

Validierung am Beispiel der Vorbildfunktion

(Auszug aus: Persönlichkeit gewinnt – Fokussiertes Recruiting mit dem berufsbezogenen Persönlichkeitstest, Schäffer-Pöschel-Verlag 2023)

Erklärung der Gütekriterien (Reliabilität, Validität und Normierung) am Beispiel der Vorbildfunktion von Führungskräften:

Der Fachbegriff für die Kriterien zur Beurteilung der Qualität von Testverfahren heißt Gütekriterien. Bei der Beurteilung der Vorbildfunktion geht es um drei zentrale Fragen:

1. Was ist ein Vorbild?
2. Vorbildfunktion: Nur wünschenswerte, weil imagefördernde Kompetenz der Führungskraft oder für den Geschäftserfolg notwendig?
3. Frage der Normierung: Wie kann man die Testergebnisse verschiedener Personen vergleichen?

1. Was ist ein Vorbild?

Wenn man zehn Menschen fragt, bekommt man vermutlich mehr als zehn verschiedene Antworten. Wenn jeder unter einem Vorbild etwas anderes versteht, ist es nicht möglich festzustellen, ob oder inwiefern eine Person die Vorbildfunktion in der Praxis tatsächlich wahrnimmt. Das führt bei der Beurteilung von Führungskräften zum Beispiel in einem 360-Grad-Feedback in der Regel zu gravierenden Irritationen und Konflikten.

Zur Lösung dieses Problems ist es notwendig, den Begriff „Vorbild“ mit konkreten Verhaltensbeschreibungen (Fachbegriff: Items) eindeutig und für andere nachvollziehbar zu definieren. Diese Items können aus einer (validierten) Theorie kommen und sind folglich eine Operationalisierung dieser Theorie. Ein solches Vorgehen der Itementwicklung bezeichnet man als *deduktiv*. Man kann die Verhaltensbeschreibungen aber auch *induktiv* aus einer Expertenbefragung gewinnen. Experten sind in diesem Fall keine „klugen“ Theoretiker oder Akademiker, sondern Betroffene wie zum Beispiel erfahrene Führungskräfte. Die zentrale Frage lautet daher, inwiefern die Verhaltensbeschreibungen für erfolgreiche Führungskräfte repräsentativ (typisch) sind. Wenn das der Fall ist, dann hat das Konstrukt „Vorbild“ eine hohe repräsentative oder inhaltliche Validität (auch Inhaltsvalidität genannt).

Eine induktiv oder deduktiv gewonnene Definition ist noch nicht ausreichend. Zusätzlich muss man anhand einer (möglichst repräsentativen) Stichprobe prüfen, inwiefern andere Menschen dieser

Definition, die durch Verhaltensbeschreibungen operationalisiert wurde, zustimmen. *Beispiele für getestete Items* (sechs von acht) sind:

- „Er meint, was er sagt.“
- „Steht für klare Wertvorstellungen.“
- „Setzt klare Ziele und Erwartungen.“
- „Verfügt über profunde Fachkenntnisse.“
- „Ist offen für Kritik und neue Ideen.“
- „Verhält sich in einer Weise, die Respekt verdient.“

In unserem Beispiel „Vorbild“ wurden 23.698 Fach- und Führungskräfte befragt. Sie bestätigten, dass alle Verhaltensbeschreibungen typisch für eine Führungskraft sind, die als Vorbild wahrgenommen wird. Dazu gibt es eine Kennzahl. Sie heißt Cronbach- α -Koeffizient oder kurz Cronbachs Alpha (s. Anhang E, Abschnitt „Gütekriterium 2: Reliabilität“) und hat in diesem Fall einen Wert von 0,96. Das bedeutet, dass die Items die Vorbildfunktion fast perfekt (1,0) beschreiben (erfassen). Mit anderen Worten: Die Vorbildfunktion ist intern *konsistent*: Alle Items passen zueinander. (Genau genommen, passt die Vorbildfunktion auch zu den anderen transformationalen Führungskompetenzen [Pelz, 2016].)

Bei den Items könnte es sich (trotz der hohen internen Konsistenz) um verschiedene Formulierungen des gleichen Inhalts handeln. Das wirft weitere Fragen auf: Überschneiden sich die Items? Sind einige überflüssig? Auch dafür gibt es eine Kennzahl: die *Trennschärfe*. In unserem Beispiel liegt die durchschnittliche Trennschärfe bei 0,63 (Korrelationskoeffizient¹ der Items). Das bedeutet, dass alle sechs Items der Vorbildfunktion sich sinnvoll ergänzen. (Eine maximale Trennschärfe von 1,0 würde auf mögliche Fehler hindeuten.)

Wenn man allerdings das Item „Fachkenntnisse“ weglässt, verbessert sich die Trennschärfe deutlich von 0,63 auf 0,86. Das bedeutet, dass Fachkenntnisse nicht unbedingt zur Vorbildfunktion gehören. Ob sie dazugehören, ist abhängig von der Branche, Hierarchieebene und Funktion der getesteten Person. Beispielsweise haben Fachkenntnisse für einen Chefarzt in der Neurologie eine völlig andere Bedeutung als für den Chef eines Betriebs, der Getränke abfüllt. Deshalb sollte man auch die Normierung (Frage 3, siehe weiter unten) beachten. Eine rein formale Betrachtung kann irreführend sein.

Fazit

Cronbachs Alpha (als Maß der internen Konsistenz) und die Trennschärfe sind die wichtigsten Kennzahlen für die *Reliabilität*, also die Zuverlässigkeit des Tests. Die Testreliabilität kann man vergleichen mit der Eichung einer Waage.

