

Mehr Perspektiven, mehr Wahrheit? 360°-Feedback auf dem Prüfstand

Eine systematische Bestandsaufnahme zu Begriff, Messqualität, Wirkung und Rechtsrahmen – mit Instrumentenvergleich, Faktencheck und einem Leitfaden für den Mittelstand

Dr. Walter Lieberei

28.10.2026

Kaum ein Instrument der Führungskräfteentwicklung genießt einen so guten Ruf wie das 360°-Feedback. Die Idee leuchtet ein: Wer eine Führungskraft nicht nur von ihren Vorgesetzten, sondern auch von Mitarbeitenden, Kolleginnen und Kollegen oder Kunden einschätzen lässt, erhält ein vollständigeres Bild – und die Führungskraft Hinweise, die ihr im Alltag kaum jemand offen gibt. Die Forschung zeichnet allerdings ein differenzierteres Bild: Die Urteile der Beurteilergruppen stimmen nur mäßig überein, die Effekte auf das spätere Verhalten sind im Durchschnitt klein und reichen im Einzelfall von deutlich positiv bis negativ, und für Deutschland gibt es nicht einmal eine belastbare Zahl dazu, wie viele Unternehmen das Verfahren nutzen.

Dieser Beitrag zieht eine systematische Bilanz. Seine Leitthese: Die Qualität eines 360°-Feedbacks ist keine Eigenschaft des Fragebogens, sondern des gesamten Prozesses – von der Zweckklärung über die Auswahl der Beurteilenden bis zur Umsetzung der Entwicklungsziele (Bracken & Rotolo, 2019; Fleenor, 2019). Wer das Verfahren einführt, entscheidet also nicht über ein Werkzeug, sondern über ein kleines Organisationssystem.

Praxisfall: Die Brenner Antriebstechnik GmbH (fiktiv)

Der Fall ist erfunden, aber typisch; er begleitet uns durch den Beitrag. Die Brenner Antriebstechnik GmbH ist ein familiengeführter Hersteller von Getriebemotoren mit 650 Beschäftigten an zwei Standorten, 48 Führungskräften auf drei Ebenen und einem Betriebsrat. Nach einer Mitarbeiterbefragung mit schwachen Werten zur Führung beschließt die Geschäftsführung: „Wir machen ein 360°-Feedback für alle Führungskräfte.“ Mittelfristig sollen die Ergebnisse auch in die Nachfolgeplanung für zwei Bereichsleitungen einfließen. Die Personalleiterin soll in drei Monaten starten.

1. Einleitung: Warum eine Bestandsaufnahme jetzt?

Erstens verschiebt sich der Verwendungszweck. 360°-Feedback ist als Entwicklungsinstrument groß geworden, wird aber zunehmend für Personalentscheidungen genutzt. In den US-Benchmarkerhebungen der 3D Group stieg der Anteil der Organisationen, die 360°-Daten im Performance Management verwenden, von 27 Prozent (2003) auf 48 Prozent (2013) (Bracken et al., 2016). Die Herausgeber des maßgeblichen Handbuchs halten die alte Frage „Entwicklung oder Beurteilung?“ deshalb für überholt; offen sei nur, wie Entscheidungsnutzung gelingt (Church et al., 2019a). Für Unternehmen in Deutschland ist das mehr als eine Nuance: Mit dem Zweck ändern sich Mitbestimmungsrechte, Datenschutzanforderungen und das Antwortverhalten der Beurteilenden.

Zweitens verändern Technologie und Künstliche Intelligenz (KI) die Praxis. Online-Plattformen haben den Aufwand gesenkt; Feedback-Apps, Kalibrierungssoftware und Textanalyse versprechen mehr Daten in kürzeren Abständen (Hunt et al., 2019; Fink & Sinar, 2019). Zugleich stuft die EU-KI-Verordnung KI-Systeme für Leistungsbewertung und Beförderung als Hochrisiko-Anwendungen ein.

Drittens klappt eine Lücke zwischen Versprechen und Evidenz. Häufig ist von „größerer Objektivität“ oder „ehrlicheren Antworten“ die Rede – Aussagen, die in ihrer Pauschalität nicht gedeckt sind. Umgekehrt kursieren Zahlen gegen das Verfahren, die aus anderen Forschungsfeldern stammen (Bracken et al., 2016).

Grundlage dieser Bestandsaufnahme sind das *Handbook of Strategic 360 Feedback* mit 31 Fachbeiträgen (Church et al., 2019a), der Leitartikel von Bracken, Rose und Church (2016), der Review zur Selbst-Fremd-Übereinstimmung von Fleenor et al. (2010), Yukls Taxonomie des Führungsverhaltens (2012), zwei deutsche Validierungsstudien sowie Meta-Analysen und Erhebungen bis 2026. Der Beitrag knüpft an unsere Fachbeiträge Nr. 26 (BARS), Nr. 35 (Kompetenzmodelle rechtskonform entwickeln), Nr. 37 (Anforderungsanalyse) und Nr. 52 (Selbstentwicklungs-Loop) an.

2. Begriff, Formen und Abgrenzung

2.1 Was 360°-Feedback ist

Der Begriff ist nicht geschützt und wird großzügig verwendet. Als Referenz hat sich die Definition von Bracken, Rose und Church (2016, S. 764) etabliert: 360°-Feedback ist ein Prozess, der Beobachtungen von Personen aus dem Arbeitsumfeld über eine Person sammelt, quantifiziert und berichtet. Er soll erstens zeigen, in welchem Ausmaß bestimmte Verhaltensweisen gezeigt werden, zweitens Vergleiche ermöglichen – zwischen Beurteilten, zwischen Beurteilergruppen und über die Zeit – und drittens eine nachhaltige Veränderung von Verhalten bewirken, das die Organisation wertschätzt. Daraus folgt: Es geht um *Verhalten*, nicht um Persönlichkeit, Werte oder Wissen; die Erhebung ist *quantitativ*; Beurteilende stammen aus dem Arbeitskontext; und die Daten müssen *genutzt* werden – sonst entsteht bloße Information, aber kein Feedback.

2.2 Das Formenspektrum

Tabelle 1: Formen des Mehrquellenfeedbacks im Vergleich

Form	Perspektiven	Typischer Zweck	Stärken	Grenzen
90°-Beurteilung	Vorgesetzte (+ Selbst)	Leistungsbeurteilung	einfach, etabliert	eine Perspektive, „Was“ statt „Wie“
180° / Upward Feedback	Mitarbeitende (+ Selbst)	Führungsfeedback	gut beobachtbare Kernsicht	Kooperation und Außenwirkung fehlen
270°	Vorgesetzte, Mitarbeitende, Kollegen (+ Selbst)	Entwicklung	breites internes Bild	Kunden- und Partnersicht fehlt
360°	zusätzlich interne/externe Kunden, Partner	Entwicklung, Talentmanagement	vollständigstes Bild	Aufwand, Zahl der Beurteilenden
Team-360	Teammitglieder, Teamleitung, Schnittstellen	Teamentwicklung	direktes Teammaß, weniger bedrohlich	keine Einzeldiagnose
Stakeholder-Interview	ausgewählte Schlüsselpersonen	Topebene, Gremien	Tiefe, Kontext, Nachfragen	nicht quantitativ, interviewerabhängig

Quellen: Bracken et al. (2016); Mockler et al. (2019); Stomski (2019); Winum (2019); eigene Zusammenstellung.

Wie wichtig die saubere Begriffswahl ist, zeigt PepsiCo: Das jährliche Führungsinstrument, in das nur direkte Mitarbeitende einbezogen werden, ist „technisch“ ein Upward-Feedback – und wird bewusst von einem vollständigen 360°-Feedback für Entwicklung und Nachfolge getrennt (Bracken & Church, 2013).

2.3 Was kein 360°-Feedback ist

Mit der Digitalisierung sind zahlreiche Verfahren entstanden, die das Etikett „360°“ tragen, ohne dessen Merkmale zu erfüllen: Apps, in denen Kolleginnen und Kollegen einander frei Lob schicken, Befragungen mit Persönlichkeitsfragen oder Marketingvarianten wie „720°“. Bracken et al. (2016) fassen sie als *Alternative Forms of Feedback* zusammen. Sie haben eigene Stärken, liefern aber kaum Belege für nachhaltige Verhal-

tensänderung; in einem Fortune-500-Unternehmen bestanden über 80 Prozent der Einträge eines solchen Systems aus Lob und Glückwünschen (Rose, 2019).

