

Werner Sarges und Heinrich Wottawa

TESTVERFAHREN FÜR DAS PERSONAL- MANAGEMENT ¹

In der Personalarbeit werden zunehmend leistungsfähige und kosteneffiziente Verfahren zur Einschätzung neuer oder schon vorhandener Mitarbeiter benötigt. Nur auf dieser Grundlage kann die Personalauswahl, die Personalplatzierung (Wer passt am besten auf welche Position?) und die Personalentwicklung (Bei wem ist was entwickelbar? Wie können knappe Personalentwicklungsressourcen am sinnvollsten genutzt werden?) optimal für das Unternehmen gestaltet werden. Treffsichere Einschätzung von Personen ist aber nicht nur für das Unternehmen, sondern auch für die Mitarbeiter und Bewerber wichtig. Das Scheitern von beruflichen Ambitionen, sei es durch den Erhalt einer überfordernden beruflichen Position oder die Teilnahme an einer zu anspruchsvollen Weiterbildungsmaßnahme, ist für Betroffene nicht nur belastend, sondern kann auch deren weiteren Lebenslauf stark negativ beeinflussen.

Es kann kein Zweifel daran bestehen, dass hinsichtlich des Verhältnisses von Aufwand und Ertrag gute Testverfahren im Vergleich zu anderen Techniken der Eignungseinschätzung hervorragend abschneiden. Immerhin sind solche Tests, die von Fachleuten nach psychometrischen Kriterien sorgfältig konstruiert wurden, das messtechnisch Anspruchsvollste, was die psychologische Eignungsdiagnostik zu bieten hat. Die Methodologie der Testkonstruktion ist weit entwickelt, und man hat in einer langen Tradition eine Fülle wertvoller Erfahrungen sammeln können. Schließlich haben Theorie und Praxis der Testkonstruktion sogar die methodischen Standards für die anderen diagnostischen Verfahren – wie Interview, Tätigkeitssimulationen oder Assessment-Center – geliefert.

In Abgrenzung zu den anderen Verfahren der Eignungsdiagnostik verstehen wir unter Tests nur Instrumente, mit Hilfe derer wir psychometrisch vergleichbare und gültige Informationen über Verhalten und Erleben einzelner Personen erhalten können – was häufig bei Zeugnisnoten oder Interviews so nicht der Fall ist. Gut konstruierte Tests müssen den Gütekriterien Standardisierung, Reliabilität und Validierung genügen. In der Regel werden psychometrische Tests von an Wissenschaftsinstitutionen arbeitenden Psychologen entwickelt und von Testverlagen oder Beratungsfirmen vertrieben. Doch entwickeln und validieren gelegentlich auch Psychologen in größeren Unternehmen oder Organisationen eigene Tests, um ihren ganz spezifischen Zielgruppen und Zwecken zu genügen, beispielsweise Tests für Anlageberater, Offiziere oder Piloten.

¹ erschienen in: Sarges, W. & Wottawa, H. (Hrsg.). (2004). *Handbuch wirtschaftspsychologischer Testverfahren – Band I: Personalpsychologische Instrumente* (2., überarb. u. erw. Aufl.; S. VII–XII). Lengerich: Pabst.

Nur am Rande sei erwähnt, dass die Validität vieler Tests im Hinblick auf Berufserfolg (d.h. als substantieller Vorhersagefaktor dafür) in den verschiedensten beruflichen Bereichen auch durch Forschungszusammenfassungen der jüngsten Zeit eindrucksvoll belegt wird. Dabei haben sich vor allem zwei testmäßige Prädiktoren (Vorhersagefaktoren) als durchgängig bedeutsam in allen möglichen Berufsfeldern etabliert: Intelligenz (aus dem Leistungsbereich) und Integrität (aus dem Persönlichkeitsbereich).

Trotz dieser – in Fachkreisen unumstrittenen – Vorteile von Testverfahren ist der Einsatz solcher Verfahren in der deutschen Wirtschaft, insbesondere auch im Vergleich zum westlichen Ausland, noch immer etwas zurückhaltend. Falls Tests bisher herangezogen wurden, dann vor allem im Rahmen der Auswahl von Auszubildenden oder für Berufe mit mittlerem Anforderungsniveau, weniger auch für qualifizierte Experten oder Führungsnachwuchskräfte bzw. Führungskräfte selbst. In etlichen anderen europäischen Ländern und in den USA ist man diesbezüglich deutlich aufgeschlossener, bis hin zur Besetzung oberster Management-Positionen: Carly Fiorina, Chefin von Hewlett-Packard, musste sich vor ihrer Nominierung einer zweistündigen Prüfung mit 900 Ja/Nein-Fragen unterziehen. In dieser Hinsicht ist Deutschland also noch Entwicklungsland. Organisationen und Bewerber sind allerdings oft gar nicht einmal so schwer davon zu überzeugen, dass auch Tests zum selbstverständlichen Bestandteil guter Eignungsdiagnostik gehören.

