

# Gütekriterien: Qualität von Befragungen und Interviews

Von Prof. Dr. Waldemar Pelz

## A. Gütekriterien: Warum „Qualität“ wichtig ist

Die Ergebnisse einer Befragung können in der **Praxis** nur so **nützlich** sein wie die Qualität (Objektivität, Validität und Reliabilität = Gütekriterien) des verwendeten Fragebogens. Das gilt für die meisten Beurteilungsmethoden: vom [360-Grad-Feedback](#) über die [Mitarbeiter- und Kundenbefragung](#), das Mitarbeitergespräch bis hin zum Einstellungs-Interview, einem Assessment Center oder einem Management-Audit. Der Grundsatz lautet: Es praktisch nicht möglich, aus Fragebögen mit **mangelhaften** Gütekriterien **praxisrelevante** Empfehlungen abzuleiten. In einem Forschungsprojekt an unserem Institut haben wir rund 300 Fragebögen kritisch analysiert und die häufigsten Gründe dafür gefunden, warum Befragungen scheitern oder sogar Schaden anrichten. Dazu einige Beispiele und Empfehlungen:

## B. Folgen untauglicher Befragungen

Ein häufiges Problem sind **mehrdeutige**, widersprüchliche, abstrakte oder nicht relevante Fragen (Items). Das führt oft zu Konflikten, Irritationen und Akzeptanzproblemen. Manche Führungskräfte können sehr "sensibel" reagieren, wenn das **Fremdbild** (zum Beispiel aus einem 360-Grad-Feedback) ihren Vorstellungen nicht entspricht, und sie möglicherweise ihre "**Karriere**" gefährdet sehen oder finanzielle **Nachteile** befürchten. Das gleiche gilt für Potentialträger, die sich durch das Assessment Center „falsch“ beurteilt fühlen. Oft werden geeignete Bewerber abgelehnt und **ungeeignete** Kandidaten **eingestellt**. Nahezu peinlich waren gescheiterte Mitarbeiterbefragungen bei einigen DAX-Unternehmen, die in den sozialen Medien ausführlich diskutiert wurden.

Ein Beispiel für **Mehrdeutigkeit**: Zu der Aussage "nimmt faire und sachliche Kritik an" gibt es mindestens zehn verschiedene **Interpretationsmöglichkeiten**. Im Fragebogen eines anderen DAX-Unternehmens soll das Item "arbeitet kooperativ mit Anderen zusammen" ein Merkmal der Kompetenz "Partnerschaft leben" darstellen; als Antwortmöglichkeiten sind angegeben "Entwicklungsbedarf" bis "persönliche Stärke". In vielen Fragebögen findet man Items, die nicht **beobachtbar** sind und sehr viel **Spekulation** oder sogar Hellschere erfordern. So heißt es zum **Beispiel** "berücksichtigt Ängste und Vorbehalte Anderer", "betreibt Selbstreflexion", „lebt die Unternehmensvision“ oder "fühlt sich persönlich für den Kunden verantwortlich". Wie können Feedback-Geber „wissen“, was andere Menschen „denken“ oder „fühlen“? Wie ist so etwas möglich? Eine Antwort liefert Viktor Lau, der den Begriff „**Management-Esoterik**“ geprägt hat. Siehe dazu die Grafik: [Probleme der Personalarbeit](#) in der Praxis.

## C. Empfehlungen

**Das Kernproblem**: Vielen Entwicklern von Fragebögen ist – trotz guter Absichten – nicht bewusst, dass man **Mehrdeutigkeiten**, **Widersprüche** und mangelnde **Praxis-Relevanz** im Wesentlichen nur am **Antwortverhalten der Befragten** erkennen kann. Und dazu ist eine Überprüfung anhand einer Stichprobe notwendig. Der Grund: Bei jedem Menschen können die gleichen Begriffe oder Formulierungen völlig unterschiedliche Assoziationen auslösen. Deswegen ist eine Mindestanzahl von Teilnehmern notwendig (Stichprobenumfang). Nur so ist es möglich, statistische Testverfahren anzuwenden, mit denen man derartige „Fehler“ entdecken kann.

