

Tests in der Managementdiagnostik

Persönlichkeitstests unter der Lupe

Persönlichkeitstests im Personalmanagement können wertvolle Informationen für die Auswahl und Entwicklung von Führungs- und Führungsnachwuchskräften liefern. Die Verfahren dazu müssen grundlegende Anforderungen an eine fundierte Testentwicklung erfüllen, klassischen Gütekriterien entsprechen und verständlich sein. Nach Überzeugung der Autorin rechnet sich ihr Einsatz auf lange Sicht. Sie hat fünf Persönlichkeitstests mit Blick auf diese Aspekte untersucht.



Dr. Daniela Eisele ist Professorin für Personalmanagement an der Hochschule Heilbronn und berät Unternehmen bei personalpolitischen Themen.

Persönlichkeitstests werden bei der Auswahl, Platzierung, Potenzialeinschätzung und Entwicklung von Führungskräften in Deutschland im Vergleich zu anderen Ländern nach wie vor zögerlich genutzt (Sarges 2000; Schuler 2008, 20). Für eine weitere Verbreitung sprechen gute Gründe (vgl. Kimmere / Neef 2005, 31 f.). Es stehen dem jedoch zahlreiche Vorbehalte gegenüber, wie zum Beispiel Nachwei und Schermuly (2009) anschaulich beschreiben.

Persönlichkeitsstruktur- und Typentests

Bedenken gegenüber projektiven Verfahren wie dem Rorschach-Test (RT) oder dem Thematischen Apperzeptionstest (TAT) sind in der beruflichen Eignungsdiagnostik durchaus angebracht. Diese Verfahren, die ein möglichst umfassendes Bild der Persönlichkeit liefern sollen, sind schon wegen rechtlicher Bedenken ungeeignet. Je weniger Fragen mit dem Arbeitsplatz in Zusammenhang stehen, desto eher sind sie als unzulässig einzustufen, denn das berechtigte Informationsinteresse des Arbeitgebers wird durch das allgemeine Persönlichkeitsrecht des Bewerbers beschränkt. Anders sind dagegen psychometrische Persönlichkeitstests zu beurteilen, zu denen die Persönlichkeitsstruktur- und die Typentests gehören. Die Persönlichkeitsstrukturtests bilden Eigenschaftsausprägungen in Relation zu den Ergebnissen einer Normstichprobe mittels Zahlenwerten und Diagrammen ab (vgl. Hossiep / Mühlhaus 2005, 30). Aus meist umfangreichen Abfragen werden differenzierte Aussagen abgeleitet. Die Typentests ordnen dagegen die Teilnehmer meist anhand weniger ausgewählter Skalen verschiedenen Typen zu. Abbildung 1 zeigt die Vielfalt der Einsatzbereiche von Persönlichkeitstests im Personalmanagement (vgl. Hossiep / Mühlhaus 2005, 94 f.). Nicht jeder Test eignet sich jedoch für jedes Ziel gleich gut (vgl. Sarges / Wottawa 2004).

Tests vor dem Einsatz: US-Soldaten werden vor einem Kampfeinsatz getestet, um herauszufinden, wer besonders gefährdet ist, ein posttraumatisches Stress-Syndrom zu entwickeln.



In einem Seminar im Master-Studiengang Unternehmensführung an der Hochschule Heilbronn wurden folgende Verfahren exemplarisch geprüft:

- Persönlichkeitsstrukturtests, darunter das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP), CAPTain (Computer Aided Personnel Test answers inevitable) und der Reflector Big Five Personality, vormalig Workplace Big Five Personality, und
- als Vertreter der Typentests das persolog®-Persönlichkeits-Profil mit den Verhaltensdimensionen dominant, initiativ, stetig und gewissenhaft, besser bekannt als D-I-S-G, und das Team Management Profil (TMP).

Prüfungsanforderungen

Eingeteilt in vier Kategorien wurden zahlreiche Kriterien gecheckt, wobei auch die Anforderungen der DIN 33430 in die Prüfung einfließen. Geprüft wurden

- die fundierte Konzeption des Verfahrens,
- die angemessene Durchführung, Auswertung und Interpretation des Tests,
- die Erfüllung der klassischen Gütekriterien der Psychologie und
- inwieweit die Verfahren den betrieblichen Anforderungen wie guter Praktikabilität, Relation von Aufwand versus (zusätzlichem) Nutzen, Akzeptanz und Rechtstreue entsprachen.