Im Übrigen werden alle Verfahren der Eignungsdiagnostik, und damit auch und insbesondere Tests, umso wichtiger, je mehr der Lebenslauf und die Zeugnisse generell oder in spezifischen Kandidatenstichproben speziell an Aussagekraft verlieren – was seit geraumer Zeit der Fall ist.

Testverfahren können nun auf viele verschiedene Arten klassifiziert werden. Inzwischen hat sich zumindest die Grobklassifikation in *Leistungstests* und *Persönlichkeitstests* durchgesetzt. Zwar gehört im alltäglichen Sprachgebrauch auch der Bereich der Leistung zur Persönlichkeit, die von der Psychologie getroffene Unterscheidung von Leistungs- vs. Persönlichkeitstests wird aber dadurch sinnvoll, dass jeweils unterschiedliche Reaktionsweisen gefordert sind: Leistungstests sollen maximales Verhalten (z.B. im Merkmalsbereich der Intelligenz), Persönlichkeitstests typisches (übliches) Verhalten (z.B. im Merkmalsbereich der Extraversion) erfassen.

Bei der noch immer zu geringen Anwendungshäufigkeit von Tests im Personalmanagement in Deutschland werden außer reinen Leistungstests inzwischen vermehrt auch Persönlichkeitstests eingesetzt. Dadurch möchte man nichtleistungsmäßige Eigenschaften im fachlichen und außerfachlichen Bereich, die aber die Arbeitsleistung im Einzelfall sehr stark mitbedingen können, ebenfalls erfassen (z.B. Emotionale Stabilität, Kooperationsbereitschaft oder Leistungsmotivation).

Dies ist besonders zu begrüßen, sind doch Persönlichkeitsmerkmale für die Prognose der beruflichen Leistung kaum weniger wichtig als Leistungsmerkmale. Dies haben jüngst erst wieder Forschungsbefunde gezeigt, die einen beachtlichen Zusammenhang gefunden haben zwischen Berufserfolg und den sog. "big five", den fünf großen Persönlichkeitsfaktoren, nach denen Menschen andere Menschen auf einer generellen Ebene einschätzen, nämlich Extraversion, Emotionale Stabilität, Offenheit für neue Erfahrungen, Gewissenhaftigkeit und Verträglichkeit. Manche Facetten beruflicher Leistung, wie Führung, Engagement oder Disziplin, lassen sich durch Persönlichkeitsmerkmale sogar besser prognostizieren als durch kognitive oder andere Leistungsmerkmale. Experten schätzen sogar, dass über 90 % der Beschäftigungsverhältnisse, die scheitern, nicht aufgrund von fachlichen oder sonstigen Könnens-

Kompetenzen des betreffenden Mitarbeiters beendet werden (die man offenbar leichter feststellen kann), sondern wegen Unstimmigkeiten von Merkmalen seiner Persönlichkeit mit den Anforderungsbedingungen der Position (was man bisher wohl nur unzureichend in Erfahrung bringen konnte). Zum Beispiel: Eine Führungskraft, die das Führen anderer Personen nicht als reizvoll und motivierend erlebt, wird unter Umständen über lange Jahre hinweg genau diejenigen Situationen meiden, in denen aktiv geführt werden müsste. Welche Konsequenzen hieraus für ein Unternehmen erwachsen können, lässt sich leicht ausmalen. Die tatsächlichen Effekte solcher und vieler anderer zeitlich relativ stabiler Verhaltensdispositionen, die im Bereich der Persönlichkeit eines Mitarbeiters angesiedelt sind, werden noch vielfach unterschätzt.