Die Gütekriterien geben Auskunft über diese „Fehlerfreiheit“. Leider existieren keine Patenzrezepte für die Formulierung von Items - aber einige Tipps und Empfehlungen:

■ **Achten Sie auf Trennschärfe und die Anzahl von Items!**

Viele Aussagen (Items) zur Beschreibung (Operationalisierung) einer Fähigkeit haben eine zu geringe **Trennschärfe** und passen nicht zu der Kompetenz: sie "messen" etwas völlig anderes, die antwortende Person hat die Frage anders verstanden als beabsichtigt, oder die Items sind bei gleichem Inhalt nur unterschiedlich formuliert. Auch die **Anzahl der Items** (Verhaltensbeschreibungen), die eine Kompetenz repräsentieren, ist oft zu gering um praxisrelevante Ergebnisse zu liefern. So ist es praktisch unmöglich, zum Beispiel „Konfliktfähigkeit“ mithilfe von drei bis vier Items einzuschätzen (weil es mehr als vier verschiedene Konfliktarten und dazu passende Lösungsmöglichkeiten gibt).

■ **Verhindern Sie Esoterik und Trivialität!**

Viele Items klingen plausibel, sind aber zu abstrakt und „messen“ **triviale** Selbstverständlichkeiten wie es bei der Sterndeutung und Wahrsagerei der Fall ist. Dazu gehören Fragen wie zum Beispiel „mein Vorgesetzter lobt mich oft“, „mein Vorgesetzter kann gut zuhören, ...zeigt Verständnis, ...setzt sich für Kunden ein“, „sieht Chancen, wo Andere nur Probleme sehen“, „identifiziert sich mit dem Unternehmen“, „unterstützt unkonventionelle Ideen“, „lebt die Unternehmensvision“, „ermächtigt seine Mitarbeiter“ etc. Hier haben wir es eher mit Esoterik als mit Kompetenzdiagnose zu tun (siehe dazu den [Barnum-Effekt](#)).

■ **Vermeiden Sie Gemeinplätze!**

In vielen Fragebögen wird „**Teamfähigkeit**“ erhoben. Damit kann gemeint sein, dass jemand im Team „konstruktiv mitarbeitet“, „mit Anderen gut auskommt“, „sich in eine Gruppe gut einfügen kann“, „für gute Laune sorgt“, „kollegial“, „konstruktiv“, „sensibel“, „umsichtig“, „freundlich“, „kompromiss- und kritikfähig“ ist. Derartige Gemeinplätze haben, abgesehen davon, dass es reine **Worthülsen** sind, nicht viel mit dem Erfolg eines Teams zu tun. Damit ein Team erfolgreich arbeitet, muss jedes Teammitglied seine spezifische Aufgabe (Team-Rolle) erfüllen. Die oben genannten Items sind für den Teamerfolg völlig irrelevant und – bestenfalls – Folge von Teamfähigkeit.

■ **Was Sie bei vorhandener Validität sonst noch beachten sollten!**

Auch bei validierten Skalen ist Vorsicht geboten. Beispielsweise misst das „Bochumer Inventar zur Berufsbezogenen Persönlichkeitsbeschreibung“ die **Teamorientierung** und meint damit, wie stark jemand Teamarbeit bevorzugt (operationalisiert mit 13 Items). Tatsächlich kommt es auf die „Bevorzugung von Teamarbeit“ gar nicht an. Vielmehr ist es wichtig, wie gut jemand im Team seine individuelle, für den Teamerfolg wichtige Rolle wahrnimmt. Einen Torwart misst man nicht daran, wie viele Tore er schießt.

