

Kompetenzmessung mit verhaltensverankerten Einstufungsskalen

Dr. Walter Lieberei 29.07.2025

In der heutigen Personalentwicklung klafft eine Lücke zwischen Anspruch und Wirklichkeit: Während wir datengestützte Entscheidungen fordern, dominieren in der Praxis oft vage Kriterien und subjektive Eindrücke bei der Auswahl von Bewerbern und der Beurteilung von Mitarbeitern. Eine Beförderungsentscheidung zwischen zwei exzellenten Kandidaten wird nicht selten mit dem Satz "Am Ende war es ein Bauchgefühl" begründet. Die Folge sind teure Fehlbesetzungen, ungenutztes Mitarbeiterpotenzial und eine Personalentwicklung, die eher zufällig als strategisch erfolgt.

Das Kernproblem liegt oft in der Art und Weise, wie wir Kompetenz messen. Abstufungen nach dem Muster "erfüllt die Erwartungen selten - manchmal - oft - immer" sind zwar einfach, aber in der Praxis oft unbefriedigend. Sie sind vage, subjektiv und bieten keine konkrete Grundlage für Entwicklungsgespräche oder strategische Entscheidungen. Was genau unterscheidet ein "häufiges" von einem "sehr häufigen" Zeigen von Teamfähigkeit? Und was muss ein Mitarbeiter konkret tun, um sich zu verbessern?

In einer immer komplexeren Arbeitswelt ist dieser Ansatz nicht nur unpräzise, sondern kann zu einem Geschäftsrisiko werden. Die Lösung liegt in einem systematischen, transparenten und fairen Ansatz: einem durchdachten Unternehmenskompetenzmodell. Das Herzstück eines solchen Modells ist ein klar definiertes **Kompetenzniveaumodell**, das beschreibt, *wie gut* eine Person eine bestimmte Fähigkeit beherrscht.

Dieser Artikel ist Ihr Leitfaden zur Gestaltung von **Best practice-Kompetenzniveaumodellen**. Wir tauchen gleichermaßen in die Wissenschaft und die Praxis von **Behaviorally Anchored Rating Scales (BARS)** und ****Kompetenzrastern** (bzw. „Rubrics“)** ein, um Ihnen zu zeigen, wie Sie die "Blackbox" der Personalentscheidungen erhellen und eine strategische Grundlage für den Erfolg Ihres Unternehmens schaffen.

1. Das Problem mit der Subjektivität: Warum "Häufig" nicht ausreicht

Die traditionelle Kompetenz- und Leistungsbeurteilung beruht in der Praxis häufig auf unpräzisen, invaliden Messinstrumenten. Skalen, die auf Frequenz oder Intensität basieren ("trifft voll zu", "trifft eher nicht zu"), lassen bei der Verhaltensbeurteilung oft massiven Interpretationsspielraum. Zwei Führungskräfte können dasselbe Verhalten eines Mitarbeiters völlig unterschiedlich bewerten, weil ihre internen Maßstäbe verschieden sind.

Dieser Ansatz führt zu: * **Mangelnder Vergleichbarkeit:** Bewertungen sind nicht über Abteilungen und Führungskräfte hinweg standardisiert. * **Geringer Akzeptanz:** Bewerber und Mitarbeiter empfinden die Beurteilung oft als unfair und willkürlich. * **Fehlender Entwicklungsorientierung:** Die Beurteilung liefert keine konkreten, beobachtbaren Verhaltensweisen, an denen ein Mitarbeiter arbeiten kann.

Die zentrale Frage lautet also: Wie können wir die Verhaltensanker – die konkreten Beschreibungen für jede Kompetenzstufe – so gestalten, dass sie objektiv, messbar,

verständlich und handlungsorientiert sind? Die Antwort liegt in der Kombination aus rigoroser Methodik und Prinzipien aus Disziplinen wie der Logik, Messtheorie und Psychologie.

2. Die Lösung: BARS und Rubrics als Werkzeuge der Präzision

Sowohl BARS als auch Rubrics sind Instrumente, die darauf abzielen, Leistung anhand konkreter, beobachtbarer Verhaltensweisen zu bewerten. Sie ersetzen vage Adjektive durch spezifische Verhaltensanker.