Trotz der Leistungs*bezüglichkeit* vieler Persönlichkeitsmerkmale lassen sich diese aber leider kaum leistungs*mäßig* (also über maximale Leistung in irgendeinem Bereich) erfassen, sondern praktisch nur über Angaben des Probanden mit Bezug auf sein normalerweise übliches (d.h. typisches) Verhalten in dieser oder jener Situation. Von daher wird häufig beklagt, dass Testanden Persönlichkeitstests (im Gegensatz zu Leistungstests) leicht in eine als positiv angenommene Richtung verfälschen könnten, so dass die erhaltenen Informationen eigentlich invalide und damit unbrauchbar seien. Dem aber ist dreierlei entgegen zu halten:

- Erstens kann eine beschönigende Selbstdarstellung, wenn man sie unkorrigiert zulässt, für Kontaktberufe sogar validitätserhöhend sein.
- Zweitens ist die Verfälschungswahrscheinlichkeit dann geringer, wenn Testanden mit einer grundsätzlichen Überprüfbarkeit ihrer Angaben in nachfolgenden Interviews, Assessment-Centern o.ä. rechnen müssen.
- Drittens lassen sich, wenn man Normwerte aus echten Bewerberstichproben gewinnen kann, die Trennwerte für Ablehnungsentscheidungen bei Negativauswahl entsprechend adjustieren.

Weiterhin ist zu bedenken, dass die Forschung zur Verfälschbarkeit von Tests nach wie vor gut im Gange ist und weitere praxisdienliche Erkenntnisse liefert: Z.B. kann man bei durch Verfälschung gefährdeten Items (Frageninhalte und -formate) durch besondere Antworttechniken (z.B. durch Zwangswahl gleich attraktiv erscheinender Antwortalternativen – "forced choice" – als Antwortformat) derartigen Tendenzen entgegenwirken.

Aber nicht nur wegen der Unterschätzung der Bedeutung von mit Tests besonders gut messbaren Merkmalen für den Berufserfolg werden solche Verfahren noch zu selten herangezogen, sondern auch wegen einer mit Besorgtheit für betroffene Mitarbeiter bemäntelten ideologischen Ablehnung. Die wesentlichen Argumente dieser Meinungsmacher gegen Tests lauten:

- unberechtigtes Eindringen in die Privatsphäre des Kandidaten oder Mitarbeiters,
- Möglichkeit von Missbrauch bzw. Weitergabe erhobener Daten,
- erhaltene Persönlichkeitsbilder bleiben unvollständig,
- häufig wenig augenfälliger Bezug zum Arbeitsinhalt,
- Nichtberücksichtigung von situativen Besonderheiten der späteren Arbeit,
- Vortäuschen von Sicherheit, obwohl man im Einzelfall auch falsch liegen kann.

Doch sind derartige Argumente gegen Tests gleichermaßen zutreffend für die meisten anderen Verfahren der praktizierten Eignungsdiagnostik, man denke nur an die oft wenig profes-

sionelle Handhabung von Interviews. Allerdings ist ein weiteres Argument in besonderer Weise testspezifisch, nämlich das gegen Tests als quantitative Methode: In seiner allgemeinsten Form lautet es, der Mensch sei nicht messbar. Dies ist in der Tat zutreffend. Indes: Auch ein Stein ist nicht messbar, aber man kann Eigenschaften an ihm messen, z.B. sein Gewicht, seine Größe, seine Farbe. Entsprechend kann man berufsrelevante Eigenschaften auch bei einem Menschen messen, wenn die betreffenden Merkmale nur die notwendige Bedingung der Stabilität über die Zeit erfüllen, z.B. seine Integrität, seine Dominanz, seine Toleranz für Unbestimmtheit.

Die Grenzen, die jede dieser Messungen hinsichtlich der Genauigkeit und Wiederholbarkeit hat, werden aufgrund der anspruchsvollen teststatistischen Auswertungen bei Testverfahren deutlicher als z.B. bei der Durchsicht von Bewerbungsunterlagen, bei Interviews oder bei Verhaltensbeobachtungen (in Assessment-Centern etwa). Dies erweckt dann allzu leicht den Eindruck, als hätten nur oder vor allem Testverfahren in diesem Bereich Mängel, obwohl gerade sie den genannten anderen Verfahren hinsichtlich Objektivität, Messgenauigkeit und Validität in den meisten Einsatzbereichen überlegen sind. Die sorgfältige wissenschaftliche Analyse, die bei Testverfahren selbstverständlich ist (oder, wenn man an manche nicht ganz so hervorragende Angebote denkt, doch selbstverständlich sein sollte), macht die von irrealen Idealvorstellungen abweichenden Messeigenschaften nur sichtbar, bei nicht standardisierten Instrumenten bleiben sie in den meisten Fällen verborgen und werden durch subjektive „Sicherheiten“ (sog. unmittelbare Evidenz) ersetzt.

