

PERSONAL- PSYCHOLOGISCHE SCHRIFTEN

Modelle, Instrumente, Studien

Herausgeber: *Prof. Dr. Jens Nachtwei, Sebastian Uedelhoven, Dr. Charlotte von Bernstorff & Doreen Liebenow*



**Handbuch zum
IQP-Testsystem[©]:
Intelligenz- und
Persönlichkeitsdiagnostik für
Personalauswahl und -entwicklung**

(1. Auflage)

Band 1/2015

Herausgeber

Prof. Dr. Jens Nachtwei, Sebastian Uedelhoven, Dr. Charlotte von Bernstorff & Doreen Liebenow

Autoren

Udo Faust, Katharina Kille, Tobias Lattke, Doreen Liebenow, Prof. Dr. Jens Nachtwei, Sebastian Uedelhoven, Dr. Charlotte von Bernstorff, Thomas Wessendorf

Personalpsychologische Schriften, Band 1/2015

ISSN 2366-3928

IQP – Privat-Institut für Qualitätssicherung in Personalauswahl und -entwicklung GmbH

Potsdamer Platz 10

10785 Berlin

E-Mail: info@iqp-germany.de

Web: www.iqp-germany.de

Inhaltsverzeichnis

Vorwort	5
Kurzzusammenfassung	6
1. Einleitung	
von Bernstorff, C. & Nachtwei, J.	7
1.1. Personalauswahl in Deutschland – Bestandsaufnahme	7
1.2. Personalauswahlverfahren im Vergleich – Funktion und Güte	9
1.3. Eignungstests – Anwendung und Nutzen	12
1.4. Anforderungsanalyse – Notwendigkeit und Umsetzung	15
2. Der Persönlichkeitstest	
Liebenow, D. & Nachtwei, J.	19
2.1. Persönlichkeit – Rolle in der Personalauswahl	19
2.2. Persönlichkeit – Zentrale Bewerbermerkmale	20
2.3. Persönlichkeitstest – Qualität und Nutzen	26
3. Der Intelligenztest	
Kille, K. & Nachtwei, J.	31
3.1. Intelligenz – Rolle in der Personalauswahl	31
3.2. Intelligenz – Zentrale Bewerbermerkmale	32
3.3. Intelligenztest – Qualität und Nutzen	35
4. Der Weg: Testkonstruktion	
Uedelhoven, S. & Nachtwei, J.	37
4.1. Einleitung – Strategien und Methoden	37
4.2. Konstruktion der Persönlichkeitsskalen	38
4.3. Konstruktion der Intelligenzskalen	40
4.4. Messfehler und Verfälschungstendenzen – Gegenmaßnahmen	43

5. Die Qualität: Testgüte	
Uedelhoven, S. & Nachtwei, J.	48
5.1. Objektivität	48
5.2. Normierung	48
5.3. Zuverlässigkeit.	51
5.4. Gültigkeit	53
5.5. Ökonomie und Nutzen	57
6. Das Tool: IQP-Testsystem®	
Faust, U., Lattke, T. & Wessendorf, T.	58
6.1. Vorbereitung der Testdurchführung	60
6.2. Testdurchführung	61
6.3. Auswertung und Interpretation der Testergebnisse	62
Glossar	64
Literaturverzeichnis	67
Index	75

Vorwort

Das vorliegende Testhandbuch vermittelt zunächst einen Einblick in die wissenschaftlichen Erkenntnisse der beruflichen Eignungsdiagnostik. Es fasst die Erkenntnisse der Jahre 2005 bis 2009 zusammen. Eine aktualisierte und vollständig überarbeitete Fassung wird im Jahr 2016 erscheinen. Der Intelligenz- und Persönlichkeitstest als spezifische Verfahren der Potentialanalyse werden detailliert dargestellt. Das IQP-Testverfahren[®] verwendet beide Testformen in spezieller Weise. Entsprechend beziehen sich alle Angaben auf die Art der Verwendung durch IQP. Darüber hinaus wird das Vorgehen bei der Konstruktion des Tests erläutert und durch Aussagen zur Qualität bzw. Güte desselben abgerundet. Abschließend wird die Funktionalität des IQP-Testsystems[®] als Softwarelösung erläutert.

Das Testhandbuch enthält mehr als 100 Quellenangaben. Diese Menge an Forschungsbefunden ist nötig, um die Nachvollziehbarkeit der Argumentation zu sichern. Die Quellenangaben sind mit Zahlen gekennzeichnet und in den Fußnoten in verkürzter Form aufgeführt. Im Literaturverzeichnis ist jede Quelle vollständig dokumentiert.

Einzelne Begrifflichkeiten werden seit jeher in der Eignungsdiagnostik verwendet und lassen sich schlecht umgangssprachlich verwenden. Diese werden an entsprechender Stelle im Text kurz und verständlich definiert sowie in einem Glossar (siehe unterstrichene Begriffe) festgehalten.

Die folgende Seite bietet Ihnen die zentralen Informationen des Testhandbuchs in verkürzter Form. Wir empfehlen jedoch, das gesamte Buch zu lesen. Zur Unterstützung der Lesbarkeit wird auf eine geschlechtsspezifische Differenzierung, wie z.B. Bewerber/Innen, verzichtet und nur die maskuline Bezeichnung verwendet. Die entsprechend verwendeten Begriffe gelten jedoch im Sinne des Allgemeinen Gleichbehandlungsgesetzes stets für beide Geschlechter.

Wir wünschen Ihnen eine angenehme und aufschlussreiche Zeit beim Lesen dieses Handbuchs.

Prof. Dr. Jens Nachtwei, Sebastian Uedelhoven,
Dr. Charlotte von Bernstorff und Doreen Liebenow

Gliederung des
Handbuchs

Hinweise zur Ver-
wendung

Kurzzusammenfassung

Die Erfassung der beruflichen Eignung durch Intelligenz- und Persönlichkeitstests ist seit einigen Jahren auch in Deutschland auf dem Vormarsch. Der finanzielle Nutzen des Einsatzes jener Tests lässt sich nachvollziehbar berechnen. Dabei werden freie Interviews und Assessment Center hinsichtlich Kosten, Nutzen und Güte bereits geschlagen, wenn die Erfassung von Intelligenz und Persönlichkeit per Test kombiniert wird. Jedoch sollte ein Test nur als ergänzende Maßnahme im Methoden-Mix eingesetzt werden.

Auch der beste Test liefert nur dann interpretierbare Ergebnisse, wenn die Anforderungen der vakanten Stelle wissenschaftlich fundiert und systematisch erhoben werden. Definierte Anforderung und Testergebnis werden verglichen, um die Eignung eines Bewerbers zu beurteilen. IQP bietet sowohl den Eignungstest als auch die Anforderungsanalyse – in höchster Qualität und softwaregestützt im IQP-Testsystem®.

Das IQP-Testverfahren® wurde nach etablierten Theorien und mit Hilfe fundierter statistischer Analysen entwickelt, um die Berufseignung von Fach- und Führungskräften schnell und zuverlässig zu identifizieren. Folgende zentrale Merkmale der Intelligenz werden erfasst: Abstrakt-Analytische Intelligenz, Numerische Intelligenz sowie Verbale Intelligenz. Zusätzlich werden sieben Persönlichkeitsmerkmale gemessen: Emotionale Belastbarkeit, Extraversion, Geistige Flexibilität, Gewissenhaftigkeit, Leistungsmotivation, Risikoneigung und Teamorientierung. Die Relevanz der genannten Merkmale für den Berufserfolg konnte nachgewiesen werden und wird im Handbuch ausführlich dokumentiert.

Die Anwendung des IQP-Testverfahrens® ermöglicht eine zuverlässige Personalauswahl im Sinne der DIN 33430 und der Bestimmungen des Allgemeinen Gleichbehandlungsgesetzes (AGG).

Eignungstests als wichtige Ergänzung bestehender Personalauswahlinstrumente

Das IQP-Testverfahren® – wissenschaftlich fundierte Personalauswahl

1. Einleitung

von Bernstorff, C. & Nachtwei, J.

1.1. Personalauswahl in Deutschland – Bestandsaufnahme

Im Sommer 2006 machte eine Pressemitteilung der Technischen Universität Dresden publik, was viele Personalverantwortliche schon lange wussten: „Unzulängliche Personalauswahl verursacht in Deutschland Milliardenverluste“.¹ Lange fand die Diskussion um Aufwand und Nutzen von Personalauswahlverfahren fast ausschließlich im Kreis von Forschern statt. Die Eignungsdiagnostik hat eine jahrzehntelange Tradition und dennoch sind viele Erkenntnisse in der Praxis nie angekommen². Die Gründe dafür sind vielfältig: Verwendung von Fachjargon und mangelnde Verbreitung der wissenschaftlichen Literatur sind Hürden, mit denen Praktiker zu kämpfen haben. Hinzu kommt, dass viele Methoden der Eignungsdiagnostik nicht so bewertet werden, wie es in der Wirtschaft nötig wäre – konkrete Angaben zu zeitlichen und finanziellen Kosten fehlen häufig. Die Effizienz von Verfahren ist dadurch schwer zu beurteilen.³

Die Entwicklung von eigenen Testverfahren bedeutet hohen Aufwand für ein Unternehmen; die Verwendung externer Verfahren birgt jedoch ebenfalls Risiken: Wie wurde das Verfahren konstruiert? Welche Qualifikationen haben die Testautoren und welche Qualität das entwickelte Verfahren? Fragen, die häufig unbeantwortet bleiben, da viele Verfahren die nötige Transparenz vermissen lassen. Ein Handbuch wie das vorliegende ist ein Weg, Einblicke in den Eignungstest und seine Entstehung zu erhalten, sowie ein Verständnis für die Güte des Tests zu entwickeln.

Seit 2002 wird versucht, den gesamten Prozess der Personalauswahl durch eine Norm, die DIN 33430, zu formalisieren. Zusammengefasst empfiehlt die Norm (verbindlich ist sie nicht), den gesamten Prozess der Auswahl von Bewerbern nach wissenschaftlich fundierten und transparenten Kriterien zu strukturieren. Dies beginnt bei einer systematischen Anforderungsanalyse der zu besetzenden Stelle, bezieht die Qualifikation der Verantwortlichen (Testentwickler, Testleiter,

Forschungserkenntnisse müssen in die Praxis finden

Normung bei der Personalauswahl: die DIN 33430

1 Pressemitteilung: 21.06.2006 (idw) – Technische Universität Dresden

2 vgl. Kersting (2006)

3 vgl. Ryan & Tippins (2004)

Praktiker im Unternehmen) mit ein und verlangt eine ausreichende Dokumentation, um die Qualität der verwendeten Verfahren beurteilen zu können. Abschließend wird ein belegbares und nachvollziehbares Urteil über den einzelnen Bewerber gefordert.⁴

Dass nicht jeder Personalverantwortliche ein Experte auf dem Gebiet der beruflichen Eignungsdiagnostik sein kann, versteht sich von selbst. Die Aneignung fundierten Wissens auf diesem Gebiet ist zeit- und aufwendig und bedeutet die Beschäftigung mit einer Vielzahl unbekannter Konzepte aus der Psychologie und Statistik. Hier bietet es sich an, Experten am Auswahlprozess zu beteiligen und sich als Praktiker auf ureigene Kompetenzen zu berufen.

Eine dieser Kompetenzen betrifft die juristischen Aspekte der Personalarbeit. Seit 2006 ist das Allgemeine Gleichbehandlungsgesetz (AGG) rechtswirksam⁵. Aus dem Gesetz ergibt sich für den Arbeitgeber die Pflicht, Arbeitsplätze „diskriminierungsfrei“ auszuschreiben. Die Diskriminierung wird u.a. hinsichtlich Alter, Geschlecht, Behinderungen und Herkunft definiert. Dabei wird zwischen der unmittelbaren und mittelbaren Diskriminierung unterschieden. Ersteres tritt auf, wenn beispielsweise in der Stellenanzeige von vornherein Bewerber einer bestimmten Altersgruppe aus dem Auswahlprozess ausgeschlossen werden. Zweitgenanntes, also die mittelbare Diskriminierung, tritt z.B. auf, wenn ein Eignungstest durchgeführt wird, welcher für die Stellenanforderungen überhaupt nicht relevant ist. Somit würde das Testergebnis keine Aussage zur Eignung für die spezifische Stelle treffen. Kurzum: Alle Bewerber müssen mit demselben Eignungstest und identischem Vergleichsmaßstab konfrontiert werden.

Der Eignungstest selbst muss zentrale berufsrelevante Merkmale erfassen und (im Gegensatz zu demografischen Merkmalen) zur Einstellungsentscheidung herangezogen werden. An dieser Stelle sei gesagt, dass das IQP-Testverfahren[®] diese Anforderungen erfüllt. Es werden keine unterschiedlichen Normen (also Vergleichsmaßstäbe) nach Alter oder Geschlecht zur Interpretation des Testwertes herangezogen und ausschließlich Merkmale erfasst, die für jede Fach- und Führungskraft relevant hinsichtlich des Berufserfolgs sind.

Nicht jeder Personalverantwortliche ist Diagnostiker

Beachtung des Allgemeinen Gleichbehandlungsgesetzes bei der Personalauswahl

**IQP Testverfahren[®]
AGG-konform**

⁴ vgl. Westhoff et al. (2005)

⁵ Gesetzestext unter <http://bundesrecht.juris.de/agg/>

Die beschriebene DIN 33430 sowie das AGG legen nahe, nicht den „Bauch“ über die Einstellung zukünftiger Mitarbeiter entscheiden zu lassen.⁶ Nutzenanalysen zeigen: Nichts wiegt schwerer als der vollständige Verzicht auf zuverlässige Methoden (vgl. Abschnitt 1.3). Noch 1999 stand Deutschland bei der Verwendungshäufigkeit von Intelligenz- und Persönlichkeitstests im europäischen Vergleich an vorletzter Stelle.⁷ Dies hat sich in den letzten Jahren glücklicherweise geändert. Und dennoch bleibt eine starke Unsicherheit bei der Frage bestehen, welches Auswahlverfahren wann angewendet werden soll. Der nächste Abschnitt stellt die gängigsten Verfahren kurz gegenüber.

Entscheidungen aus dem Bauch heraus sind teuer

1.2. Personalauswahlverfahren im Vergleich – Funktion und Güte

Um die Masse von Verfahren zur Potentialanalyse zu strukturieren, wird zunächst eine Grobunterscheidung vorgenommen in biographie-, simulations- und eigenschaftsorientierte Verfahren.⁸

biographie-, simulations-, eigenschaftsorientierte Verfahren

In der Praxis wird erstgenannter Messansatz immer verwendet: Die Sichtung von Bewerbungsunterlagen gehört ebenso dazu, wie die Verwendung von Einstellungsinterviews. In jedem Fall geht es um die Erfassung und Bewertung von Informationen aus der Biographie des Bewerbers. Die Idee dahinter: Von vergangenem Verhalten (z.B. häufigem Stellenwechsel) kann auf zukünftiges im eigenen Unternehmen geschlossen werden.

Mithilfe von simulationsorientierten Verfahren hingegen soll aktuelles Verhalten in konkreten, dem Arbeitskontext sehr ähnlichen Situationen erfasst werden. Dies wird häufig mit Hilfe von verschiedenen Übungen in Assessment Centern umgesetzt. Vom dort beobachteten Verhalten wird versucht, auf allgemeingültiges Verhalten und dahinter liegende Eigenschaften zu schließen. In beiden Verfahrensgruppen gibt es – ungeachtet der vielen Vorteile – ein Problem: die Verbindung zwischen Messung und Realität. Beim biographieorientierten Verfah-

Informationen aus Biographie und Simulation müssen die Realität widerspiegeln

6 vgl. Ryan & Tippins (2005)

7 vgl. Ryan et al. (1999)

8 vgl. Schuler (2001)

ren wird versucht, vom Vergangenen auf die Zukunft zu schließen. Bei Simulationen soll der Schluss vom in der (simulierten) Situation gezeigten Verhalten auf allgemeine Verhaltensweisen bzw. Eigenschaften gelingen. Sind die jeweiligen Verfahren, wie strukturiertes Interview oder Assessment Center, gut konstruiert und werden fachmännisch durchgeführt, fällt diese Problematik weniger ins Gewicht.

Eines gelingt jedoch nicht: eine bessere Vorhersage von Berufserfolg als dies Verfahren des letztgenannten, eigenschaftsorientierten Ansatzes vermögen. Bei diesem Messansatz geht es um die Erfassung von Eigenschaften, die zum einen zeitlich stabil sind und zum anderen Verhalten in einer Vielzahl von Situationen (so auch im beruflichen Kontext) zuverlässig vorhersagen. Intelligenz und Persönlichkeit sind hierbei die zentralen Eigenschaftsbereiche, in welchen sich Personen voneinander unterscheiden.⁹

Das Forschungsinteresse an Intelligenz und Persönlichkeit ist seit vielen Jahrzehnten ungebrochen.¹⁰ So verwundert es kaum, dass sich gerade im Bereich der Eigenschaftsmessung Standards herausgebildet haben. In Wissenschaft und Praxis wird meist auf dieselben theoretischen Überlegungen und entsprechend vergleichbare Messansätze zur Intelligenz und Persönlichkeit zurückgegriffen. Bei Kenntnis von Theorien und Konstruktionsprinzipien der Tests ist es demnach möglich, Berufseignung mit Hilfe des eigenschaftsorientierten Ansatzes auf höchstem Niveau zu betreiben. Studien, welche Verfahren der drei erläuterten Messansätze vergleichen, kommen zum einheitlichen Ergebnis, dass der Intelligenztest (neben Arbeitsproben) den Berufserfolg am genauesten vorhersagt.¹¹ Der Persönlichkeitstest sagt ebenfalls wichtige Aspekte der Leistung im Beruf vorher¹², sollte jedoch nicht unreflektiert verwendet werden. Durch die Kombination von Intelligenz- und Persönlichkeitstest lässt sich die Vorhersage von Berufserfolg maximieren. Dies ist von den zuvor genannten Einzelverfahren (Interview, Assessment Center) nicht zu übertreffen.¹³

Tradition und Know-how bei der Messung von Eigenschaften

Intelligenz- und Persönlichkeitstest in Kombination sagen Berufserfolg am besten vorher

9 vgl. Asendorpf (1999)

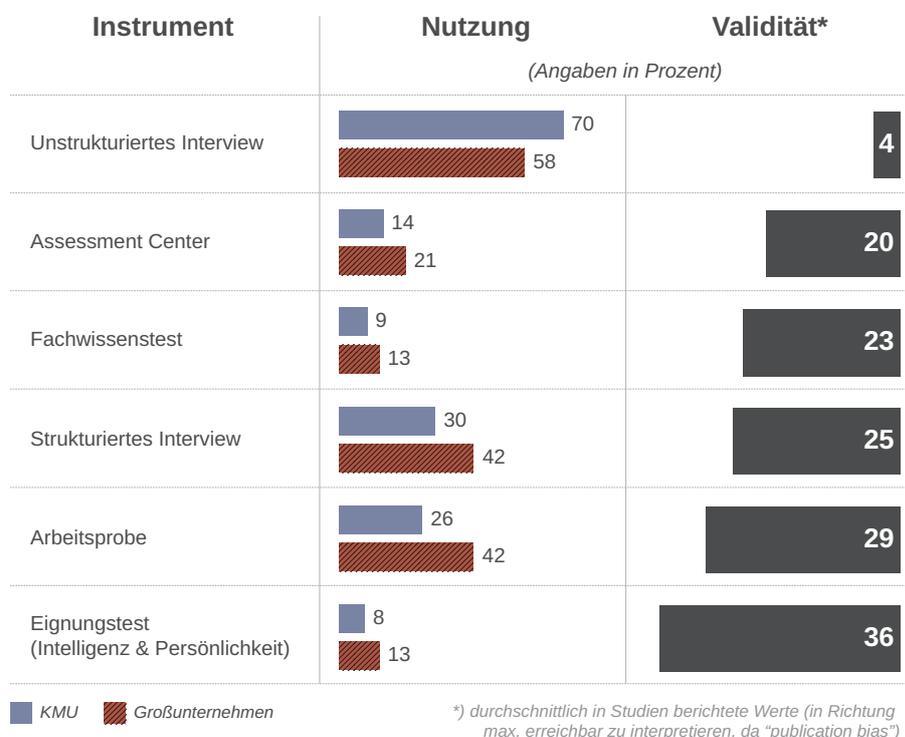
10 vgl. Plate (2006)

11 vgl. Kersting (2004)

12 vgl. Hogan & Holland (2003)

13 vgl. Schmidt & Hunter (1998)

Wir haben in einer eigenen Studie für den Harvard Business Manager¹⁴ die Nutzungshäufigkeit der genannten Instrumente in 113 deutschen Unternehmen sowie deren in der Literatur verbriefte Validität bei der Personalauswahl (= Vorhersage von Berufserfolg, hier umgerechnet in %) gegenübergestellt. Der zentrale Befund: Das freie Einstellungsinterview wird zwar sehr häufig genutzt, weist mit 4% Validität jedoch die geringste Qualität auf. Eignungstests hingegen werden (zumindest in Deutschland) sehr selten herangezogen, besitzen aber mit 36% die im Vergleich höchste Validität.



Nutzung vs. Validität von Personalauswahlinstrumenten in Prozent

Jedes Verfahren hat Vor- und Nachteile, so dass sich in jedem Fall ein Methoden-Mix anbietet: Beginnend bei den Bewerbungsunterlagen, über eine Vorauswahl per Intelligenz- und/oder Persönlichkeitstest, bis hin zum abschließenden Interview oder Assessment Center. Alternativ sind auch integrierte Lösungen aus Interview, Assessment Center und Eignungstest denkbar.