Konzeption Bei der Frage nach der Konzeption des Verfahrens ging es darum, ob der Test für den beruflichen Kontext, zielgruppenspezifisch und für konkrete Aufgaben (Selektion, Platzierung, Laufbahnplanung, Teamentwicklung oder Beratung) entwickelt wurde. Von Interesse waren auch die erfassten Konstrukte und deren

theoretische und empirische Fundierung. Außerdem spielten die Erfahrungen der Autoren und die Referenzen der Anbieter bei der Untersuchung eine Rolle.

Angemessener Ablauf Umfang und Bestandteile des Testmaterials, ergänzende Angebote, wie zum Beispiel Schulungen, sowie elektronische und manuelle Auswertungsformen standen beim Betrachten der Durchführung, Auswertung und Interpretation der Tests im Mittelpunkt. Augenmerk lag dabei jeweils auf genauen Instruktionen. Zudem wurde recherchiert, ob und welche zusätzlichen (objektiven) Informationen zu den Tests am deutschen Markt verfügbar sind. Mit Blick auf den Test selbst interessierten neben dessen Aufbau die Anzahl der erfassten Skalen in Relation zur Zahl der Items (Fragen und Aufgaben), eine verständliche und berufsnahe Formulierung der Fragen und Aufgaben sowie die Verwendung gängiger Antworttechniken. Auch eine sorgfältige Auswahl der Items anhand von Verteilungen, Item-Schwierigkeit, Trennschärfe und Homogenität wurde geprüft. Dabei sollten möglichst Fragen und Aufgaben, deren Ergebnisverteilung extrem weit oder extrem wenig streute und deren Ergebnisse anderen Items der Skala zuwiderliefen, eliminiert werden.

Klassische Gütekriterien Basis der klassischen Gütekriterien ist eine hohe Objektivität bei der Durchführung, Auswertung und Interpretation der Tests. Eine entscheidende Rolle spielten auch hier Vollständigkeit und Anschaulichkeit der Instruktionen. Nur ein objektiver Test kann auch zuverlässig, also reliabel, sein. Dies wird anhand von möglichst hohen Reliabilitätskoeffizienten ($> .7$) festgestellt. Ermittelt werden die Werte mittels Retest (Testwiederholung), Paralleltest, Testhalbierung oder Berechnung der inneren Konsistenz beziehungsweise Homogenität.

Der Aussagegehalt eines Tests hängt letztlich von seinem konkreten Einsatz ab.

Über die generelle Aussagekraft können Validitätsziffern Auskunft geben ($> .25$). Unterschieden werden auch hier verschiedene statistische Verfahren: Im Vergleich mit einer (zukünftigen) Fremdeinschätzung wird von (prognostischer) Kriteriumsvalidität oder Übereinstimmungsvalidität gesprochen. Im Rahmen der (konvergenten oder divergenten) Konstruktvalidität werden die Ergebnisse des Verfahrens mit den Werten anderer Verfahren verglichen. Auf qualitativen Einschätzungen von Laien oder Experten beruhen dagegen Augenschein- beziehungsweise Inhaltsvalidität, die ebenfalls eingehen.

Zur fundierten Interpretation von Testergebnissen ist außerdem der Abgleich mit einer Norm oder vorformulierten Anforderungen notwendig. So ist beispielsweise das Ergebnis einer Zuverlässigkeit von drei erst aussagekräftig, wenn ein Durchschnittswert von vier oder gar eine Forderung nach fünf Punkten bekannt ist. Daher wurde geprüft, ob ausreichend große und passende Normgruppen vorhanden sind oder ob zumindest Aussagen zu Zielnormen gemacht werden. Ein weiterer Punkt war der Umgang mit einer möglichen Verfälschung der Ergebnisse durch die Teilnehmer hinsichtlich der Tendenz, sozial erwünscht zu antworten. Daher wurde untersucht, ob sich hierzu Aussagen finden und welche Strategien zur Minimierung vorgeschlagen werden.