Selbst ein früher sehr wichtiges Argument gegen den praktischen Testeinsatz verliert zunehmend seine Gültigkeit: Um aussagekräftige Ergebnisse zu erzielen, muss bei Testverfahren in besonderem Maße die richtige und verständliche Instruktion, die Sicherstellung des Verständnisses der Testaufgaben, die genaue Erhebung der Antworten der getesteten Personen sowie deren Auswertung und Verrechnung gewährleistet sein. Während dies früher den Einsatz von psychologischem Fachpersonal erforderlich machte, wird heute die Objektivität und Fehlerfreiheit der Erhebung durch die Vorgabe und Auswertung von Testverfahren in computergestützter Form realisiert.

Allerdings haben Testverfahren naturgemäß auch ihre Begrenzungen. Das Interview ist den Testverfahren hinsichtlich seiner großen Bandbreite bei den erhobenen Informationen überlegen, Verhaltensübungen zeigen wesentlich besser das aktuelle Sprachvermögen, wichtige Facetten der Sozialen Kompetenz und auch die Sympathie bzw. „Ausstrahlung“ eines Bewerbers, und oft sind die Fachkenntnisse aus dem Lebenslauf besser abzuschätzen als durch „Prüfungen“. Nicht das missionarische Beharren auf *einer* Methode, sondern der für die jeweilige Fragestellung passend zusammengestellte Methodenmix („Multimethodalität“) stellt den optimalen Ansatz in der Personalarbeit dar.

Insofern sollte man skeptisch sein gegenüber jedweder Einseitigkeit, wie beispielsweise für die Auswahl von Bewerbern nur Interviews zu empfehlen (was noch sehr häufig in der Praxis geschieht) oder bei Assessment-Centern nur Tätigkeitssimulationen zuzulassen. Tests sollten in eignungsdiagnostischen Arrangements nie die einzige, vielmehr nur *eine* Verfahrensklasse neben anderen sein – natürlich nur, wenn es irritationsfrei geht.

ZIELSETZUNG DES VORLIEGENDEN HANDBUCHS

Für die sinnvolle Gestaltung des Methodenmix leistungsfähiger Eignungsdiagnostik steht heute eine Vielzahl von Testverfahren zur Verfügung, die sich hinsichtlich ihrer speziellen

Zielsetzung und auch hinsichtlich ihrer Qualität und Kosten sehr stark unterscheiden. Für den HR-Manager ist es daher notwendig, sich im Bedarfsfall schnell einen breiten Überblick über die verfügbaren Methoden zu verschaffen und die Passung eines bestimmten Testverfahrens für seine Fragestellung näher beurteilen zu können. Dazu benötigt er übersichtlich zusammengestellte Informationen. Genau dabei will das vorliegende Handbuch behilflich sein.

Die Vielzahl der allein im deutschen Sprachraum verfügbaren Testverfahren (weit über 1000) macht es notwendig, eine Auswahl für die Aufnahme in die hier vorliegende Übersicht vorzunehmen. Es wurden zunächst nur Verfahren ausgewählt, die einen eindeutigen Bezug zu Fragestellungen der Personalarbeit in der Wirtschaft haben, also keine „klinischen“ oder nur im schulisch-pädagogischen Bereich einsetzbaren Instrumente. Darüber hinaus wurden bevorzugt jene aufgenommen, die einen gewissen Bekanntheits- und Verbreitungsgrad im deutschen Sprachraum haben oder erwarten lassen. Und schließlich konnten nur jene Verfahren berücksichtigt werden, deren Autoren bzw. Vertreter sich bereit erklärten, Angaben zu allen in unserem Beschreibungsraster vorgesehenen Aspekten zu machen.

Trotz allem Bemühen um eine größtmögliche Vollständigkeit dieser Zusammenstellung mussten über die genannten Aspekte hinaus weitere Restriktionen vorgenommen werden, um die Übersichtlichkeit einer Darstellung in Buchform nicht zu gefährden. So wurde gerade von Anbietern mit einer größeren Anzahl von für die Wirtschaft brauchbaren Instrumenten nur eine kleinere, möglichst typische Auswahl aufgenommen. Auch war es manchmal notwendig, Testverfahren von der Aufnahme auszuschließen, wenn die Darstellung durch die Autoren zu „werbend“ war oder zu wenig seriös und plausibel hinsichtlich der Leistungsversprechen der Verfahren. Und manche der von uns angesprochenen Testvertreiter haben erst gar nicht geantwortet, meist wahrscheinlich aus gutem Grund.