Der Methoden-Mix – Kombination der Vorteile verschiedener Auswahlverfahren

14 vgl. Nachtwei & Schermuly (2009)

Im folgenden Abschnitt wird auf die Anwendung und den Nutzen von Eignungstests eingegangen, welche nach dem erläuterten eigenschaftsorientierten Ansatz konstruiert sind. Das IQP-Testverfahren[®] wurde nach diesem Ansatz entwickelt, so dass diese Ausführungen übertragbar sind.

1.3. Eignungstests – Anwendung und Nutzen

Eine umfangreiche Studie zur Verwendung verschiedener Personalauswahlverfahren aus dem Jahr 2006¹⁵ kam zu dem ernüchternden Ergebnis: Eignungstests werden (v.a. vor dem Hintergrund ihres nachweislich hohen Nutzens) zu selten eingesetzt. Im Hinblick auf die oben zitierte, eigene Studie, scheint sich dieser Zustand bisher nicht geändert zu haben. Dies ist häufig der Einstellung verschuldet, dass ein Test nicht die Komplexität einer Person beschreiben kann; dies sei nur im direkten Kontakt mit dem Bewerber möglich. Es ist richtig, dass ein persönliches Zusammentreffen von Arbeitgeber und Bewerber vor einer endgültigen Personalentscheidung unverzichtbar ist. Falsch ist jedoch, jedem Bewerber diese Möglichkeit einzuräumen. Gerade in Zeiten hoher Bewerberzahlen und/oder Kostendrucks ist es äußerst unökonomisch, die Bewerberflut ungefiltert in Interviews oder Assessment Centern zu begutachten.

An dieser Stelle kommt der Vorteil eines Eignungstests als Instrument der Vorselektion von Bewerbern voll zum Tragen: Standardisierung. Jeder Bewerber muss sich genau denselben Testfragen stellen; die Durchführung am Computer erspart den Einsatz von Personal. Die Testdauer ist begrenzt, die Auswertung rechnerbasiert und somit objektiv und schnell. Kurzum: Ein gut konstruierter Eignungstest ist in der Lage, mit wenig Aufwand eine große Zahl von Bewerbern (oder auch Mitarbeitern) hinsichtlich berufsrelevanter Merkmale standardisiert zu prüfen.

Jedoch dürfen Eignungstests nicht unüberlegt eingesetzt werden. Zum einen sollten sie theoretisch fundiert und methodisch sauber konstruiert sein. Zum anderen spielt aber auch die Form der Durchführung

Eignungstests: hoher Nutzen, jedoch geringe Verbreitung

Standardisierung als großer Vorteil von Tests

Von der Testdurchführung im Internet ist abzuraten

15 vgl. Hell et al. (2006)

eine große Rolle. So werden diese Tests häufig im Internet durchgeführt – der Bewerber kann über einen Login den Test zu Hause absolvieren. Das Problem dabei ist, dass die Standardisierung als Vorteil des Eignungstests an dieser Stelle verloren geht. Die Nachteile von Internettests sind augenscheinlich: Es kann kaum kontrolliert werden, wer den Test eigentlich durchführt¹⁶ – die Login-Daten sind schließlich übertragbar und kein Testleiter überwacht, wer den Test bearbeitet. Ferner sind die Testbedingungen hier sehr unterschiedlich: Ein ruhiger Testraum ohne Ablenkungen steht nicht jedem Bewerber zur Verfügung, wodurch die Testdaten z.B. durch Ablenkungen verzerrt werden.¹⁷ Und auch der Einsatz von Hilfsmitteln (z.B. Taschenrechner bei Intelligenztests) entzieht sich der Kontrolle des Arbeitgebers. Ein weiterer Nachteil der Internetdiagnostik bezieht sich auf die Bewerberakzeptanz. Ohne Betreuung und persönliches Feedback wirkt das Massenmedium Internet wenig wertschätzend, was sich letztlich negativ auf die Wahrnehmung des Images des Unternehmens beim Bewerber auswirken kann. Das Fazit lautet: Ein qualitativ hochwertiger Test kann sein Potential nur unter kontrollierten Durchführungsbedingungen entfalten. Eine Internetversion eines Persönlichkeitstests ist vertretbar, sofern eine gute Vorabinformation sowie eine direkte Ergebnismeldung erfolgen.

Neben Vorteilen in der praktischen Anwendung und der statistisch nachweisbaren Vorhersagegenauigkeit von Berufserfolg ist der finanzielle Nutzen von Eignungstests das Kernkriterium schlechthin. Jedem Personalverantwortlichen ist klar, dass sich die Leistung der Mitarbeiter auf den Unternehmenserfolg auswirkt. Weit weniger intuitiv ist der Zusammenhang zwischen Testeinsatz und Unternehmenserfolg. Dieser besteht nur dann, wenn der Test Berufsleistung auch wirklich vorhersagen kann und außerdem die Kosten für den Einsatz des Verfahrens unter dem potentiellen Nutzen bleiben.

Es gibt jedoch inzwischen eine sowohl detaillierte als auch praktisch nutzbare Formel zur Schätzung, welchen finanziellen Nutzen ein Testeinsatz hat.¹⁸ Danach ergibt sich der Nutzen durch folgende Varia-

Der finanzielle Nutzen eines Tests als Kernkriterium

Der finanzielle Nutzen kann berechnet werden

16 vgl. Chuah et al. (2006)

17 vgl. Johnson (2005)

18 vgl. Schuler & Höft (2004)

blen, die miteinander multipliziert werden:

- Anzahl der ausgewählten Bewerber N_A
- Betriebszugehörigkeit dieser Bewerber in Jahren T
- Zuwachs an Vorhersagegenauigkeit durch den Testeinsatz r_{xy}
- mittlere Leistung der Bewerber im Test Z_x
- geschätzter Unterschied der Bewerber im Berufserfolg SD_y

Der Nutzen ist demnach maximal, wenn viele Bewerber eine Stelle erhalten (N_A), die neuen Mitarbeiter lange im Unternehmen verbleiben (T), der Test stark zur Verbesserung der Vorhersage von Berufserfolg beiträgt (r_{xy}), die Bewerber eher hohe Testergebnisse erzielen (Z_x) und sich im Berufserfolg auch voneinander unterscheiden (SD_y ; also nicht von vornherein alle denselben Berufserfolg haben, da z.B. die Anforderungen der Stelle sehr niedrig sind). Das Produkt der genannten Variablen ergibt den Nutzen des Testeinsatzes. Davon müssen die Kosten abgezogen werden, welche durch Multiplikation der folgenden Variablen berechnet werden:

- direkte Kosten durch Verwendung des Tests C
- Anzahl der getesteten Bewerber N_B

Fazit: Netto-Nutzen $\Delta U = N_A \times T \times r_{xy} \times Z_x \times SD_y - C \times N_B$

In konkreten Zahlen: Ein Bewerber wird für fünf Jahre eingestellt. Der Test macht die Vorhersage des Berufserfolgs um 4% (= Zusammenhang von Test und Berufserfolg als Korrelation r_{xy} von .20) genauer. Dies ist eine pessimistische Schätzung; 10-35% sind zu erreichen, wenn zuvor beispielsweise nur ein unstrukturiertes Einstellungsinterview genutzt wurde¹⁹. Die Bewerber zeigen eine mittlere Leistung im Einstellungstest und unterscheiden sich im Berufserfolg geschätzt um 40% eines mittleren Jahresgehalts von ca. 52.000 € (etablierter Schätzer für Berufserfolg in €²⁰).

Das Ergebnis: ein Brutto-Nutzen für das Unternehmen von 20.800 € über den Zeitraum von fünf Jahren. Diese Modellrechnung bestätigt die Arbeit von vielen Personalverantwortlichen in Unternehmen, die

Der maximale Nutzen hängt von mehreren Faktoren ab

Ein Brutto-Nutzen eines Testeinsatzes von 4000 € pro Bewerber und Jahr ist keine Seltenheit

19 vgl. Schmidt & Hunter (1998); Schuler (2001); Kersting (2006)

20 vgl. Schuler & Höft (2004)

bereits eine systematische Bewerberauswahl unter Einsatz von Eignungstests durchführen. Sie zeigt aber auch, dass die Qualität des Tests selbst darüber entscheidet, wie hoch der finanzielle Vorteil für das Unternehmen letztlich ist. Ist ein Eignungstest gut konstruiert und sauber durchgeführt, wird in jedem Fall ein maßgeblicher Netto-Nutzen für das Unternehmen erzielt. In Kapitel 2 und 3 wird die Qualität der Vorhersage (= prädiktive Validität) von Berufserfolg durch Persönlichkeits- und Intelligenztests noch einmal gesondert und im Detail dargestellt.

1.4. Anforderungsanalyse – Notwendigkeit und Umsetzung

Die Notwendigkeit einer systematischen Analyse der Anforderungen einer zu besetzenden Stelle zeigt sich am besten im Titel eines Artikels der FAZ²¹: „Der Beste ist häufig nicht der Geeignetste“. Auf einen Nenner gebracht: Nicht jede Stelle erfordert einen stressresistenten, nahezu zwanghaft gewissenhaften „Mix aus Goethe und Einstein“. Das bedeutet, dass nicht die maximale Ausprägung von Persönlichkeit und Intelligenz entscheidet über Erfolg im Beruf, sondern die Passung der Eigenschaft zur Anforderung der Stelle. Die Phänomene der Über- und Unterforderung aufgrund mangelnder Passung sind schon lange als nachteilig für den Berufserfolg nachgewiesen.²²

Das Flow-Modell²³ (siehe Abbildung) verdeutlicht diese Konzeption und beschreibt Flow als einen Zustand, in welchem eine Person in ihrer Arbeit aufgeht und innere Befriedigung aus der Zielerreichung aufgrund einer idealen Passung von Anforderung der Stelle und Merkmal der Person erreicht.

Der Weg, um diesen in der Abbildung dargestellten Flow-Kanal auszuloten, ist die Anforderungsanalyse, also die Beschreibung der Stellenanforderungen und ihre Verbindung mit angestrebten Merkmalsausprägungen der Bewerber. Als Ergebnis entsteht ein Anforderungsprofil, um die Testergebnisse der Bewerber hinsichtlich der

Erst ein Anforderungsprofil der Stelle macht Testergebnisse interpretierbar

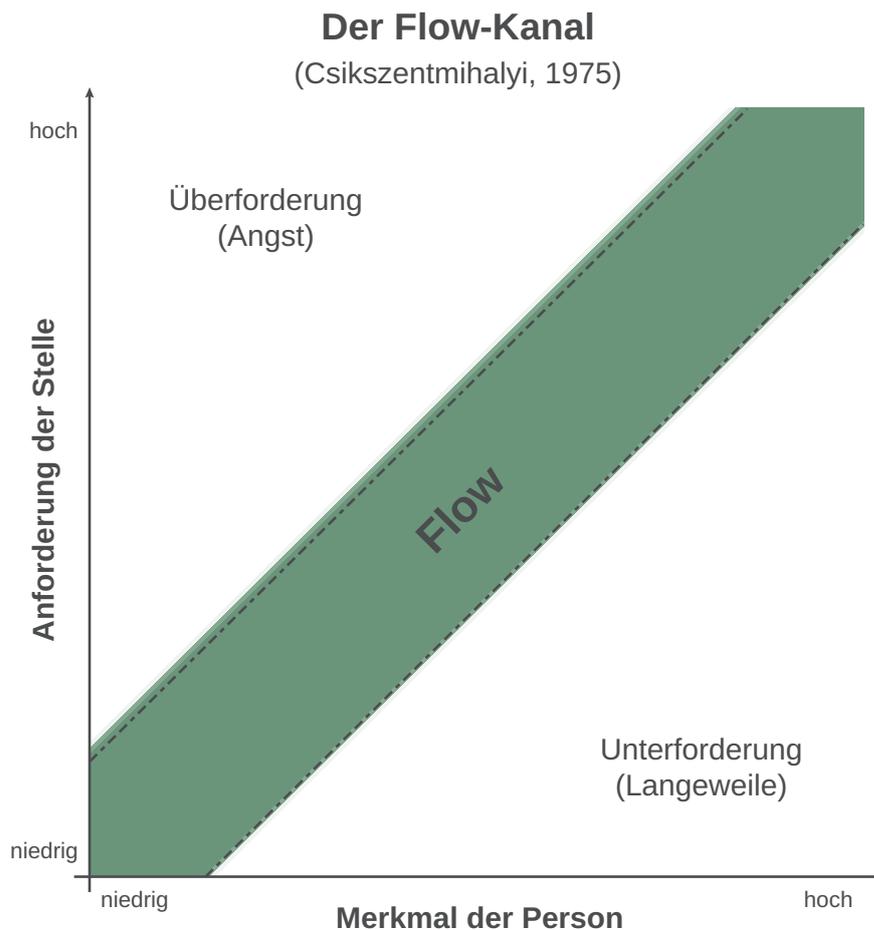
Das Flow-Modell – Visualisierung von Idealzustand, Über- und Unterforderung

21 vgl. Artikel, Frankfurter Allgemeine Zeitung, FAZJob.net (15.04.2005)

22 vgl. Kristof-Brown et al. (2005)

23 vgl. Csikszentmihalyi (1975)

Passung zur Stelle interpretieren zu können. Dabei muss sich eine Anforderungsanalyse immer direkt oder indirekt auf die Merkmale beziehen, die im Eignungstest auch erfasst werden. Die Anforderung „hohe Kreativität“ zu definieren, macht keinen Sinn, wenn der Eignungstest dieses Merkmal gar nicht erfasst.



Der Flow-Kanal als Zielbereich: Pas-sung von Eignung und Anforderung

Dass saubere Personalauswahl nur durch den Einsatz zuverlässiger Anforderungsanalysen funktionieren kann, beschreibt schon die oben zitierte DIN 33430.²⁴ Die Norm trifft allerdings keine verbindliche Aussage darüber, welches anforderungsanalytische Verfahren in welchem Bereich zum Einsatz kommen sollte. Der Anwender hat die Qual der Wahl und das Problem, dass Standardverfahren eventuell nicht auf den verwendeten Eignungstest passen.

Zum IQP-Testverfahren[®] gibt es aus diesem Grund ein eigens ent-

Wissenschaftlich fundierte Anforderungsanalyse für das IQP-Testverfahren[®]

²⁴ vgl. Westhoff et al. (2005)

wickeltes Anforderungsanalyseverfahren, welches aus bis zu drei Stufen besteht.

1. Expertenevaluation (Experten = Fachvorgesetzter/ Führungskraft, Personalmanager, langjähriger Leistungsträger aus der Belegschaft, Eignungsdiagnostiker) unter Berücksichtigung der Auswirkung von Merkmalsausprägungen auf die jeweilige Stelle
2. strukturierte Anforderungsanalyse auf Basis mathematischer Gewichtung von spezifischen Anforderungen an die Stelle (auf Basis von Literaturbefunden und Expertenevaluationen von Psychologen)
3. empirische Ermittlung des Anforderungsprofils basierend auf den Testergebnissen von Stelleninhabern im Unternehmen

Dabei ist zu beachten, dass sich die Vorgehensweisen keinesfalls ersetzen. Vielmehr ist es für die Erstellung eines möglichst präzisen Anforderungsprofils wichtig, die eher flexible und freie Vorgehensweise der Expertenevaluation mit der empirischen Ermittlung (falls im Unternehmen umsetzbar) und der strukturierten Anforderungsanalyse zu kombinieren.

Den wissenschaftlichen Hintergrund der strukturierten Analyse in Stufe 2 bildet der „Fragebogen zur Arbeitsanalyse (FAA)“ von Frieling & Hoyos (1978).²⁵ Das Instrument wird nicht verwendet, bietet jedoch einen guten Rahmen für eine umfassende und zugleich ökonomische Klassifikation von Stellenanforderungen. Dabei wird die vakante Stelle anhand von drei Dimensionen beschrieben:

- Aufnahme und Verarbeitung arbeitsrelevanter Informationen (z.B.: Welche Denk- und Entscheidungsprozesse sind notwendig?)
- Beziehungen zu arbeitsrelevanten Personen (z.B.: Wie hoch ist der Umfang persönlicher Kontakte?)
- Arbeitsrelevante Randbedingungen (z.B.: Wie stark ist die Arbeit vorstrukturiert?)

Zu jeder dieser Dimensionen gehört eine Menge von Merkmalen,

²⁵ vgl. Frieling & Hoyos (1978)

Die Kombination der drei Stufen führt zum präzise definierten Anforderungsprofil

Beschreibung der Stellenanforderung über Stellenmerkmale

um die vakante Stelle zu beschreiben (siehe Beispiele in Klammern). Jedem dieser Merkmale wiederum sind die Skalen des IQP-Testverfahrens[®] (z.B. Gewissenhaftigkeit) zugeordnet. So werden anhand von Expertenurteilen und Befunden aus der Literatur einzelne konkrete Merkmale der Stelle mit Intelligenz- und Persönlichkeitsmerkmalen verbunden, die ein Bewerber aufweisen muss, um diesem Stellenmerkmal gerecht zu werden.

Zum Beispiel: In der Dimension „Aufnahme und Verarbeitung arbeitsrelevanter Informationen“ der Anforderungsanalyse wird definiert, ob Arbeitsinformationen vorwiegend aus schriftlichen Dokumenten versus Gesprächen und Meetings abzuleiten sind. Die Anforderung an das im IQP-Testverfahren[®] erfasste Merkmal Abstrakt-Analytische Intelligenz ist bei Meetings insofern höher, da sofort reagiert und dem Gedankengang anderer in Echtzeit gefolgt werden muss. Schriftliche Dokumente dagegen erlauben eine längere Beschäftigung mit den Inhalten sowie häufigeres Nachlesen und -fragen bei Unklarheiten. Mitarbeiter, die vorwiegend mit schriftlichen Dokumenten arbeiten, sollten nicht übermäßig extrovertiert sein, während ein Mindestmaß davon für das Bestehen in Meetings sowie die (Arbeits-)Zufriedenheit mit dieser Arbeitsform unabdingbar ist. Ob jemand teamorientiert ist, spielt umso weniger eine Rolle, desto weniger soziale Kontakte diese Person zur Aufgabenbewältigung haben muss.

Diese Einschätzung hinsichtlich Intelligenz- und Persönlichkeitsmerkmalen erfolgt über insgesamt 16 Stellenmerkmale. Im Ergebnis wird die Anforderung der Stelle an den Bewerber anhand von Intelligenz- und Persönlichkeitsmerkmalen beschrieben.

Mit Hilfe etablierter Dimensionen der Anforderungsanalyse und einer Kombination mit den spezifischen Skalen des IQP-Testverfahrens[®] liegt somit ein maßgeschneidertes Verfahren vor, um die Anforderungen der Stelle an die Intelligenz und Persönlichkeit von Fach- und Führungskräften systematisch zu definieren.

Genauere Analyse der mit der Stelle verbundenen Aufgaben und Randbedingungen

Beschreibung der Anforderung über 16 Stellenmerkmale

2. Der Persönlichkeitstest

Liebenow, D. & Nachtwei, J.

2.1. Persönlichkeit – Rolle in der Personalauswahl

Eine zentrale Dimension zur Beschreibung von Bewerbern ist ihre Persönlichkeit. Ein Persönlichkeitsmerkmal wird dabei als zeitlich (zumindest mittelfristig) stabiles Merkmal beschrieben, welches das Verhalten einer Person in unterschiedlichen Situationen bestimmt.²⁶ Dass jene Merkmale überhaupt zur Beurteilung von Personal herangezogen werden, war nicht immer selbstverständlich. Lange wurde der Sinn von Persönlichkeitstests zur Aufklärung von Berufserfolg nicht gesehen. Das liegt vor allem daran, dass vorwiegend mit kleinen Personengruppen und sehr vielen verschiedenen Persönlichkeitsmerkmalen gearbeitet wurde. Verallgemeinerbare Aussagen waren dementsprechend sehr schwer ableitbar.²⁷ Diese wurden mit der Nutzung sogenannter Metaanalysen, welche die Ergebnisse verschiedener Studien statistisch integrieren, jedoch möglich. Es konnte gezeigt werden, dass Persönlichkeitsmerkmale relativ unabhängig vom jeweiligen Beruf wichtige Aspekte wie Leistung, Zufriedenheit, Fluktuation etc. vorhersagen können.²⁸

Jedoch vor allem die Möglichkeit, die Persönlichkeit sparsam und zugleich umfassend zu umschreiben, eröffnete neue Wege der ökonomischen Personalauswahl. Die sogenannten „Big Five“ (Gewissenhaftigkeit, Extraversion, Neurotizismus, Offenheit, Verträglichkeit) konnten als die fünf zentralen Persönlichkeitsmerkmale identifiziert werden.²⁹ Aufgrund der extensiven Forschung zum „Big Five Modell“ der Persönlichkeit ist jeder Praktiker gut beraten, einen Test zu nutzen, dessen Konstruktionsprinzip an diesem Modell angelehnt ist.

IQP folgt ebenfalls diesem Ansatz und passt ihn durch Abstimmung des Tests auf den beruflichen Kontext sowie die Erweiterung durch zwei zusätzliche Merkmale (Leistungsmotivation und Risikoneigung) an. Im Zuge der Anpassung wurden Merkmale praxistauglicher formuliert, ohne dabei den Kern des Merkmals zu verändern

Relevanz von Persönlichkeit für Auswahlentscheidung heute höher denn je

Die „Big Five“ als „Periodensystem der Elemente“ der Persönlichkeitsdiagnostik

26 vgl. Asendorpf (1999)

27 vgl. Borkenau et al. (2005)

28 vgl. Hunter & Hunter (1984)

29 vgl. Woods & Hampson (2005)

(Neurotizismus = Emotionale Belastbarkeit, Offenheit = Geistige Flexibilität, Verträglichkeit = Teamorientierung). Um ökonomisch zu bleiben, ist es jedoch zentral, Tests nicht künstlich durch eine Vielzahl von spezifischen Merkmalen aufzublähen. Dies führt ab einem bestimmten Punkt dazu, dass Redundanzen entstehen, d.h., Merkmale auch durch eine Kombination anderer Merkmale beschreibbar und somit im Grunde ersetzbar sind.