Kasten 1: Sechs Prüffragen – echtes 360°-Feedback oder nicht?

1. Werden beobachtbare Verhaltensweisen quantitativ erhoben?
2. Sind die Beurteilergruppen klar und sinnvoll voneinander abgegrenzt?
3. Sind die Inhalte an Verhalten gekoppelt, das für die Organisation wichtig ist?
4. Wird das Feedback von genügend qualifizierten Beurteilenden aktiv *eingeholt* (nicht nur angeboten)?
5. Zeigt der Bericht Unterschiede zwischen *und* innerhalb der Beurteilergruppen?
6. Werden die Ergebnisse so geteilt und genutzt, wie es vorab festgelegt wurde?

Nach Bracken et al. (2016, S. 768).

2.4 Eine kurze Entwicklungsgeschichte

Als Vorläufer gelten die Offiziersauswahl der Wehrmachtpsychologie, frühe Formen des Upward Appraisal und Blake und Moutons *Managerial Grid* der 1960er-Jahre, das Selbst- und Mitarbeitendenurteile in einem Seminar zusammenführte (Church & Burke, 2019). In den 1970er-Jahren folgten große Unternehmensprogramme, etwa bei Citibank und NASA, in den 1990er-Jahren der Boom. Online-Erhebungen ließen die Rücklaufquoten von rund 60 Prozent bei Papierbögen auf häufig über 90 Prozent steigen (Bracken et al., 2016).

3. Verbreitung und Bedeutung: Was belegt ist – und was nicht

Dass 360°-Feedback verbreitet ist, steht außer Frage. Wie weit, ist schwerer zu sagen, als viele Darstellungen vermuten lassen: Die verfügbaren Zahlen stammen überwiegend aus den USA, aus Befragungen bereits aktiver Anwender oder aus kleinen Stichproben besonders entwicklungsstarker Großunternehmen.

Tabelle 2: Datenlage zur Verbreitung

Quelle	Bezugsgruppe	Befund	Was sich daraus ableiten lässt
3D Group (2003/2013), zit. n. Bracken et al. (2016)	US-Benchmarkbefragungen	27 % → 48 % nutzen 360° im Performance Management	Trend zur Entscheidungsnutzung unter Anwendern
Church & Rotolo (2013), zit. n. Church (2019)	84 Unternehmen mit führender Talententwicklung	66 % nutzen 360° bei High Potentials, 60 % bei Senior Executives	Standardinstrument in Spitzenunternehmen
Shields et al. (2023)	165 australische Organisationen (2021)	40,5 % nutzen 360°-Feedback	einzige peer-reviewte Organisationsquote; geringer Rücklauf
CIPD (2016)	113 britische HR-Leitungen	34 % nennen 360° Standardpraxis	historischer Anhaltspunkt
3D Group (2025)	Anwenderorganisationen	5 % setzen KI im 360°-Prozess ein	KI noch in der Frühphase

Eigene Zusammenstellung.

Zwei Folgerungen sind wichtig. Die oft zitierte Aussage, 90 Prozent der Fortune-500-Unternehmen nutzten 360°-Feedback, ist durch keine nachvollziehbare Erhebung gedeckt. Und für Deutschland und den DACH-Raum gibt es keine repräsentative, aktuelle Nutzungsquote; auch die internationale CRANET-Erhebung weist keine verwertbare Zahl aus. Diese Lücke sollte man benennen, statt sie mit Schätzungen zu füllen.

Aussagekräftiger ist das typische Design: 80 Prozent der Anwender nutzen eine Fünferskala, die meisten Fragebögen haben unter 70 Aussagen, etwa die Hälfte führt das Verfahren jährlich durch, rund zwei Drittel bie-

ten ein Auswertungsgespräch an. Eine Gruppe wird fast überall erst ab drei Antworten ausgewiesen. Verpflichtende Schulungen für Beurteilende sind dagegen selten (Bracken & Rotolo, 2019; Rose, 2019).

4. Die Zweckfrage: Entwicklung, Entscheidung oder beides?

4.1 Fünf mögliche Zwecke

Bracken et al. (2016) sehen in der übersprungenen Zweckdiskussion die Hauptursache gescheiterter 360°-Prozesse. Sie unterscheiden fünf Zwecke: reine Beurteilung, reine Entwicklung, Beurteilung *und* Entwicklung, Organisationswandel sowie Programmevaluation. Jeder verlangt andere Antworten auf dieselben Fragen: Wer sieht die Ergebnisse? Wie viele Beurteilende sind nötig? Welche Skala?

4.2 Die „Great Debate“

Seit den 1990er-Jahren wird gestritten, ob 360°-Feedback ausschließlich der Entwicklung dienen sollte. Für eine strikte Trennung sprechen solide Befunde: In der Meta-Analyse von Smither, London und Reilly (2005) verbesserten sich die Werte bei reinem Entwicklungszweck stärker ($d = 0,25$) als bei administrativem Zweck ($d = 0,08$), wobei der zweite Wert auf wenigen Studien beruht (London & Smither, 2019). Urteile für administrative Zwecke fallen um etwa ein Drittel einer Standardabweichung milder aus (Jawahar & Williams, 1997, zit. n. Fleenor et al., 2010), und gut ein Drittel der Beurteilenden gab an, bei Entscheidungsnutzung anders zu urteilen (McCauley & Brutus, 2019).

Die Gegenseite verweist darauf, dass folgenloses Feedback kaum Veränderung auslöst. In einer Längsschnittstudie mit 232 südkoreanischen Unternehmen hing der Einsatz von 360°-Feedback mit späterer Produktivität zusammen – am stärksten bei kombinierter Nutzung für Entwicklung und Entscheidung, die allerdings nur 13 Prozent der Firmen praktizierten (Kim et al., 2016, zit. n. London & Smither, 2019). Zudem gilt: Sobald jemand außer der Führungskraft die Ergebnisse sieht, beeinflussen sie dessen Urteil.

4.3 Vom Entweder-oder zum Kontinuum

Dahling und Chau (2019) ersetzen die Dichotomie deshalb durch ein Kontinuum der Konsequenzen.

Abbildung 1: Das Kontinuum der Konsequenzen



Nach Dahling & Chau (2019) und Bracken & Church (2013); eigene Darstellung.

Organisationen bewegen sich auf diesem Kontinuum oft unbeabsichtigt. Bei einem US-Finanzdienstleister blieb die Regel „nur Entwicklung“ formal bestehen, doch immer mehr Führungskräfte teilten ihren Bericht mit ihren Vorgesetzten – das Feedback floss so faktisch in deren Urteil ein (Shepherd, 2019). Bei einem Pharmakonzern wurde eine als Bestandsaufnahme angekündigte Erhebung später auch zur Talentauswahl genutzt (Boyce & Linderbaum, 2019). Diese schleichende Zweckverschiebung („Purpose Drift“) beschädigt Vertrauen nachhaltig und ist in Deutschland zudem rechtlich problematisch (Kapitel 10).

4.4 Bewährte Lösungsmuster

Vier Regeln lassen sich ableiten. *Erstens* wird der Zweck vorab verbindlich festgelegt und nie rückwirkend geändert; Church berichtet, sich als Berater aus einem Projekt mit über 180 Topführungskräften zurückgezogen zu haben, als ein verdeckter Entscheidungszweck bekannt wurde (Church et al., 2019b). *Zweitens* erhalten unterschiedliche Zwecke unterschiedliche Instrumente – bei PepsiCo fließt ein jährliches Upward-Feedback in die Bewertung der „People Results“ ein, die 50 Prozent der Leistungsbewertung ausmachen, während ein separates 360°-Feedback der Entwicklung dient (Bracken & Church, 2013). *Drittens* werden Konsequenzen schrittweise eingeführt, idealerweise nach einem folgenlosen Pilotdurchgang. *Viertens* sind 360°-Daten nie alleinige Entscheidungsgrundlage, sondern ein gewichteter Baustein neben Ergebnissen und weiteren Verfahren (Church, 2019; Pulakos & Rose, 2019).

Praxisfall Brenner (1): Die Zweckklärung

In einem Workshop mit Geschäftsführung, Personalleitung und Betriebsrat zeigt sich, dass zwei Zwecke vermischt wurden. Man einigt sich: Der erste Durchgang dient ausschließlich der Entwicklung; der Bericht gehört der Führungskraft, die ihn mit ihrer Vorgesetzten und einem Coach bespricht. Für die Nachfolgeplanung wird ein separates, offen angekündigtes Verfahren für die sechs Kandidatinnen und Kandidaten entwickelt. Beides regelt eine Betriebsvereinbarung.