1. Vorbildfunktion: Nur wünschenswerte, imagefördernde Kompetenz der Führungskraft oder für den Geschäftserfolg notwendig?

Diese zweite zentrale Frage gibt Auskunft über die *Validität*. Diese Kennzahl kann man als Maß des Praxisbezugs interpretieren. Anders formuliert lautet die Frage:

Ist die Vorbildfunktion „nur“ eine „wünschenswerte“, moralisch hoch angesehene Kompetenz und Modeerscheinung, die dem Image der Führungskraft dient („nice to have“), oder ist sie für den Geschäftserfolg notwendig? Misst das Konstrukt „Vorbild“ vielleicht „nur“ die Beliebtheit einer Führungskraft?

Die Antwort: Die Vorbildfunktion ist für den Geschäftserfolg notwendig, weil sie typisch ist für herausragende Unternehmerpersönlichkeiten. Dies wurde durch andere (unabhängige) Studien und die eigene Untersuchung des Verhaltens von Geschäftsführern mittelständischer Weltmarktführer (Hidden Champions) ermittelt und in der Befragung von 23.598 Fach- und Führungskräften bestätigt. Die Kennzahl für den Zusammenhang von Erfolg und Vorbild heißt *Kriteriumsvalidität oder kriteriumsbezogene Validität*. Sie misst die Korrelation zwischen unternehmerischem Erfolg und Vorbildfunktion (Korrelationskoeffizient = 0,93). Den unternehmerischen Erfolg könnte man auch anders messen, zum Beispiel mit dem Einkommen, dem Marktanteil, der Rendite oder der Kunden- und Mitarbeiterzufriedenheit (Frühindikatoren der Rendite). Diese Zusammenhänge könnte man auch so interpretieren, dass Führungskräfte, die die Vorbildfunktion erfüllen, ein höheres persönliches Einkommen oder eine höhere Produktivität und Rentabilität im Unternehmen erzielen. Derartige statistische Zusammenhänge werden in der Regel in weiteren Forschungsprojekten anhand von Hypothesen (Kausalanalyse) empirisch überprüft. Ein statistischer Zusammenhang ist nicht automatisch ein kausaler Zusammenhang.

Es gibt noch andere Arten der Validität; die Kriteriumsvalidität bietet aber den größten Nutzen in der Praxis, weil sie Antworten auf alltägliche Fragen des Managements liefert. Beispiele für solche Fragen sind:

- Besteht ein Zusammenhang zwischen Leistungsmotivation (gemessen mit einem Test) und tatsächlicher Leistung?
- Korreliert ein positives Arbeitsklima mit Produktivität?
- Haben empathische Verkäufer bessere Kundenbeziehungen?
- Erleben resiliente Manager weniger Stress?
- Besteht ein Zusammenhang zwischen transformationaler Führung und Rentabilität?

2. Frage der Normierung: Wie kann man die Testergebnisse verschiedener Personen vergleichen?

Wenn die Vorbildfunktion einer Führungskraft mit 3,7 (auf einer Skala von 1 bis 5) bewertet wurde, kann das sehr gut oder sehr schlecht sein. Man benötigt Vergleichswerte, Testergebnisse anderer Personen, die typisch (repräsentativ) für diese Personengruppe (Führungskräfte) sind. Wenn die zu beurteilende Person weiblich ist, über maximal 5 Jahre Führungserfahrung verfügt, eine kaufmännische Ausbildung aufweist und in einem Großunternehmen der Finanzwirtschaft arbeitet, dann müssen ihre Testergebnisse mit Personen verglichen werden, die die gleichen demografischen Merkmale haben. Ein Vergleich der Führungskompetenz dieser Dame mit einem selbstständigen Unternehmer im Technologiebereich würde keinen Sinn machen.

Im Idealfall stammen die Vergleichswerte von Wettbewerbern. Wenn zum Beispiel die Führungskompetenz der Potenzialträger eines Unternehmens auf einer Skala von 1 bis 5 bei 3,1 liegt und die Nachwuchskräfte vergleichbarer Wettbewerber den Wert 4,6 erzielen, deutet das auf dringenden Handlungsbedarf hin.

Fazit

Die Ergebnisse eines Testverfahrens sind nur mit Daten einer Vergleichsgruppe sinnvoll zu interpretieren. Umstritten ist dabei, ob eine Normierung nach der klassischen Testtheorie für die Praxis sinnvoll und hilfreich ist. Im Kontext des berufsbezogenen Persönlichkeitstests (BPT) verneinen wir diese Frage eindeutig.

ⁱ Der Korrelationskoeffizient ist ein Maß für die Häufigkeit des gemeinsamen Auftretens von mindestens zwei Merkmalen wie zum Beispiel Intelligenz und Schulnoten (intelligente Schüler haben gute Noten) oder Selbstvertrauen und Lebenszufriedenheit (Menschen mit starkem Selbstvertrauen sind zufrieden mit ihrem Leben). Der Korrelationskoeffizient kann die Werte zwischen + 1 (starker positiver) und – 1 (starker negativer Zusammenhang) annehmen. In unserem Beispiel wäre der Zusammenhang stark negativ, wenn sehr intelligente Schüler zugleich sehr schlechte Noten, und Menschen mit sehr schwachem Selbstvertrauen äußerst zufrieden wären. Ist kein Zusammenhang (in einer untersuchten Stichprobe) erkennbar, ist der Korrelationskoeffizient gleich null. Inwiefern ein Zusammenhang auf Ursachen und Wirkungen schließen lässt, muss in zusätzlichen Analysen geprüft werden.