Betriebliche Anforderungen In der betrieblichen Realität mindestens ebenso wichtig ist die praktische Handhabung der Tests, die durch Befragung einzelner Anwender sowie im Selbstversuch eingeschätzt wurde. Zudem wurden Kosten und Zeitaufwand für Vorbereitung, Durchführung, Auswertung und Ergebnisinterpretation in Relation zur Informationsausbeute gesetzt. Dabei wurde der (zusätzliche) Nutzen generell eingeschätzt, weshalb dazu nur Tendenzangaben möglich waren. Ein beson-



Standardisiert für den Masseneinsatz: 1 560 angehende Studenten absolvieren im französischen Poitiers die Aufnahmeprüfung für Logopäden.

ders kritischer Punkt ist die Akzeptanz der Testverfahren durch die Probanden, aber auch durch die Unternehmensvertreter (vgl. Nachtwei / Schermuly 2009, 8). Um diesen Aspekt abzubilden, wurden Angaben zur Verbreitung des Tests und konkrete Aussagen zur Akzeptanz ermittelt. Hinter dem Begriff der Rechtstreue, der Compliance, stehen beim Testeinsatz auf Managementebene in erster Linie die Wahrung von Persönlichkeitsrechten und die Vermeidung von Benachteiligungen im Sinne des Allgemeinen Gleichbehandlungsgesetzes.

Jeweils ein studentisches Team prüfte diese Kriterien anhand des Informationsmaterials der Autoren und Anbieter, der Testunterlagen und der Verfahrensdurchläufe und durch ergänzende Literaturrecherche und Befragungen. Die Studierenden hatten mit Ausnahme des BIP Zugang zu allen Verfahren (teilweise online). Der BIP wurde aus Studiengebühren erworben. Alle deutschen Testanbieter (bzw. im Falle des BIP das Projektteam Testentwicklung) standen den Studierenden für Fragen zur Verfügung. Zudem konnten sie sich in fast allen Fällen mit Anwendern über deren teils langjährige Erfahrungen unterhalten.

Die Tabelle (siehe Abb. 2) bietet einen ersten Überblick über besondere Stärken, aber auch Grenzen der genannten Tests. Zu betonen ist dabei, dass in dem Seminar der Einsatz der Tests in der betrieblichen Manage-

mentdiagnostik im Mittelpunkt stand. Unbenommen ist, dass einige der Verfahren ihre eigentlichen Stärken besser in anderen Bereichen entfalten können (z. B. CAPTain im Vertrieb oder D-I-S-G im persönlichen Coaching).

Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)

Das BIP wurde in den 1990er-Jahren, basierend auf umfangreichen Recherchen und umfassenden Expertenbefragungen, unter der Leitung des promovierten Psychologen Rüdiger Hossiep erstellt. Seit dieser Zeit beforcht, entwickelt und ergänzt das Projektteam Testentwicklung an der Ruhr-Universität Bochum das von Hogrefe vertriebene Verfahren. Mithilfe von 210 Items werden 14 Skalen und vier Zusatzindizes in den folgenden Bereichen abgefragt:

- berufliche Orientierung, darunter Leistungs-, Gestaltungs- und Führungsmotivation,
- Arbeitsverhalten, darunter Gewissenhaftigkeit, Flexibilität und Handlungsorientierung,
- soziale Kompetenz, darunter Sensitivität, Soziabilität, Kontaktfähigkeit Teamorientierung und Durchsetzungsstärke sowie
- psychische Konstitution, darunter emotionale Stabilität, Belastbarkeit und Selbstbewusstsein.



Nichts als die Wahrheit... verspricht der Einsatz des Lügendetektors. Hier lernen Studenten am Polygraph Institute in Bogotá den Einsatz der umstrittenen Technik, die in Kolumbien recht verbreitet ist.

Den geschlossenen Statements (Beispiel: „Ich rücke nur ungerne von einem festen Tagesablauf ab“) ist auf einer sechsstufigen Skala mehr oder weniger zuzustimmen. Die damit verbundene Gefahr der Testverfälschung greifen die Autoren auf. Es finden sich Gegenmaßnahmen wie eine offene Kommunikation und die Einbindung der Testergebnisse in den weiteren Prozess sowie deren zielorientierte Ergänzung. Die Tendenz zur sozialen Erwünschtheit können sie damit jedoch nicht in allen Fällen ausschließen. Das kann die Aussagekraft des Tests im Einzelfall einschränken. Die mit verschiedenen Verfahren (Retest, Testhalbierung und innere Konsistenz) ermittelte Reliabilität wird als gut angegeben, auch die Validität ist entsprechend beforscht. Das gilt vor allem für die Konstruktvalidität. Die Zusammenhänge mit Skalen des Eysenck-Persönlichkeits-Inventars (EPI), des NEO-Fünf-Faktoren Inventars (NEO-FFI) und der revidierten Fassung des 16-Persönlichkeits-Faktoren-Tests (16-PF-R) werden als hoch bis sehr hoch angegeben. Die Kriteriumsvalidität erzielt ausreichende Werte.