2.2. Persönlichkeit – Zentrale Bewerbermerkmale

Im vorherigen Abschnitt wurden bereits wichtige Persönlichkeitsmerkmale genannt. Es ist wichtig, genau zu verstehen, was sich hinter diesen Merkmalen verbirgt, welche Zusammenhänge zwischen ihnen bestehen und welchen Einfluss sie auf berufsrelevante Aspekte haben. Da letztgenannter Zusammenhang das höchste Gewicht hat, erfolgt eine detaillierte Darstellung im nachfolgenden Abschnitt. Im Folgenden wird jedes Merkmal entsprechend des wissenschaftlichen Konsens und der Verwendung als Skala im IQP-Testverfahren[®] definiert.

**Die Persönlichkeitsmerkmale im IQP-Testverfahren[®]:
Definition und Bezug zum Beruf**

Emotionale Belastbarkeit

Die Skala Emotionale Belastbarkeit (teils als „Emotionale Stabilität“ bezeichnet) erfasst, wie gut Personen mit Stress umgehen und hohe Anforderungen emotional bewältigen können. Diese Eigenschaft bestimmt zu einem großen Anteil die Leistungsfähigkeit in Zeiten hoher Belastungen. Ausgeglichenheit und Stabilität gehören ebenfalls zu den Kernbereichen der Skala.

Stressresistenz fördert Rationalität und Führungskompetenz

Emotionale Belastbarkeit ist in hohem Maß dort notwendig, wo gesteigertes Arbeitsaufkommen und schwierige Arbeitsverhältnisse über einen längeren Zeitraum zusammenkommen. Tätigkeiten mit hohem Termindruck und nur kurzen oder unregelmäßigen Regenerationsphasen erfordern ebenfalls ein überdurchschnittliches Maß an Emotionaler Belastbarkeit. So ist Personen mit sehr geringen Werten auf dieser Skala von einer Karriere z.B. im schnelllebigen Projektgeschäft abzuraten. Eine sehr geringe Ausprägung auf dieser Skala kann in den meisten Berufen das Risiko für negative Beanspruchungsfolgen

erhöhen. Burnout und vergleichbare Nebenerscheinungen können gehäuft auftreten. In vielen Berufsfeldern ist eine mittlere Emotionale Belastbarkeit ausreichend, da höhere Anforderungen an die Stresstoleranz meist nicht über längere Zeiträume bestehen bleiben. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Emotionale Belastbarkeit ebenfalls hoch ausgeprägt ist (und andersherum)³⁰:

- Rationalität beim Argumentieren
- leistungsförderliches Verhalten in Teams
- Führungskompetenzen
- Aufgabenorientierung

Extraversion

Die Skala Extraversion erfasst die Tendenz von Personen, sich auf ihr soziales Umfeld zu konzentrieren. Ferner wird das Interesse an Geselligkeit und der Darstellung der eigenen Person gemessen. Die Skala bestimmt auch, wie hoch die Stimulationen aus der Umwelt sein müssen, um optimal ausgelastet zu sein.

Ein hohes Maß an Extraversion ist in allen sozialen Berufen sowie bei Tätigkeiten mit intensivem Kundenkontakt von immenser Wichtigkeit. Personen mit hohen Werten auf dieser Skala sollten darüber hinaus keine Karriere in Arbeitsbereichen anstreben, in denen streckenweise hohe Monotonie und Routine herrschen. Abwechslung in der Tätigkeit ist für jene Personen sehr wichtig. Ein zumindest durchschnittliches Maß an Extraversion ist für die meisten Führungspositionen erforderlich, wobei hier eine zu hohe Ausprägung durchaus hinderlich wirken kann. Personen mit einer unterdurchschnittlichen Extraversion sollten vor allem in Bereichen mit seltenerem Sozialkontakt und höherer Strukturierung der Arbeitsinhalte eingesetzt werden. Ein zu geringes Maß an Extraversion kann dazu führen, dass die betroffene Person nur unzureichend in der Lage ist, eigene Begabungen und Kompetenzen zu vermarkten. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Extraversion ebenfalls hoch ausgeprägt ist (und andersherum)³¹:

- Managementleistung in Berufen mit hoher Autonomie

30 vgl. Rothstein & Goffin (2006)

31 vgl. Rothstein & Goffin (2006)

**Geselligkeit und
Selbstdarstellung fördern
Inspirations- und
Managementfähigkeit**

- Führungskompetenzen
- Leistungsmotivation
- Fähigkeit, andere zu inspirieren

Geistige Flexibilität

Die Skala Geistige Flexibilität (auch als „Offenheit für (neue) Erfahrungen“ bezeichnet) erfasst den Stellenwert, den Personen Ideenreichtum und dem Streben nach mehr Wissen einräumen. Die Offenheit gegenüber Neuem sowie eine positive Einstellung zur Unkonventionalität im Problemlösen werden ebenfalls mit dieser Skala gemessen.

Offenheit für Neues erleichtert kreative Leistungen und den Umgang mit anderen

Geistige Flexibilität spielt vor allem in den Bereichen Forschung und Kunst eine große Rolle. In Berufen mit selten wechselnden Anforderungen und hoher Routine sollten keine Beschäftigten mit sehr hohen Werten auf dieser Skala eingesetzt werden. Die resultierende Fehlbeanspruchung kann sich negativ auf die Leistungsfähigkeit ausüben. Es ist auch davon abzuraten, Personen mit niedriger Geistiger Flexibilität mit Aufträgen zu betrauen, welche Kulturgrenzen überschreiten. Die Anpassung an Unbekanntes kann hier schwerer fallen. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Geistige Flexibilität ebenfalls hoch ausgeprägt ist (und andersherum)³²:

- Soziale Fertigkeiten
- Effektivität bei Führungsaufgaben
- Kreative Leistungen
- Trainingsleistung bei interkulturellen Weiterbildungsprogrammen

Gewissenhaftigkeit

Die Skala Gewissenhaftigkeit erfasst das Ausmaß, in welchem sich Personen Regeln verpflichtet fühlen und sich bestehenden Normen anpassen. Sie misst weiterhin den Grad der Sorgfalt, den Personen bei der Erledigung von Aufgaben walten lassen. Dazu gehört ferner die Einstellung zu Pünktlichkeit und Pflichterfüllung.

Gewissenhaftigkeit fördert das Streben nach Sorgfalt und Zuverlässigkeit

Berufe, in denen kleine Fehler große Auswirkungen haben können, erfordern ein hohes Maß an Gewissenhaftigkeit. Dies betrifft alle Überwachungs-, Wartungs- und Controlling-Tätigkeiten sowie Bereiche, in

³² vgl. Rothstein & Goffin (2006)

welchen mit hohen finanziellen Risiken gearbeitet wird (Bankwesen etc.). Gewissenhaftigkeit spielt darüber hinaus in allen exakten Disziplinen (Forschung, Ingenieurwesen etc.) eine übergeordnete Rolle. Eine sehr niedrige Gewissenhaftigkeit wirkt sich auf die Leistung in nahezu jedem Beruf negativ aus. In einigen Fällen, vor allem, wenn es darum geht, in kurzer Zeit viele Ideen mit geringem Detailgrad zu generieren, kann Gewissenhaftigkeit schwächer ausgeprägt sein. Eine zu hohe Ausprägung auf dieser Skala kann jedoch auch nachteilig sein, wenn sie aufgrund von übersteigertem Perfektionsstreben zu einem verzögerten Abschluss von Aufgaben führt. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Gewissenhaftigkeit ebenfalls hoch ausgeprägt ist (und andersherum)³³:

- Unfallfreiheit
- Leistung in Wartungsaufgaben
- Leistungsmotivation
- Zuverlässigkeit bei Service-Aufgaben mit Kundenkontakt

Leistungsmotivation

Die Skala Leistungsmotivation erfasst das Bestreben, Ergebnisse zu erzielen, die mit Leistungskriterien im Zusammenhang stehen. Der Wert von Feedback durch andere (z.B. Vorgesetzte) als Indikator für die eigene Leistung sowie das Bemühen, die eigenen Fertigkeiten auszubauen und sich mit anderen Personen zu messen, werden ebenfalls erfasst.

Die Leistungsmotivation sollte bei all jenen Personen überdurchschnittlich ausgeprägt sein, bei denen es um eindeutig messbare Ergebnisse in der Arbeit geht. Diese Personen werden mit größter Anstrengung versuchen, die definierten Ziele zu erreichen. Dabei kann es jedoch auch zu Eskalationen mit Kollegen kommen, wenn der Wettbewerbsgedanke über allem steht. Ferner ist zu beachten, dass Personen mit einer sehr hohen Ausprägung regelmäßiges Feedback benötigen und dieses eventuell häufiger einfordern. Eine zumindest durchschnittlich ausgeprägte Leistungsmotivation ist in nahezu jedem Beruf notwendig. Bei sehr geringen Ausprägungen auf dieser Skala kann es dazu kommen, dass die betroffene Person wenig zielorien-

**Leistungsmotivation
fördert die Karriere
und den unternehmerischen Erfolg**

³³ vgl. Rothstein & Goffin (2006)

tiert, konfus oder schlicht zu wenig arbeitet. Jene Personen benötigen in den meisten Fällen Unterstützung durch Vorgesetzte und Kollegen, die sie motivieren und ihr Vorankommen stetig überprüfen. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Leistungsmotivation ebenfalls hoch ausgeprägt ist (und andersherum):

- Unternehmerischer Erfolg³⁴
- Karrierefortschritt³⁵
- Gewissenhaftigkeit³⁶
- Akademische Leistungen³⁷

Risikoneigung

Die Skala Risikoneigung erfasst die Tendenz, Handlungen eher mit hoher Geschwindigkeit und zu Lasten der Genauigkeit auszuüben. Aufgaben werden schnell erledigt und das Verüben von Fehlern wird als weniger wichtig erachtet. Die Skala misst auch die Tendenz, sich über etablierte Regeln und Normen hinwegzusetzen.

Die Risikoneigung sollte in Berufen stärker ausgeprägt sein, in welchen es auf schnelle Entscheidungen ohne Rücksicht auf eventuelle Gefahren ankommt. Eine extrem hohe Ausprägung ist jedoch nur in sehr wenigen Situationen zweckmäßig, da in diesem Fall jegliche Gefahrenabschätzung kaum noch funktioniert und Material und/oder Personen unnötigen Risiken ausgesetzt werden. Auf der anderen Seite sollte dieses Merkmal nicht im unteren Bereich extrem ausgeprägt sein. In jenem Fall würde die betreffende Person mit ihren Aufgaben kaum fertig, da ein sehr starkes Fehlervermeidungsmotiv am Abschluss hindert. Hier bestünde die Gefahr zwanghaften und rigiden Verhaltens. In den meisten Berufsfeldern ist eine mittlere Ausprägung der Risikoneigung, als Spagat zwischen Geschwindigkeit und Genauigkeit in der Aufgabenerledigung, am zweckmäßigsten. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Risikoneigung ebenfalls hoch ausgeprägt ist (und andersherum):

- Unfallrate (z.B. im Straßenverkehr)³⁸

Risikoneigung fördert schnelles, aber auch fehlerintensives Verhalten

34 vgl. Collins et al. (2004)

35 vgl. McClelland (1987)

36 vgl. Rothstein & Goffin (2006)

37 vgl. Busato et al. (2000)

38 vgl. Grey et al. (1989)

- Fluktuation³⁹
- Impulsivität⁴⁰
- Handlungsorientierung⁴¹
- Missachtung von Sicherheitsvorschriften/ Instruktionen⁴²

Teamorientierung

Die Skala Teamorientierung erfasst die Stärke der Anpassung von Personen an definierte soziale Standards und Konfliktvermeidung. Dazu zählt die wahrgenommene Verbindlichkeit von Normen und Gruppenregeln sowie die mehr oder minder stark ausgeprägte Tendenz, sich allgemein anerkannten Standards zu widersetzen.

Ein hohes Maß an Teamorientierung ist überall dort notwendig, wo stark ausdifferenzierte Regelwerke den Arbeitsalltag in vielen Bereichen bestimmen. Dies bezieht sich vor allem auf Regeln im Umgang mit Kollegen, Kunden und Vorgesetzten bzw. die Akzeptanz von Hierarchien. Wenn es eher auf die Leistung des Teams als auf die Leistungen des Einzelnen ankommt, ist ebenfalls eine überdurchschnittliche Ausprägung dieser Eigenschaft von Vorteil. Ein zumindest mittleres Maß an Teamorientierung ist hingegen in fast allen Berufen notwendig. Es gilt jedoch auch: Je geringer der notwendige berufliche Kontakt der Person zu anderen Personengruppen und je niedriger die Abhängigkeit anderer von der Kooperation dieser Person, desto geringer kann die Teamorientierung ausgeprägt sein, ohne negative Folgen nach sich zu ziehen. Folgende berufsrelevante Aspekte sind hoch ausgeprägt, wenn Teamorientierung ebenfalls hoch ausgeprägt ist (und andersherum)⁴³:

- Soziale Fertigkeiten
- Empathie
- Kooperation und Hilfeleistung
- Vermeiden von Druckaufbau gegenüber anderen

**Teamorientierung
fördert Harmonie in
Gruppen und Ko-
operationswillen**

39 vgl. Vardaman et al. (2008)

40 vgl. Dickman & Meyer (1988)

41 vgl. Kuhl (1983)

42 vgl. Nachtwei (2008)

43 vgl. Rothstein & Goffin (2006)

2.3. Persönlichkeitstest – Qualität und Nutzen

Die Qualität eines Tests und sein Nutzen stehen in direktem Zusammenhang. In Abschnitt 1.3 zeigt sich dies in der Formel zur Ermittlung des Netto-Nutzens hinsichtlich des Einsatzes eines Personalauswahlverfahrens. Umso höher die Vorhersagequalität, desto höher der finanzielle Vorteil. Die Vorhersagequalität wird in der Diagnostik prädiktive Validität genannt. Dieses Qualitätsmaß kann in einer Zahl ausgedrückt werden, um zu beziffern, wie gut der Test den Berufserfolg vorhersagen kann. Diese Zahl ist gemeinhin die Korrelation, ein statistisches Maß für die Abhängigkeit zweier Variablen. Jene Variablen sind das Testergebnis (= Prädiktor) und Berufserfolgs-/ Leistungsindikatoren (= Kriterium).

Die Korrelation variiert zwischen -1 und 1. Ein Zusammenhang von -1 würde bedeuten, dass alle Bewerber, die im Test gut abschneiden, sich im Beruf sehr schlecht bewähren. Es besteht also ein umgekehrter Zusammenhang zwischen Prädiktor (Testergebnis) und Kriterium (Berufserfolg). Die 0 hingegen sagt aus, dass keinerlei Zusammenhang besteht. Wohingegen positive Korrelationen ausdrücken, dass mit besserem Testergebnis auch Berufserfolg steigt. Dabei ist jedoch die 1, also eine 100%-ige Validität, nicht zu erreichen, da es zu viele anderweitige Einflüsse (Zufall, Wirtschaftslage, Beziehungen) gibt, welche den Berufserfolg mit beeinflussen. Eignungsdiagnostische Verfahren, auch wenn sie kombiniert werden, haben eine prädiktive Validität von 0,7 noch nicht überschreiten können.⁴⁴

Ein Problem der Verwendung von Korrelationen besteht jedoch darin, dass per se keine Aussage über die Richtung des Zusammenhangs getroffen wird. Die Korrelation an sich sagt also nur aus, dass zwei Variablen zusammenhängen und nicht, welche Variable die andere beeinflusst.⁴⁵ Theoretisch ist es also möglich, dass nicht (nur) die Gewissenhaftigkeit auf die Berufsleistung wirkt, sondern auch andersherum. Denkbar wäre z.B., dass eine hohe Berufsleistung die Gewissenhaftigkeit vermindert, da der Person klar wird, dass sie auch mit ungenau-

Die prädiktive Validität sagt aus, wie gut ein Test den Berufserfolg vorhersagt

Es muss klar sein, dass ein Personenmerkmal den Berufserfolg vorhersagt und nicht andersherum

44 vgl. Schmidt & Hunter (1998)

45 vgl. Bredenkamp & Feger (1983)

erer Arbeit ihre angestrebten Ziele erreicht. Diese Problematik kann jedoch durch sogenannte Längsschnittuntersuchungen behoben werden: Das Persönlichkeits- (oder Intelligenz-)merkmal wird lange vor der Erfassung der Berufsleistung gemessen. Somit wäre es unlogisch anzunehmen, dass eine zu einem späteren Zeitpunkt erfasste Variable (Berufserfolg) auf die frühere Messung (z.B. der Persönlichkeit) wirkt. Viele der im Folgenden berichteten Zusammenhänge beruhen auf einem solchen Messansatz.

Das IQP-Testverfahren[®] orientiert sich, wie bereits diskutiert, am meist etablierten und bewährten Messansatz der Persönlichkeitsforschung durch die „Big Five“.⁴⁶ Dieser wird ergänzt durch die Messung zweier zusätzlicher, in der beruflichen Eignungsdiagnostik unentbehrlicher, Merkmale (Leistungsmotivation und Risikoneigung). Im Folgenden werden detaillierte Ergebnisse zur Vorhersagequalität jener insgesamt sieben Skalen aufgeführt. Auf die Angabe der genauen Höhe der Korrelationen (= Korrelationskoeffizienten) wird verzichtet, da es sich bei den genannten Skalen zwar um denselben theoretischen und messmethodischen Ansatz, jedoch nicht um identische Skalen (da eigene Entwicklung durch IQP) handelt. Eine Übertragbarkeit ist demnach problemlos möglich, aber nicht 1:1 sinnvoll. Ergänzend werden natürlich auch eigene Validierungsstudien durchgeführt (konkrete Ergebnisse dazu finden sich in Kapitel 5).

Die Skalen des IQP-Testverfahrens[®] sagen eine Reihe zentraler berufsrelevanter Aspekte vorher

Emotionale Belastbarkeit

Hohe im Test erfasste Emotionale Belastbarkeit (vgl. Emotionale Stabilität) sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hohe Arbeitsleistung bei Managern (Supervisor-Bewertung)⁴⁷
- hohe Arbeitszufriedenheit⁴⁸ und hohe Karrierezufriedenheit⁴⁹
- seltenere Jobwechsel⁵⁰
- vermindertes Burnout-Risiko (bei hoher Extraversion)⁵¹

46 vgl. Woods & Hampson (2005)

47 vgl. Salgado (1997)

48 vgl. Decker & Borgen (1993)

49 vgl. Seibert & Kraimer (2001)

50 vgl. Boudreau et al. (1998)

51 vgl. Bellani et al. (1996)

- höheres Gehalt (als Indikator für Berufsleistung)⁵²
- hohe Führungsqualitäten (z.B. Effektivität)⁵³
- hohe Güte bei Problemlöseaufgabe im Team⁵⁴

Extraversion

Hohe im Test erfasste Extraversion sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hohe Arbeitsleistung bei Managern (Supervisor-Bewertung)⁵⁵
- mehr Fehltage, wenn dazu geringe Gewissenhaftigkeit⁵⁶
- hohe Karrierezufriedenheit⁵⁷
- hohe Verkaufs-, Management- und Trainingsleistung⁵⁸
- höhere Fluktuation durch von Mitarbeiter initiierte Kündigung⁵⁹
- hohe Führungsqualitäten (z.B. Effektivität)⁶⁰

Geistige Flexibilität

Hohe im Test erfasste Geistige Flexibilität (vgl. Offenheit für neue Erfahrungen) sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- stärkere Suche nach Alternativen zum aktuellen Job⁶¹
- hohe Führungsqualitäten (z.B. Auftreten)⁶²
- hohe allgemeine Intelligenz⁶³

Gewissenhaftigkeit

Hohe im Test erfasste Gewissenhaftigkeit sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hohe Berufsleistung über alle Berufssparten⁶⁴

52 vgl. Gelissen & deGraaf (2006)

53 vgl. Judge et al. (2002)

54 vgl. Kichuk & Wiesner (1997)

55 vgl. Salgado (1997)

56 vgl. Judge et al. (1997)

57 vgl. Seibert & Kraimer (2001)

58 vgl. Barrick et al. (2002)

59 vgl. Boudreau et al. (1998)

60 vgl. Judge et al. (2002)

61 vgl. Boudreau et al. (1998)

62 vgl. Judge et al. (2002)

63 vgl. Gignac et al. (2004)

64 vgl. Barrick et al. (2002)

- hohe Präferenz einer ergebnisorientierten Unternehmenskultur und geringe Präferenz einer innovativen Unternehmenskultur⁶⁵
- wenn gepaart mit hoher Extraversion, dann Präferenz von Managerposition als Karriereziel⁶⁶ und höhere Leistung in Managerpositionen bei hoch autonomer Stelle⁶⁷
- hohe Arbeitsleistung bei Managern (Supervisor-Bewertung)⁶⁸
- hohe Verkaufsleistung (vermittelt über Streben nach Erfüllung von Zielen)⁶⁹
- hohe Führungsqualitäten (z.B. Effektivität)⁷⁰

Leistungsmotivation

Hohe im Test erfasste Leistungsmotivation sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hoher unternehmerischer Erfolg⁷¹
- schneller Karrierefortschritt⁷²
- hohe Gewissenhaftigkeit⁷³
- hohe akademische Leistungen⁷⁴

Risikoneigung

Hohe im Test erfasste Risikoneigung sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hohe Unfallrate (z.B. im Straßenverkehr)⁷⁵
- hohe Impulsivität⁷⁶
- hohe Handlungsorientierung⁷⁷
- höhere Missachtung von (Sicherheits-)Instruktionen⁷⁸

65 vgl. Judge & Cable (1997)

66 vgl. Garden (1997)

67 vgl. Barrick & Mount (1993)

68 vgl. Salgado (1997)

69 vgl. Salgado (1997)

70 vgl. Judge et al. (2002)

71 vgl. Collins et al. (2004)

72 vgl. McClelland (1987)

73 vgl. Rothstein & Goffin (2006)

74 vgl. Busato et al. (2000)

75 vgl. Grey et al. (1989)

76 vgl. Dickman & Meyer (1988)

77 vgl. Kuhl (1983)

78 vgl. Nachtwei (2008)

Teamorientierung

Hohe im Test erfasste Teamorientierung (vgl. Verträglichkeit) sagt folgende berufsrelevante Aspekte vorher (positive Korrelationskoeffizienten):

- hohe Empathie und Hilfsmotivation⁷⁹
- hohe Arbeitsleistung bei Managern (Supervisor-Bewertung)⁸⁰
- hohe Leistung in Jobs, die Teamarbeit erfordern⁸¹
- höhere Chance von Selbstaufgabe als Reaktion auf Stress⁸²
- hohe Güte bei Problemlöseaufgabe im Team⁸³

79 vgl. Costa & McCrae (1992)

80 vgl. Salgado (1997)

81 vgl. Mount et al. (1998)

82 vgl. Costa, Zonderman & McCrae (1991)

83 vgl. Kichuk & Wiesner (1997)

3. Der Intelligenztest

Kille, K. & Nachtwei, J.