Definitionen:

- **Kompetenzraster (Rubric):** Ein Rubric ist ein Bewertungsinstrument in Tabellenform, das die Erwartungen an eine Leistung oder Kompetenz transparent macht. Es besteht typischerweise aus drei Teilen:
 1. **Kriterien/Dimensionen:** Die Aspekte, die bewertet werden (z. B. "Kommunikation", "Problemlösung").
 2. **Leistungsniveaus/Stufen:** Eine Skala zur Abstufung der Leistung (z. B. "Novize", "Kompetent", "Experte").
 3. **Verhaltensanker/Deskriptoren:** Eine spezifische Beschreibung des beobachtbaren Verhaltens für jedes Kriterium auf jeder Stufe.
- **Behaviorally Anchored Rating Scale (BARS):** BARS sind eine spezielle, hochentwickelte Form von Rubrics. Ihre Besonderheit liegt im empirischen und partizipativen Entwicklungsprozess. Die Verhaltensanker werden nicht einfach von Experten formuliert, sondern durch die "Critical Incident Technique" von den tatsächlichen Stelleninhabern und Führungskräften gesammelt, gruppiert und bewertet. Dies gewährleistet eine extrem hohe Praxisrelevanz und Validität.

Anwendungsfelder in der Personalentwicklung:

Diese Instrumente sind das Rückgrat moderner, kompetenzbasierter Personalarbeit und werden eingesetzt in: * **Strukturierten Interviews:** Um die Antworten von Bewerbern objektiv zu bewerten. * **Assessment & Development Centern:** Zur Beobachtung und Bewertung von Verhalten in Simulationen. * **Leistungsbeurteilungssystemen:** Als Grundlage für faire und transparente Mitarbeitergespräche. * **Feedbacksystemen & 360-Grad-Feedbacks:** Um konkretes, handlungsorientiertes Feedback zu ermöglichen. * **Nachfolgeplanung:** Zur Identifikation von Potenzialträgern und Entwicklungsbedarf. * **Personalentwicklung:** Zur Erstellung gezielter Entwicklungspläne.

3. Die Wissenschaft optimaler Verhaltensanker: Regeln aus Logik, Messtheorie und Psychologie

Um wirklich hochwertige BARS und Rubrics zu entwickeln, müssen wir über das bloße Beschreiben von Verhalten hinausgehen. Wir müssen sicherstellen, dass unsere Skalen logisch konsistent, psychometrisch valide und sprachlich präzise sind. Hierfür ziehen wir Prinzipien aus verschiedenen wissenschaftlichen Disziplinen heran.

3.1 Prinzipien aus Logik und Sprachwissenschaft: Die Macht der präzisen Sprache

Die formale Logik, insbesondere die Prädikatenlogik, lehrt uns, Aussagen so zu strukturieren, dass sie eindeutig und überprüfbar sind. Daraus leiten wir folgende Regeln für Verhaltensanker ab:

- **Regel 1: Handlung statt Eigenschaft formulieren (Prädikatenlogik).**
Beschreiben Sie, was eine Person *tut*, nicht was sie *ist*.
 - *Schlecht*: "Ist ein guter Kommunikator." (Eigenschaft, vage)
 - *Gut*: "Fasst die wichtigsten Ergebnisse eines Meetings schriftlich zusammen UND stellt sicher, dass alle Teilnehmer die nächsten Schritte verstanden haben." (Beobachtbare Handlungen)
- **Regel 2: Eindeutigkeit und Zweiwertigkeit anstreben.** Ein Verhalten trat entweder auf oder nicht. Vermeiden Sie vage Adverbien ("oft", "angemessen", "gut"), die subjektiver Interpretation unterliegen.
 - *Schlecht*: "Kommuniziert regelmäßig mit dem Team."
 - *Gut*: "Führt wöchentliche Team-Meetings durch und versendet im Anschluss ein Protokoll."
- **Regel 3: Logische Verknüpfungen (UND/ODER) bewusst einsetzen.**
Komplexe Verhaltensweisen lassen sich durch logische Junktoren präzise beschreiben. "UND" signalisiert, dass mehrere Bedingungen erfüllt sein müssen, was typisch für höhere Kompetenzstufen ist.
 - *Beispiel (hohe Stufe)*: "Analysiert die Finanzdaten UND leitet daraus drei strategische Empfehlungen für den Vorstand ab."
- **Regel 4: Quantoren zur Skalierung nutzen.** Begriffe wie "alle", "einige", "mindestens eine/r" (Quantoren) ermöglichen eine klare Differenzierung zwischen den Stufen.
 - *Beispiel (mittlere Stufe)*: "Bezieht *einige* Teammitglieder in die Entscheidungsfindung ein."
 - *Beispiel (hohe Stufe)*: "Stellt sicher, dass *alle* relevanten Stakeholder vor einer Entscheidung konsultiert werden."