Für eine Grundausstattung ist mit Investitionen in Höhe von 500 Euro (schriftliche Version) beziehungsweise 2 000 Euro (PC-Version) zu rechnen. Bei hohen Fallzahlen kostet der einzelne Test dann allerdings nur noch wenige Euro. Hier ist die schriftliche Auswertung

mithilfe einer Schablone jedoch unpraktisch, obwohl das Vorgehen an sich dabei Schritt für Schritt nachvollziehbar wird. Dagegen geht die Beantwortung trotz des Frageumfangs mit rund 30 Minuten einigermaßen zügig. Die Ergebnisse werden als (relative) Zahlenwerte in einem Profilblatt anschaulich dargestellt. Interpretieren muss sie der Anwender allerdings selbst. Dazu werden ihm vielfältige Informationen und Interpretationshilfen angeboten. Dabei ist mit zusätzlichem Aufwand für die Einarbeitung zu rechnen, außerdem werden Grundlagenkenntnisse vorausgesetzt. Positiv ist hervorzuheben, dass diese intensive Auseinandersetzung förderlich für die Qualität des Gesamtprozesses ist.

Computer Aided Personnel Test answers inevitable (CAPTain)

Das Verfahren CAPTain liegt in verschiedenen computerbasierten Versionen vor, die unterschiedliche Skalen umfassen. Unter anderem werden eine Managementversion und sogar die Anpassung an spezifische Anforderungen angeboten. CAPTain baut auf der Persönlichkeitstheorie von Henry Murray (1938) auf und wurde von der promovierten Psychologin Ulla Petäjävaara in den

1980er-Jahren in Finnland konzipiert. Christoph Nagler, Psychologe und geschäftsführender Gesellschafter der CNT Gesellschaft für Personal- und Organisationsentwicklung GmbH, übernahm den Test ins Deutsche und entwickelte ihn weiter. Er ist Alleinanbieter dieses Tests, liefert zahlreiche Informationen dazu und offeriert auch Schulungen. Die im Test-Retestverfahren gewonnenen Reliabilitätswerte sind demnach ausreichend; die Validitätswerte werden als gut angegeben. Aufgeführt sind neben Konstrukt- auch hier Kriteriumsvaliditäten. Während der Test zwar in einigen Abschlussarbeiten und Dissertationen beforscht wird, wird er in deutschsprachigen Veröffentlichungen nur selten erwähnt.

Die Bereiche Arbeitsleistung, Führungseigenschaften, Entscheidungsfreude, Persönlichkeit, Teamfähigkeit und Potenziale werden durch maximal 38 Skalen abgebildet. Dem Bereich Führungseigenschaften sind zum Beispiel wiederum fünf Skalen zugeordnet, darunter Führungsstärke, Autoritätsorientierung, Einflussnahme und Delegation. Alle 38 Skalen werden zusammengenommen durch 183 Aussagenpaare wiedergegeben. Nach dem Forced-Choice-Verfahren ist von jeweils zwei Aussagen (Beispiel: „Mir fällt es leicht, andere zu aktivieren“ und „Ich

komme pünktlich zu Terminen“) die jeweils zutreffende zu wählen. Dieses Vorgehen wirkt der Tendenz zu sozialer Erwünschtheit sehr gut entgegen. Die Bezeichnung als Messung des „objektiven Verhaltensstils“ erscheint dennoch etwas irreführend. Auch die für den Anwender nicht nachvollziehbare Auswertung wirkt sich zwar positiv auf eine mögliche Verfälschung, aber negativ auf die Transparenz aus.

Dieser Teil des Tests wird durch einen zweiten, die subjektiven Idealvorstellungen, ergänzt. Hier gibt der Proband auf einer zehnstufigen Skala mit fünf Verhaltensankern bis zu 38-mal seine Selbsteinschätzung ab (z. B. von „Detailarbeit meide ich“ bis hin zu „Details sind das Wichtigste; nach diesem Grundsatz handle ich“).