3.1. Intelligenz – Rolle in der Personalauswahl

Über eine Definition zur Intelligenz wird seit langem in der Fachwelt gestritten. Eine ausführliche Diskussion macht an dieser Stelle wenig Sinn. Die Darstellung dessen, was bisher als Konsens angenommen wird, scheint zielführender. Demnach ist Intelligenz nicht nur Wissen aus Büchern, sondern die breite Fähigkeit, seine Umwelt und deren Komponenten in einen Sinnzusammenhang zu bringen.⁸⁴ Auch werden gemeinhin Problemlösen und Lernfähigkeit (oder allgemeiner die Fähigkeit, sich anzupassen) als Kern von Intelligenz definiert.⁸⁵ Intelligenz hat dabei noch weitere Eigenschaften: Die Intelligenzleistung stabilisiert sich während der Kindheit und verändert sich im Erwachsenenalter nicht mehr bedeutsam.⁸⁶ Intelligenz wird zu 40-80% vererbt; der Rest sind Umwelteinflüsse (Erziehung, Schulbildung).⁸⁷ Andere Autoren gehen von jeweils ca. 50% genetischem und umweltbedingtem Einfluss aus.⁸⁸

Die im Kontext Potentialanalyse wichtigste Aussage zur Intelligenz lautet jedoch: Die allgemeine kognitive Leistungsfähigkeit sagt die Leistung in allen Berufssparten vorher.⁸⁹ Hierbei stellt sich allerdings die Frage, was genau mit „allgemeiner kognitiver Leistungsfähigkeit“ gemeint ist. Die erste diesbezügliche systematische Betrachtung ist bereits mehr als 100 Jahre alt. Spearman unterschied 1904 einen g-Faktor und s-Faktoren der Intelligenz. Der g-Faktor wird dabei definiert als allgemeine Intelligenzleistung, die s-Faktoren gelten als spezielle Intelligenzleistungen (wie z.B. der Umgang mit Zahlen oder Worten).⁹⁰ Eine weitere Unterscheidung nimmt Cattell 1963 vor: fluide vs. kristalline Intelligenz. Demnach ist fluide Intelligenz all dies, wo Vorwissen und Vorerfahrungen keinen Einfluss haben. Kristalline Intelligenz hingegen wird beeinflusst von Bildung, Vorerfahrungen und

Eine Definition von Intelligenz ist schwierig, aber möglich

Verschiedene theoretische Modelle zur Intelligenz wurden von Carroll 1993 integriert

84 vgl. Gottfredson (1997)

85 vgl. Carlstedt et al. (2000)

86 vgl. Gottfredson (1997)

87 vgl. Gottfredson (1997)

88 vgl. Chipuer et al. (1990)

89 vgl. Hunter (1986)

90 vgl. Frazier & Youngstrom (2007)

kulturellen Einflüssen. Hier spielt die fluide Intelligenz als Basis jedoch auch mit hinein.⁹¹ Jene beiden Sichtweisen auf Intelligenz von Spearman und Cattell wurden 1993 von Carroll in einem dreistufigen hierarchischen Modell der kognitiven Fähigkeiten integriert.⁹² Ganz oben in dieser Hierarchie steht der g-Faktor, also die allgemeine Intelligenzleistung. Darunter werden breite Intelligenzleistungen unterschieden, zu welchen auch fluide und kristalline Intelligenz gehören. Auf der untersten Abstraktionsstufe finden sich spezifische Intelligenzleistungen wie z.B. numerische und verbale Intelligenz.

Für die Personalauswahl macht diese Unterscheidung insofern Sinn, als dass zwar die allgemeine Intelligenz (g-Faktor) in jeder Berufssparte den Erfolg eines Bewerbers vorhersagt. Da jede Stelle jedoch speziellere Anforderungen als beispielsweise einfach nur „hohe allgemeine Intelligenz“ an den Bewerber stellt, sollte zumindest zwischen verbaler und numerischer Intelligenz unterschieden werden. Diese für die meisten Berufe relevante Differenzierung wird auch im IQP-Testverfahren[®] vorgenommen. Hinzu kommt, dass z.B. kristalline Intelligenz über den Einfluss von (Vor-)Wissen definiert wird. Um diese Form der Intelligenz genau zu messen, müsste für jede Stelle genau definiert werden, welches Wissen zur Bewältigung der Anforderungen nötig ist. Abgesehen vom immensen Aufwand ist eine vollständige Erfassung des gesamten Wissens nicht möglich. Darüber hinaus ist Wissen immer vermittelbar, während fluide Intelligenz oder Abstraktes Problemlösen kaum erlernbar sind. Die hauptsächliche Konzentration auf vorwissensunabhängige (fluide) Intelligenzleistungen ist demnach viel versprechender und somit die überwiegende Grundlage des IQP-Testverfahrens[®].

„Allgemeine Intelligenz“ als Anforderung an eine Stelle reicht nicht aus

3.2. Intelligenz – Zentrale Bewerbermerkmale

Entsprechend der Diskussion im obigen Abschnitt werden drei übergeordnete Bewerbermerkmale hinsichtlich der Intelligenz unterschieden: Abstrakt-Analytische, Numerische und Verbale Intelligenz.

Zentrale Intelligenzmerkmale: Abstrakt-Analytische, Numerische und Verbale Intelligenz

91 vgl. Johnson & Bouchard (2005)

92 vgl. te Nijenhuis et al. (2007)

Zentral ist hierbei, dass diese Merkmale durch zeitbegrenzte Aufgaben erfasst werden, um die kognitive Effizienz der Bewerber zu erfassen. Es ist demnach nicht allein interessant, ob jemand eine Aufgabe lösen kann, sondern ob dies auch in angemessener Zeit gelingt (ein wichtiges Kriterium in der Praxis). Im Folgenden wird jedes Merkmal entsprechend dem wissenschaftlichen Konsens und der Verwendung als Skala im IQP-Testverfahren® definiert.

Abstrakt-Analytische Intelligenz

Die Skala Abstrakt-Analytische Intelligenz erfasst die Fähigkeit, Probleme mit komplexer und nicht auf Anhieb transparenter Struktur schnell zu analysieren und richtig zu lösen. Dazu gehört das in Beziehung setzen sowie das flexible Umstrukturieren abstrakter Informationen, ohne das ein Rückgriff auf Vorwissen möglich ist.

In einigen Berufen spielt Abstrakt-Analytische Intelligenz eine untergeordnete Rolle. Dies trifft vor allem auf jene Berufe zu, bei denen stets dieselbe Routine zu bewältigen ist und kaum neue Anforderungen an den Beschäftigten gestellt werden. Sind die Aufgaben eher einfach und überschaubar, so kann eine hohe Abstrakt-Analytische Intelligenz sogar kontraproduktiv wirken, da hier mit akuter Unterforderung und negativen Folgen für die Leistungsfähigkeit und Arbeitszufriedenheit zu rechnen ist. In diesen Fällen sollte also von einer zu hohen Definition der Anforderung an die Abstrakt-Analytische Intelligenz Abstand genommen werden.

Es gilt die Faustregel: Je komplexer die gestellten Aufgaben sowie je häufiger und schneller die Problemstellungen wechseln, desto höher sollte die Abstrakt-Analytische Intelligenz ausgeprägt sein. Sind keine Standardvorgehensweisen zur Lösung von Problemen vorhanden und/ oder anwendbar, so ist eine überdurchschnittliche Abstrakt-Analytische Intelligenz unbedingt notwendig. Dies trifft vor allem auf Tätigkeiten mit Projektcharakter und Forschung zu. In den meisten Fällen bedeutet eine geringe Abstrakt-Analytische Intelligenz nicht, dass das Problem unlösbar ist. Es ist jedoch mit einem höheren zeitlichen Aufwand für die Lösung und eventuell mit Einbußen in der Lösungsgüte zu rechnen. Abstrakt-Analytische Intelligenz lässt sich nur schwer trainieren.

Abstrakt-Analytische Intelligenz fördert die Bewältigung komplexer und schnell wechselnder Problemstellungen

Numerische Intelligenz

Die Skala Numerische Intelligenz erfasst das Beherrschen grundlegender Rechenoperationen sowie die Fähigkeit, Regelmäßigkeiten in Zahlenmaterial zu entdecken. Erstgenannte Fähigkeit ist vorwiegend das Ergebnis von (Schul-)Bildung während das Ableiten von Regeln in höherem Maße begabungsbedingt und vorwissensunabhängig ist.

Ein Mindestmaß an Numerischer Intelligenz ist in nahezu jeder Berufsgruppe für eine erfolgreiche Lösung der gestellten Anforderungen notwendig. Tätigkeiten mit kaufmännischen Komponenten oder dem Anspruch, große Datenmengen zu organisieren und zu bewerten, sind nur Personen mit einer mindestens durchschnittlichen Numerischen Intelligenz zu empfehlen.

Für ingenieurbezogene Berufe oder Tätigkeiten mit mathematisch-physikalischen Problemstellungen sollte die Numerische Intelligenz überdurchschnittlich ausgeprägt sein. Es kann nachteilig sein, eine zu hohe Numerische Intelligenz als Anforderung für einen Beruf anzugeben, der diese im Grunde nicht benötigt. Dies kann dazu führen, dass die entsprechend ausgewählten Personen in diesem Beruf unterfordert sind, was Leistungsfähigkeit und Arbeitszufriedenheit in Mitleidenschaft zieht. Eine unterdurchschnittliche Numerische Intelligenz ist jedoch kein alleiniges Ausschlusskriterium für Berufe, bei denen der Umgang mit Zahlen ohnehin eine untergeordnete Rolle spielt. Denn diesen geringeren Anforderungen kann ein Bewerber durch Training und Weiterbildung durchaus gerecht werden.

Verbale Intelligenz

Die Skala Verbale Intelligenz erfasst die Fähigkeit, die innere Struktur von verbalen Informationen zu analysieren und folgerichtig auf gegebene Probleme anzuwenden. So wird zum einen die eher wissensbedingte richtige Ergänzung von Sätzen sowie die vorwiegend durch Begabung geleitete Fähigkeit, verbale Analogien zu finden, geprüft.

Gewandtheit im Umgang mit Sprache ist in den meisten Berufssparten ein notwendiges Rüstzeug. Für einige Tätigkeiten, wie z.B. handwerkliche, sonstige eher auf körperliche Leistungsfähigkeit orientierte Tätigkeiten sowie Berufe, in denen es fast ausschließlich um mathematisch-logische Befähigung geht, spielt verbale Intelligenz nur

Numerische Intelligenz fördert den flexiblen Umgang mit Zahlen und der logischen Verbindung dieser

Verbale Intelligenz fördert die Gewandtheit im Umgang mit Sprache

eine untergeordnete Rolle. Die Anforderungen an diese Eigenschaft sollten deshalb, sofern es sich um einen solchen Beruf mit geringen Anforderungen an Sprachgefühl und Ausdrucksfähigkeit handelt, nicht zu hoch gesetzt werden. Die Gefahren sind Unterforderung sowie Schwierigkeiten bei der Zusammenarbeit mit anderen Vertretern dieser Berufsgruppen. Ein zumindest durchschnittliches Maß an verbaler Intelligenz ist hingegen in jedem Beruf nötig, bei dem es darum geht, sich mit anderen abzustimmen und Vorschläge sowie Arbeitsergebnisse zu präsentieren. Jede Tätigkeit mit direktem Kundenkontakt oder Sprache in Wort und Schrift als gängigem Arbeitsmittel erfordert eine überdurchschnittliche Verbale Intelligenz.

3.3. Intelligenztest – Qualität und Nutzen

In den vorhergehenden Kapiteln (Abschnitt 1.3 und 2.3) wurde bereits dargestellt, dass die Vorhersagequalität (prädiktive Validität) eines Tests bzw. einer Skala direkt mit dem finanziellen Nutzen hinsichtlich des Testeinsatzes zusammenhängt. Da alle verwendeten Intelligenzskalen in nahezu jedem Berufszweig eine Vielzahl von Indikatoren des Berufserfolgs vorhersagen, macht eine Detailaufschlüsselung der Vorhersagequalität nach den drei verwendeten Skalen wenig Sinn. Es folgt eine allgemeinere Darstellung:

Aufgaben zur Numerischen und Verbalen Intelligenz sowie zu abstraktem Problemlösen (vgl. Abstrakt-Analytische Intelligenz) gelten als die zuverlässigsten Prädiktoren der Leistung in einer Vielzahl von Berufen.⁹³ Allgemeine Intelligenz (vgl. g-Faktor) kann als eine Verrechnung spezifischer Skalen (wie z.B. Verbale und Numerische Intelligenz) betrachtet werden.⁹⁴ Andererseits kann sie auch mit bestimmten Aufgaben direkt gemessen werden. Diese Aufgaben werden als Matrizenaufgaben bezeichnet und gehen ursprünglich auf Raven zurück.⁹⁵ Unter anderem wird die Güte der Problemlösung bei Teamaufgaben von dieser allgemeinen Intelligenz beeinflusst.⁹⁶ Im Bereich

Vorhersagequalität eines Tests und finanzieller Nutzen des Testeinsatzes sind eng verbunden

93 vgl. Sternberg (1997)

94 vgl. Colom et al. (2000)

95 vgl. Carpenter et al. (1990)

96 vgl. Kichuk & Wiesner (1997)

der beruflichen Weiterbildung sagt die allgemeine Intelligenz die Leistung der Teilnehmer im Training vorher.⁹⁷ Allgemein gilt: Je höher die intellektuellen Anforderungen der Stelle und je weniger strukturiert der Arbeitsbereich, desto höher der Vorteil einer hohen Intelligenz.⁹⁸ Kurzum: Intelligenz bestimmt direkt, wie korrekt und schnell eine Person ihre Aufgaben lösen kann. Die Art und Weise der Aufgabenlösung (im Sinne von Ausdauer, Genauigkeit, Einbezug von Kollegen etc.) dagegen hängt eher von der Persönlichkeit der Person ab.

Aktuelle Theorien gehen außerdem von einer Wechselwirkung von Persönlichkeit und Intelligenz aus: Demnach beeinflusst die Persönlichkeit einerseits, wie und wo Personen ihre intellektuellen Fähigkeiten einsetzen. Andererseits bietet Intelligenz die intellektuelle Basis für die Entwicklung von persönlichkeitsbezogenen Merkmalen wie Interessen, Vorlieben, Einstellungen und Orientierungen zu bestimmten Aktivitäten.⁹⁹

**Wechselwirkung
von Persönlichkeit
und Intelligenz**

97 vgl. te Nijenhuis et al. (2007)

98 vgl. Gottfredson (1997)

99 vgl. Demetriou et al. (2003)

4. Der Weg: Testkonstruktion

Uedelhoven, S. & Nachtwei, J.

4.1. Einleitung – Strategien und Methoden

Die Konstruktion eines Tests im Allgemeinen bzw. einer Skala im Speziellen bedarf einer ausgefeilten Methodik sowie Kenntnis statistischer Verfahren. Jede Testentwicklung zeichnet sich dabei durch einen mehrstufigen Prozess aus theoretischen Überlegungen, Datenerhebungen und statistischen Berechnungen aus. In diesem Kapitel wird versucht, dieses Vorgehen möglichst allgemeinverständlich (siehe dazu auch die Definitionen im Glossar) und sparsam zu beschreiben.

Zunächst kann zwischen drei sich im Idealfall ergänzenden Strategien der Testkonstruktion unterschieden werden¹⁰⁰:

- rational/ theoriegeleitet: Konstruktion anhand des Forschungsstandes der psychologischen Diagnostik
- induktiv: Konstruktion durch statistische Auswahl der Testaufgaben
- external: Konstruktion durch Vergleich der Testergebnisse anhand verschiedener Gruppen (z.B. nach Schulabschluss)

Bei der Entwicklung des IQP-Testverfahrens[®] wurden je Skala alle drei Konstruktionsstrategien verwendet: Nach einer anfänglichen Sammlung der Erkenntnisse aus der wissenschaftlichen Literatur wurden Aufgaben (sogenannte Items) entwickelt. Im Anschluss wurden diese Items hinsichtlich ihrer Binnenstruktur durch Faktorenanalysen, Itemschwierigkeits- sowie Reliabilitäts- und Trennschärfeanalysen an einer Selektionsstichprobe geprüft. Im Ergebnis dieser Berechnungen wurden einige Items entsprechend ihrer statistischen Kennwerte ausgewählt. Im letzten Schritt fand anhand von Varianz- und Korrelationsanalysen eine Überprüfung durch den Vergleich der Testergebnisse für verschiedene Untergruppen statt. Betrachtet man das Vorgehen bei der Konstruktion etwas genauer, so ergeben sich Unterschiede zwischen der Entwicklung von Persönlichkeits- und Intelligenzskalen. Aus diesem Grund werden die Darstellungen in zwei getrennten Abschnitten behandelt.

Es gibt viele Wege einen Test zu konstruieren; ihre Kombination ist sinnvoll

Verwendung aller drei Konstruktionsstrategien bei der Entwicklung des IQP-Testverfahrens[®]

100 vgl. Amelang & Schmidt-Atzert (2006)

Auf eine genaue Darstellung aller Feinheiten der statistischen Analysen wird aus Gründen der Übersichtlichkeit und Ökonomie verzichtet. Die Kernpunkte werden genannt, um die Vorgehensweise (v.a. für Experten auf diesem Gebiet) transparent zu machen, jedoch nicht weiter ausgeführt. Bei Interesse verweisen wir gern auf die einschlägige Literatur zu diesen Themen.¹⁰¹

4.2. Konstruktion der Persönlichkeitsskalen

Grundlage der Konstruktion jeder der insgesamt sieben Persönlichkeitsskalen bildete eine große Menge an aus theoretischen Überlegungen und Forschungsbefunden abgeleiteten Items. Von diesen anfänglich 137 Items für alle Skalen verblieben am Ende des Analyse- und Selektionsprozesses 83 Items in der Endversion des Tests. Dies entspricht einer fast 40%-igen Ausdünnung der Itemmenge und zeigt, dass nur mit Hilfe statistischer Analysen fundierte Aussagen über die Qualität einzelner Items getroffen werden können. Im Einzelnen wurde bei der Auswahl der Items für die Endversion des Tests wie folgt vorgegangen:

Mit Hilfe von Faktorenanalysen wurde geprüft, ob sich die einzelnen Skalen in Unterskalen ausdifferenzieren. Beispielsweise kann Leistungsmotivation zum einen aus einer Hoffnung auf Erfolg und zum anderen aus einer Angst vor Misserfolg bestehen. Die Faktorenanalyse kann dazu beitragen, derartige Differenzierungen zu erkennen. Dabei wird zunächst eine explorative Faktorenanalyse berechnet. Diese zeigt, wie viele Faktoren (oder Unterskalen) bei einer Skala unterschieden werden können. Drei Kriterien werden benutzt, um über die Faktorenzahl zu entscheiden (Eigenwert, Screeplot und Faktorladung). Im Anschluss wurde an einer weiteren Stichprobe eine konfirmatorische Faktorenanalyse durchgeführt. Hierbei wurden die zuvor gefundenen Faktoren angegeben und rechtwinklig rotiert (Varimax-Rotation), um die Binnenstruktur der jeweiligen Skala aufzudecken. Die Rotationsmethode (Annahme: Faktoren sind unabhängig voneinander) wurde gewählt, da möglichst wenig Redundanz zwischen den Faktoren einer

Eine Vielzahl statistischer Berechnungen hilft bei der Beurteilung, welche Items in der Skala verbleiben

Die Faktorenanalyse als Methode zur Vermeidung von Redundanzen

101 vgl. Field (2005)

Skala entstehen sollte.