3.2 Prinzipien aus der Messtheorie und Psychologie: Die Struktur der Skala

Die Psychometrie gibt uns Werkzeuge an die Hand, um sicherzustellen, dass unsere Skalen psychologisch sinnvoll und fair sind.

- **Regel 5: Psychologische Äquidistanz der Stufen anstreben.** Der wahrgenommene "Sprung" in der Kompetenz von Stufe 1 zu 2 sollte ungefähr so groß sein wie der von Stufe 2 zu 3. Empirische Studien, wie die von Casper et al. (2020), liefern Daten zur wahrgenommenen Distanz zwischen verbalen Anker, die als Leitfaden dienen können.
- **Regel 6: Minimale Überlappung der Anker sicherstellen.** Die Verhaltensbeschreibungen für zwei benachbarte Stufen müssen so trennscharf sein, dass ein bestimmtes Verhalten klar einer Stufe zugeordnet werden kann. Die

empirische Bewertung der Überlappung von Anker (z.B. mit "Tilton's measure of overlap") ist hierfür ein wissenschaftlicher Ansatz.

- **Regel 7: Kontextabhängigkeit berücksichtigen.** Das gleiche Verhalten kann in unterschiedlichen Kontexten eine andere Bedeutung haben. Die Anker sollten daher so formuliert sein, dass sie im relevanten beruflichen Kontext Sinn ergeben.

3.3 Prinzipien für das Gesamtsystem: Die Architektur des Modells

Ein wissenschaftliches Klassifikationssystem muss bestimmten übergeordneten Regeln folgen, um robust und nützlich zu sein.

- **Regel 8: Vollständigkeit.** Alle relevanten Ausprägungen einer Kompetenz sollten auf der Skala abgebildet werden können.
- **Regel 9: Disjunktheit (Gegenseitiger Ausschluss).** Jedes beobachtbare Verhalten sollte idealerweise nur einer Stufe zugeordnet werden können.
- **Regel 10: Operationalisierbarkeit und Transparenz.** Die Kriterien müssen so klar sein, dass verschiedene Bewerter bei der Beobachtung desselben Verhaltens zu einem ähnlichen Ergebnis kommen (Interrater-Reliabilität).

4. Die Prinzipien in Aktion: Kompetenzniveaus praktisch gestalten

Wenden wir diese Regeln nun auf zwei praxisrelevante Kompetenzen an. Wir verwenden ein 5-stufiges, expertiseorientiertes Modell (angelehnt an Dreyfus), um die Lernprogression abzubilden.

Beispiel 1: Überfachliche Kompetenz "Analysieren & Interpretieren"

Niveaustufe	Verhaltensanker (basierend auf den abgeleiteten Prinzipien)
1. Novize	Folgt einer detaillierten Checkliste, um Daten aus einer vorgegebenen Quelle zu sammeln UND trägt diese in eine vorbereitete Vorlage ein.
2. Fortgeschrittener Anfänger	Erkennt auf Basis früherer, ähnlicher Aufgaben einfache Muster in den Daten. Erstellt Standard-Reports UND identifiziert offensichtliche Abweichungen von der Norm.
3. Kompetent	Führt Daten aus mehreren Quellen eigenständig zu einem Gesamtbild zusammen. Analysiert komplexe Ursache-Wirkungs-Beziehungen UND entwickelt erste Lösungsoptionen.
4. Gewandt	Erfasst eine komplexe Situation ganzheitlich, identifiziert das Kernproblem intuitiv UND bewertet die Risiken verschiedener Szenarien, um eine klare Handlungsempfehlung zu geben.
5. Experte	Antizipiert zukünftige Trends auf Basis schwacher Signale in den Daten. Entwickelt neue analytische Modelle UND berät die Unternehmensführung bei strategischen Weichenstellungen.

