

Checkliste zur Qualität eines Fragebogens für ein 360 Grad Feedback

Mit dieser Checkliste lässt sich beurteilen, ob ein Fragebogen wissenschaftlich begründet, zuverlässig und für die Führungskräfteentwicklung praktisch nutzbar ist. Besonders wichtig ist die Kriteriumsvalidität: Sie zeigt, ob die erfassten Kompetenzen tatsächlich mit erfolgreicher Führung zusammenhängen.

Bewertungsschlüssel

☐ erfüllt	☐ teilweise erfüllt	☐ nicht erfüllt	☐ noch nicht geprüft
Nachweis oder klare Begründung liegt vor	nur teilweise belegt oder eingeschränkt	Anforderung wird nicht erreicht	Daten oder Nachweise fehlen noch

Hinweis: Die Prüffragen können jeweils direkt im Dokument markiert und durch eigene Notizen ergänzt werden.

1 Inhaltsvalidität

Prüffrage: Erfassen die Items alle wesentlichen Verhaltensweisen, die zu der jeweiligen Führungskompetenz gehören?

Begründung: Eine Kompetenz darf nicht nur ausschnittsweise gemessen werden. Zur Vorbildfunktion gehören beispielsweise Glaubwürdigkeit, Verantwortungsübernahme, konsequentes Handeln sowie die Übereinstimmung von Worten und Taten.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

2 Verständlichkeit und Beobachtbarkeit

Prüffrage: Sind die Aussagen eindeutig formuliert, und können die Beurteilenden das beschriebene Verhalten tatsächlich beobachten?

Begründung: Abstrakte Eigenschaften wie Charisma oder Dominanz werden sehr unterschiedlich verstanden. Konkrete Verhaltensbeschreibungen ermöglichen dagegen nachvollziehbare Bewertungen.

Beispiel: Statt „Diese Führungskraft besitzt Charisma“ besser: „Diese Führungskraft vermittelt eine überzeugende Vorstellung davon, wohin sich unser Bereich entwickeln soll.“

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

3 Konstruktvalidität

Prüffrage: Messen die Items tatsächlich die Kompetenz, der sie zugeordnet wurden?

Begründung: Ein Item zur Delegation sollte beispielsweise nicht hauptsächlich Freundlichkeit, Leistungsorientierung oder Kommunikation erfassen. Anderenfalls erhält die Skala zwar einen Namen, misst aber inhaltlich etwas anderes.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

4 Faktorielle Validität

Prüffrage: Bestätigen empirische Analysen die angenommene Struktur des Fragebogens?

Begründung: Die Items einer Kompetenz sollten statistisch zusammengehören und sich hinreichend von anderen Kompetenzen unterscheiden. Wenn Items mehreren Skalen gleichzeitig zugeordnet werden könnten, sind die Ergebnisse schwer interpretierbar.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

5 Reliabilität

Prüffrage: Werden die einzelnen Kompetenzen ausreichend zuverlässig gemessen?

Begründung: Eine Skala sollte bei vergleichbaren Bedingungen zu stabilen Ergebnissen führen. Kennzahlen wie Cronbachs Alpha oder McDonalds Omega zeigen, wie konsistent die Items einer Skala zusammenwirken.

Wichtig: Eine hohe Reliabilität allein beweist noch nicht, dass die richtige Kompetenz gemessen wird. Sehr ähnliche oder fast gleichlautende Items können die Reliabilität künstlich erhöhen.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

6 Trennschärfe der Items

Prüffrage: Trägt jedes einzelne Item erkennbar zur Messung seiner Kompetenz bei?

Begründung: Ein trennscharfes Item unterscheidet Personen mit hohen und niedrigen Ausprägungen der jeweiligen Kompetenz. Items mit geringer Trennschärfe sind häufig missverständlich, unpassend oder inhaltlich zu allgemein.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

7 Kriteriumsvalidität

Besondere Priorität im Gesamturteil

Prüffrage: Hängen die gemessenen Führungskompetenzen nachweislich mit tatsächlichem Führungserfolg zusammen?

Begründung: Ein Fragebogen ist nicht schon deshalb gut, weil seine Items verständlich formuliert sind, statistisch zusammenpassen und zuverlässige Skalen ergeben. Entscheidend ist, ob hohe Kompetenzwerte auch mit relevanten Erfolgsmerkmalen einhergehen.

Beispiel: Wenn eine Skala Delegation und Empowerment misst, sollte geprüft werden, ob hohe Werte tatsächlich mit größerer Selbstständigkeit, Verantwortungsübernahme oder Leistungsfähigkeit der Mitarbeitenden zusammenhängen.

Geeignete Außenkriterien sind zum Beispiel:

- Zielerreichung und Ergebnisqualität
- Mitarbeiterbindung und Fluktuation
- Engagement und Leistungsbereitschaft
- Teamleistung und Zusammenarbeit
- Kundenzufriedenheit
- Veränderungs- und Umsetzungserfolg
- ein empirisch validierter Erfolgsindex

Wichtig: Ohne Kriteriumsvalidität ist lediglich bekannt, dass etwas zuverlässig gemessen wird. Es bleibt jedoch offen, ob dieses Merkmal für erfolgreiche Führung überhaupt relevant ist. Ein Fragebogen kann somit hoch reliabel sein und trotzdem wenig praktischen Nutzen haben.