In einem weiteren Schritt erfolgte die Berechnung der Internen Konsistenz. Dieses Maß sagt aus, wie homogen eine Skala ist, d.h. ob alle Items sinngemäß dasselbe erfassen. Berechnet wird dabei die mittlere Korrelation aller Items einer Skala miteinander. Dieses Homogenitätsmaß trifft eine Aussage über die Reliabilität (Zuverlässigkeit) einer Skala. Je höher die Reliabilität, desto weniger fehlerbehaftet ist die Messung und somit umso höher die Wahrscheinlichkeit, relevante Kriterien, wie den Berufserfolg, vorherzusagen. Items, welche zu einem starken Abfall der Internen Konsistenz führten, wurden ausselektiert. Dabei ist zu beachten, dass eine zu hohe Homogenität zu einem Abfall der Vorhersagequalität (prädiktive Validität) der Skala führt. Das liegt daran, dass eine zu hohe Homogenität (Interne Konsistenz) die Vielfalt der Merkmalsmessung einschränkt (alle Items messen mehr oder minder genau dasselbe), das Kriterium Berufserfolg im Gegensatz dazu jedoch sehr komplex und heterogen ist.

Im Anschluss wurde eine Trennschärfeanalyse durchgeführt. Die Trennschärfe ist die Korrelation eines Items mit dem Skalenwert (gebildet aus den restlichen Items). Sie sagt aus, wie gut ein Item in der Lage ist, zwischen Personen mit einer hohen vs. niedrigen Merkmalsausprägung zu trennen. Da es bei der beruflichen Eignungsdiagnostik letztlich um eine Unterscheidung von Personen geht, ist dieses Maß sehr zentral. Items, deren Trennschärfeindex unter einem kritischen Wert lagen, wurden ausselektiert.

Zusätzlich zur Analyse der Trennschärfe wurde auch die Schwierigkeit der Items geprüft. Zwischen beiden statistischen Kennwerten besteht ein Zusammenhang derart, dass vor allem mittelschwere Items hoch trennscharf sind. Denn beispielsweise könnte nahezu jeder ein extrem leichtes Item beantworten, wodurch eine Trennung zwischen Personen mit hoher vs. niedriger Ausprägung für dieses Item nicht möglich ist. Schwierigkeit wird dabei definiert als die Stärke des Merkmals und gemessen über den Anteil derjenigen Personen, die das Item bejahen. Je weniger Personen es bejahen, desto schwieriger ist es. Der Anspruch besteht hierbei darin, viele mittelschwere Items und einige wenige sehr leichte und sehr schwere Items in den Test aufzunehmen. Dies erlaubt eine Differenzierung von Personen auf der

Berechnung der Internen Konsistenz um die Zuverlässigkeit der Skalen zu bestimmen

Die Analyse von Trennschärfe und Schwierigkeit der Items als statistisches Verfahren zur Optimierung der Vorhersagequalität des Eignungstests

gesamten Bandbreite der Skala.

Flankierend zu den genannten statistischen Berechnungen wurden umfangreiche deskriptive Analysen (Spannweite, Standardabweichungen), Ausreißerbetrachtungen (vgl. Häufigkeiten in z-Verteilung und Boxplots) und Verteilungsprüfungen (vgl. Kolmogorov-Smirnoff-Test auf Normalverteilung) durchgeführt. Diese Analysen vermitteln einen zusätzlichen Eindruck zur Beschaffenheit der Skala.

Je besser eine Skala konstruiert ist, desto höher ist auch ihre letztendliche Güte. Diese Betrachtung wird in Bezug auf die Persönlichkeitsskalen des IQP-Testverfahrens[©] in Abschnitt 5.2 dargestellt.

4.3. Konstruktion der Intelligenzskalen

Da sich die Intelligenzskalen strukturell mehr voneinander unterscheiden als dies bei den Persönlichkeitsskalen der Fall ist, wird im Folgenden für jede Skala einzeln auf die Konstruktionsprinzipien eingegangen. Dabei ist das statistische Vorgehen zwar dasselbe, jedoch unterscheidet sich die inhaltliche Herangehensweise von Skala zu Skala. Für alle Skalen gilt hinsichtlich Durchführung und statistischer Auswertung:

Die Aufgaben sind ohne Hilfsmittel oder Nebenrechnungen und unter Zeitbegrenzung zu lösen. Dabei ist jedes Item zeitlich begrenzt und kann nur einmal bearbeitet werden (kein Zurückblättern), um kognitive Effizienz messen zu können und um das Verwenden unterschiedlicher Lösungsstrategien zu vermeiden (z.B. ein Item sehr lange bearbeiten vs. zunächst nach lösbaren Items suchen und diese zuerst lösen). Die Hauptanforderung bestand darin, jene Aufgaben auszuwählen, welche zu leicht (alle Personen lösen die Aufgabe) oder zu schwer (keine Person löst die Aufgabe) waren. Die übrigen Items wurden im Sinne einer sogenannten Guttman-Skala nach aufsteigender Schwierigkeit sortiert. Der Bewerber beginnt demnach mit leichten Aufgaben und arbeitet sich bis zu sehr schweren voran.

Da jede Intelligenzskala bzw. ihre Unterskalen vom Aufgabentyp sehr homogen und zudem relativ kurz sind, kann auf Datenreduktionsverfahren wie die Faktorenanalyse verzichtet werden. Jedoch sind

Jeder Intelligenzskala liegt inhaltlich ein anderes Konstruktionsprinzip zugrunde

Zuverlässigkeitsanalysen der Skalen sinnvoll. Diese wurden durch die sogenannte Split-Half Reliabilität berechnet. Dabei wird die Skala in zwei Hälften geteilt und die Items beider Hälften korreliert. Anschließend wird per Spearman-Brown-Formel auf die Reliabilität des Gesamttests hochgerechnet. Items, die zu einem bedeutsamen Abfall der Reliabilität führten, wurden ausselektiert.

Skala Abstrakt-Analytische Intelligenz

Diese Skala arbeitet mit geometrischen Formen, um möglichst inhalts- und somit bildungsunabhängig zu testen. Diese geometrischen Formen stehen in einer bestimmten Beziehung zueinander, wobei eine Form fehlt, um das logische Gesamtbild zu komplettieren. Es gilt, diese Beziehung bzw. Regel zu erkennen und eine geometrische Form aus einem Angebot an Formen zu wählen, um sie zu vervollständigen. Die Konstruktion der einzelnen Aufgaben erfordert die Beachtung der Komplexität der geometrischen Formen (Anzahl und Art der verwendeten Grafikelemente) sowie deren Verknüpfung nach bestimmten Regeln (Addition, Subtraktion und Multiplikation im Sinne von Überdeckung von Grafikelementen). Je nachdem, wie komplex die Grafikelemente sind und wie hoch die Anzahl der Verknüpfungsregeln ist, lässt sich die Schwierigkeit der Aufgabe beeinflussen.

Daneben spielt die Art der falschen Antwortalternativen (Nähe zur richtigen Lösung, Art der Abwandlung z.B. im Sinne von Rotation) für die Schwierigkeit eine Rolle. Die Leistung bei jenen Aufgaben wird zum einen durch die Leistung des Arbeitsgedächtnisses¹⁰² (Planung, Koordination und Überwachung von Zielen und Unterzielen) und zum anderen durch die Wahrnehmungsfähigkeit bestimmt.¹⁰³ Letzteres steht jedoch eigentlich nicht im Fokus der Messung. Der Einfluss der Wahrnehmungsfähigkeit auf die Aufgabenlösung lässt sich durch Anheben der Aufgabenschwierigkeit minimieren, was bei dieser Skala entsprechend umgesetzt wurde.

Abstrakt-Analytische Intelligenz misst unabhängig vom Vorwissen

Die Gestaltung der Antwortalternativen wirkt sich auf die Schwierigkeit der Aufgabe aus

102 vgl. Buehner et al. (2005)

103 vgl. Mackintosh & Bennett (2005)

Skala Numerische Intelligenz

Diese Skala ist in zwei Unterskalen unterteilt: Rechenfähigkeit und Zahlenreihen. Bei der Unterskala Rechenfähigkeit wurde inhaltlich im Einklang mit der Fachliteratur darauf geachtet, dass zentrale Rechenoperationen und -fertigkeiten mit Hilfe der Items möglichst repräsentativ abgebildet wurden. Dies sind Bruchrechnung, Klammerregeln, Dreisatz, Exponential- und Wurzelrechnung sowie Sachaufgaben.

Bei der Unterskala Zahlenreihen geht es darum, aus einer Reihe von Zahlen eine Regel zu erkennen, um diese Reihe sinnvoll fortsetzen zu können. Die Regeln, nach welchen die Zahlenreihen zusammengesetzt sind, können sehr unterschiedlich und komplex sein. Jede Zahl ist dabei mit einer oder mehreren anderen Zahlen durch eine Rechenregel (Multiplikation, Division, Addition, Subtraktion) verbunden. Diese Beziehung kann entweder zum direkten Nachbarn oder weiter entfernten Zahlen bestehen. Darüber hinaus können eine oder mehrere Rechenregeln innerhalb einer Skala verwendet werden. Je mehr Rechenregeln es gibt und je weiter die aufeinander bezogenen Zahlen voneinander entfernt sind, desto schwieriger ist es, die Regel zu erkennen und entsprechend logisch fortzusetzen.

Zwei Unterskalen der Numerischen Intelligenz: Rechenfähigkeit und Zahlenreihen

Skala Verbale Intelligenz

Auch diese Skala ist in zwei Unterskalen unterteilt: Analogien und Satzergänzungen. Bei der Unterskala Analogien wird die Fähigkeit erfasst, Beziehungen zwischen Worten zu erkennen und auf andere Kontexte zu übertragen. Dazu werden drei Worte vorgegeben, wovon die ersten beiden in einer bestimmten Beziehung zueinander stehen. Zum dritten Wort muss ein viertes Wort gefunden werden, welches eine ähnliche Beziehung zwischen diesen zwei Worten (3 und 4) bildet wie sie zwischen Wort 1 und 2 besteht. Dabei kann aus einer Anzahl von Antwortalternativen ausgewählt werden, von denen nur eine die richtige Lösung darstellt. Bei der Konstruktion ist darauf zu achten, dass verschiedene Arten von Beziehungen (Gegensätze, Zusammenhänge, zeitliche Abhängigkeiten etc.) sowie verschiedene Kontexte (Alltag, Natur, Technik etc.) in den einzelnen Aufgaben verwendet werden. Die Schwierigkeit einer Aufgabe definiert sich dabei auch über die Beschaffenheit der falschen Antwortalternativen (sogenannte Dis-

Zwei Unterskalen der Verbalen Intelligenz: Analogien und Satzergänzungen

traktoren). Je näher diese Alternativen der richtigen Lösung scheinen, desto höher die Anforderung an die Differenzierungsfähigkeit der Person.

Bei der Unterskala Satzergänzungen wird die Fähigkeit abgeprüft, Sätze korrekt zu vervollständigen. Sowohl ein großer Wortschatz als auch logisches Denkvermögen sind dabei wichtig. Die Schwierigkeit entsteht durch die Verwendung von Randbedingungen (z.B. immer und manchmal) sowie das Erkennen logischer Abhängigkeiten (z.B.: Welche Bedingung ist notwendig, um ein bestimmtes Ziel zu erreichen und welche ist nur hinreichend?). Auch hier werden Antwortalternativen vorgegeben, von denen nur eine richtig ist. Je näher die falschen Alternativen an der richtigen, desto schwieriger die Aufgabe.

4.4. Messfehler und Verfälschungstendenzen – Gegenmaßnahmen

Ein Messfehler ist jede Ungenauigkeit einer Messung, welche dazu führt, dass ein gemessener Wert (das Testergebnis zum Merkmal X) vom sogenannten wahren Wert (das tatsächliche Merkmal X) abweicht. Den wahren Wert gibt es nur theoretisch, da er praktisch nie erfassbar ist – es wird immer irgendeinen (und sei er noch so klein) Messfehler geben, welcher das 100%-ig genaue Auffinden des wahren Wertes unmöglich macht. Die Herausforderung der psychologischen Eignungsdiagnostik besteht nun also darin, den Messfehler klein zu halten und sich damit dem wahren Wert möglichst stark anzunähern.¹⁰⁴ Einige Forschungsbefunde zu diesem Thema beziehen sich auf Fragebögen in Papierform, sind jedoch auf computerbasiertes Testen übertragbar.¹⁰⁵

Im Folgenden werden vier mögliche Messfehler und die Maßnahmen zu ihrer Minimierung beschrieben.

**Fehlerfreies Messen
ist eine Illusion – eine
Annäherung an das
Optimum ist jedoch
immer möglich**

104 vgl. Bortz (1999)

105 vgl. Booth-Kewley et al. (2007)

*Random Response Error (RRE)*¹⁰⁶

Dieser Fehler entsteht durch Ablenkungen oder Schwankungen in der Aufmerksamkeit. Der Fehler tritt zufällig während eines Testlaufes auf und kann zu anderen Zeitpunkten völlig verschieden ausgeprägt sein. Dieser Messfehler ist durch kontrollierte Testbedingungen im Labor minimierbar. Ablenkungen kann dort wirksam durch den Testleiter entgegengewirkt werden (was bei einer Internet-Testung am heimischen PC nicht möglich wäre).

Kontrollierte Testbedingungen reduzieren Fehler durch Schwankungen in der Aufmerksamkeit

*Specific Factor Error (SFE)*¹⁰⁷

Dieser Fehler tritt nur bei bestimmten Testaufgaben auf. Er ist unabhängig vom Testzeitpunkt oder der getesteten Person. Die Ursache liegt bei der Art der Aufgabe (z.B. Fragen, die uneindeutig gestellt sind). Mit Hilfe einer Homogenitätsanalyse (wie oben beschrieben) können jene Items ausgesondert werden.

Saubere Testkonstruktion verhindert fehlerhafte Testaufgaben

*Transient Error (TE)*¹⁰⁸

Dieser Fehler tritt in einem Testlauf aufgrund spezifischer emotionaler und mentaler Zustände (Ängste, Stress) der Testperson auf. Der Fehler beeinflusst den gesamten Testlauf, kann jedoch zu einem späteren Zeitpunkt völlig anders geartet (kleiner vs. höher) sein. Analysen der Skalenzuverlässigkeit (Reliabilität) beschränken sich meist nur auf einen Messzeitpunkt und sind deshalb durch diesen Fehler (der über verschiedene Messzeitpunkte variiert) verunreinigt. Die wiederholte Messung (Test-Retest-Reliabilität) würde Abhilfe schaffen, ist jedoch im Bereich der Eignungsdiagnostik nicht immer zweckmäßig. Einzig ein geübter Testleiter (welcher Ängste oder auch krankheitsbedingte Unaufmerksamkeit erkennt und evtl. den Test verschiebt) und das Testen unter kontrollierten Bedingungen im Labor können diesem Messfehler entgegenwirken.

Geschulte Testleiter und kontrollierte Testbedingungen reduzieren Ängste und Stress bei der Testperson

*Soziale Erwünschtheit (SE)*¹⁰⁹

Dieser Fehler beschreibt die Tendenz einer Person, sich in einer

Ein innovativer Ansatz misst bewusste Verfälschungstendenzen

106 vgl. Schmidt & Hunter (1999)

107 vgl. Hunter & Schmidt (2004)

108 vgl. Heggstad et al. (2006)

109 vgl. Hough (1998)

Testsituation bewusst besser darzustellen, als sie tatsächlich ist. Möglich ist dies naturgemäß nur bei Persönlichkeitsmerkmalen (denn im Intelligenztest kann man die richtige Lösung entweder wissen oder nur raten). In der Praxis wird diese Tendenz einer Person als eindimensional betrachtet und nicht weiter ausdifferenziert.¹¹⁰ Während einige Forscher dieser Neigung eine negative Betonung (im Sinne von Lügen, um eingestellt zu werden) geben, sehen andere diese als die Fähigkeit, Anforderungen einer Stelle richtig vorherzusehen/ zu deuten.¹¹¹ In beiden Fällen stellt die Soziale Erwünschtheit jedoch eine Verfälschung der Testdaten und somit Abweichung vom „wahren Wert“ dar.¹¹²

Eine Reihe von Studien kommt jedoch auch zu dem Ergebnis, dass diese Verfälschungstendenz letztlich die Vorhersagegüte des Persönlichkeitstests hinsichtlich des Berufserfolgs nicht vermindert.¹¹³ Letztlich ist der Vergleich von Bewerbern zentral, d.h. ihre relative Position zueinander. Diese wird jedoch kaum durch die Soziale Erwünschtheit verändert.¹¹⁴ Auch wenn die Wissenschaft sich noch uneins über das Verfälschungspotential dieser „Beschönigungstendenz“ ist, sollte alles unternommen werden, um ihren Einfluss zu minimieren oder zumindest messbar zu machen.

Die klassischen „Lügenskalen“ (vgl. Unlikely Virtues Scales)¹¹⁵ (z.B. „Ich lüge nie“, „Ich sage immer die Wahrheit“) haben sich dabei als ungeeignet erwiesen, da sie durchschaubar und damit die Antworten darauf selbst auch verfälschbar sind. Dem Bewerber ist also vollkommen klar, dass er sich durch die Aussage, er lüge nie, verdächtig macht. Im IQP-Testverfahren[®] wurde dieses Verfahren modifiziert. Es wurden Items konstruiert, welche sich inhaltlich in die jeweilige Persönlichkeitsskala (z.B. Gewissenhaftigkeit) einfügen. Diese sind jedoch so extrem formuliert, dass eine maximale Zustimmung zum Item bei ehrlichem Antwortverhalten sehr unwahrscheinlich ist. Beispielsweise wird es kaum Menschen geben, die vor jeder Freizeitaktivität alle an-

Soziale Erwünschtheit wirkt wahrscheinlich weniger schwer als oft vermutet

Das IQP-Testverfahren[®] verwendet speziell konstruierte Items statt klassischer Lügenskalen zur Erfassung von Verfälschungstendenzen

110 vgl. Helmes & Holden (2003)

111 vgl. Martin et al. (2002)

112 vgl. Mueller-Hanson et al. (2003)

113 vgl. Hough et al. (1990)

114 vgl. Martin et al. (2002)

115 vgl. Hough (1998)

stehenden Aufgaben erledigt haben. Dies ist zwar wünschenswert und wäre ein Zeichen hoher Gewissenhaftigkeit; dennoch ist es sehr unrealistisch. Da jedoch nicht direkt nach Lügen, sondern im Kontext der Gewissenhaftigkeit gefragt wird, fällt dieses Item als Kontrollinstanz der Sozialen Erwünschtheit weniger auf. Natürlich wird die Antwort auf jenes Item am jeweiligen Skalenwert relativiert. Denn ist eine Person tatsächlich sehr gewissenhaft (gemessen über die normalen Items zur Gewissenhaftigkeit), so darf eine starke Zustimmung zum genannten Kontrollitem nicht so stark ins Gewicht fallen. Denn dieses zielt ja in gewisser Weise auch auf hohe Gewissenhaftigkeit ab.

Dieses Vorgehen wird für jede der sieben Persönlichkeitsskalen praktiziert. Dies ergibt sieben Items zur Sozialen Erwünschtheit, aus welchen ein entsprechender Skalenwert gebildet wird. Dieser Skalenwert hilft bei der Einschätzung, ob eine Person sozial erwünscht geantwortet hat oder nicht. Eine 100%-ige Sicherheit gibt es jedoch (wie in jedem Berufseignungstest) nicht, weshalb dieser Wert im IQP-Testverfahren[®] für den Personalverantwortlichen zwar einsehbar ist, jedoch nur als Hinweis zu verstehen sein sollte.

Eine weitere wirksame Möglichkeit zur Verminderung sozial erwünschter Antworten ist die Warnung der Bewerber. Studien zeigen: Wird vor Konsequenzen (z.B. Ausschluss aus Bewerbungsverfahren) von verfälschten Angaben gewarnt oder angekündigt, dass die Verfälschungstendenz identifizierbar ist, vermindert sich diese Tendenz.¹¹⁶ Dabei zeigt sich, dass der Fehler nicht nur im Mittel abnimmt (alle Bewerber verfälschen weniger), sondern dass auch der Anteil von Bewerbern mit Falschaussagen abnimmt.¹¹⁷ Letzteres ist wichtiger, da hierdurch die Rangfolge der Bewerber beeinflusst wird (was bei einer vergleichbaren Zu- oder Abnahme von Werten bei allen Personen nicht der Fall ist). Jedoch können übertriebene Warnungen die Bewerber einschüchtern und eventuell dazu führen, dass Bewerber ohne Verfälschungstendenz sehr konservativ antworten.¹¹⁸ Im IQP-Testsystem[®] wird aus diesem Grund mit einer weniger starken Warnung gearbeitet: Der Bewerber wird zu Beginn informiert, dass Verfälschungstenden-

**Moderate Warnungen
führen zu ehrlicheren
Angaben im Test**

116 vgl. Hough et al. (1990)

117 vgl. Griffith et al. (2006)

118 vgl. Griffith et al. (2006)

zen identifizierbar sind und dem Stellenanbieter kommuniziert werden. Auf konkrete Konsequenzen wird nicht eingegangen.

Diese Auflistung von potentiellen Fehlerquellen muss dennoch nicht entmutigen. Es wurde im IQP-Testverfahren[®] alles getan, um die aus der Forschung bekannten Fehlerquellen durch entsprechende Maßnahmen zu minimieren (Laborbedingung, geschulter Testleiter, Instruktionen) und zu kontrollieren (Skala zur Sozialen Erwünschtheit).

Hoher Aufwand wurde betrieben, um Messfehler zu minimieren bzw. zu kontrollieren

5. Die Qualität: Testgüte

Uedelhoven, S. & Nachtwei, J.