Beispiel 2: Fachliche Kompetenz “Mitarbeiterführung” (für Führungskräfte)

Niveaustufe	Verhaltensanker (basierend auf den abgeleiteten Prinzipien)
1. Novize	Weist Aufgaben auf Basis eines bestehenden Plans zu UND kontrolliert deren Erledigung anhand einer Checkliste.
2. Fortgeschrittener Anfänger	Gibt einzelnen Mitarbeitern regelmäßiges, aufgabenbezogenes Feedback zu ihrer Leistung. Formuliert klare Erwartungen für die nächsten Arbeitsschritte.
3. Kompetent	Führt mit allen Teammitgliedern strukturierte Entwicklungsgespräche. Vereinbart individuelle Ziele UND identifiziert konkrete Entwicklungsmaßnahmen (z. B. Schulungen).
4. Gewandt	Schafft ein Umfeld, in dem das Team proaktiv Feedback gibt UND voneinander lernt. Coacht Mitarbeiter bei der Überwindung von Leistungsblockaden UND delegiert Verantwortung.
5. Experte	Entwickelt gezielt Nachfolger für Schlüsselpositionen im Team. Agiert als Mentor für andere Führungskräfte im Unternehmen UND schafft eine Kultur der Spitzenleistung, die über das eigene Team hinauswirkt.

5. Das Vorgehensmodell: In 6 Schritten zu wissenschaftlich fundierten Verhaltensankern

Wie gelangt man in der Praxis zu solch hochwertigen Anker? Der Goldstandard ist der Entwicklungsprozess der Behaviorally Anchored Rating Scales (BARS).

Schritt 1: Identifikation erfolgskritischer Kompetenzen Führen Sie eine gründliche Arbeitsplatzanalyse durch (z. B. durch Interviews, Workshops), um die Kompetenzen zu identifizieren, die für eine bestimmte Rolle oder Rollenfamilie wirklich entscheidend sind.

Schritt 2: Sammlung von Verhaltensbeispielen (Critical Incidents) Bitten Sie eine Gruppe von Fachexperten (Subject Matter Experts, SMEs – z. B. erfahrene Mitarbeiter und deren Vorgesetzte), konkrete Beispiele für besonders effektives und ineffektives Verhalten in diesen Kompetenzbereichen zu beschreiben. Wichtig ist die Formel: Situation – Verhalten – Ergebnis.

Schritt 3: Definition der Verhaltensanker Eine zweite, unabhängige Gruppe von Experten verdichtet und formuliert die gesammelten Beispiele zu standardisierten, präzisen Verhaltensankern. Sie entfernen Redundanzen und stellen sicher, dass die Prinzipien der klaren Sprache (siehe Abschnitt 3) eingehalten werden.

Schritt 4: Re-Translation (Rückübersetzung) Eine dritte Gruppe von Experten erhält die Liste der Verhaltensanker und die Definitionen der Kompetenzen. Ihre Aufgabe ist es, jeden Anker wieder der passenden Kompetenz zuzuordnen. Anker, bei denen keine hohe Übereinstimmung erzielt wird, werden aussortiert. Dies stellt sicher, dass die Anker eindeutig sind.

Schritt 5: Skalierung der Leistungsniveaus Die verbliebenen, eindeutigen Verhaltensanker werden von den Experten nun auf einer Leistungsskala (z. B. von 1 bis 5) bewertet, um ihre Effektivität einzuschätzen. Nur Anker, bei denen sich die Experten über das Leistungsniveau einig sind (geringe Standardabweichung), werden für die finale Skala ausgewählt.