Zur Einschätzung der „optimalen“ Teamarbeit gibt es verschiedene (validierte) Konzepte. Ein Beispiel ist das von Meredith Belbin. Eine wichtige Teamrolle dabei ist die des **Bewerter** (Kritikers). Diese Personen machen sich im Team häufig sehr unbeliebt, sind aber für den Teamerfolg unerlässlich. In vielen Assessment-Centern wird diesen Personen dennoch mangelnde Teamfähigkeit bescheinigt. Siehe hierzu den Beitrag:

<https://www.managementkompetenzen.de/teamfaehigkeit.html>

■ **Orientieren Sie sich am aktuellen Stand des Wissens!**

In vielen Fällen stammen die Items aus Modellen und Theorien, die zwar weit verbreitet

und beliebt sind, aber keinen Praxisbezug haben. Beispiele sind die „Theorie des Situativen Führens“, das „[DISG-Modell](#)“, das „Reiss-Profil“ oder das „[Systemische Coaching](#)“. Diese oft plausibel klingenden Aussagen haben tatsächlich die Qualität von Horoskopen oder Binsenweisheiten. Siehe dazu als Beispiel das Diskussionspapier „[Kritik des Reiss-Profils](#)“ und den Beitrag von *Thompson und Vecchio*: Situational leadership theory: A test of three versions, in: *The Leadership Quarterly* 20/2009).

Zu Persönlichkeitsmerkmalen gibt es zahlreiche validierte Konzepte wie zum Beispiel das „Big-Five-Modell“ (Fünf-Faktoren-Modell) mit den Dimensionen Gewissenhaftigkeit, Offenheit, Verträglichkeit, Neurotizismus und Extraversion. Die Verwendung derartiger Merkmale führt zu anderen Problemen: Zum einen sind Persönlichkeitsmerkmale äußerst stabil und mit „vernünftigem“ Aufwand praktisch nicht veränderbar oder *erlernbar* (bzw. „*verlernbar*“). Zum anderen ist es ethisch problematisch, die Persönlichkeit eines Menschen verändern (manipulieren) zu wollen. Schließlich besteht kaum ein Zusammenhang zwischen Persönlichkeitsmerkmalen und aufgabenspezifischen Fähigkeiten, die notwendig sind, bestimmte Ergebnisse (Resultate) zu erzielen. Diese aufgabenbezogenen Fähigkeiten und Erfahrungen nennt man Kompetenzen.

Zur Unterscheidung zwischen Charaktereigenschaften, Wertvorstellungen und Kompetenzen empfehlen wir die Seite [360°-Beurteilung](#).

Die gleiche Problematik wie im Falle der Persönlichkeitsmerkmale gilt auch für Motive und Führungsstile. Beispielsweise hat eine Studie von Albert Bandura von der Stanford University gezeigt, dass praktisch kein Zusammenhang zwischen der (im Test gemessenen) Leistungsmotivation und der tatsächlichen Leistung besteht. Bislang ist es auch nicht gelungen, einen Erfolg versprechenden Führungsstil zu finden. Abgesehen davon, ist es kaum möglich, den „Stil“ eines anderen Menschen nachzuahmen. Im Gegenteil: Individualität und Authentizität haben sich in unserer Studie über die Erfolgsfaktoren mittelständischer Weltmarktführer ([Hidden Champions](#)) als wesentlich wichtiger erwiesen als irgendwelche (theoretischen) Stile.

■ **Legen Sie zuerst die Ziele der Befragung oder Beurteilung fest!**

Die wichtigste Aufgabe ist es, eine Schnittmenge zwischen dem Persönlichkeits- und Kompetenzprofil des Kandidaten und der Unternehmenskultur herzustellen. Beides muss zueinander passen und auf die Zukunft ausgerichtet sein. Ferner sollte man klären, ob eine bestimmte Fähigkeit oder Charaktereigenschaft eine Stärke oder Schwäche darstellt. Dies ist in erster Linie von den Zielen Abhängig. Erfolge und überzeugende Ergebnisse wird man in der Regel nur mit Stärken erreichen. In Standardwerken findet man bis zu 80 Management- und Führungskompetenzen. An allen Kompetenzen zu arbeiten ist weder möglich noch sinnvoll. Aus der Perspektive des Kandidaten und des Unternehmens ist somit die Setzung von „richtigen“ Prioritäten notwendig. Nur so findet man beiderseitige Vorteile, die langfristig tragfähig sind.

Beispiele für validierte Fragebögen unter [www.managementkompetenzen.de](http://www.managementkompetenzen.de)

*Waldemar Pelz, Bad Soden am Taunus, im August 2019*