Als aus Probandensicht kritisch beschrieben wird der Umfang der Befragung in Kombination mit der Forced-Choice-Methode: Zu viele, teilweise als unabhängig empfundene Aussagenpaare reihen sich aneinander. Dafür enthält der sieben Reports umfassende Auswertungsbericht neben Skalenbeschreibungen und Interpretationshinweisen eine Zusammenfassung, die Ergebnisse zu den beiden Teilen, eine Gegenüberstellung sowie den konkreten Bezug auf den Einsatzbereich. Vor allem Letzteres ist positiv hervorzu-

heben. Um mit der umfangreichen Auswertung zurechtzukommen, wird die Vollversion im Paket mit einer Schulung angeboten. Die Vielzahl der kombinierbaren Angebote ist attraktiv, erschwert es dem Anwender aber, Kosten und Nutzen gegenüberzustellen. Die Preisspanne für den Test reicht von 60 Euro für die kurze Version bis zu mehreren Tausend Euro für die Einrichtung eines kundengerechten Komplettpakets inklusive Schulungen.

Reflector Big Five Personality (RBFP)

Das Ende der 90er-Jahre vom US-amerikanischen Beratungs- und Weiterbildungsinstitut Centacs entwickelte Vorgängermodell des Reflector Big Five Personality beruht auf dem wohl bekanntesten Verfahren zur Persönlichkeitsabbildung, dem Fünf-Faktoren-Modell (FFM). Die fünf Faktoren, die im Test durch 24 Skalen wiedergegeben werden, sind:

- Neurotizismus / Emotionale Instabilität
- Extraversion
- Offenheit für Erfahrungen / Kultur / Intellekt
- Verträglichkeit / Soziabilität / Liebenswürdigkeit
- Gewissenhaftigkeit



In Deutschland vertreibt die Firma Wildenmann Tools & Services GmbH & Co. KG das Verfahren webbasiert in Kooperation mit der niederländischen PI Company. Es umfasst 144 Aussagen im Stil von „Unterstützt andere nur dann, wenn er / sie selbst etwas davon hat“ und fünfstufige Skalen von -2 (trifft gar nicht zu) bis +2 (trifft vollkommen zu). Die Einschätzung aller Aussagen benötigt circa 20 Minuten. Auch wenn die Aussagen in der dritten Person und teilweise negativ formuliert sind, sind Tendenzen zu sozialer Erwünschtheit nicht vollständig zu vermeiden. Leider macht der Anbieter hierzu keine Angaben. Bezüglich der Reliabilität führt er lediglich sehr hohen Wert der inneren Konsistenz an. Mit enormen Korrelationen glänzen die Vergleiche mit der revidierten Fassung des NEO-PI-R (revidierte Fassung des NEO-Persönlichkeitsinventars) und mit dem „Schwestertest“ Connector P. Außerdem werden dessen Ergebnisse zur Kriteriumsvalidität als weitgehend übertragbar auf den RBFP angeben.

Zumindest im deutschen Sprachraum finden sich sonst kaum weitere Informa-

tionen zu dem Verfahren. Neben dem normierten und sehr anschaulichen Persönlichkeitsbericht, der zehn Seiten umfasst, gehört zur Auswertung als Besonderheit ein 20-seitiger Kompetenzbericht. Dieser gibt für 43 ausgewählte Kompetenzen der Bereiche Management und Führung, Business, Analyse und Entscheidung, Kommunikation, Persönlichkeit und Motivation an, wie leicht oder schwer dem Kandidaten die jeweilige Kompetenzentwicklung aufgrund des ermittelten Profils fällt. Während die Auswertung des Persönlichkeitsprofils nachvollziehbar ist, bleibt die Übertragung auf die Kompetenzen intransparent. Es wird lediglich auf eine fundierte Experteneinschätzung verwiesen. Die Ergebnisse interpretieren typischerweise Mitarbeiter von Wildenmann oder zertifizierte Trainer in einem ausführlichen und als sehr konstruktiv empfundenen (telefonischen) Gespräch. Dies schlägt sich auch in den Kosten (bei Wildenmann rund 150 Euro) nieder. Eine eigene Interpretation durch den Anwender setzt ein Training voraus, das nur einmalig einen höheren Aufwand verursacht.

persolog® Persönlichkeits-Modell oder D-I-S-G

Das persolog® Persönlichkeits-Profil unterscheidet vier Verhaltensgrundtypen, denen die Farben Rot für Dominanz, Gelb für Initiative, Grün für Stetigkeit und Blau für Gewissenhaftigkeit zugeordnet sind. Es wurde in den 1970er-Jahren vom amerikanischen Verhaltenswissenschaftler John G. Geier im Rahmen seiner Dissertation entwickelt und basiert auf den Erkenntnissen von William M. Marston (1928). Dieser entwickelte das Verfahren ursprünglich zum Zweck der Selbsteinschätzung zur Beurteilung einer weiteren Hochschullaufbahn. Das sehr eingängige Modell wurde weit darüber hinaus bekannt. Zielgruppe und Einsatzbereiche des persolog® Persönlichkeits-Profiles werden heute äußerst breit beschrieben, ausgehend von seinem ursprünglichen Zweck der Selbsteinschätzung bis zur Beurteilung einer weiteren Hochschullaufbahn. Neben dem berufsbezogenen Einsatz wird das Verfahren vor allem auch im privaten Bereich angewendet. Dies ist mit ein Grund für einen sehr hohen Bekanntheitsgrad des D-I-S-G.