5. Was gemessen werden soll: Inhalte und Instrumente

5.1 Verhalten statt Eigenschaften

Die wichtigste inhaltliche Regel ist zugleich die am häufigsten verletzte. In der Literaturlauswertung von Campion, Campion und Campion (2019) über 221 Quellen gehören jobbezogene Items (58 Belege) und die Erfassung beobachtbaren Verhaltens statt Eigenschaften (41 Belege) zu den am besten gestützten Gestaltungsregeln. Formulierungen wie „versteht“, „weiß“ oder „ist belastbar“ verlangen Schlüsse auf innere Zustände, die niemand beobachten kann (Bracken & Rotolo, 2019). Hilfreich ist die Unterscheidung von *Kompetenz* (zugeschriebene Fähigkeit), *Verhalten* (gezeigtes Handeln) und *Aufgabenerfüllung* (erwarteter Beitrag samt Ergebnis). 360°-Feedback eignet sich vor allem für das Verhalten; ein niedriger Wert bei „Strategie“ kann fehlende Fähigkeit bedeuten – oder schlicht fehlende Zuständigkeit.

5.2 Ein evidenzbasierter Referenzrahmen

Als inhaltliche Landkarte eignet sich Yukls (2012) hierarchische Taxonomie, die fünf Jahrzehnte Forschung zu Führungsverhalten in vier Meta-Kategorien mit 15 Verhaltensweisen ordnet. Sie ist breiter als einzelne Führungsverhaltensstilmodelle, weil sie auch operative Steuerung und das Handeln über Teamgrenzen hinweg einschließt.

Abbildung 2: Yukls Taxonomie des Führungsverhaltens und ihre Beobachtbarkeit

Nach Yukl (2012, Tab. 1); Spalte „Gut beobachtbar durch“ als eigene Ableitung aus Yukl (2012, S. 68) und Bracken & Rotolo (2019).

In einer Studie von Yukl, Mahsud, Prussia und Hassan (2019) mit 235 Beschäftigten erklärten Aufgaben-, Beziehungs- und Veränderungsverhalten gemeinsam 47 Prozent der Varianz in der wahrgenommenen Führungseffektivität. Am stärksten hing *Probleme lösen* mit der Effektivität zusammen ($r = .65$), während für die Arbeitszufriedenheit allein das Beziehungsverhalten bedeutsam war ($\beta = .58$). Da Verhalten und Kriterium von derselben Person eingeschätzt wurden, sind diese Werte nach oben verzerrt; mit unabhängigen Leistungsmaßen fallen Zusammenhänge deutlich schwächer aus (Yukl, 2012). Für die Praxis wichtig: Externe Orientierung, kollektives Lernen und Problemlösen fehlen in vielen Unternehmensfragebögen.

5.3 Spezifisch vor breit – und das „Zuviel des Guten“

Für Entwicklungsfeedback ist die Ebene der einzelnen Verhaltensweisen informativer als die der Meta-Kategorien: Ein niedriger Wert in „Aufgabenorientierung“ sagt nicht, ob jemand zu wenig plant, Ziele unklar formuliert oder Probleme liegen lässt (Yukl, 2012). Zudem kann fast jedes Verhalten über- wie unterdosiert sein. Führungskräfte werden fünfmal häufiger als „übertreibend“ in ihren Stärkenbereichen eingeschätzt als anderswo, und der Höchstwert einer Fünferskala bildet in etwa einem Drittel der Fälle ein Zuviel ab (Kaiser & Chamorro-Premuzic, 2019). Antwortformate wie „zu wenig – angemessen – zu viel“ machen solche Übertreibungen sichtbar.

5.4 Instrumente im Vergleich

Einen Testsieger gibt es nicht; geeignet ist, was zu Zweck, Zielgruppe und Sprache passt.

Tabelle 3: Ausgewählte Instrumente für Führungsfeedback

Instrument	Struktur	Sprache	Validierungsstand	Eignung für 360°	Nutzungsrechte
German Managerial Practice Survey (GMPS)	4 Skalen, 15 Items (je ein Item pro Yukl-Verhalten)	Deutsch	N = 328, nur Mitarbeitendensicht; $\alpha = .70-.81$	für Upward-/360°-Feedback vorgesehen; nur Skalenebene	CC BY-NC-SA (nicht kommerziell)
Managerial Practices Survey (MPS G-16-4)	16 Skalen, 64 Items	Englisch	etabliertes Forschungsinstrument	für Mitarbeitendenurteil konzipiert	Beratungseinsatz zustimmungs- und vergütungspflichtig
LEaD-Kompetenzmodell	5 Faktoren, 18 Dimensionen, 54 Items	Deutsch	4 Studien, N = 720; $\alpha = .79-.91$	ausdrücklich für 90°- bis 360°-Feedback	kommerziell
Benchmarks (CCL)	Führungskompetenzen	Englisch, Übersetzungen	große Normdatenbasis	etabliertes 360°-Instrument	kommerziell
Leadership Practices Inventory (LPI)	5 Praktiken, 30 Items	mehrsprachig	große Datenbasis	bei Nutzung des Modells „Five Practices“	kommerziell
Multifactor Leadership Questionnaire (MLQ 5X)	Führungsstile; 45 Items	mehrsprachig	umfangreich beforscht	theoriegebundene Stildiagnose	kommerziell
Leadership Versatility Index (LVI)	Balance gegensätzlicher Verhaltensweisen	Englisch	Forschung mit Mehrquellendaten	Skala „zu wenig/zuviel“	kommerziell

Quellen: Fischer (2021); Yukl (2012); Schmidt-Huber et al. (2014); Posner (2016); Vergauwe et al. (2022); Furr et al. (2024); Anbieterangaben; eigene Bewertung.

Drei Befunde verdienen Aufmerksamkeit. Der **GMPS** ist transparent dokumentiert und mit rund vier Minuten Bearbeitungszeit äußerst ökonomisch; Beteiligung und Delegation, die bei Yukl zum Befähigen gehören, deckt er jedoch nicht ab, und geprüft wurde er bislang nur aus Mitarbeitendensicht (Fischer, 2021). Das **LEaD-Modell** ist unter den hier betrachteten deutschsprachigen Modellen das einzige, das theoretisch breit hergeleitet und an Außenkriterien validiert ist ($r = .53$ bis $.77$ mit der eingeschätzten Effektivität); gegenüber dem MPS erklärt es aber nur ein bis fünf Prozentpunkte zusätzliche Varianz, und mit Fehlzeiten oder variabler Vergütung hängt es kaum zusammen (Schmidt-Huber et al., 2014). Und **Nutzungsrechte** sind ein unterschätztes Kriterium: Die Lizenz ist zu klären, bevor die erste Einladung verschickt wird.

5.5 Standardinstrument oder Maßanfertigung?

Rund 85 Prozent aller Kompetenzmodelle decken konzeptionell denselben Inhalt ab (Church, 2014, zit. n. Church & Burke, 2019). Unterscheidungskraft gewinnt ein 360°-Feedback daher durch den Zuschnitt auf Strategie, Werte und Rolle; alle sieben Unternehmensfallstudien des Handbuchs arbeiten mit eigenen Instrumenten (Adler, 2019). Bewährt hat sich der Weg über eine Anforderungsanalyse (Fachbeitrag Nr. 37) zu verhaltensverankerten Aussagen (Fachbeitrag Nr. 26) und einem kompakten Instrument mit 40 bis 60 Items und höchstens 30 Minuten Bearbeitungszeit (Fleenor, 2019; Bracken & Rotolo, 2019). Ein guter Kompromiss: ein evidenzbasierter Kern wie die Yukl-Kategorien, ergänzt um wenige unternehmensspezifische Aussagen.

6. Messqualität: Wie verlässlich sind Mehrquellenurteile?

6.1 Übereinstimmung und Reliabilität

Die vielleicht wichtigste und am wenigsten bekannte Tatsache über 360°-Feedback: Die Übereinstimmung der Beurteilenden ist gering – innerhalb der Gruppen wie zwischen ihnen.