Schritt 6: Erstellung und Validierung der finalen Skala Aus den validierten und skalierten Ankern wird das endgültige Kompetenzraster bzw. die BARS erstellt. Dieses Instrument sollte anschließend pilotiert und validiert werden, bevor es unternehmensweit ausgerollt wird. Eine Schulung der Anwender (insbesondere der Führungskräfte) ist für eine konsistente Anwendung unerlässlich.

6. Fazit: Vom strategischen Werkzeug zur gelebten Transformation

Ein differenziertes Kompetenzniveaumodell ist weit mehr als ein bürokratisches HR-Instrument. Es ist ein **strategisches Werkzeug zur Unternehmenstransformation**. Es schafft eine gemeinsame Sprache und eine transparente, faire Grundlage für alle Personalentscheidungen.

Durch die Abkehr von vagen Frequenzskalen und die Hinwendung zu einem wissenschaftlich fundierten, verhaltensbasierten Ansatz erreichen Sie: * **Klarheit:** Jeder Mitarbeiter und jede Führungskraft weiß, was exzellente Leistung konkret bedeutet. * **Fairness:** Beurteilungen werden objektiv, nachvollziehbar und frei von subjektiven Verzerrungen. * **Entwicklung:** Feedback wird zu einem präzisen Werkzeug, das Mitarbeitern hilft, ihre Stärken auszubauen und an Schwächen zu arbeiten. * **Strategie:** Die Personalentwicklung wird vom administrativen Verwalter zum strategischen Partner, der gezielt die Kompetenzen aufbaut, die das Unternehmen für die Zukunft braucht.

Der Weg von einem vagen "Bauchgefühl" zu einem klaren, datengestützten Kompetenzmanagement ist eine Investition. Aber es ist eine Investition, die sich in Form von besseren Mitarbeitern, geringerer Fluktuation und letztlich einem nachhaltigeren Unternehmenserfolg mehr als auszahlt. Es ist der entscheidende Schritt, um die "Blackbox" der Personalarbeit endgültig zu erhellen und die wichtigste Ressource Ihres Unternehmens – Ihre Mitarbeiter – systematisch zum Erfolg zu führen.

Literaturverzeichnis

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, 57(1), 15-22. <https://doi.org/10.1037/h0034185>

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50(3), 655-702. <https://doi.org/10.1111/j.1744-6570.1997.tb00709.x>

Casper, W. C., Edwards, B. D., Wallace, J. C., Landis, R. S., & Fife, D. A. (2020). Selecting response anchors with equal intervals for summated rating scales. *Journal of Applied Psychology*, 105(4), 390-409. <https://doi.org/10.1037/apl0000444>

- Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). *Development of the Behaviorally Anchored Rating Scales for the Skills Demonstration and Progression Guide* (Research Report No. RR-18-24). Educational Testing Service. <https://doi.org/10.1002/ets2.12210>
- Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M., & Reese, C. (2016). Behaviorally anchored rating scales: An application for evaluating teaching practice. *Teaching and Teacher Education, 59*, 414–419. <https://doi.org/10.1016/j.tate.2016.07.026>
- Salgado, J. F., & Moscoso, S. (2019). Meta-Analysis of Interrater Reliability of Supervisory Performance Ratings: Effects of Appraisal Purpose, Scale Type, and Range Restriction. *Frontiers in Psychology, 10*, 2281. <https://doi.org/10.3389/fpsyg.2019.02281>
- Schwab, D. P., Heneman, H. G., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology, 28*(4), 549–562. <https://doi.org/10.1111/j.1744-6570.1975.tb01392.x>
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*(2), 149–155. <https://doi.org/10.1037/h0047060>
- Diagnostik- und Testkuratorium (Hrsg.). (2018). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Springer. <https://doi.org/10.1007/978-3-662-53772-5>
- Wylie, E. C., & Lyon, C. J. (2020). Developing a Formative Assessment Protocol to Support Professional Growth. *Educational Assessment, 25*(4), 314–330. <https://doi.org/10.1080/10627197.2020.1766956>