

Master-Thesis

Analyse des inkrementellen Nutzens eines Persönlichkeitsfragebogens in der Managementdiagnostik

eingereicht am 31.05.2023 über den eCampus

Prüfer **Prof. Dr. Manfred Mühlfelder**

SRH Fernhochschule

Modul: Master-Thesis

Studiengang: Wirtschaftspsychologie (M. Sc.)

von

Dagmar Rostek

Klosterstr. 12

52511 Geilenkirchen

Studiengang: Wirtschaftspsychologie (M. Sc.)

Matrikelnummer: 2004169

Abstract

Diese Masterarbeit untersucht den Mehrwert eines Persönlichkeitsfragebogens in der Managementdiagnostik und evaluiert die Effizienz und Prädiktionsfähigkeit diverser Führungskräfte-Auswahlmethoden, einschließlich Assessment-Centern, Selbstauskunftsbögen und künstlicher Intelligenz (KI)-gestützten Systemen. Ziel der Arbeit ist es, durch umfassende Analyse und Vergleich der genannten Methoden eine Forschungslücke zu adressieren und zur Verbesserung der Auswahlverfahren beizutragen. Es wurden teilweise Übereinstimmungen zwischen Selbst- und Fremdeinschätzungen bezüglich bestimmter Führungskompetenzen festgestellt. Gleichwohl blieb die Klarheit des Mehrwerts zusätzlicher Daten, die durch die Kombination von Assessment-Centern und Selbstauskunftsmethoden gewonnen wurden, aus. Erste Untersuchungen deuten darauf hin, dass KI-basierte Systeme eine bestimmte Vorhersagegenauigkeit für die Eignung von Führungskräften besitzen. Weitere Forschung ist notwendig, um diese Befunde zu untermauern und weiter zu vertiefen.

Schlüsselwörter: Managementdiagnostik, Persönlichkeitsfragebogen (LeadIM), Assessment-Center, KI-gestützte Systeme

ABBILDUNGSVERZEICHNIS	5
TABELLENVERZEICHNIS	7
GENDER-HINWEIS	8
1. EINLEITUNG	9
1.1 PROBLEMSTELLUNG	9
1.2 ZIELSETZUNG DER ARBEIT UND LEITFRAGEN	10
1.3 ÜBERSICHT ÜBER DIE ARBEIT	11
2. THEORETISCHER TEIL	13
2.1 EIGNUNGSDIAGNOSTIK	13
2.1.1 DER EINSATZ VON KÜNSTLICHER INTELLIGENZ IN DER EIGNUNGSDIAGNOSTIK 15	
2.2 ASSESSMENT-CENTER	18
2.2.1 ENTWICKLUNG EINES ASSESSMENT CENTERS	21
2.2.2 ENTWICKLUNG UND AUSWAHL DER AUFGABEN	21
2.3 PRÄSENTATIONEN	22
2.4 DAS INTERVIEW	23
2.5 DAS ROLLENSPIEL	26
2.5.1 ROLLENSPIELARTEN	28
2.5.2 ROLLENSPIELKONZEPTION	28
2.6 BEOBACHTUNG	30
2.6.1 BEOBACHTUNG VON ROLLENSPIELEN	34
2.6.2 ABLAUF DER VERHALTENSBEOBACHTUNG	34
2.7 DURCHFÜHRUNG EINES ASSESSMENT CENTERS	37
2.8 GESAMTURTEIL	38
2.8.1 RÜCKMELDUNG AN DIE TEILNEHMER	39
2.8.2 ASSESSMENT CENTER VERSUS DEVELOPMENT CENTER	40
2.9 SELBSTAUSKUNFTSFRAGEBÖGEN	42
2.9.1 LEADIM FRAGENBOGEN	43
2.10 ERGEBNISSE FORSCHUNGSPROJEKT	44
2.11 EIGENE FRAGESTELLUNGEN	46

2.11.1	PRÄZISIERUNG DER ZIELSETZUNG	46
3.	METHODE UND VORGEHEN	47
3.1	FORSCHUNGSDESIGN	48
3.2	QUALITATIVES EXPERTENINTERVIEW	48
3.3	KONZEPTION DES ROLLENSPIELS.....	50
3.4	TEAMANSPRACHE.....	52
3.4.1	STICHPROBE.....	53
3.5	ASSESSMENT CENTER ABLAUF (2 SEITEN).....	54
4.	ERGEBNISSE.....	56
4.1	DESKRIPTIVE STATISTIK ZU DEN ERHOBENEN VARIABLEN.....	56
4.2	REPLIKATION DER ERGEBNISSE DES FORSCHUNGSPROJEKTES.....	59
4.3	BEANTWORTUNG DER LEITFRAGEN	64
4.3.1	ERGEBNISSE ZU HYPOTHESE 1	65
4.3.2	ERGEBNISSE ZU HYPOTHESE 2	69
4.3.3	ERGEBNISSE ZU HYPOTHESE 3	72
5.	DISKUSSION.....	76
5.1	INTERPRETATION ERGEBNISSE	76
5.2	QUALITÄT DES EIGENEN VORGEHENS (OBJEKTIVITÄT, RELIABILITÄT, VALIDITÄT).....	80
5.3	AUSBLICK UND WEITERES VORGEHEN.....	82
	LITERATURVERZEICHNIS.....	84
	ANHANG A BEURTEILUNGSFEHLER UND WAHRNEHMUNGSVERZERRUNGEN	89
	ANHANG B TEAMANSPRACHE.....	91
	ANHANG C MITARBEITERGESPRÄCH	93
	ANHANG D INTERVIEWLEITFADEN	94
	EIDESSTATTLICHE ERKLÄRUNG	103