Tabelle 4: Kennzahlen zur Messqualität

Kennwert	Befund	Quelle
Übereinstimmung innerhalb einer Gruppe	Vorgesetzte $\approx$.50, Kollegen $\approx$.40, Mitarbeitende $\approx$.30 (Zielwert .70–.80)	Conway & Huffcutt (1997), zit. n. Pulakos & Rose (2019)
Übereinstimmung zwischen Gruppen	.20 bis .50	Pulakos & Rose (2019)
Korrelation Selbst – Mitarbeitende	.14 (niedrigster Wert aller Paare)	Conway & Huffcutt (1997), zit. n. Fleenor et al. (2010)
Korrelation Vorgesetzte – Kollegen	.34 (höchster Wert aller Paare)	ebd.
Zusammenhang mit objektiven Leistungsmaßen	$r \approx$.20 bis Mitte .30	Pulakos & Rose (2019)
Zusammenhang 360° – Vorgesetztenbeurteilung	$r =$.22 (nicht signifikant), Fall General Mills	Maylett (2019)

Unbrauchbar wird das Verfahren dadurch nicht. Ein Teil der geringen Übereinstimmung ist echte Perspektivenvielfalt – Mitarbeitende erleben eine Führungskraft in anderen Situationen als deren Vorgesetzte –, ein anderer Teil Messfehler durch unterschiedliche Beobachtungsgelegenheiten, Maßstäbe oder Sympathie. Beides lässt sich im Einzelfall kaum trennen (Furnham, 2019). Vor jeder Mittelwertbildung sollte deshalb geprüft werden, ob die Urteile innerhalb einer Gruppe hinreichend übereinstimmen (Fleenor et al., 2010).

6.2 Wie viele Beurteilende braucht man?

Die übliche Mindestzahl von drei Beurteilenden pro Gruppe ist eine Vertraulichkeits-, keine Präzisionsregel. Für eine verlässliche Reliabilität wären etwa vier Vorgesetzte, acht Kollegen und neun Mitarbeitende nötig (Greguras & Robie, 1998, zit. n. Scott et al., 2019). Fleenor (2019) empfiehlt, alle direkten Mitarbeitenden und in den übrigen Gruppen vier bis sechs Personen einzubeziehen. Ebenso wichtig ist die Gewichtung: Wird über eine Vorgesetzte, vier Mitarbeitende und vier Kollegen ungewichtet gemittelt, zählt das Vorgesetztenurteil nur elf Prozent; Church (2019) empfiehlt gleich gewichtete Gruppen. Gruppen unter der Berichtsschwelle sollten auch nicht in einen Gesamtwert einfließen, weil sich ihre Antworten sonst zurückrechnen lassen (Bracken & Rotolo, 2019).

6.3 Verzerrungen und ihre Ursachen

Das Kernproblem fast aller 360°-Daten ist die **Milde**: Die Werte ballen sich am oberen Skalenende (Bracken & Rotolo, 2019). Bei General Mills lag der 360°-Mittelwert bei 4,4, der der Vorgesetztenbeurteilung derselben Personen bei 3,6; über 90 Prozent schnitten im 360°-Feedback besser ab (Maylett, 2019). Hinzu kommen **Halo-Effekte**, **Sympathie** – die Aufwärts- und Kollegenurteile stärker verzerrt als Urteile von oben – und **strategisches Urteilen** bis hin zu Absprachen, wenn Ergebnisse über Beförderungen mitentscheiden (Fleenor et al., 2010; Pulakos & Rose, 2019). Die Bereitschaft zu ehrlichen Urteilen hängt vor allem von Zynismus gegenüber dem Management und der Angst vor Vergeltung ab (Smith & Fortunato, 2008, zit. n. Bracken & Rotolo, 2019). **Geschlechterstereotype** zeigen sich weniger in Gesamtwerten als in einzelnen Dimensionen: Frauen erhalten mitunter gleich gute Leistungs-, aber schlechtere Beförderungsurteile (Valerio & Sawyer, 2019). Wer 360°-Daten für Beförderungen nutzt, sollte diese Dimensionen gezielt prüfen.

6.4 Gestaltungsregeln für Items und Skalen

Kasten 2: Zehn Regeln für Items und Skalen

1. Jedes Item beschreibt genau ein beobachtbares Verhalten.
2. Keine negativ formulierten Items; sie bilden eigene Faktoren.
3. Items thematisch gruppieren statt zufällig mischen.
4. Den Beobachtungszeitraum begrenzen, in der Regel auf zwölf Monate.
5. Fünf bis sieben Skalenstufen, jede sprachlich beschriftet; die Mitte als „solide, erfüllt die Erwartungen“ verankern.
6. „Nicht beobachtet“ als eigene Option anbieten – nie mit dem niedrigsten Skalenwert zusammengelegt.
7. Keine Zustimmungsskalen („stimme zu“) für Verhaltensurteile.
8. Für Entscheidungszwecke keine reinen Häufigkeitsskalen: Häufig ist nicht gleich wirksam.
9. Beurteilende mit einem kurzen Frame-of-Reference-Training auf gemeinsame Maßstäbe einstimmen.
10. Beurteilenden bei der Eingabe Rückmeldung zu auffälligen Mustern geben (z. B. überall derselbe Wert); das senkt nachweislich die Milde.

Nach Bracken & Rotolo (2019, S. 263–285); Fleenor (2019); Furnham (2019).

6.5 Prädiktor oder Kriterium?

360°-Urteile messen in erster Linie die heutige Leistung in der heutigen Rolle. Als Prädiktor für eine künftige Rolle taugen sie nur, soweit sich die Anforderungen überschneiden – und externe Bewerbende haben ohnehin keine 360°-Daten (Pulakos & Rose, 2019). Für Nachfolgeentscheidungen gehören sie deshalb mit zukunftsbezogenen Verfahren wie Simulationen und strukturierten Interviews kombiniert. Ihren Eigenwert zeigt der General-Mills-Fall: Vorgesetzte bewerteten vor allem das *Was* – Kennzahlen, Sicherheit, Regeltreue –, Kolleginnen und Kollegen das *Wie*. Eine Gruppe mit guter Vorgesetztenbeurteilung, aber schwachem 360°-Feedback erreichte ihre Produktionsziele auf Kosten von Beziehungen und langfristiger Wirksamkeit (Maylett, 2019).

Praxisfall Brenner (2): Kleine Teams, kleine Gruppen

Viele Meisterinnen und Meister bei Brenner führen Teams mit drei bis fünf Beschäftigten; Kolleginnen und Kollegen auf gleicher Ebene gibt es oft nur zwei. Statt Gruppen mit zwei Antworten auszuweisen, werden Kollegen und sonstige Beurteilende zu „weitere Beurteilende“ zusammengefasst. Das Vorgesetztenurteil wird gesondert gezeigt, die Gruppen werden gleich gewichtet. Die Betriebsvereinbarung hält fest, dass Gruppen unter drei Personen nie ausgewiesen werden.

7. Selbstbild und Fremdbild richtig deuten

Der Abgleich von Selbst- und Fremdbild gilt als Herzstück des Verfahrens – und wird häufig überinterpretiert.

Das Fremdurteil ist keine objektive Wahrheit. Unterschiede zwischen den Quellen sind Information, aber auch das gemittelte Urteil der anderen ist keine „wahre“ Leistung; in einer klassischen Längsschnittstudie ging der verlässliche Anteil der Urteilsvarianz überwiegend auf die Beurteilenden zurück (Fleenor et al., 2010). Ernst nehmen sollte die Führungskraft das Fremdbild trotzdem: Wahrnehmungen prägen Zusammenarbeit und Vertrauen.

Selbstüberschätzung folgt einem Muster. Sie nimmt mit Hierarchieebene und Alter zu und ist bei Männern häufiger (Fleenor et al., 2010). Führungskräfte, die sich deutlich besser einschätzen als ihr Umfeld, werden von Mitarbeitenden am niedrigsten bewertet und reagieren auf Kritik häufiger mit Abwehr; wer sich eher unterschätzt, scheitert seltener in der Laufbahn (Nowack, 2019; Furnham, 2019).

Für Deutschland gilt ein bemerkenswerter Befund. Atwater, Waldman, Ostroff, Robie und Johnson (2005) untersuchten 2.732 Führungskräfte aus europäischen Ländern und verglichen die Befunde mit US-Daten. In den USA hing die Selbst-Fremd-Übereinstimmung mit der Effektivität zusammen; in Deutschland, Dänemark, Italien und Frankreich sagte allein das Fremdurteil die Effektivität voraus (Fleenor et al., 2010; Nowack, 2019). Für deutsche Anwender heißt das: Das Fremdbild verdient die Aufmerksamkeit, die Differenz zum Selbstbild ist vor allem ein Reflexionsanlass.