Die Testgüte ist unmittelbar von der Testkonstruktion abhängig. Je höher der Aufwand, um einen Test auf eine solide, wissenschaftlich fundierte Basis zu stellen, desto höher wird die Testgüte ausfallen. Dabei lassen sich zur Beurteilung der Qualität eines Tests verschiedene Gütekriterien verwenden. Gemeinhin werden dementsprechend Aussagen zur Objektivität eines Tests, zur Normierung, Zuverlässigkeit (Reliabilität) und Gültigkeit (Validität) sowie zu Ökonomie und Nutzen getroffen. Jene Kriterien sind jeweils noch einmal differenzierter zu betrachten. Im Folgenden werden die Kriterien kurz erläutert und angegeben, inwiefern das IQP-Testverfahren[®] das jeweilige Kriterium erfüllt.

Die Testgüte lässt sich über verschiedene Gütekriterien ermitteln

5.1. Objektivität

Ein Test ist dann objektiv, wenn die Testergebnisse unabhängig von der Person sind, die den Test leitet und/oder in Auftrag gibt. Ein Beispiel für sehr geringe Objektivität ist ein freies Gespräch – je nach Tagesform und Sympathie des Interviewers wird kein Gespräch dem anderen gleichen. Je subjektiver ein Test, desto mehr Fehler entstehen bei der Messung. Demnach leidet bei geringer Objektivität auch die Zuverlässigkeit des Tests (siehe Abschnitt 5.3). Jeder Test, dessen Durchführung, Auswertung und Ergebnisinterpretation am bzw. durch den Computer realisiert wird, weist maximale Objektivität auf. Dies ist beim IQP-Testverfahren[®] vollends gegeben.

Computerbasiertes Testen gewährleistet hohe Objektivität

5.2. Normierung

Erst die Normierung macht Testergebnisse von Bewerbern wirklich interpretierbar. Eine Norm vergleicht das Ergebnis eines Bewerbers mit den Ergebnissen anderer Bewerber. Somit kann die Position dieses Bewerbers relativ zu anderen (der Normstichprobe) abgelesen werden. Diese Position wird in Form eines Prozentranges dargestellt:

Normen dienen der Vergleichbarkeit von Testergebnissen

Der Prozentrang (ein Wert zwischen 0 und 100) gibt Auskunft über die Position des Bewerbers innerhalb einer Vergleichsgruppe (= Normstichprobe). Je größer der Wert, desto höher ist das Testergebnis im Vergleich zu dieser Gruppe. Beispielsweise bedeutet ein Prozentrang von 80 in der Skala Gewissenhaftigkeit, dass 20% der Normstichprobe höhere Werte auf jener Skala haben als der betreffende Bewerber.

Zentral sind Anzahl und Demographie der Personen in der Normstichprobe. Das IQP-Testverfahren[®] wurde zur Auswahl von Fach- und Führungskräften konzipiert. Entsprechend besteht die Normstichprobe vorwiegend aus Personen mit Hochschulreife oder höheren Abschlüssen (Bachelor, Master, Diplom etc.); Für die Merkmale Verbale und Numerische Intelligenz existiert eine gesonderte Normstichprobe für Mittlere Reife. Getrennte Alters- oder Geschlechtnormen gibt es getreu den Vorgaben des Allgemeinen Gleichbehandlungsgesetzes (AGG, 2006) nicht.

Die Altersverteilung entspricht der üblichen Verteilung im beruflichen Kontext (siehe Abbildung). Die Normstichprobe besteht ausschließlich aus Personen, die unter kontrollierten Bedingungen den Test absolviert haben. Eine erste Normstichprobe enthielt knapp 250 deutschsprachige Personen zwischen 18 und 62 Jahren mit mindestens Hochschulreife. In den letzten Jahren erfolgten Erweiterungen der Normstichprobe. Aktuell umfasst diese 2.000 Personen (Version 5.0), wovon 242 anonymisiert und daher ohne soziodemographische Daten erhoben wurden. Einen detaillierten Überblick zur Alters- und Geschlechtsverteilung gibt die nachfolgende Abbildung.

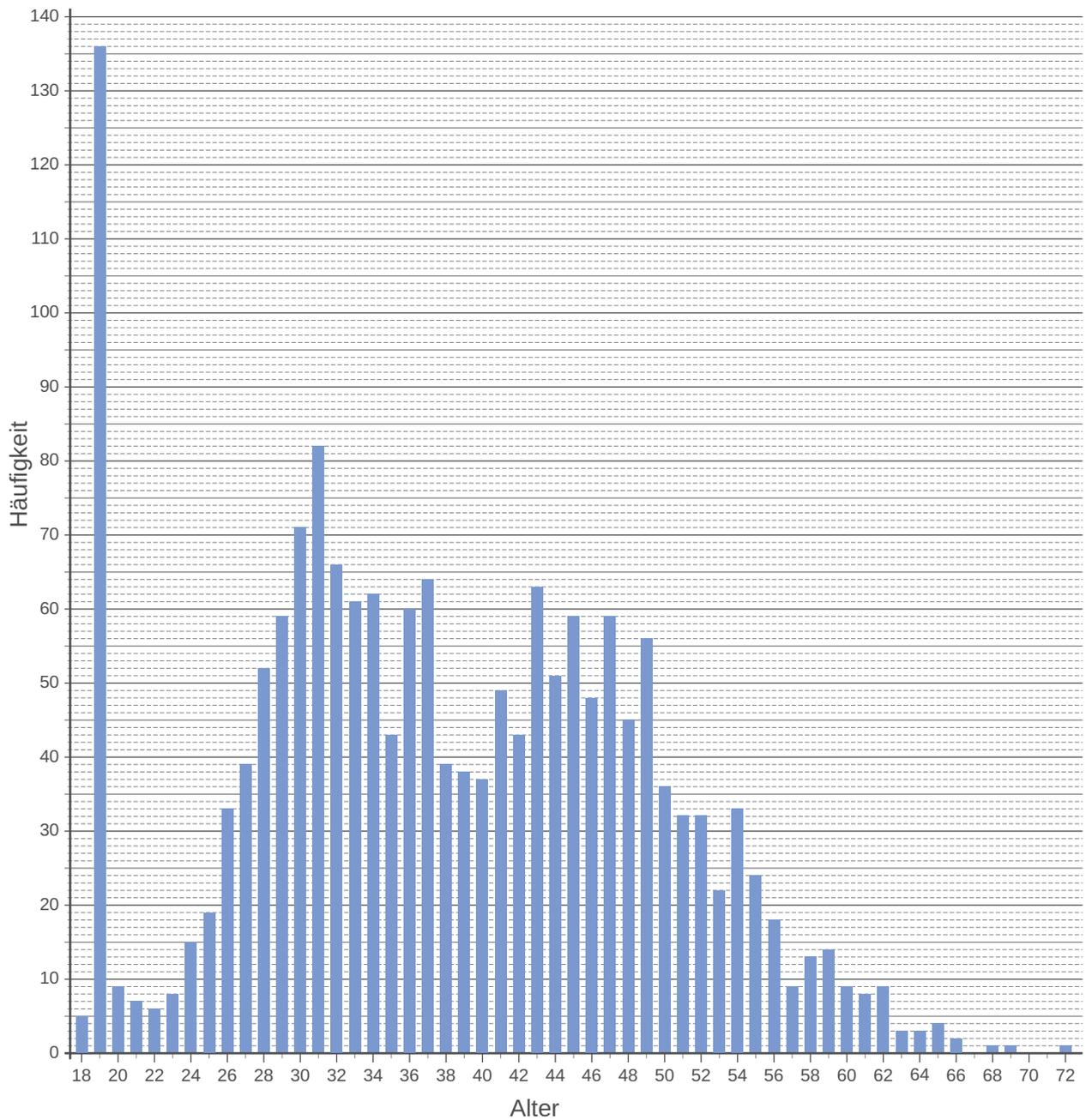
Es gilt: Je mehr Personen die Normstichprobe enthält, desto genauer ist die Berechnung der Position eines Bewerbers relativ zu dieser Gruppe. Die Normstichprobe wird daher stetig durch Bewerber aus dem Praxiseinsatz des Verfahrens erweitert.

Dennoch ist nicht die schiere Größe der Stichprobe entscheidend. Wichtig ist neben der Datenqualität vor allem die Repräsentativität. So macht es weder Sinn, einen Bewerber für eine Stelle als Bauingenieur mit einer Stichprobe von Personen ohne Abitur, noch mit einer Stichprobe von knapp 70-jährigen zu vergleichen. Der Vergleich mit Abiturienten bzw. Hochschulabsolventen jungen bis mittleren Alters hingegen ist in diesem Fall sinnvoll.

Die Normstichprobe besteht ausschließlich aus Personen, die unter kontrollierten Bedingungen getestet wurden

Fokus auf Datenqualität – Personen in der Normstichprobe sind geeignet für den Vergleich mit Fach- und Führungskräften

Übersicht Altersverteilung



Statistiken:

Alter		Geschlecht	
Mittelwert:	38,08	Männlich:	65,9 %
Standardabweichung:	10,898	Weiblich:	34,1 %
	Minimum: 18		
	Maximum: 72		

Übersicht zur Alters- und Geschlechtsverteilung in der Normstichprobe

Anmerkung: Die Normstichprobe umfasst weitere 242 Personen, die anonymisiert (ohne soziodemographische Daten) erhoben wurden.

5.3. Zuverlässigkeit

Die Zuverlässigkeit (= Reliabilität) sagt aus, wie gut ein Test, eine Skala oder ein Item in der Lage ist, den sogenannten „wahren Wert“ einer Person zu messen. Dabei wird davon ausgegangen, dass der wahre Wert nicht vollends erreichbar ist, da jede Messung fehlerbehaftet ist. Je höher die Zuverlässigkeit, desto geringer der Messfehler und entsprechend näher die Messung am wahren Wert.

Zwei gängige Ansätze zur Zuverlässigkeitsberechnung werden unterschieden: Interne Konsistenz (Homogenitätsmaß) und Test-Retest-Reliabilität (Zusammenhang zwischen Erst- und Folgemessung). Pauschal gilt für diese Maße: je näher zur 1, desto höher die Zuverlässigkeit. Für die Interne Konsistenz gilt jedoch auch, dass eine zu homogene Skala (Werte ab .90) die Vorhersagegültigkeit (prädiktive Validität) vermindert.¹¹⁹ Das liegt daran, dass in diesem Fall versucht würde, ein heterogenes Merkmal (Berufserfolg) durch ein sehr homogenes Merkmal (gemessen über die jeweilige Skala) zu erfassen. Hohe Homogenität bedeutet zugleich auch hohe Redundanz: Viele Items messen genau dasselbe und bringen keinen zusätzlichen Erkenntnisgewinn; dies macht den Testeinsatz unökonomisch. Im IQP-Testverfahren® wird diesem für die Praxis essentiellen Umstand Rechnung getragen – die Skalen wurden möglichst facettenreich konstruiert, um der Vielschichtigkeit der beruflichen Bewährung von Bewerbern gerecht zu werden. Demnach werden mittlere (bzw. in jedem Fall unter .90 liegende) Interne Konsistenzen angestrebt, welche – entgegen der veralteten Lehrmeinung – einer guten Vorhersage des Kriteriums (hier Berufserfolg) keinen Abbruch tun.¹²⁰

Naturgemäß ist bei den Intelligenzskalen mit einer niedrigeren Internen Konsistenz als bei den Skalen zur Erfassung der Persönlichkeit zu rechnen. Denn der Facettenreichtum von Intelligenzmerkmalen drückt sich nicht nur inhaltlich aus (beispielsweise sind eine Vielzahl verschiedener Mechanismen nötig, um Sprache richtig zu verstehen und anwenden zu können), sondern auch in der Auswahl der Aufgaben, um das Merkmal zu messen. So fanden möglichst viele verschie-

Die Zuverlässigkeit ist die Freiheit von Messfehlern

Die Skalen im IQP-Testverfahren® sind praxisnah konstruiert

119 vgl. Lienert & Raatz (1998)

120 vgl. Schmitt (1996)

dene Aufgaben Eingang in die jeweilige Skala, um eine maximale Anzahl von Aspekten messen zu können.

Die Interne Konsistenz der Skalen des IQP-Testverfahrens® wird bei jeder Normstichprobenrevision neu berechnet. Ferner wird die Trennschärfe angegeben. Die Trennschärfe sagt aus, wie gut die Skala zwischen Personen mit niedriger vs. hoher Merkmalsausprägung unterscheidet. Für Persönlichkeits- und Intelligenzskalen gelten hinsichtlich der empfohlenen Höhe der Trennschärfe unterschiedliche Konventionen: .40 für Persönlichkeits- und .20 für Intelligenzskalen. Die mittlere Interne Konsistenz der Persönlichkeitsskalen beträgt für die aktuelle Normstichprobe .73 bei einer mittleren Trennschärfe von .39. Die Intelligenzskalen weisen eine durchschnittliche Interne Konsistenz von .51 bei einer mittleren Trennschärfe von .22 auf. Die Zielvorgaben einer mittleren bis leicht erhöhten Internen Konsistenz und einer der Konvention entsprechenden Trennschärfe sind erreicht.

Die statistischen Zielvorgaben sind erreicht

Anzumerken ist, dass die Internen Konsistenzen maßgeblich von der Anzahl der Aufgaben/ Items einer Skala abhängen. Je mehr Items eine Skala hat, desto höher ist die Interne Konsistenz.¹²¹ In einem Screening-Verfahren wie dem IQP-Testverfahren®, welches mit Hilfe einer überschaubaren Itemzahl Merkmale ökonomisch misst, sind die Internen Konsistenzen naturgemäß nicht so hoch, wie in mehrstündigen Tests. Für weitere Besonderheiten von Kurzskalen sei auf weiterführende Literatur verwiesen¹²².

Die Berechnung einer Test-Retest-Reliabilität ist bei Intelligenzskalen schwierig, da mit Übungseffekten gerechnet werden muss. Für Persönlichkeitsskalen hingegen sind derartige Berechnungen angebracht, jedoch nicht einfach zu realisieren. Denn dafür müssen Personen den Test zweimal durchlaufen – eine erneute Behelligung von Bewerbern nach Beendigung des eigentlichen Tests ist im Standardfall nicht ratsam und in der Unternehmenspraxis nur schwer umsetzbar. Das IQP hat von 2009 bis 2013 eine eigene Studie zur Überprüfung der zeitlichen Stabilität der 10 gemessenen Merkmale aufgesetzt. Der Test wurde im Abstand von im Mittel 20 Monaten in bei derselben Stichprobe von Berufstätigen eingesetzt. Alle Skalen wiesen trotz des

121 vgl. Cortina (1993)

122 Kain & Nachtwei (2009).

hohen zeitlichen Abstands von bis zu 41 Monaten zwischen den Messzeitpunkten eine hohe bis sehr hohe zeitliche Stabilität auf ($r \geq .60$).

5.4. Gültigkeit

Die Gültigkeit – oder Validität – eines Tests besagt, ob dieser in der Lage ist, das interessierende Merkmal auch zu messen. Dabei wird zwischen der so genannten Konstrukt- und Kriteriumsvalidität unterschieden.

Die Gültigkeit beschreibt den inhaltlichen Gehalt einer Skala

Konstruktvalidität

Die Konstruktvalidität sagt aus, ob die Messung das Merkmal erfasst, welches erfasst werden soll.¹²³ Da die Skalen des IQP-Testverfahrens® streng nach dem aktuellen wissenschaftlichen Konsens entwickelt wurden, kann eine hohe Konstruktvalidität als gegeben angenommen werden. Dennoch wurde diese auch rechnerisch nachgewiesen: Dies lässt sich vor allem darüber prüfen, ob die Messung per Skala den theoretischen Vorstellungen zum jeweiligen Merkmal entspricht. Beispielsweise nimmt die Theorie der „Big Five“ an, dass die fünf zentralen Persönlichkeitsmerkmale in der Lage sind, Persönlichkeit sparsam (also mit möglichst wenig Redundanz) zu beschreiben. Entsprechend dürften die Skalen zur Messung der „Big Five“ nicht stark miteinander zusammenhängen. Denn wären z.B. die Skalen Gewissenhaftigkeit und Extraversion absolut gegenläufig, so könnte man sich eine der beiden sparen. Man wüsste dann, dass jeder Gewissenhafte automatisch wenig extrovertiert ist und andersherum. Die relative Unabhängigkeit der „Big Five“-Skalen sowie der Skalen Leistungsmotivation und Risikoneigung konnte im IQP-Testverfahren® eindeutig nachgewiesen werden. Die gemeinsame Varianz der 7 Skalen beträgt lediglich 6,47%.

Die Konstruktvalidität kann theoretisch abgeleitet und rechnerisch nachgewiesen werden

Eine weitere theoretische Annahme ist, dass sich Skalen, die nicht zu den „Big Five“ gehören, zwar aus diesen zusammensetzen, jedoch auch einen erheblichen eigenen Anteil messen (welcher nicht durch die „Big Five“ messbar ist). Dies wird überprüft, indem der Anteil der

¹²³ vgl. Asendorpf (1999)

„Big Five“ Skalen an der betreffenden Skala berechnet wird (multiple lineare Regression). Dabei zeigt sich, dass die beiden Nicht-„Big Five“ Skalen Risikoneigung und Leistungsmotivation mindestens 50% einen eigenen, nicht durch „Big Five“ erklärbaren, Anteil aufweisen. Die Konstruktvalidität der Persönlichkeitsskalen konnte auch in diesem Fall auch rechnerisch nachgewiesen werden.

Eine ähnliche Vorgehensweise bot sich auch bei den Intelligenzskalen an. Wie bereits in Kapitel 3 beschrieben, orientiert sich das IQP-Testverfahren[®] am dreistufigen hierarchischen Modell der kognitiven Fähigkeiten von Carroll (1993).¹²⁴ Ganz oben in dieser Hierarchie steht der g-Faktor, also die allgemeine Intelligenzleistung, gemessen durch die Skala Abstrakt-Analytische Intelligenz. Auf der untersten Abstraktionsstufe finden sich dann spezifische Intelligenzleistungen wie Numerische und Verbale Intelligenz, gemessen durch die gleichnamigen Skalen. Entsprechend diesem Modell müsste es jeweils moderate Zusammenhänge der Skalen Numerische und Verbale Intelligenz zur Skala Abstrakt-Analytische Intelligenz geben. Denn letztere symbolisiert eine allgemeine Intelligenz, die für jede spezifischere mehr oder minder notwendig ist. Die Skalen Verbale und Numerische Intelligenz dürften ebenfalls miteinander nicht übermäßig stark zusammenhängen, da sie sonst keine unabhängigen Dimensionen auf derselben Hierarchieebene bilden würden. Die beschriebenen Annahmen konnten rechnerisch bestätigt (gemeinsame Varianz der drei Skalen von 15%) und damit eine Konstruktvalidität auch für die Intelligenzskalen nachgewiesen werden.

Weitere Studien zur ausführlicheren Prüfung der Konstruktvalidität durch Zusammenhangsmessungen mit anderen Merkmalen sind geplant bzw. bereits in der Durchführung.

Kriteriumsvalidität

Die Kriteriumsvalidität sagt (im hier zu betrachtenden Fall) aus, ob ein Test bzw. eine Skala (= Prädiktor) in der Lage ist, ein relevantes Kriterium (wie z.B. den Berufserfolg) vorherzusagen. Dieser Zusammenhang wird auch prädiktive Validität genannt.

Die Höhe der Kriteriumsvalidität hängt zum einen vom Test (= Prä-

124 vgl. te Nijenhuis et al. (2007)

Das IQP-Testverfahren[®] orientiert sich am dreistufigen hierarchischen Modell der kognitiven Fähigkeiten von Carroll (1993)

Die Kriteriumsvalidität ist gegeben

dikator) ab. Je höher die Reliabilität und Konstruktvalidität des Tests (siehe oben), desto höher die maximal erreichbare Kriteriumsvalidität. Dies ist nachvollziehbar, da die genannten Merkmale des Tests etwas über seine „innere oder strukturelle Güte“ aussagen. Nur wenn die innere Struktur gesichert ist, kann etwas außerhalb des Tests stehendes (das Kriterium Berufserfolg) vorhergesagt werden. Zum anderen hängt die Kriteriumsvalidität aber auch vom Kriterium selbst ab. Schließlich gibt es nicht den Berufserfolg, sondern je nach Anwendungskontext (Berufsfeld, Firma, Stelle) unterschiedliche Indikatoren des Berufserfolgs (von der Zahl verkaufter Versicherungen bis zur Verminderung der Ausfallrate unterstellter Mitarbeiter).

Da nicht von vornherein all diese Erfolgsindikatoren messbar und mit den Testergebnissen in Zusammenhang zu bringen sind, wird sich auf die Ergebnisse aus der Literatur berufen. Diese Studien beschreiben den Zusammenhang zwischen Tests, die nach einem vergleichbaren theoretischen sowie statistisch-methodischen Grundprinzip konstruiert wurden, und verschiedenen Berufserfolgsindikatoren. Eine ausführliche Darstellung findet sich in den Kapiteln 1 bis 3. Bei streng wissenschaftlichem und theorieorientiertem Vorgehen ist eine Vorhersage von Berufserfolg demnach gegeben.