Der einzige Anbieter der letzten Version des Persönlichkeitsprofils von Geier im deutschsprachigen Raum, die persolog GmbH, ist seit Anfang der 1990er-Jahre auf dem Markt und hat bereits über 3 000 Trainer für die Anwendung zertifiziert. Um das Verfahren weiter zu etablieren wird es vom Anbieter regelmäßig beforscht und weiterentwickelt. Seit mehr als einem Jahrzehnt werden Reliabilitätskennzahlen im guten oder sogar sehr guten Bereich veröffentlicht. Auch die Konstruktvalidität (mit den Skalen des NEO-PI-R) bewertet der Anbieter als gut. Mehr als generalisierende Aussagen und Verweise von dritter Seite sowie Werbebeiträge der Trainer finden sich dagegen im deutschen Sprachraum nicht.

Stärken und Schwächen der Tests im Überblick			
Test	Stärken	Schwächen	Empfehlung
BIP	<ul style="list-style-type: none"> Wissenschaftliche Begleitung, Forschung, aktuelle Normierung und konstante Weiterentwicklung Speziell für die (deutsche) Berufseignungsdiagnostik entwickelt (mittlerweile auch international) 	<ul style="list-style-type: none"> Relativ umfangreiche Befragung Tendenz zur sozialen Erwünschtheit Anschaulich normiertes Profil, aber kein Auswertungsbericht; allgemeine Interpretationshilfen sind vom Anwender fallweise zu übertragen 	Fundierte Unterstützung für verschiedene Zwecke des (internationalen) HRM, insbesondere bei höheren Fallzahlen; für Einsatz in Trainings zur Teamentwicklung nicht das richtige Verfahren
CAPTain	<ul style="list-style-type: none"> Komfortable Handhabung Test / Auswertung vielsprachig Verschiedene Versionen mit unterschiedlichem Umfang und Möglichkeit der modularen Anpassung an ein konkretes Profil 	<ul style="list-style-type: none"> Auswertung und Bezeichnung „Messung des objektiven Verhaltensstils“ schwer durchschaubar Hoher Umfang der Vollversion in Kombination mit dem Forced-Choice-Verfahren verringert die Akzeptanz 	Einsatz der unterschiedlichen Versionen für verschiedene Zwecke des internationalen HRM, insbesondere wenn Anwendungszahlen die Anpassung an ein konkretes Profil lohnenswert erscheinen lassen
Reflector Big Five	<ul style="list-style-type: none"> Basis bildet das akzeptierte Persönlichkeitsmodell Big Five Neben einem (spezifisch) normierten Persönlichkeitsprofil gibt es einen Kompetenzbericht Ausführliches Feedbackgespräch 	<ul style="list-style-type: none"> Auswertung des Kompetenzberichts intransparent Fehlende Aussagen zur Testverfälschung Zeitlicher und materieller Aufwand des Feedbacks 	Empfehlenswert erscheint der Einsatz aufgrund des Kompetenzberichts insbesondere für die Karriereberatung und Laufbahnplanung; für Trainings eher weniger geeignet
D-I-S-G	<ul style="list-style-type: none"> Hoher Bekanntheitsgrad Nachvollziehbarkeit der Auswertung, die (mühsam) auch selbst vorgenommen werden kann Anschaulichkeit der simplen Vier-Felder-Typologisierung 	<ul style="list-style-type: none"> Fehlende Aussagen zur Testverfälschung Verhältnis von geringem Testumfang und Auswertungslänge unstimmig Auswertungsbericht ist teilweise platitudenhaft formuliert 	In der Entwicklung und im Coaching insbesondere zur Selbstreflexion empfehlenswert; insbesondere für die Auswahl nicht geeignet
TMP	<ul style="list-style-type: none"> Anschaulichkeit des Team Management Systems Umfangreiche Normierung Gute Nachvollziehbarkeit der Auswertung 	<ul style="list-style-type: none"> Fehlende Aussagen zur Testverfälschung Validitätswerte werden nicht berichtet Verhältnis von geringem Testumfang und Auswertungslänge unstimmig 	Besonders sinnvoll für die Teamzusammenstellung und -entwicklung; insbesondere für die Auswahl nicht geeignet