Methodische Konsequenzen. Einfache Differenzwerte und Etiketten wie „Überschätzer“ oder „Unterschätzer“ gelten als methodisch überholt, weil sie die Beiträge beider Urteile vermischen (Fleenor et al., 2010); in Berichte gehören sie nicht. Als Faustregel gilt eine Abweichung von mehr als einem Punkt auf einer Fünfer-skala als Hinweis auf eingeschränkte Selbstwahrnehmung – und damit als Anlass für ein klärendes Gespräch, nicht als Diagnose.

8. Wirkt 360°-Feedback? Die Evidenzlage

8.1 Was die Meta-Analysen zeigen

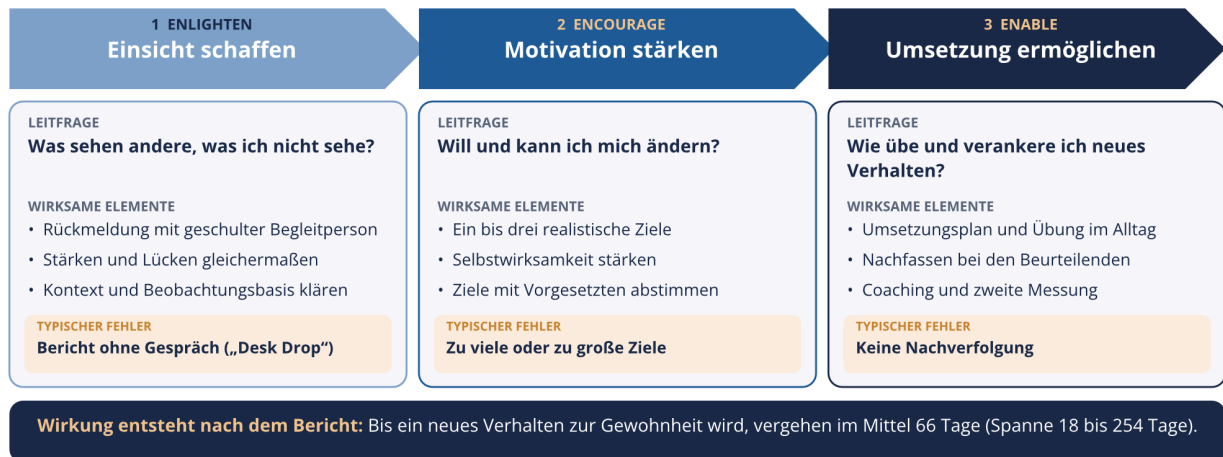
Die meistzitierte Meta-Analyse von Smither, London und Reilly (2005) über 24 Längsschnittstudien zeigt: Nach einem 360°-Feedback verbessern sich die späteren Fremdurteile im Durchschnitt – aber nur geringfügig und mit großen Unterschieden zwischen Personen und Programmen. Neuere Synthesen bestätigen das nüchterne Bild. In Führungstrainings waren Programme mit 360°-Feedback nicht signifikant wirksamer als solche mit Feedback aus nur einer Quelle (Lacerenza et al., 2017). In der Coaching-Meta-Analyse von Jones et al. (2016) fielen die Effekte in Studien mit Mehrquellenfeedback sogar kleiner aus – ein Hinweis, dass mehr Information nicht automatisch mehr Wirkung bedeutet. Eine Campbell-Übersichtsarbeit ist seit Mai 2026 als Protokoll registriert; Ergebnisse stehen aus (Wietrak et al., 2026). Fazit: Der durchschnittliche Effekt ist klein, die Streuung groß – und genau darin liegt der Gestaltungsspielraum.

8.2 Wann Feedback schadet

Mehrere Studien fanden keine messbare Veränderung oder sogar negative Effekte auf Engagement und Leistung (Nowack, 2019). Smither und Walker (2004) zeigten an 176 Führungskräften: Wenige, verhaltensbezogene kritische Kommentare führten zu Verbesserungen, sehr viele negative Kommentare zu einer Verschlechterung (zit. n. Nowack, 2019). Von über 1.300 untersuchten Freitexten waren nur 32 Prozent konkret genug, um Handeln anzuleiten (Rose, 2019). Und auf erstes kritisches Feedback reagieren viele mit Schock, Ärger und Ablehnung, bevor sie es annehmen (Fleenor, 2019). Ein Bericht ohne Begleitung kann daher mehr schaden als nützen.

8.3 Wie Wirkung entsteht

Abbildung 3: Von der Einsicht zur Veränderung – die Wirkungskette



Nach Nowack (2019); McCauley & Brutus (2019); eigene Darstellung.

Die Befunde stützen diese Logik. Feedback ohne Maßnahmenplanung bleibt praktisch wirkungslos (Church & Burke, 2019). Führungskräfte, die nach dem Feedback mit ihren Beurteilenden sprechen, Ziele benennen und später nachfragen, verbessern sich deutlich häufiger (Goldsmith & Morgan, 2004, zit. n. Rose, 2019). Coaching hilft, Einsichten in Ziele zu übersetzen (Nowack, 2019). Und Verhaltensänderung braucht Zeit: Bis ein neues Verhalten zur Gewohnheit wird, vergehen im Mittel 66 Tage, mit einer Spanne von 18 bis 254 Tagen (Lally et al., 2010, zit. n. Nowack, 2019) – genau der Übungszyklus, den unser Fachbeitrag Nr. 52 als Selbstentwicklungs-Loop beschreibt.

8.4 Faktencheck: Verbreitete Behauptungen

Tabelle 5: Behauptungen und Befunde im Vergleich

Verbreitete Behauptung	Befund der Forschung
„360°-Feedback ist objektiver als ein Vorgesetztenurteil.“	Es ist <i>breiter</i> , nicht objektiver: Die Übereinstimmung zwischen Gruppen liegt bei .20–.50; Milde und Sympathie verzerren gerade Aufwärts- und Kollegenurteile.
„Anonymität garantiert ehrliche Antworten.“	Anonymität fördert Offenheit, garantiert sie aber nicht; Ehrlichkeit hängt von Vertrauen und Angst vor Vergeltung ab. In kleinen Gruppen ist Anonymität kaum herstellbar.
„Je ein Drittel der Studien zeigt Verbesserung, Verschlechterung und keinen Effekt.“	Die Zahl stammt aus einer Meta-Analyse zu Feedback <i>allgemein</i> (Kluger & DeNisi, 1996), nicht zu 360°-Feedback (Bracken et al., 2016).
„Die Leistung erklärt nur 21 Prozent der Urteilsvarianz.“	Das entspräche einer Korrelation von etwa .46 – ein guter Wert für ein Personalverfahren; zudem war das Leistungskriterium aus den Urteilen selbst gebildet (Bracken et al., 2016).
„90 Prozent der Fortune-500-Unternehmen nutzen 360°-Feedback.“	Keine nachvollziehbare Erhebung; belastbare Quoten sind deutlich vorsichtiger (Kapitel 3).
„360°-Feedback plus Coaching ist immer überlegen.“	Coaching wirkt im Mittel positiv, ein Zusatznutzen gerade des 360°-Feedbacks ist nicht belegt (Jones et al., 2016; de Haan & Nilsson, 2023).

9. Erfolgsfaktoren: Den Prozess wirksam gestalten

Fünf typische Gründe für das Scheitern von 360°-Prozessen sind ein unklarer Zweck, eine nicht vorbereitete Organisation, ein ungeeignetes Instrument, eine schlechte Logistik und fehlende Nachbereitung (Fleenor,

Taylor & Chappelow, 2008, zit. n. Fleenor, 2019). Gescheiterte Prozesse scheitern also aus vorhersehbaren Gründen. Tabelle 6 ordnet die wichtigsten Entscheidungen den sechs Prozessphasen zu.