IQP betreibt jedoch auch eigene Langzeitstudien, um die prädiktive Validität des IQP-Testverfahrens® für ganz spezifische Anwendungsfälle nachzuweisen. Exemplarisch sollen an dieser Stelle zunächst interessante, statistisch abgesicherte Befunde zur Kriteriumsvalidität des IQP-Testverfahrens® dargestellt werden. So lässt sich beispielsweise zeigen, dass Personen ohne eine psychologische oder psychiatrische Behandlung eine höhere Emotionale Belastbarkeit im Test aufweisen. Zudem ließen sich durch Verbale Intelligenz zum einen der Redean teil in einem Assessment Center und zum anderen der Berufserfolg in einem beratungsnahen Berufsfeld vorhersagen. Personen mit höherer Extraversion im Test zeigten sich auch dominanter in Gruppendiskussionen mit berufsbezogenen Themen (Verhandlung um ein Budget), wohingegen Personen mit hoher Teamorientierung in derartigen Situationen freundlicher auftraten (jeweils Verhaltensbeurteilung durch geschulte Beobachter).

Weiterhin zeigte sich in Studien bei unseren Kunden, dass bei-

Studien beschreiben den Zusammenhang zwischen Tests und Berufserfolgsindikatoren

IQP betreibt eigene Langzeitstudien

spielsweise in den Merkmalen Numerische und Abstrakt-Analytische Intelligenz signifikante Unterschiede zwischen Mitarbeitern mit durchschnittlicher vs. hoher Leistung bestehen (höhere Werte bei Leistungsträgern). Zudem konnten US-amerikanische Studien repliziert werden, welche den Zusammenhang zwischen Leistungsmotivation und Professionalität im Außenauftritt (gemessen an der Seriosität der persönlichen E-Mailadresse) ausweisen. Weiterhin konnte die Zuverlässigkeit im Sinne der Termintreue durch das Merkmal Risikoneigung (welches mit Impulsivität zusammenhängt) vorhergesagt werden: Personen mit einer sehr hohen Risikoneigung hielten einen schriftlich fixierten Termin ca. 7 mal seltener (51,7% Nicht-Erscheinen) ein als Personen mit einer sehr geringen Risikoneigung (6,9% Nicht-Erscheinen).

Daneben ist IQP im Bereich Diagnostik von Gründerpotentialen vertreten. Auch hier zeigen sich hypothesenkonforme Ergebnisse, wie beispielsweise die höhere Abstrakt-Analytische Intelligenz und höhere Geistige Flexibilität bei Gründern im Vergleich zu Festangestellten.

Eine Validierungsstudie in einem Call Center Unternehmen zeigte, dass sich erfolgreiche Call Agents durch eine im Vergleich zu nur durchschnittlich performanten Kollegen höhere Extraversion und Risikoneigung sowie eine geringere Teamorientierung auszeichneten. Eine Studie mit Projektmanagern verschiedener Unternehmen zeigte, dass bei extern arbeitenden Projektmanagern Emotionale Belastbarkeit und Projektmanagementenerfolg positiv korreliert waren. Zudem hing die Risikoneigung bei Projektmanagern in der IT-Branche signifikant negativ mit Projekterfolg zusammen. Es zeigte sich ferner, dass bei weiblichen Projektmanagern das Merkmal Teamorientierung positiv mit dem Projekterfolg korreliert war.

Eine weitere Studie untersuchte Führungskräfte in einem Industrieunternehmen der Bau- und Energiebranche. Es zeigte sich unter anderem, dass die Verbale Intelligenz der Führungskräfte die Leistung (ermittelt über Beobachter-Urteil zur Kommunikationsfähigkeit im AC) in der Gruppendiskussion und im Rollenspiel Mitarbeitergespräch aufklärte. Zudem bestand ein signifikanter positiver Zusammenhang zwischen Geistiger Flexibilität und beurteilter Problemlösefähigkeit im Mitarbeitergespräch. Weitere aktuelle Studien des IQP erfassen den

**Einsatz des IQP-
Testverfahrens®
zur Potentialanalyse
bei Gründern**

Zusammenhang der Testskalen mit konkret per logfile erfasstem Verhalten und der Leistung in einem computerbasierten Problemlöszenario, den Zusammenhang zwischen Persönlichkeit und unternehmerischem Erfolg, zwischen Persönlichkeit und Vertriebs Erfolg sowie Intelligenz und Persönlichkeit und Erfolg in einem Unternehmen der Wohnungswirtschaft.

Die meisten Ergebnisse sowie methodische Aspekte der Testkonstruktion werden im Rahmen von Kooperationen mit Arbeitsgruppen an der HU Berlin (Prof. Scholl), im Rahmen des Zentrums für empirische Evaluationsmethoden (Prof. Krause) oder Projektgruppen an der TU Berlin veröffentlicht.

5.5. Ökonomie und Nutzen

Gerade für die Praxis gilt die Ökonomie hinsichtlich Durchführung und Auswertung eines Tests als zentrales Qualitätskriterium. Zwei Ziele müssen dabei in Einklang gebracht werden: zuverlässige, gültige Skalen, welche gleichzeitig hinreichend kurz sind. Dass beides vereinbar ist, konnte z.B. anhand von Intelligenzskalen gezeigt werden (vgl. Abstrakt-Analytische Intelligenz).

Wie kurz ein Test tatsächlich sein kann, hängt aber auch von zwei anderen Fragestellungen ab: Wofür wird er eingesetzt und bei welcher Personengruppe? Das IQP-Testverfahren[®] ist ein Screening-Verfahren, welches zwischen Sichtung der Bewerbungsunterlagen und Interview bzw. Assessment Center geschaltet oder mit den genannten Verfahren in einer Lösung kombiniert wird. Eine Einstellungsentscheidung beruht also nicht allein auf den Ergebnissen dieses Tests. Eine mehrstündige Testdiagnostik ist demnach hier nicht zweckmäßig und schlichtweg zu teuer. Das diagnostische Gesamtbild einer Person ergibt sich ohnehin aus mehreren Quellen (Unterlagen, Test, Gespräch). Die zweite Frage bezieht sich auf die Personengruppe, bei welcher der Test zum Einsatz kommt.

Es wird klar: Der Nutzen eines Tests für die Auswahl von Bewerbern ist maximal, wenn er mit hoher Zuverlässigkeit und Gültigkeit misst und zugleich nicht zu lang und somit zu teuer ist.

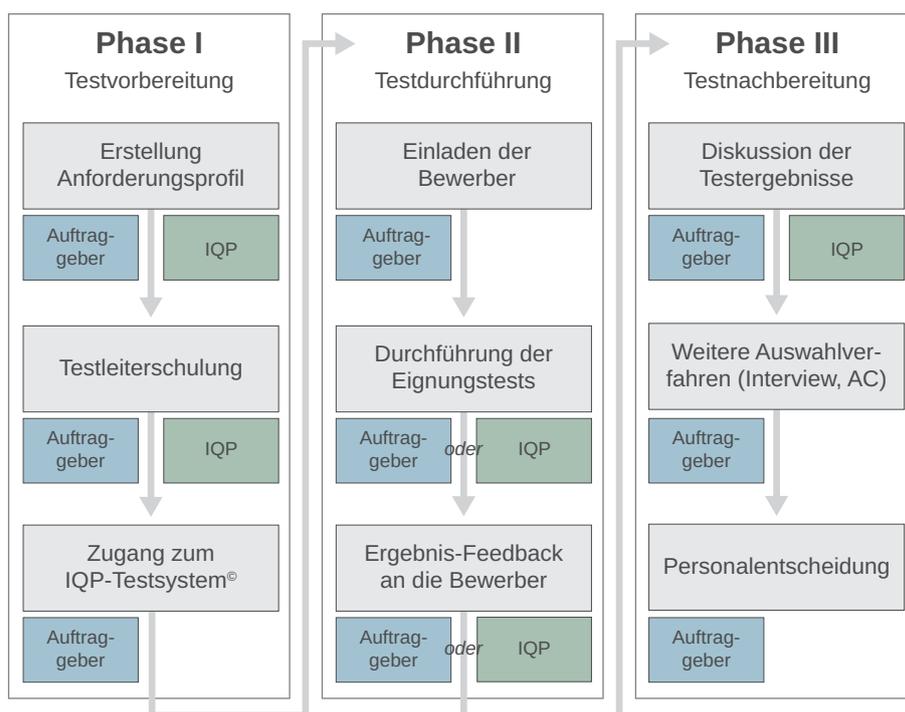
In der Praxis sind kurze Tests gefragt – länger als 60 Minuten sollte im Screening nicht getestet werden

Das IQP-Testverfahren[®] erfüllt die Anforderungen an einen praxistauglichen und wissenschaftlich fundierten Eignungstest

6. Das Tool: IQP-Testsystem®

Faust, U., Lattke, T. & Wessendorf, T.

Das IQP-Testsystem® bringt die Anforderungen der Stelle und die Merkmale des Bewerbers zusammen und begleitet damit den gesamten eignungsdiagnostischen Prozess (siehe Abbildung) aus technischer Sicht.



Der eignungsdiagnostische Prozess im Überblick

Es besteht aus einer Server-Lösung, auf die durch den Personalverantwortlichen zugegriffen wird, sowie aus dem eigentlichen Eignungstest (Client-Lösung), welcher vom Bewerber absolviert wird. Bei der technischen Realisierung des IQP-Testsystems® wurde sehr darauf geachtet, die hohe Funktionalität von Server- und Client-Lösung so benutzerfreundlich wie möglich zu gestalten.

Das IQP-Testsystem® – Hohe Funktionalität benutzerfreundlich realisiert

So erfolgt der Zugriff auf die Server-Lösung über eine Internetanwendung – benötigt werden dafür neben den persönlichen Zugangsdaten (Nutzername und Passwort) lediglich ein Rechner mit Internetzugang sowie ein gewöhnlicher Browser zur Anzeige der Webinhalte. Das Anlegen und Verwalten von Stellen, Anforderungsprofilen und Bewerbern sowie die Auswertung der Testergebnisse lassen sich dadurch

mit wenigen Klicks realisieren. Selbstverständlich sind beim Umgang mit derart sensiblen Daten alle Internetverbindungen zum IQP-Server gesondert verschlüsselt.

Die Client-Lösung wurde so konzipiert, dass alle für den Ablauf des Eignungstests erforderlichen Programme in einer eigenen Laufzeitumgebung auf einem USB-Stick ausgeführt werden. Für den Personalverantwortlichen bedeutet dies, dass der IQP-Client USB-Stick an jedem handelsüblichen Windows-PC eingesetzt werden kann – und zwar ohne Installation von Software. Die erforderliche Anzahl an USB-Sticks wird durch IQP vorbereitet und dem Unternehmen zur Verfügung gestellt.

Sowohl die Server- als auch die Client-Lösung lassen sich daher problemlos in die bestehenden Arbeitsprozesse und die vorhandene Infrastruktur integrieren.

Um die Qualität des Verfahrens zu gewährleisten, werden die (Personal-) Verantwortlichen im Unternehmen ausführlich im Umgang mit dem IQP-Testsystem® geschult. Neben dem sicheren Umgang mit dem System stehen vor allem die Fähigkeit der Auswertung und Interpretation der erhaltenen Testergebnisse sowie die Gestaltung des Feedbacks für den Bewerber als wichtige Lerninhalte im Vordergrund.

Das umfangreiche und detaillierte Schulungsmaterial verbleibt bei den Schulungsteilnehmern im Unternehmen. Ständig durch IQP aktualisiert, beinhaltet es neben den im Rahmen der Schulung vermittelten Inhalten zahlreiche aktuelle Hintergrundinformationen zur Thematik wie beispielsweise rechtliche Hintergründe zum Einsatz von Eignungstests im Unternehmen, Empfehlungen zur Durchführung und Organisation von externen (Bewerber) und internen (Mitarbeiter) Testungen sowie Vordrucke für Einverständniserklärungen. Zusammen mit dem persönlichen IQP-Support sichert das Schulungsmaterial so nachhaltig die Qualität des Verfahrens im Unternehmen.

Nachfolgend werden die grundlegende Funktionsweise und der Einsatz des IQP-Testsystems® im Unternehmen vorgestellt. Spezifische Hinweise zur Benutzung sind in der Software selbst in Form eines Hilfesystems integriert und werden im Rahmen der Schulung persönlich vermittelt.

Problemlose Integration in die bestehenden Arbeitsprozesse

Schulung und persönlicher IQP-Support sichern nachhaltig die Qualität des Verfahrens im Unternehmen

6.1. Vorbereitung der Testdurchführung

Erstellen eines Anforderungsprofils

Der erste Schritt ist die Erstellung eines Anforderungsprofils für die vakante Stelle. Die Stellenanforderung wird hinsichtlich der 10 Skalen des IQP-Testverfahrens® vom Personalverantwortlichen, Fachvorgesetzten und im Idealfall langjährigen Mitarbeiter gemeinsam mit einem IQP-Berater (und Diplom-Psychologen) definiert. Im Rahmen einer Anforderungsanalyse werden dabei die idealen Ausprägungen der Merkmale in Form von Prozentrangwerten festgelegt (siehe Abschnitt 5.2).

Das Anforderungsprofil für die vakante Stelle wird auf dem IQP-Server über Schieberegler eingepflegt.

Sind die Einstellungen für alle 10 Skalen vorgenommen, wird das entstandene Profil als sogenannte Haupt-Anforderungsprofil der vakanten Stelle auf dem IQP-Server abgespeichert. Im Anschluss ist es auch möglich, jederzeit Neben-Anforderungsprofile zu erstellen. Diese werden jedoch ohne Unterstützung eines geschulten Beraters vom Personalverantwortlichen selbst definiert. Die Testergebnisse der Bewerber (welche sich auf dieselben 10 Skalen beziehen) werden am Ende mit dem definierten Anforderungsprofil verglichen, um die spezifische Eignung für die vakante Stelle zu erfassen. Diese Übereinstimmung wird in Prozent ausgedrückt, wobei sehr geringe Abweichungen von der definierten Anforderung weniger stark gewichtet werden als sehr hohe Abweichungen. Diese unterschiedliche Gewichtung erklärt sich dadurch, dass ab einer bestimmten Abweichung nicht nur quantitativ die Nähe zur Anforderung fehlt, sondern auch qualitativ (z.B. kann extrem hohe Gewissenhaftigkeit schnell zu Zwang werden und ist somit nicht nur zahlenmäßig weit von einer mittleren Gewissenhaftigkeit entfernt).

Einpflegen der Bewerberdaten

Ist das Anforderungsprofil angelegt, können die Bewerber mit ihren Stammdaten (Name, Geburtstag etc.) in die entsprechenden Masken des Systems eingepflegt werden. Bis zu diesem Zeitpunkt hat nur der Personalverantwortliche der beauftragenden Firma Kontakt zu den Bewerbern und die Vorauswahl anhand der Bewerbungsunterlagen

Zusammen mit IQP werden die Anforderungen der Stelle präzise in einem Anforderungsprofil definiert

Das Einpflegen der Bewerberdaten erfolgt über intuitive Eingabemasken im IQP-Testsystem®

vorgenommen. Entsprechend werden die Bewerberdaten von ihm oder Mitarbeitern seiner Abteilung eingepflegt. Im Anschluss erfolgt die Einladung der Bewerber für die Testdurchführung.

6.2. Testdurchführung

Um einen beliebigen Windows-PC für die Testdurchführung vorzubereiten, muss lediglich der IQP-Client USB-Stick an den entsprechenden Rechner gesteckt und das Programm auf dem USB-Stick gestartet werden. Der zu testende Bewerber wird im Programm ausgewählt und im Rahmen der Testung hinsichtlich aller 10 Skalen des IQP-Testverfahrens® auf seine Eignung für die vakante Stelle getestet.

Die hohe Kompatibilität des Tests ermöglicht den Einsatz an nahezu jedem Windows-PC



Funktionsübersicht des IQP-Client

Im Anschluss an den Test besteht die Möglichkeit, auf dem Stick ein Testzertifikat der Testergebnisse im pdf-Format zu generieren. Neben Informationen zum Test und den verwendeten Skalen werden die Test-

ergebnisse des Bewerbers als Prozentrangwerte abgebildet – eine Passung zur Stelle wird dabei im Zertifikat nicht berücksichtigt.

Für die Gestaltung des Kurzfeedbacks für den Bewerber kann das Testzertifikat unmittelbar vor Ort nach dem Test ausgedruckt oder später per Mail versendet werden.

An einem Rechner mit Internetverbindung können mit wenigen Klicks die Testergebnisse auf dem IQP-Client USB-Stick mit dem IQP-Server synchronisiert werden und stehen dort im Anschluss dem Personalverantwortlichen für Auswertungen zur Verfügung.

Ein Zertifikat mit den Testergebnissen der Person wird automatisch vom IQP-Testsystem® generiert

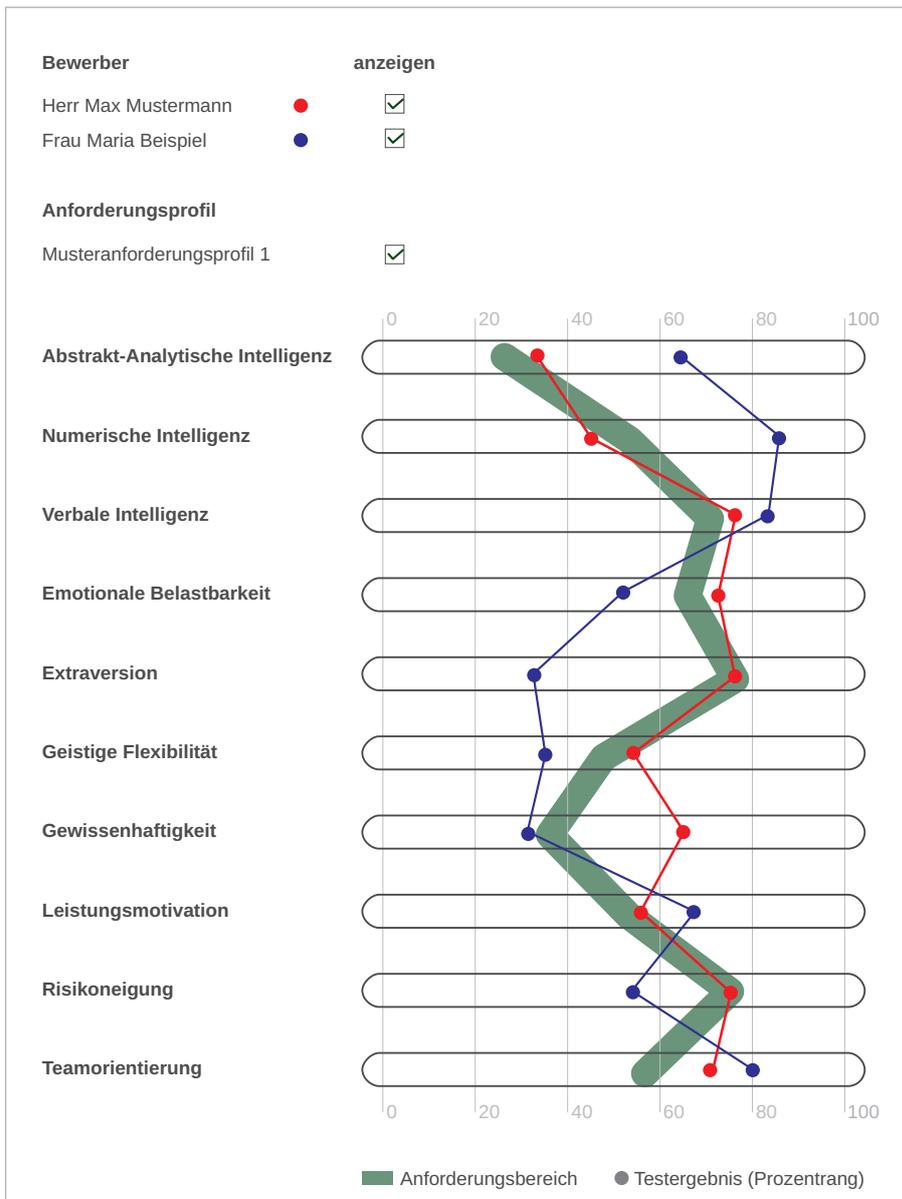
6.3. Auswertung und Interpretation der Testergebnisse

Nach der Synchronisierung des IQP-Client USB-Sticks mit dem IQP-Server stehen die Testergebnisse aller getesteten Bewerber dem jeweiligen Personalverantwortlichen in seinem Login auf dem Server zur Verfügung. Die Ergebnisse können auf zweierlei Weise ausgewertet werden: Prozentrangwert und Prozentwert. Während der Prozentrangwert die Position des Bewerbers in der Normstichprobe beschreibt, zeigt der Prozentwert die Passung des Bewerbers zur Stellenanforderung an. Grundsätzlich sollte die Entscheidung für und wider einen Bewerber anhand des Prozentwertes gefällt werden, da dieser einen direkten Bezug zur Stellenanforderung aufweist. Es ist prinzipiell möglich, das Testergebnis mit verschiedenen Neben-Anforderungsprofilen zu vergleichen. Empfohlen wird jedoch das mit Unterstützung durch den Berater definierte Haupt-Anforderungsprofil.

Unterschiedliche Darstellungsformen der Testergebnisse ermöglichen eine komfortable Auswertung der Testergebnisse

Für beide Auswertungsformen der Testergebnisse (Prozentrangwert vs. Prozentwert) kann neben der tabellarischen Darstellung ein grafischer Profilvergleich der Werte mehrerer Bewerber erstellt werden (siehe Abbildung).

Grafischer Vergleich von Ergebnisprofilen zweier Bewerber in Relation zum Anforderungsprofil auf dem IQP-Server®



Nach Beendigung aller Testläufe und Auswertung der Testergebnisse erfolgt die Entscheidung, welche Bewerber in die nächste Auswahlrunde (z.B. Interview) gelangen sollen.

