta, A. & Blumenschein, S. (2018). Performance Management: Auf dem Prüfstand. *HR Today*, 31. Januar 2018. Abgerufen am 2. Juni unter: <https://www.hrtoday.ch/de/article/auf-dem-pruefstand-0>
- Schnelli, S. (2013). Zu gut ist schlecht. *HR Today*, 19. April 2013. Abgerufen am 2. Juni 2018 unter: <http://www.hrtoday.ch/de/article/zu-gut-ist-schlecht>
- Scullen, S. E. & Mount, M. K. (2000). Understanding the latent structure of job performance ratings, *Journal of Applied Psychology*, 85 (6), S. 956-970.
- Seibt, T., Hormel, R. & Kröger, S. (2017). Personalbeurteilung und Personalentwicklung. In: J. Stierle, K. Glasmachers & H. Siller (Hrsg.), *Praxiswissen Personalcontrolling: Erfolgreiche Strategien und interdisziplinäre Ansätze für die Resource Mensch* (S. 251-304). Wiesbaden: Springer Gabler.
- Studer, J. (2009). *Personalbeurteilung: Sich qualifizieren für morgen*. In: HRM Dossier (44. Ausgabe). Zürich: SPEKTRAMedia.

Anhang

Abbildung 5: Beispiel einer alphabetischen Rating-Skala

A	Konstant ausgezeichnete Leistung (übertrifft das Ziel deutlich/mit Leichtigkeit)
B	Überdurchschnittliche + regelmässig professionelle Leistung (übertrifft das Ziel klar)
C	Gute , abgerundete Leistung (Ziel erreicht, arbeitet selbstständig)
D	Verbesserungsfähige Leistungen (erreicht das Ziel nicht oder nur mit Hilfe)
E	Mangelhafte , ungenügende Leistungen (erreicht das Ziel klar nicht und ist viel auf Hilfe angewiesen)

(Quelle: Studer, 2009, S. 18)

Abbildung 6: Beispiel einer Rating-Skala mit Leistungsbeschreibungen

hervorragend	Für Mitarbeitende mit aussergewöhnlichen und ausgezeichneten Leistungen. Die Leistungen liegen deutlich über den Erwartungen. Diese Stufe wird selten von Mitarbeitern in vergleichbarer Position erreicht.
sehr gut	Für Mitarbeitende, die konstant herausragende Leistungen in mehreren Bereichen erbringen. Die überdurchschnittlichen Leistungen qualitativ und quantitativ hochstehend. Die Leistungen liegen über den hohen Erwartungen.
gut	Für Mitarbeitende, welche die qualitativen und quantitativen Anforderungen in den Schlüsselbereichen ihrer Funktion erfüllen. Die Leistungen sind gut und qualitativ hochstehend. Die Mitarbeitenden arbeiten effektiv und professionell. Die Erwartungen werden konstant/in wesentlichen Teilen erfüllt.
befriedigend	Die Leistungen und/oder das Verhalten der Mitarbeitenden sind knapp genügend. Mehrheitlich werden die Anforderungen erfüllt, aber in wesentlichen Bereichen nicht. Es besteht ein Entwicklungsbedarf.
unbefriedigend	Der Mitarbeitende erfüllt die Anforderungen in Bezug auf Leistung und Verhalten mehrheitlich nicht. Es sind dringend Lösungen zu suchen.

(Quelle: Studer, 2009, S. 18-19)