Tabelle 6: Die sechs Phasen eines 360°-Prozesses

Phase	Schlüsselentscheidungen	Häufige Fehler
1. Klärung und Readiness	Zweck, Zielgruppe, Datenzugriff, Sponsor; Vertrauensklima und Zeitpunkt prüfen; Topmanagement als Pilotgruppe	Start in Restrukturierungsphasen; vager Zweck
2. Design	Inhalte aus Strategie und Anforderungsanalyse; Skala; Berichtsschwellen; Gewichtung; Regeln für Freitexte	Katalog „von der Stange“; zu lang; Zustimmungsskala
3. Durchführung	Beurteilende: Vorschlag durch die Führungskraft, Freigabe durch Vorgesetzte; mindestens sechs Monate Zusammenarbeit; höchstens acht bis zehn Fragebögen pro Person; Kurztraining	unterschiedliche Regeln für verschiedene Führungskräfte; Überlastung
4. Auswertung und Bericht	Gruppen getrennt, mit Fallzahl und Streuung; Rücklauf unter 75 % als Warnsignal	Ampelfarben, Scheinpräzision, rückrechenbare Kleinstgruppen
5. Rückmeldung	Gespräch mit geschulter Begleitperson nach eigener Lektüre; Stärken und Lücken; ein bis drei Ziele	Bericht ohne Gespräch; Fixierung auf Schwächen; Suche nach „dem schlechten Beurteiler“
6. Umsetzung und Evaluation	Dreiergespräch mit Vorgesetzten und Coach; Nachfassen bei Beurteilenden; Wiederholung nach 12 bis 18 Monaten	keine Nachverfolgung; Veränderungswerte bei wechselnden Beurteilenden überinterpretiert

Nach Fleenor (2019); Bracken & Rotolo (2019); McCauley & Brutus (2019); Campion et al. (2019); eigene Darstellung.

Vier Faktoren kosten wenig und bewirken viel. **Einheitliche Spielregeln:** Wird die eine Führungskraft von allen Mitarbeitenden beurteilt, während die andere ihre Beurteilenden frei wählt, gilt das Verfahren als unfair und scheitert (Fleenor, 2019). **Begleitung statt Bericht:** In einem Großunternehmen haben 27 Prozent von über 4.000 Führungskräften ihren Bericht nie heruntergeladen (Bracken et al., 2016). Begleitung muss nicht teuer sein; drei einstündige Coachingtermine über sechs Monate haben sich bewährt, ein einzelner Termin wurde an der London Business School als zu wenig verworfen (McCauley & Brutus, 2019; Conger, 2019). **Gestaffelter Rollout:** Ein rollierendes Verfahren – etwa ein Viertel der Führungskräfte pro Quartal – entlastet Beurteilende und Coaches (McCauley & Brutus, 2019). **Saubere Wiederholungsmessung:** Veränderungswerte sind nur interpretierbar, wenn überwiegend dieselben Personen erneut urteilen; liegt die Überschneidung unter 75 Prozent, sind Differenzen irreführend (Fleenor, 2019).

Praxisfall Brenner (3): Der Fahrplan

Brenner startet mit der Geschäftsführung und den sieben Bereichsleitungen als Pilotgruppe; die übrigen Führungskräfte folgen in drei Wellen im Abstand von drei Monaten. Das Instrument umfasst 42 Aussagen: 30 aus den vier Yukl-Kategorien und zwölf zu den Brenner-Führungsleitlinien, darunter Arbeitssicherheit. Vor der Eingabe sehen alle Beurteilenden ein zehnminütiges Video mit Beispielen für jede Skalenstufe. Jede Führungskraft erhält ihren Bericht zwei Tage vor dem ersten von drei Coachingterminen; das Gespräch mit dem eigenen Team ist Pflicht, die Weitergabe des Berichts nicht.

10. Rechtlicher und ethischer Rahmen in Deutschland

Die Handbuchbeiträge zum Recht beziehen sich auf die USA, ihre Kernbotschaft lässt sich aber übertragen: Sobald 360°-Daten Personalentscheidungen beeinflussen, werden sie zu einem Beurteilungs- und Auswahlverfahren und müssen entsprechend fair, nachvollziehbar und valide sein (Scott et al., 2019). In Deutschland

treten Betriebsverfassungs-, Datenschutz- und Antidiskriminierungsrecht hinzu. Die folgende Übersicht ist eine fachliche Einordnung und ersetzt keine Rechtsberatung.

Betriebsverfassungsrecht. Ein standardisierter 360°-Fragebogen mit Beurteilungsdimensionen berührt regelmäßig das Mitbestimmungsrecht bei allgemeinen Beurteilungsgrundsätzen (§ 94 Abs. 2 BetrVG); eine Online-Plattform ist eine technische Einrichtung, die zur Überwachung von Verhalten oder Leistung geeignet ist (§ 87 Abs. 1 Nr. 6 BetrVG). Fließen Ergebnisse in Beförderungen ein, kommen Auswahlrichtlinien nach § 95 BetrVG in Betracht. Hinzu kommen Erörterungs-, Einsichts- und Beschwerderechte der Beschäftigten (§§ 82–84 BetrVG). Ein späterer Wechsel vom Entwicklungs- zum Entscheidungszweck ist daher ohne Anpassung der Betriebsvereinbarung nicht möglich (vgl. Fachbeitrag Nr. 35).

Datenschutzrecht. Der Grundsatz der Zweckbindung (Art. 5 Abs. 1 lit. b DSGVO) deckt sich mit der fachlichen Regel, den Zweck nie rückwirkend zu ändern. Als Rechtsgrundlage kommen Art. 6 und 88 DSGVO in Verbindung mit § 26 BDSG oder einer Betriebsvereinbarung in Betracht; die Tragfähigkeit von § 26 BDSG ist seit dem EuGH-Urteil vom 30. März 2023 (C-34/21) umstritten. Einwilligungen sind im Arbeitsverhältnis nur eingeschränkt freiwillig. Das Auskunftsrecht der Beurteilten (Art. 15 DSGVO) ist mit dem Schutz der Beurteilenden abzuwägen; externe Anbieter brauchen einen Vertrag zur Auftragsverarbeitung (Art. 28 DSGVO), bei umfangreichen Verfahren ist eine Datenschutz-Folgenabschätzung (Art. 35 DSGVO) zu prüfen.

Antidiskriminierung und KI. Systematische Unterschiede zulasten einer Gruppe können bei Beförderungsnutzung eine mittelbare Benachteiligung nach § 3 Abs. 2 AGG begründen; 360°-Urteile sind dafür nach US-Studien nicht weniger anfällig als Vorgesetztenurteile (Scott et al., 2019). Die Organisationspflichten des § 12 AGG sprechen für verpflichtende Schulungen der Beurteilenden. KI-Systeme zur Bewertung von Leistung und Verhalten oder für Beförderungen gelten nach Anhang III der Verordnung (EU) 2024/1689 als Hochrisiko-Systeme.

Ethik. Macey und Barbera (2019) raten, nicht pauschal „Anonymität“ zu versprechen, sondern präzise festzulegen, was vertraulich bleibt; eine teilweise Anonymität gibt es nicht. Besondere Risiken bergen die Berichtsphase, etwa rückrechenbare Teilgruppen, und Freitexte, deren Stil die Verfassenden verraten kann. Ändern sich die Umstände, etwa durch einen Eigentümerwechsel, darf die ursprüngliche Vereinbarung nicht einseitig umgedeutet werden.

Kasten 3: Was eine Betriebsvereinbarung zum 360°-Feedback regeln sollte

1. Zweck und ausdrücklich ausgeschlossene Verwendungen
2. Teilnehmende, Freiwilligkeit und Rhythmus
3. Inhalte und Instrument (als Anlage) sowie Verfahren bei Änderungen
4. Auswahl der Beurteilenden und Mindestgruppengrößen
5. Vertraulichkeit: wer welche Daten in welcher Form sieht
6. Umgang mit Freitexten
7. Rückmeldung, Begleitung und Entwicklungsmaßnahmen
8. Speicherdauer, Löschung und Nutzung für Benchmarks
9. Externe Dienstleister und Auftragsverarbeitung
10. Ausschluss automatisierter Entscheidungen
11. Erörterungs- und Beschwerderecht
12. Evaluation des Verfahrens und Information des Betriebsrats

Eigene Zusammenstellung auf Basis von Macey & Barbera (2019) und Scott et al. (2019).

11. Varianten, Alternativen und Trends

Team-360. Hier bewertet ein Team seine eigene Zusammenarbeit. Whirlpool erhob Team-, Leitungs- und Mitgliedersicht zu fünf Merkmalen von Hochleistungsteams, besprach die Ergebnisse in einem moderierten Workshop, hielt Vereinbarungen in einem „Team Compact“ fest und maß nach sechs Monaten erneut (Mockler et al., 2019). Team-360 ist weniger bedrohlich als Einzelfeedback und eignet sich auch für kleine Einheiten.