Abb. 2

Der Test selbst ist äußerst übersichtlich. Anzukreuzen ist 24-mal ein Begriff von vier (z. B. freundlich, überzeugend, bescheiden oder einfallsreich), der am ehesten auf einen zutrifft. Ebenfalls 24-mal ist die Aussage zu kennzeichnen (z. B. ich bin schüchtern, ich bin entschlossen, es fällt mir leicht, andere zu begeistern, und ich gebe leicht nach), die am wenigsten zu einem passt. Dadurch ist die Beantwortung intuitiv und zügig. Das Verfahren selbst kostet in der schriftlichen Version mithilfe eines Buches weniger als 35 Euro. Der Anwender muss den Test allerdings selbst aufwendig auswerten und interpretieren. Dies erfordert Zeit für die Einarbeitung und Durchführung. Nutzt er die Onlineversion, erhält er den elektronischen Bericht direkt zugesandt. Dieser erscheint in Relation zum Fragenkatalog als sehr umfangreich und bleibt damit (gezwungenermaßen) zum Teil oberfläch-

lich. Er wird zudem in einem (telefonischen) Gespräch von einem Trainer erläutert, wobei die Kosten in Abhängigkeit von Trainer, Einsatzgebiet und weiteren Faktoren zwischen 50 und 200 Euro pro Stunde stark variieren. Das Beratungs- und Weiterbildungsunternehmen bietet außerdem viele Informationen und zahlreiche (kostenpflichtige) Zusatzleistungen rund um das Persönlichkeitsprofil an.

Team Management Profil (TMP)

Das Team Management System (TMS) und das darauf aufbauende Team Management Profil (TMP) wurden Mitte der 1980er-Jahre von den britisch-australischen Wissenschaftlern Charles Margerison und Dick McCann entwickelt. Im deutschsprachigen Raum vertreibt das TMS-Zentrum unter der Leitung des Ökonomen Marc Tscheuschner das Verfahren. Ziel des Tests ist es, die Arbeitspräfe-

renzen von Mitgliedern in einem Team zu definieren, sodass die entsprechenden Teamrollen bestmöglich ausgefüllt werden können. Dies sind: Beraten (informierter Berater), Innovieren (kreativer Innovator), Promoten (entdeckender Promotor), Entwickeln (auswählender Entwickler), Organisieren (zielstrebigem Organisator), Umsetzen (systematischer Umsetzer), Überwachen (kontrollierender Überwacher) oder Stabilisieren (unterstützender Stabilisator). Die Präferenzen werden mithilfe von vier sogenannten Schlüsselbereichen ermittelt:

- Beziehungen: extrovertiert vs. introvertiert,
- Informationen: praktisch vs. kreativ,
- Entscheidungen: analytisch vs. gegründet auf Überzeugungen,
- Organisation: strukturiert vs. flexibel.

Jeder dieser Schlüsselbereiche wird durch 15 Items repräsentiert. Bei einem Aussagenpaar (z. B. „Bei mir ist Logik das Wichtigste“ vs. „Bei mir sind Gefühle das Wichtigste“) ist eine mittlere oder starke Präferenz in die eine oder andere Richtung anzugeben. Die insgesamt 60 Aussagenpaare sind in rund 15 Minuten schriftlich oder online zu bearbeiten. Jeder Präferenzkombination, zum Beispiel IPAS (introvertiert, praktisch, analytisch, strukturiert), wird plakativ eine Rolle (hier: „systematischer Umsetzer“) zugeschrieben. Anschaulich ist auch der Auswertungsbericht. Mit 30 Seiten erscheint er allerdings trotz Visualisierung der (normierten) Ergebnisse und ergänzender Informationen, zum Beispiel zum Kommunikations- und Führungsverhalten sowie zur Arbeitsweise des Typs, in Relation zum Testumfang etwas lang.

