

# Qualität von Befragungen und Interviews

Von Prof. Dr. Waldemar Pelz

## A. Warum „Qualität“ wichtig ist

Die Ergebnisse einer Befragung können nur so gut sein wie die Qualität (Objektivität, Validität und Reliabilität) des verwendeten Fragebogens. Das gilt für die meisten Erhebungen: vom 360-Grad-Feedback über die Mitarbeiter- und Kundenbefragung, das Mitarbeitergespräch bis hin zum Biographischen Interview im Assessment Center. Es gilt der Grundsatz, wonach es praktisch nicht möglich ist, aus Fragebögen, die den genannten Standards nicht entsprechen, praxisrelevante Empfehlungen abzuleiten. In unserem Forschungsprojekt haben wir rund 300 Fragebögen kritisch analysiert und die häufigsten Gründe dafür gefunden, warum viele Befragungen scheitern. Aus diesen Erkenntnissen lassen sich zahlreiche Regeln und Empfehlungen ableiten. Dazu einige Beispiele:

## B. Folgen untauglicher Befragungen

Eine häufige Folge Fragebögen sind mehrdeutige, widersprüchliche oder nicht relevante Ergebnisse. Das führt oft zu Streitigkeiten, 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 unfair beurteilt fühlen. Nahezu peinlich waren viele gescheiterte Mitarbeiterbefragungen bei großen Unternehmen, die erheblichen Schaden angerichtet haben. Darüber hat die Wirtschaftspresse ausführlich berichtet.

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 messbar sind und sehr viel Spekulation oder sogar Hellscherei erfordern. So heißt es zum Beispiel "berücksichtigt Ängste und Vorbehalte Anderer", "betreibt Selbstreflektion", „lebt die Unternehmensvision“ oder "fühlt sich persönlich für den Kunden verantwortlich". Wie können Feedback-Geber „wissen“, was andere Menschen „fühlen“?

## C. Empfehlungen

### ■ Das Kernproblem

Hinter den skizzierten Beispielen für typische „Fehler“ verbergen sich gut gemeinte Absichten und durchaus sinnvolle Themen. Vielen Entwicklern von Fragebögen ist aber nicht bewusst, dass man Mehrdeutigkeiten, Widersprüche und mangelnde Relevanz im Wesentlichen nur am *Antwortverhalten der Befragten* erkennen kann. Bei jedem Menschen können die gleichen Begriffe oder Formulierungen völlig unterschiedliche Assoziationen auslösen. Deswegen ist je nach Ziel der Befragung und Thema eine Mindestanzahl von Teilnehmern notwendig (Stichprobenumfang). Nur so ist es möglich, statistische Testverfahren anzuwenden, mit denen man derartige „Fehler“ entdecken kann. Leider gibt es hierzu keine Patenzrezepte, aber einige Empfehlungen:

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

Viele Aussagen (Items) zur Beschreibung einer Fähigkeit haben eine zu geringe Trennschärfe und passen somit nicht zu der Kompetenz, die sie definieren sollen; sie "messen" also etwas völlig anderes, oder die antwortende Person hat die Frage anders verstanden als beabsichtigt. Auch die Anzahl der Items (Verhaltensbeschreibungen), die eine Kompetenz repräsentieren, ist oft zu gering um aussagekräftige 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 gibt und jede andere Fähigkeiten erfordert).

■ **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.

■ **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. Ohne diese Voraussetzung können Teamgeist und Effektivität in der Regel gar nicht entstehen. Die oben genannten Items sind für den Teamerfolg völlig irrelevant, wie es das nachfolgende Beispiel zeigt.

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

Selbst 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 ja auch nicht daran, wie viele Tore er schießt oder einen Stürmer daran, wie viele Torschüsse er gehalten hat. Zur Einschätzung der „optimalen“ Teamarbeit gibt es verschiedene (validierte) Konzepte. Ein Beispiel ist das von Meredith Belbin. Eine wichtige Teamrolle dabei ist die des Bewerbers (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.

■ **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“ oder das „Reiss-Profil“. Diese oft plausibel klingenden Aussagen haben tatsächlich die Qualität von Horoskopen oder Binsenweisheiten.

Siehe dazu als Beispiel unser Diskussionspapier „Kritik des Reiss-Profiles“ unter <http://www.management-innovation.com/download/Reiss-Profil.pdf> 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 unter anderem zu folgenden Problemen: Zum einen sind Persönlichkeitsmerkmale äußerst stabil und mit „vernünftigem“ Aufwand praktisch nicht veränderbar. Zum anderen ist es ethisch problematisch, die Persönlichkeit eines Menschen verändern zu wollen. Schließlich besteht nur ein schwacher Zusammenhang zwischen Persönlichkeitsmerkmalen und aufgabenspezifischen Fähigkeiten, die notwendig sind, bestimmte Ergebnisse (Resultate) zu erzielen. Diese aufgabenbezogenen Fähigkeiten und Erfahrungen nennt man Kompetenzen. Diese sollten grundsätzlich jeder Befragung zugrunde liegen (mit Ausnahmen – siehe weiter unten).

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. Dennoch kann es notwendig sein, **Werte, Persönlichkeitsmerkmale** oder Führungsstile (**Führungskultur**) einzuschätzen. In solchen Fällen ist die prognostische Validität der zugrunde liegenden Items entscheidend für den praktischen Nutzen des Fragebogens.

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

Fazit: Auch bei validierten Fragebögen sollte man darauf achten, dass sie keine Persönlichkeitsmerkmale, Motive oder Führungsstile messen, sondern beobachtbare, aufgabenspezifische Verhaltensweisen und Erfahrungen (Kompetenzen), die die zu beurteilende Person (der Feedback-Nehmer) relativ einfach und schnell erlernen oder verändern kann (mit Ausnahme von Items mit prognostischer Validität). Beachten Sie ferner das Problem der „richtigen“ Auswahl von Kompetenzen. Ob eine bestimmte Fähigkeit eine Stärke oder Schwäche darstellt, ist davon abhängig, welche Ziele der Feedback-Nehmer verfolgt. Es sind berufliche (Karriere), geschäftliche (Kennzahlen) oder sonstige persönliche (familiäre, fachliche, physische und ethische) Ziele. Je nach Ziel, können völlig unterschiedliche Kompetenzen relevant sein. 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. Der Mitarbeiter muss also Prioritäten setzen und diese mit den Unternehmenszielen abstimmen. Nur so findet man beiderseitige Vorteile, die langfristig tragfähig sind.

*Waldemar Pelz, Bad Soden am Taunus, den 24. August 2015*