Stakeholder-Interviews. An der Spitze stoßen Fragebögen an Grenzen: wenige Beurteilende, kaum Anonymität, seltene, aber folgenreiche Situationen. Walmart ersetzte das quantitative 360°-Feedback für Topführungskräfte deshalb durch Interviews externer Coaches und eine Reputationsabfrage bei drei bis fünf Schlüsselpersonen (Stomski, 2019). Auch Aufsichtsgremien öffnen sich: Von 620 befragten US-Board-Mitgliedern wünschten sich 90 Prozent individuelles Feedback, nur 39 Prozent erhielten es (Winum, 2019).

Kombination mit anderen Verfahren. 360°-Feedback zeigt das *typische* Verhalten im Alltag, ein Assessment Center das *maximale* Verhalten, Persönlichkeitsdiagnostik die zugrunde liegenden *Tendenzen* (Shepherd, 2019). Kaiser und Chamorro-Premuzic (2019) verbinden 360°- und Persönlichkeitsdaten zu einer Vier-Felder-Matrix aus Stärken, Leistungsrisiken, angepassten Fähigkeiten und ungenutztem Potenzial (vgl. Fachbeiträge Nr. 8 und 17).

Kontinuierliches Feedback und Beurteilungen ohne Noten. Feedback-Apps und Kalibrierungsrunden ersetzen vielerorts jährliche Ratings (Hunt et al., 2019). Die Evidenz ist dünn: Nach Abschaffung der Noten sanken in einer CEB-Studie die Qualität der Mitarbeitergespräche um 14 Prozent und das Engagement um 6 Prozent (Rose, 2019).

KI und Datenanalyse. Textanalyse kann Freitexte verdichten, Unterschiede zwischen Beurteilergruppen können auf Risiken hinweisen, aggregierte Daten zeigen Entwicklungsbedarfe ganzer Bereiche (Fink & Sinar, 2019). Bisher nutzen erst rund fünf Prozent der Organisationen KI im 360°-Prozess (3D Group, 2025). Selbst die Handbuch-Herausgeber sind uneins: Fleenor skizziert ein vernetztes Datenökosystem bis hin zu Wearables, Bracken lehnt technische Verhaltensüberwachung ab. Gemeinsam erwarten sie mehr Entscheidungsnutzung, mehr Rechtsstreitigkeiten – und daraus folgend mehr methodische Strenge (Church et al., 2019b).

12. Leitfaden für den Mittelstand

Campion et al. (2019) halten 360°-Feedback für Kleinunternehmen und interaktionsarme Tätigkeiten oft für nicht sinnvoll. Mittelständler sollten deshalb nicht verzichten, aber das Format an ihre Größe anpassen. Die Hürden sind kleine Teams, die Mindestgruppengrößen verfehlen, persönliche Nähe zwischen Beurteilenden und Beurteilten, knappe Ressourcen – und eine Unternehmensspitze, bei der kaum ehrliches Feedback ankommt.

Tabelle 7: Entscheidungsmatrix – welches Format passt?

Situation	Zweck: Entwicklung	Zweck: Entscheidung (z. B. Nachfolge)
Teams mit unter fünf Mitarbeitenden	Upward Feedback mit zusammengefassten Gruppen oder Team-360	nicht empfohlen; Ergebnisse und strukturierte Gespräche nutzen
Breite Führungsebene (5–15 Mitarbeitende je Führungskraft)	kompaktes 360°-Feedback (30–45 Items), Coaching	erst nach zwei folgenlosen Durchgängen; nur als gewichteter Baustein
Geschäftsführung und Bereichsleitungen	interviewbasiertes Stakeholder-Feedback durch Externe	Anforderungsprofil der Zielrolle → 360° + Simulation + Interview
Führungsteams	Team-360 mit moderiertem Workshop	–
Familienunternehmen mit Beirat	Feedback für Geschäftsführung und Beirat durch Externe	Rolle definieren → Person einschätzen → Entwicklung beschleunigen (Adler, 2019)

Eigene Ableitung aus den in diesem Beitrag dargestellten Befunden.

Kasten 4: Checkliste in zwölf Schritten

1. Readiness prüfen: Vertrauen, Zeitpunkt, keine laufende Restrukturierung.
2. Zweck festlegen und schriftlich fixieren.
3. Geschäftsführung als Sponsor und Pilotgruppe gewinnen.
4. Betriebsrat frühzeitig einbinden; Betriebsvereinbarung schließen.
5. Datenschutz klären: Rechtsgrundlage, Auftragsverarbeitung, Speicherdauer.
6. Inhalte ableiten: Anforderungsanalyse, Yukl-Kern plus wenige eigene Items.
7. Instrument wählen oder bauen; Nutzungsrechte klären.
8. Gruppen und Schwellen festlegen; kleine Gruppen zusammenfassen.
9. Beurteilende kurz schulen; Spielregeln einheitlich anwenden.
10. Berichte mit Streuung und Fallzahl; Rückmeldung nur mit Begleitung.
11. Ein bis drei Ziele, Teamgespräch, drei Coachingtermine, Nachfassen.
12. Nach 12 bis 18 Monaten erneut messen und das Verfahren selbst bewerten.

Praxisfall Brenner (4): Bilanz nach 18 Monaten

Der Rücklauf lag bei 86 Prozent. Die größten Selbst-Fremd-Unterschiede zeigten sich nicht bei den Meisterinnen und Meistern, sondern bei zwei Bereichsleitungen. Unternehmenweit wurden „Probleme lösen“ und „kollektives Lernen“ am niedrigsten bewertet; Brenner führte daraufhin kurze Nachbesprechungen nach Störungen ein. Im zweiten Durchgang verbesserten sich die Werte vor allem bei den Führungskräften, die mit ihrem Team über die Ergebnisse gesprochen hatten. Der Betriebsrat stimmte der Fortführung zu; in die Nachfolgeentscheidung flossen die 360°-Ergebnisse nur mit Zustimmung der Kandidatinnen und Kandidaten ein.

13. Fazit: Sieben Thesen zum 360°-Feedback

Mehr Perspektiven, mehr Wahrheit? Die Antwort der Forschung lautet: mehr Perspektiven, ja – mehr Wahrheit nur unter Bedingungen.

1. **360°-Feedback ist ein Prozess, kein Fragebogen.** Seine Qualität entsteht aus Zweck, Design, Begleitung und Umsetzung.
2. **Der Zweck ist die wichtigste Stellschraube.** Er bestimmt Antwortverhalten, Messanforderungen und Rechtsrahmen – und wird nie rückwirkend geändert.
3. **Gemessen wird Verhalten** – spezifisch, beobachtbar, an Strategie und Rolle ausgerichtet, mit Blick auf das Zuviel wie das Zuwenig.
4. **Mehrquellenurteile sind wertvoll, aber unpräzise.** Sie verlangen Bescheidenheit bei der Deutung, genügend Beurteilende und Berichte, die Unsicherheit zeigen.
5. **Das Fremdbild zählt mehr als die Differenz.** Gerade in Deutschland sind Selbst-Fremd-Abweichungen ein Gesprächsanlass, keine Diagnose.
6. **Wirkung entsteht nach dem Bericht.** Ziele, Gespräche mit den Beurteilenden, Coaching und Zeit zum Üben machen den Unterschied.
7. **Im Mittelstand schlägt Passung den Umfang.** Ein gut dimensioniertes Format – Upward Feedback, Team-360 oder Stakeholder-Interview – wirkt mehr als ein großes 360°-Feedback, das an Gruppengrößen, Vertrauen oder Ressourcen scheitert.

Richtig eingesetzt, ist 360°-Feedback kein Urteil über eine Person, sondern eine Einladung zur Entwicklung, die eine Organisation ernst meinen muss. Ob sie das tut, zeigt sich nicht am Bericht, sondern daran, was danach geschieht.

Literaturverzeichnis

Kapitel aus dem *Handbook of Strategic 360 Feedback* (Church et al., 2019a) sind einzeln aufgeführt. Primärstudien, die nach Sekundärquellen zitiert werden („zit. n.“), sind nicht gesondert verzeichnet.