Das TMS-Verfahren wird im deutschen Sprachraum vor allem in zwei Werken des Anbieters anschaulich erläutert. Hier gibt er auch Reliabilitätswerte der inneren Homogenität, der Testhalbierung und des Retests an. Diese sind, von einer Ausnahme abgesehen, alle gut bis zufriedenstellend. Hinsichtlich der Validität finden sich dagegen nur Hinweise zur Konstruktvalidität, unter anderem mit dem Meyer-Briggs-Typenindikator, aber keine konkreten Angaben. Laut Bericht wurden die Auswertungsprofile durch Probandenfeedback sukzessiv verbessert, was der Augenscheinvalidität zugutekommt. Die Anfälligkeit für soziale Erwünschtheit wird nicht weiter erwähnt. Von Dritten wurde der Test bislang nicht geprüft, zumindest erbrachte die Recherche keine deutschsprachigen Veröffentlichungen. Die Auswertung durch TMS für 150 bis 200 Euro ist gut nachvollziehbar (teilweise in Englisch erläutert) und wird durch Mitarbeiter oder zertifizierte Trainer in einem (telefonischen) Gespräch interpretiert. Die Inanspruchnahme von Angeboten für zusätzliches Coaching

und Teamworkshops ziehen entsprechenden Mehraufwand nach sich.

Fazit

Alle begutachteten Testverfahren können einen durchaus aufschlussreichen Informationsbeitrag für verschiedene Einsatzbereiche im Personalmanagement liefern. Im optimalen Fall werden jedoch alle drei möglichen Zugänge zu einer Person kombiniert (Hossiep / Paschen 2003, 46):

- Objektive Informationen (wie Ausbildungsabschluss oder Auslandserfahrungen) bilden die Basis.
- Referenzen oder Beobachtungen (z. B. im Assessment-Center etc.) liefern das Fremdbild einer Person.
- Das Selbstbild ergibt sich durch Selbstbeschreibung im Interview und mithilfe von Testverfahren und rundet die Gesamteinschätzung ab.

Hohe (zusätzliche) Aussagekraft liefert ein Test allerdings nur auf der Basis formulierter Anwendungsziele im Unternehmen und der darauf abgestimmten gezielten Auswahl eines Verfahrens. Um die intensive Beschäftigung mit den Anforderungen und möglichen Tests für den jeweiligen Einsatz kommen Anwender nicht herum. Der vorliegende Beitrag liefert jedoch eine generelle Struktur und bietet gebündelt Informationen für eine solche Verfahrensentscheidung an.

Summary

Management Diagnostics Tests

Personality tests can be used in a wide range of personnel management applications in selecting, assigning, evaluating and developing management personnel and up-and-coming executives. However, such tests are still only being used hesitantly in Germany. Such tests can only be used in a responsible way if their intended purpose has been defined and such purpose corresponds with

the objective of the test. Personality structure tests comprise the Bochum Psychological Test Procedure to Describe Occupation-Related Personality Characteristics (Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)), CAP-Tain (Computer Aided Personnel Tests answers inevitable) and the Reflector Big Five Personality, previously known as the Workplace Big Five Personality. Personality type tests include among others the persolog Personality Model, better known as the DISG, and the Team Management Profile (TMP). Time and know-how are required to make a well-founded decision about conducting a test. Admittedly, these test procedures are not suitable for stand-alone use but they can provide valuable insights when combined with objective information and the results of a third party assessment. This applies at least insofar as the test procedures have been developed in a substantiated way and have proven to be at least satisfactory considering the classic quality criteria.

Literatur

- Berufsverband Deutscher Psychologinnen und Psychologen (Hg.) (2002): DIN 33430 Berufsbezogene Eignungsdiagnostik, Berlin
- Hossiep, R. / Paschen, M. (2003): BIP – Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung, 2. Aufl., Göttingen
- Hossiep, R. / Mühlhaus, O. (2005): Personalauswahl und -entwicklung mit Persönlichkeitstests, Göttingen
- Kimmerer, M. / Neef, M. (2005): Einsatz von Persönlichkeitstypologien in der deutschen Wirtschaft, in: Wirtschaftspsychologie aktuell, 3, 31–36
- Nachtwei, J. / Schermuly, C. (2009): Acht Mythen über Eignungstests, in: Harvard Business Manager, 4, 6–9
- Sarges, W. (Hg.) (2000): Management-Diagnostik, 3. Aufl., Göttingen et al.
- Sarges, W. / Wottawa, H. (2004): Handbuch wirtschaftspsychologischer Testverfahren, 2. Aufl., Stuttgart
- Schuler, H. (2008): Das Bewusstsein für mehr Qualität schärfen, in: Personalwirtschaft, 35 (6), 17–20