Glossar

Faktorenanalyse	Die Faktorenanalyse ist ein Verfahren zur Datenreduktion. Eine Menge von Items lässt sich auf ihre Zusammengehörigkeit hin überprüfen. Die Items werden nach Faktoren gruppiert, welche inhaltlich interpretiert werden und Teilaspekte von Skalen darstellen.
Item	Als Item wird die kleinste Einheit in einem Test bezeichnet. Ein Item kann eine Intelligenzaufgabe oder auch eine Aussage bzw. Frage sein. Jedes Item ist in der einen oder anderen Weise von der zu testenden Person zu beantworten.
Itemschwierigkeit	Die Itemschwierigkeit wird definiert über den Anteil von Personen, die das Item bejahen bzw. richtig lösen. Umso weniger Personen ein Item richtig lösen, desto schwieriger ist es (und andersherum). Die Schwierigkeit sagt somit etwas darüber aus, wie stark das Merkmal ausgeprägt ist, welches über das jeweilige Item erfasst werden soll.
Korrelation	Die Korrelation drückt einen Zusammenhang zwischen Variablen in Zahlen aus. Dieser Zahlenwert variiert zwischen -1 und 1. Negative Korrelationen sagen aus, dass die eine Variable hoch ausgeprägt ist, wenn die andere gering ausgeprägt ist. Positive Korrelationen bedeuten, dass eine hohe Ausprägung der einen Variablen mit einer hohen Ausprägung der anderen einhergeht. Eine Nullkorrelation zeigt an, dass die Variablen in keinerlei Zusammenhang stehen.
Kriterium	Das Kriterium ist ein Maß, das durch einen Test vorhergesagt wird. Kriterien mit praktischer Relevanz sind beispielsweise der Berufserfolg oder die Arbeitszufriedenheit.

Norm, Normen	Die Norm ist als eine Art Vergleichsmaßstab für Bewerber zu verstehen. Die Norm wird gebildet durch alle Personen, die einen Test bereits durchlaufen haben – dies nennt man Normstichprobe. Das Testergebnis des Einzelnen wird demnach mit den Testergebnissen der Normstichprobe verglichen. Somit hängt der Testwert von den Merkmalen der Normstichprobe (Alter, Bildung etc.) maßgeblich ab. Beispielsweise macht es keinen Sinn, Hochschulabsolventen mit einem für Hauptschüler normierten Intelligenztest zu konfrontieren. Alle Hochschulabsolventen hätten in diesem Fall stark überschätzte Intelligenztestwerte.
Normstichprobe	siehe Norm, Normen
Prädiktor	Der Prädiktor ist ein Maß, welches das Kriterium (siehe oben) vorhersagt
Prozentrang	Der Prozentrang (ein Wert zwischen 0 und 100) gibt Auskunft über die Position des Bewerbers innerhalb einer Vergleichsgruppe (= Normstichprobe). Je höher der Wert, desto höher ist das Testergebnis im Vergleich zu dieser Gruppe. Beispielsweise bedeutet ein Prozentrang von 80 in der Skala Gewissenhaftigkeit, dass 20% der Normstichprobe höhere Werte auf jener Skala haben als der betreffende Bewerber.
Regression	Die Regression kann im einfachen Fall mit einer Korrelation gleichgesetzt werden. Bei einer multiplen Regression hingegen wird ein Kriterium durch mehrere Prädiktoren vorhergesagt. Damit ist eine Einschätzung möglich, welcher Prädiktor das Kriterium relativ betrachtet am stärksten vorhersagt.
Reliabilität	Die Reliabilität sagt etwas über die Zuverlässigkeit einer Messung aus. Zuverlässig ist sie dann, wenn wenige Fehlereinflüsse bestehen bzw. der „wahre Wert“ möglichst genau durch den „gemessenen Wert“ erfasst wird. Es gibt verschiedene Formen der Reliabilität, z.B.: Interne Konsistenz (inhaltliche Homogenität), Test-Retest-Reliabilität (zeitliche Stabilität).
Selektionsstichprobe	Die Selektionsstichprobe ist jene Gruppe von Personen, an welchen die Itemselektion (Auswahl von Items einer Skala) vorgenommen wird. Diese Stichprobe darf nach abgeschlossener Itemselektion nicht mehr verwendet werden.

Skala	Eine Skala fasst inhaltlich verwandte Items zu einer Sinneinheit zusammen. Beispielsweise besteht die Skala Gewissenhaftigkeit aus mehreren Items, welche dieses Merkmal erfassen. Die Güte einer Skala drückt sich vorwiegend in ihrer Reliabilität und Validität aus.
Standardabweichung	Die Standardabweichung gibt an, wie stark Messwerte um ihren Mittelwert streuen.
Trennschärfe	Die Trennschärfe sagt aus, wie gut ein Item zwischen Personen mit unterschiedlich hoher Merkmalsausprägung unterscheidet. Dafür muss ein Item eine mittlere Schwierigkeit aufweisen, denn wenn es zu leicht (alle lösen es) oder zu schwer (niemand löst es) ist, lassen sich eigentlich vorhandene Merkmalsunterschiede nicht messen.
Validität, prädiktiv	Die prädiktive Validität bezeichnet die Vorhersagegüte eines Tests (einer Skala, eines Items). Sie variiert zwischen 0 und 1. Je höher die prädiktive Validität, desto besser sagt der Test ein Kriterium (beispielsweise den Berufserfolg) vorher.
Varianzanalyse	Die Varianzanalyse ist ein statistisches Verfahren, um Gruppenunterschiede zu identifizieren. Es wird verwendet, wenn mehr als zwei Gruppen miteinander verglichen werden.

Literaturverzeichnis

- AMELANG, M. & SCHMIDT-ATZERT, L. (2006). Psychologische Diagnostik und Intervention. Berlin: Springer.
- ASENDORPF, J. (1999). Psychologie der Persönlichkeit (2. Auflage). Berlin: Springer.
- BARRICK, M. R., & MOUNT, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111–118.
- BARRICK, M.R., PIOTROWSKI, M. AND STEWART, G.L. (2002), "Personality and job performance: test of the mediating effects of motivation among sales representatives", *Journal of Applied Psychology*, 87(1), 43-52.
- BELLANI, M. L., FURLANI, F., GNECCHI, M., PEZZOTTA, P., TROTTI, E. M., & BELLOTTI, G. G. (1996). Burnout and related factors among HIV/AIDS health care workers. *AIDS Care*, 8, 207–221.
- BOOTH-KEWLEY, S., LARSON, G.E., MIYOSHI, D.K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, 23, 463-477.
- BORKENAU, P., EGLOFF, B., EID, M., HENNING, J., KERSTING, M., NEUBAUER, A. C., SPINATH, F. M. (2005). Persönlichkeitspsychologie: Stand und Perspektiven. *Psychologische Rundschau*, 56, 271-290.
- BORTZ, J. (1999). Statistik für Sozialwissenschaftler (5.Auflage). Berlin: Springer.
- BOUDREAU, J.W., BOSWELL, W.R., JUDGE, T.A., & BRETZ, R.D., JR. (2001). Personality and cognitive ability as predictors of job search among employed managers. *Personnel Psychology*, 54, 25-50.
- BUEHNER, M., KRUMM, S., & PICK, M. (2005). Reasoning=working memory ≠ attention. *Intelligence*, 33(3), 251-272.

- BUSATO, V. V., PRINS, F. J., ELSHOUT, J. J., & HAMAKER, C. (1999). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences*, 26, 129–140.
- CARLSTEDT, B., GUSTAFSSON, J.-E., & ULLSTADIUS, E. (2000). Item sequencing effects on the measurement of fluid intelligence. *Intelligence*, 28, 145–160.
- CARPENTER, P. A., JUST, M. A., & SHELL, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- CHIPUER, H.M., ROVINE, M.J. & PLOMIN, R. (1990). Lisrel modeling: Genetic and environmental influences on IQ. revisited. *Intelligence*, 14, 11-29.
- CHUAH, S. C., DRASGOW, F., & ROBERTS, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, 40, 359-376.
- COLLINS, C. J., HANGES, P. J., & LOCKE, E. A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human Performance*, 17 (1), 95-117.
- COLOM, R., JUAN-ESPINOZA, M ABAD, F.J. & GARCIA, L.F. (2000). Negligible sex differences in general intelligence. *Intelligence*, 28, 57-68.
- CORTINA, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- COSTA, P. T., JR., & MCCRAE, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI) professional manual. Odessa, FL: Psychological Assessment Resources.
- COSTA, P. T., JR., ZONDERMAN, A. B., & MCCRAE, R. R. (1991). Personality, defense, coping, and adaptation in older adulthood. In E. M. Cummings, A. L. Greene, & K. H. Karraker (Eds.), *Life-span developmental psychology: Perspectives on stress and coping* (pp. 277–293). Hillsdale, NJ: Erlbaum.

- CSIKSZENTMIHALYI, M. (1975). *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, San Francisco: Jossey-Bass.
- DECKER, P. J., & BORGES, F. H. (1993). Dimensions of work appraisal: Stress, strain, coping, job satisfaction, and negative affectivity. *Journal of Counseling Psychology*, 40, 470–478.
- DEMETRIOU, A., KYRIAKIDES, L., & AVRAAMIDOU, C. (2003). The missing link in relations between intelligence and personality. *Journal of Research in Personality*, 37, 547–581.
- DICKMAN, S.J., MEYER, D.E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology*, 54, 274-290.
- FEGER, H. & BREDEKAMP, J. (1983). *Enzyklopädie der Psychologie, Themenbereich B, Serie I: Forschungsmethoden der Psychologie*. Göttingen: Hogrefe
- FIELD, A. (2005). *Discovering statistics using SPSS (2nd edition)*. Sage Publications.
- FRAZIER, T. W., & YOUNGSTROM, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, 35(2), 169-182.
- FRIELING, E., HOYOS, GRAF C. (1978). *Fragebogen zur Arbeitsanalyse (FAA)*. Bern: Verlag Hans Huber.
- GARDEN, A. (1997). Relationships between MBTI profiles, motivation profiles, and career paths. *Journal of Psychological Type*, 41, 3–16.
- GELISSEN, J.P.T.M., & GRAAF, P.M. DE (2006). Personality, social background, and occupational career success. *Social Science Research*, 35(3), 702-726.
- GIGNAC, G. E., STOUGH, C., & LOUKOMITIS, S. (2004). Openness, intelligence, and self-report intelligence. *Intelligence*, 32, 133–143.
- GOTTFREDSON, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, 24, 13–23.

- GREY, E.M., TRIGGS, T.J., HAWORTH, N.L. (1989). Driver Aggression – The Role of personality, social characteristics, risk and motivation. Report No. CR 81, Transport and Communications, Federal Office of Road Safety: Canberra.
- HAMEL, R. & SCHMITTMANN, V.D. (2006). The 20-Minute Version as a Predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, 66(6), 1039-1046.
- HEGGESTAD, E.D., GEORGE, E., & REEVE, C.L. (2006). Transient error in personality scores: Considering honest and faked responses. *Personality and Individual Differences*, 40, 1201-1211.
- HELL, B., SCHULER, H., BORAMIR, I. & SCHAAR, H. (2006). Verwendung und Einschätzung von Verfahren der internen Personalauswahl und Personalentwicklung im 10 Jahres-Vergleich. *Zeitschrift für Personalforschung*, 20(1), 58-78.
- HELMES, E., & HOLDEN, R. R. (2003). The construct of social desirability: one or two dimensions? *Personality and Individual Differences*, 34, 1015 - 1023.
- HOGAN, J., & HOLLAND, B. (2003). Using theory to evaluate personality and job-performance relations. *Journal of Applied Psychology*, 88, 100–112.
- HOUGH, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209–244
- HOUGH, L.M., EATON, N.K., DUNNETTE, M.D., KAMP, J.D., MCCLOY, R.A. (1990). Criterion-related validities of personality constructs and the effects of response distortion on those validities [Monograph]. *Journal of Applied Psychology*, 75, 581-595.
- HUNTER, J. E., & SCHMIDT, R. L. (2004). *Methods of meta-analysis: correcting error and bias in research findings* (second ed.). Thousand Oaks, CA: Sage Publications, Inc.
- HUNTER, J.E. & HUNTER, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.

- HUNTER, JOHN E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and. job performance. *Journal of Vocational Behavior*, 29(3), 340-362.
- JOHNSON, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 39, 103-129.
- JOHNSON, W. & BOUCHARD, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33, 393-436.
- JUDGE, T. A., & CABLE, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology*, 50, 359–394.
- JUDGE, T. A., MARTOCCHIO, J. J., & THORESEN, C. J. (1997). Five-Factor model of personality and employee absence. *Journal of Applied Psychology*, 82(5), 745-755.
- JUDGE, T.A., BONO, J.E., ILIES, R., & WERNER, M. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.
- KAIN, S. & NACHTWEI, J. (2009). Die Rolle von Kontrollvariablen in der Human Factors Forschung - Ein bewährtes Konzept in einem modernen Anwendungsfeld. In: Krause, B. und Metzler, P. (Hrsg.), *Empirische Evaluationsmethoden Band 13* (S. 5-14). ZeE Verlag: Berlin.
- KERSTING, M. (2004). Kosten und Nutzen beruflicher Eignungsbeurteilungen. In L.F. Hornke & U. Winterfeld (Hrsg.), *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 55-77). Heidelberg: Spektrum Akademischer Verlag.
- KERSTING, M. (2004). Zur Bedeutung der Validität und der sozialen Akzeptanz in der. Berufseignungsdiagnostik. *Zeitschrift für Personalpsychologie*, 3, 83-86.
- KERSTING, M. (2006). Stand, Herausforderungen und Perspektiven der Managementdiagnostik. *Personalführung*, 10, 16-27.
- KERSTING, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253.

- KICHUK, S. L., & WIESNER, W. H. (1997). The Big five personality factors and team performance: Implications for selecting successful product design teams. *Journal of Engineering and Technology Management*, 14(3), 195-221.
- KRISTOF-BROWN, L. AMY, RYAN D. ZIMMERMAN AND ERIN C. JOHNSON (2005), 'Consequences of Individuals' Fit at Work: A Meta-analysis of Person-job, Person-organization, Person-group, and Person-supervisor Fit', *Personnel Psychology*, 58 (2), 281-342 .
- KUHL, J. (1983). *Motivation, Konflikt und Handlungskontrolle*. Berlin.
- LIENERT, G. A. & RAATZ, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- MACKINTOSH, N. J., & BENNET, E., S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33(6), 663-674.
- MARTIN, B. A., BOWEN, C. C., & HUNT, S. T. (2002). How effective are people at faking on. personality questionnaires? *Personality and Individual Differences*, 32, 247-256.
- MCCLELLAND, D. C. (1987). *Human motivation*. Cambridge: Cambridge University Press.
- MOUNT, M. K., BARRICK, M. R., & STEWART, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, 11, 145-165.
- MUELLER-HANSON, R., HEGGESTAD, E. D., & THORNTON, G. C. III, (2003). Faking and selection: considering the use of personality from a select-in and a select-out perspective. *Journal of Applied Psychology*, 88, 348-355.
- NACHTWEI, J. (2008). *Ein experimenteller Ansatz in der kognitiven Ergonomie – Personenmerkmale als Kodeterminanten der Komplexität einer belebten Mikrowelt*. VDM Verlag Dr. Müller: Saarbrücken.
- NACHTWEI, J. & SCHERMULY, C. (2009). Acht Mythen über Eignungstests. *Harvard Business Manager*, 6-10.
- PLATE, T. (2006). *Evaluation der Eignungsdiagnostik bei der Personalauswahl von Unternehmensberatern*. Dissertation. Universität Mannheim, Fakultät für Sozialwissenschaften.

- ROTHSTEIN, M.G., GOFFIN, R.D., (2006). The use of. personality measures in personnel selection: What. does current research support. *Human Resource Management Review*, 16, 155-180.
- RYAN, A & TIPPINS, N. (2004). Attracting and selecting: What psychological research tells us? *Human Resource Management*, 43 (4), 305-318.
- RYAN, A.M., MCFARLAND, L. BARON, H. & PAGE, R. (1999). An international look at selection practices: nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359-391
- SALGADO, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30–43.
- SCHMIDT, F. L., & HUNTER, J. E. (1999). Theory testing and measurement error. *Intelligence*, 27, 183–198.
- SCHMIDT, F.L. AND HUNTER, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Research: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124(2), 262-274.
- SCHMITT, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 81-84.
- SCHULER, H. & HÖFT, S. (2004). Berufseignungsdiagnostik und Personalauswahl. In H. Schuler (Hrgs.). *Enzyklopädie der Psychologie, Themenbereich D, Serie III, Band 3: Organisationspsychologie - Grundlagen und Personalpsychologie*, S. 339-532. Göttingen: Hogrefe.
- SCHULER, H. (2001). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.
- SEIBERT, S. & KRAIMER, M.L. (2001). The Five-Factor Model of Personality and Career Success. *Journal of Vocational Behavior*, 58, 1-21.
- STERNBERG, R.J. (1997). Managerial intelligence: Why IQ isn't enough. *Journal of Management*, 23(3), 475–493.

- TE NIJENHUIS, J., VAN VIANEN, A.E.M., VAN DER FLIER, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, 35, 283-300.
- VARDAMAN, J. M., ALLEN, D. G., RENN, R. W. & MOFFITT, K. R. (2008). Should I stay or should I go? The role of risk in employee turnover decisions. *Human Relations*, 61(11), 1531-1563.
- WESTHOFF, K., HELLFRITSCH, L.J. , HORNKE, L.F. , KUBINGER, K.D. , LANG, F., MOOSBRUGGER, H., PÜSCHEL, A., REIMANN, G. (Hrsg.). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*. 2. Auflage. Lengerich: Pabst.
- WOODS, S. A., & HAMPSON, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality*, 19(5), 373-390.

Index

A

Abstrakt-Analytische Intelligenz 33, 41
Allgemeine Gleichbehandlungsgesetz (AGG)
8, 9, 49
Anforderungsanalyse 15, 16, 60
Anforderungsanalyse, strukturierte 17
Anforderungsprofil 15, 17, 60

B

Big Five 19, 27, 53, 54
Burnout 21

C

Client-Lösung 58

D

Dimension 17
DIN 33430 7, 9, 16
Diskriminierung
mittelbare 8
unmittelbare 8

E

eignungsdiagnostischer Prozess 58
Eignungstest 12, 15
Emotionale Belastbarkeit 20, 27
Emotionale Stabilität 20, 27
Ermittlung, empirische 17
Expertenevaluation 17
Extraversion 19, 21, 28

F

Faktor

g-Faktor 31, 32, 35

s-Faktor 31

Faktorenanalyse 37, 38

Flow-Kanal 15

Flow-Modell 15

Formel 13

G

Geistige Flexibilität 20, 22, 28

Gewissenhaftigkeit 19, 22, 28

grafischer Profilvergleich 62

Gültigkeit 53

Guttman-Skala 40

I

Intelligenz 10

Abstrakt-Analytische 33, 41

fluide 31, 32

kristalline 31, 32

Numerische 34, 42

Verbale 34, 42

Intelligenzskala 52

Intelligenztest 10

Internet 13

Internetdiagnostik 13

Internetversion 13

Item 37, 38, 39

Itemschwierigkeit 39

Itemschwierigkeitsanalyse 37

K

Konsistenz, interne 39, 52

Konstruktion 37, 38

Konstruktionsstrategie 37
Konstruktvalidität 53
Korrelation 14, 26
Korrelationsanalyse 37
Kriteriumsvalidität 54
Kurzsкала 52

L

Längsschnittuntersuchung 27
Langzeitstudie 55
Leistungsfähigkeit, kognitive 31
Leistungsmotivation 19, 23, 29
Lügenskala 45

M

Merkmale
berufsrelevante 8
demografische 8
Persönlichkeits- 18, 19, 20
Stellen- 18
Messfehler 43
Methoden-Mix 6, 11

N

Neurotizismus 19
Norm 7
Normierung 48
Normstichprobe 48, 49, 65
Normstichprobenrevision 52
Numerische Intelligenz 34, 42
Nutzen 13, 14, 26

O

Objektivität 48
Offenheit 19

P

Passung 15
Persönlichkeit 10, 19
Persönlichkeitsmerkmal 18, 19, 20
Persönlichkeitsskala 38, 52
Persönlichkeitstest 10, 19
Potentialanalyse 9, 31
Prozentrang 48, 49, 62
Prozentrangwert 60, 62
Prozentwert 62

Q

Qualität 7, 11, 15, 26
Qualitätskriterium 57

R

Random Response Error 44
Redundanz 20
Reliabilität 51
Reliabilitätsanalyse 37
Risikoneigung 19, 24, 29
Rotationsmethode 38

S

Schulung 59
Schwierigkeit 39, 40
Selektionsstichprobe 37
Server-Lösung 58
Skala 18, 20
Soziale Erwünschtheit 44, 46
Specific Factor Error 44
Standardisierung 12
Stellenanforderung 15
Stellenmerkmal 18
Studie 11

T

Teamorientierung 20, 25
Testgüte 48
Testkonstruktion 37, 48
Testzertifikat 61
Transient Error 44
Trennschärfeanalyse 37, 39

U

Überforderung 15
Unterforderung 15

V

Validierungsstudie 56
Validität 11, 26, 53
Validität, prädiktive 26, 39, 51, 54
Varianzanalyse 37
Verbale Intelligenz 34, 42
Verfahren
 biographieorientierte 9
 eigenschaftsorientierte 9
 simulationsorientierte 9
Verfälschungstendenz 45, 46
Vergleichsgruppe 49, 65
Verträglichkeit 19
Vorhersagegenauigkeit 13
Vorhersagequalität 26, 35

Z

zeitliche Stabilität 52
Zuverlässigkeit 51
Zuverlässigkeitsberechnung 51