- 3D Group. (2025, 30. Januar). *Study highlights limited usage of artificial intelligence in 360 feedback* [Pressemitteilung zur 8. Benchmarkstudie].
- Adler, S. (2019). Mitigating succession risk in the C-suite: A case study. In Church et al. (2019a), pp. 361–372.
- Boyce, C. C., & Linderbaum, B. (2019). What kind of talent do we have here? Using 360s to establish a baseline assessment of talent. In Church et al. (2019a), pp. 397–407.
- Bracken, D. W. (2019). What is "strategic 360 feedback"? In Church et al. (2019a), pp. 11–16.
- Bracken, D. W., & Church, A. H. (2013). The "new" performance management paradigm: Capitalizing on the unrealized potential of 360 degree feedback. *People & Strategy*, 36(2), 34–40.
- Bracken, D. W., Rose, D. S., & Church, A. H. (2016). The evolution and devolution of 360° feedback. *Industrial and Organizational Psychology*, 9(4), 761–794. <https://doi.org/10.1017/iop.2016.93>
- Bracken, D. W., & Rotolo, C. T. (2019). Can we improve rater performance? In Church et al. (2019a), pp. 255–289.
- Campion, E. D., Campion, M. C., & Campion, M. A. (2019). Best practices when using 360 feedback for performance appraisal. In Church et al. (2019a), pp. 19–59.
- Chartered Institute of Personnel and Development. (2016). *HR outlook: Winter 2015–16. Leaders' views of our profession*. CIPD.
- Church, A. H. (2019). Strategic 360 feedback for talent management. In Church et al. (2019a), pp. 97–121.
- Church, A. H., Bracken, D. W., Fleenor, J. W., & Rose, D. S. (Eds.). (2019a). *The handbook of strategic 360 feedback*. Oxford University Press.
- Church, A. H., Bracken, D. W., Fleenor, J. W., & Rose, D. S. (2019b). The handbook of strategic 360 feedback: Themes, prognostications, and sentiments. In Church et al. (2019a), pp. 517–530.
- Church, A. H., & Burke, W. W. (2019). Strategic 360 feedback for organization development. In Church et al. (2019a), pp. 213–235.
- Conger, J. A. (2019). Harnessing the potential of 360 feedback in executive education programming. In Church et al. (2019a), pp. 343–351.
- Dahling, J. J., & Chau, S. L. (2019). Moving beyond "the great debate": Recasting developmental 360 feedback in talent management. In Church et al. (2019a), pp. 149–158.
- de Haan, E., & Nilsson, V. O. (2023). What can we know about the effectiveness of coaching? A meta-analysis based only on randomized controlled trials. *Academy of Management Learning & Education*, 22(4), 641–661. <https://doi.org/10.5465/amle.2022.0107>
- Fink, A. A., & Sinar, E. F. (2019). Using analytics to gain more insights from 360 feedback data. In Church et al. (2019a), pp. 447–459.
- Fischer, J. A. (2021). German Managerial Practice Survey (GMPS): Fragebogen zur Messung von Führungsqualität. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. <https://doi.org/10.6102/zis296>
- Fleenor, J. W. (2019). Factors affecting the validity of strategic 360 feedback processes. In Church et al. (2019a), pp. 237–254.
- Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly*, 21(6), 1005–1034. <https://doi.org/10.1016/j.leaqua.2010.10.006>
- Furnham, A. (2019). Rater congruency: Why ratings of the same person differ. In Church et al. (2019a), pp. 291–307.
- Furr, R. M., Jeong, S. B., Fleenor, J. W., & Shrestha, S. (2024). Evaluating a novel method of scoring multirater assessments of leader competencies. *Consulting Psychology Journal*. Advance online publication. <https://doi.org/10.1037/cpb0000285>
- Hunt, S. T., Sherwood, J., & Bidwell, L. M. (2019). Technological innovations in the use of 360 feedback for performance management. In Church et al. (2019a), pp. 77–96.
- Jones, R. J., Woods, S. A., & Guillaume, Y. R. F. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, 89(2), 249–277. <https://doi.org/10.1111/joop.12119>
- Kaiser, R. B., & Chamorro-Premuzic, T. (2019). Integrating personality assessment with 360 feedback in leadership development and coaching. In Church et al. (2019a), pp. 193–212.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology*, 102(12), 1686–1718. <https://doi.org/10.1037/apl0000241>

- London, M., & Smither, J. W. (2019). Historical challenges of using 360 feedback for performance evaluation. In Church et al. (2019a), pp. 61–75.
- Macey, W. H., & Barbera, K. M. (2019). The ethical context of 360 feedback. In Church et al. (2019a), pp. 461–478.
- Maylett, T. M. (2019). The journey from development to appraisal: 360 feedback at General Mills. In Church et al. (2019a), pp. 327–341.
- McCauley, C., & Brutus, S. (2019). Application of 360 feedback for leadership development. In Church et al. (2019a), pp. 135–148.
- Mockler, S., McGourty, R., & Goudy, K. (2019). Leveraging team 360 to drive business-enhancing change across the enterprise at Whirlpool Corporation. In Church et al. (2019a), pp. 385–396.
- Nowack, K. M. (2019). From insight to successful behavior change: The real impact of development-focused 360 feedback. In Church et al. (2019a), pp. 175–192.
- Posner, B. Z. (2016). Investigating the reliability and validity of the Leadership Practices Inventory. *Administrative Sciences*, 6(4), Article 17. <https://doi.org/10.3390/admsci6040017>
- Pulakos, E. D., & Rose, D. S. (2019). Is 360 feedback a predictor or criterion measure? In Church et al. (2019a), pp. 309–324.
- Rose, D. S. (2019). 360 feedback versus alternative forms of feedback: Which feedback methods are best suited to enable change? In Church et al. (2019a), pp. 409–426.
- Schmidt-Huber, M., Dörr, S., & Maier, G. W. (2014). Die Entwicklung und Validierung eines evidenzbasierten Kompetenzmodells effektiver Führung (LEaD). *Zeitschrift für Arbeits- und Organisationspsychologie*, 58(2), 80–94. <https://doi.org/10.1026/0932-4089/a000138>
- Scott, J. C., Scott, J. M., & Foster, K. E. (2019). The legal environment for 360 feedback. In Church et al. (2019a), pp. 479–501.
- Shepherd, W. J. (2019). Integrating strategic 360 feedback at a financial services organization. In Church et al. (2019a), pp. 373–384.
- Shields, J., Kim, S., Chhetri, A., Stanton, P., & Nankervis, A. (2023). Traditional, transitional and new performance management practices in Australian organisations: Incidence, coverage and perceived effectiveness. *Asia Pacific Journal of Human Resources*, 61(3), 554–581. <https://doi.org/10.1111/1744-7941.12372>
- Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58(1), 33–66. <https://doi.org/10.1111/j.1744-6570.2005.514.1.x>
- Stomski, L. (2019). An alternative form of feedback: Using stakeholder interviews to assess reputation at Walmart. In Church et al. (2019a), pp. 353–360.
- Valerio, A. M., & Sawyer, K. (2019). Gender, diversity, and 360 feedback. In Church et al. (2019a), pp. 427–445.
- Vergauwe, J., Hofmans, J., & Wille, B. (2022). The leadership arena–reputation–identity (LARI) model: Distinguishing shared and unique perspectives in multisource leadership ratings. *Journal of Applied Psychology*, 107(12), 2243–2268. <https://doi.org/10.1037/apl0001012>
- Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz). *Amtsblatt der Europäischen Union*, L, 12.07.2024.
- Wietrak, E., Cioca, I. A., Gifford, J., & Urbano, C. (2026). PROTOCOL: Multisource feedback and work performance: A systematic review and meta-analysis. *Campbell Systematic Reviews*, 22(2). <https://doi.org/10.1177/18911803261451264>
- Winum, P. (2019). Using stakeholder input to support strategic talent development at board and senior executive levels: A practitioner's perspective. In Church et al. (2019a), pp. 123–132.
- Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives*, 26(4), 66–85. <https://doi.org/10.5465/amp.2012.0088>
- Yukl, G., Mahsud, R., Prussia, G., & Hassan, S. (2019). Effectiveness of broad and specific leadership behaviors. *Personnel Review*, 48(3), 774–783. <https://doi.org/10.1108/PR-03-2018-0100>

Hinweis: Dieser Beitrag stellt allgemeine Fachinformationen bereit und ersetzt keine individuelle arbeits- oder datenschutzrechtliche Beratung. Die Ausführungen zum Rechtsrahmen geben den Stand bei Redaktionsschluss (Oktober 2026) wieder. Die Brenner Antriebstechnik GmbH ist ein fiktives Fallbeispiel.

KI-Footer: Erstellt mit KI-Unterstützung. Geprüft und verantwortet von Dr. Walter Lieberei.