Bewertung: ☐ überzeugend nachgewiesen ☐ teilweise nachgewiesen ☐ nicht nachgewiesen ☐ offen

8 Objektivität

Prüffrage: Sind Durchführung, Auswertung und Interpretation ausreichend standardisiert?

Begründung: Das Ergebnis sollte nicht davon abhängen, wer die Befragung organisiert, die Daten auswertet oder das Feedback erläutert. Einheitliche Instruktionen, Antwortskalen, Berechnungsregeln und Ergebnisberichte erhöhen die Vergleichbarkeit.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

9 Differenzierungsfähigkeit sowie Decken und Bodeneffekte

Prüffrage: Kann der Fragebogen relevante Unterschiede zwischen Führungskräften sichtbar machen?

Begründung: Wenn fast alle Personen die besten Bewertungen erhalten, liegt ein Deckeneffekt vor. Bei fast ausschließlich schlechten Bewertungen spricht man von einem Bodeneffekt. In beiden Fällen kann der Fragebogen Stärken und Entwicklungsfelder nur unzureichend unterscheiden.

Beispiel: Aussagen wie „behandelt andere respektvoll“ erhalten häufig sehr hohe Zustimmungswerte. Sie wirken plausibel, liefern aber möglicherweise kaum zusätzliche Information.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

10 Qualität der Beurteilergruppen

Prüffrage: Beruht das Ergebnis auf ausreichend vielen Personen, die das jeweilige Verhalten tatsächlich beurteilen können?

Begründung: Einzelurteile können stark durch Sympathie, Konflikte oder einzelne Erlebnisse geprägt sein. Mehrere geeignete Beurteilende liefern ein stabileres Bild und schützen zugleich die Anonymität.

Wichtig: Unterschiede zwischen Selbstbild, Vorgesetzten, Kollegen und Mitarbeitenden sind nicht automatisch ein Qualitätsmangel. Gerade diese unterschiedlichen Perspektiven bilden den besonderen Erkenntniswert eines 360-Grad-Feedbacks.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

11 Vergleichbarkeit der Beurteilergruppen

Prüffrage: Werden die Items von Mitarbeitenden, Kollegen und Vorgesetzten in vergleichbarer Weise verstanden?

Begründung: Die Gruppen können unterschiedliche Informationen besitzen und verschieden strenge Bewertungsmaßstäbe anlegen. Deshalb sollte geprüft werden, ob Unterschiede tatsächliche Perspektiven widerspiegeln oder durch ein unterschiedliches Verständnis der Fragen entstehen.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

12 Schutz vor Urteilsverzerrungen

Prüffrage: Werden soziale Erwünschtheit, Gefälligkeitsurteile und andere Verzerrungen ausreichend berücksichtigt?

Begründung: Bewertungen können durch Freundschaft, Konflikte, Hierarchie, Angst vor Rückverfolgung oder eine allgemeine Tendenz zu guten beziehungsweise schlechten Urteilen beeinflusst werden. Anonymität, neutrale Formulierungen und geeignete Kontrollanalysen verringern dieses Risiko.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

13 Fairness und Akzeptanz

Prüffrage: Werden alle Führungskräfte anhand relevanter, beeinflussbarer und für sie nachvollziehbarer Verhaltensweisen beurteilt?

Begründung: Personen sollten nicht für Bedingungen bewertet werden, die sie kaum beeinflussen können. Ein als fair und nachvollziehbar wahrgenommenes Verfahren erhöht außerdem die Bereitschaft, das Feedback anzunehmen und für die eigene Entwicklung zu nutzen.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

14 Praktischer Entwicklungsnutzen

Prüffrage: Lassen sich aus den Ergebnissen konkrete Entwicklungsmaßnahmen ableiten?

Begründung: Ein 360-Grad-Feedback soll nicht nur Unterschiede messen, sondern Veränderungen anstoßen. Die Items müssen deshalb so konkret sein, dass die Führungskraft erkennen kann, welches Verhalten sie beibehalten, verändern oder neu erlernen sollte.

Bewertung: ☐ erfüllt ☐ teilweise erfüllt ☐ nicht erfüllt ☐ noch nicht geprüft

Gesamturteil

Besondere Stärken des Fragebogens

.....
.....
.....

Wesentliche Schwächen oder offene Prüfungen

.....
.....
.....

Vorrangiger Verbesserungsbedarf

.....
.....
.....

Nachweis der Kriteriumsvalidität

☐ überzeugend ☐ eingeschränkt ☐ fehlt ☐ noch zu prüfen

Gesamtbewertung

- ☐ wissenschaftlich und praktisch gut begründet
- ☐ mit Einschränkungen einsetzbar
- ☐ vor dem Einsatz zu überarbeiten
- ☐ für wichtige Personalentscheidungen nicht ausreichend validiert