Abbildungsverzeichnis

Abbildung 1 Architektur der DIN 33430 (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 7)	14
Abbildung 2 Grobe Einschätzung der Objektivität verschiedener Methoden (Quelle: Kanning, 2019, S. 201)	17
Abbildung 3 Verfahrenstypen (Quelle: In Anlehnung an Obermann, 2018, S.6)	19
Abbildung 4 Präsentationsthemen (Quelle: modifizierte Darstellung Obermann, 2018, S. 134)	23
Abbildung 5 Schlussfolgerung von beobachtbarem Verhalten auf ein Eignungsmerkmal (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 31)	30
Abbildung 6 Wesentliche Arbeitsschritte bei der Verhaltensbeobachtung und -beurteilung (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 35)	32
Abbildung 7 Beispiele für numerische Urteile (Operationalisierungsansatz und verhaltensverankerte Beurteilungsskalen) (Quelle: Diagnostik- u. Testkuratorium, 2017, S. 49)	33
Abbildung 8 Detailablauf einer Verhaltensbeobachtung und -beurteilung mit einer Markierung möglicher Fehlerquellen (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 52)	37
Abbildung 9 Welche Informationen erhält der Teilnehmer vor dem AC? (Quelle: Obermann et al. 2012, zitiert nach Obermann, 2018, S. 247)	38
Abbildung 10 Beispiel qualitative Ergebniszusammenfassung (Quelle: modifizierte Darstellung Obermann, 2018, S. 258)	41
Abbildung 11 Exemplarische Darstellung der Ergebnisse eines Teilnehmers, die Straffierten Felder wurden in den jeweiligen Settings nicht erhoben (Quelle: eigene Darstellung)	56
Abbildung 12 Alter der Teilnehmer (Quelle: Ausgabe SPSS)	57
Abbildung 13 höchster Bildungsabschluss (Quelle: Ausgabe SPSS)	58
Abbildung 14 Beruflicher Status (Quelle: Ausgabe SPSS)	58
Abbildung 15 Rolle (Quelle: Ausgabe SPSS)	59
Abbildung 16 Erklärte Gesamtvarianz (Quelle: Ausgabe SPSS)	60
Abbildung 17 Kommunalitäten (Quelle: Ausgabe SPSS)	62
Abbildung 18 Komponentenmatrix, nach Koeffizientengröße sortiert (Quelle: Modifizierte Darstellung Ausgabe SPSS)	63
Abbildung 19 Modell Zusammenfassung AC versus LeadIM (Quelle: Ausgabe SPSS)	70
Abbildung 20 Omnibus-Test (Quelle: Ausgabe SPSS)	73

Abbildung 21 Modellzusammenfassung (Quelle: Ausgabe SPSS)	73
Abbildung 22 Regressionskoeffizienten (Quelle: Ausgabe SPSS)	74

Tabellenverzeichnis

Tabelle 1 Matrix: Mit welchen Modulen können einzelne Kompetenzen gemessen werden? (Quelle: in Anlehnung an Obermann, 2018, S. 112)	22
Tabelle 2 Komponenten und erklärte Varianz aus 16 LeadIM Variablen (Quelle: eigene Darstellung)	61
Tabelle 3 Komponenten und erklärte Varianz aus 16 LeadIM Variablen (Quelle: Modifizierte Darstellung SPSS)	63
Tabelle 4 Rotierte Komponentenmatrix, N=130 (Quelle: Modifizierte Darstellung Ausgabe SPSS)	64
Tabelle 5 Spearman Korrelation (Quelle: Modifizierte Darstellung Ausgabe SPSS)	66
Tabelle 6 Spearman (Quelle: Modifizierte Darstellung Ausgabe SPSS)	68
Tabelle 7 AC-Schätzungen der vier Komponenten (als z-Werte) (Quelle: eigene Darstellung)	69
Tabelle 8 Regressionskoeffizienten, logistische Regression, LeadIM, AC (Quelle: modifizierte Darstellung Ausgabe SPSS)	70
Tabelle 9 Regressionskoeffizienten, logistische Regression, LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)	71
Tabelle 10 Klassifikationstabelle der Eignung für eine Teamleiterrolle AC + LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)	71
Tabelle 11 Klassifikationstabelle der Eignung für eine Teamleiterrolle LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)	72
Tabelle 12 Regressionskoeffizienten, logistische Regression, LeadIM.....	75
Tabelle 13 Klassifikationstabelle Eignung für eine Teamleiterrolle.....	75

Gender-Hinweis

Zur besseren Lesbarkeit wird in dieser Hausarbeit das generische Maskulinum verwendet. Die in dieser Arbeit verwendeten Personenbezeichnungen beziehen sich – sofern nicht anders kenntlich gemacht – auf alle Geschlechter.

1. Einleitung

In der heutigen Wirtschaftslandschaft ist die Identifizierung von Führungskräften, die sowohl auf fachlicher Ebene als auch hinsichtlich kultureller Passung und Führungspotential passen, eine essenzielle und gleichzeitig komplexe Herausforderung. Die Vielfalt der beteiligten Faktoren, die Folgen von Fehlentscheidungen und die zunehmende Bedeutung neuer technologischer Ansätze, wie der Einsatz von künstlicher Intelligenz (KI), machen die Personaleignungsdiagnostik zu einem Schlüsselthema sowohl in der wissenschaftlichen Forschung als auch in der praktischen Anwendung.

Trotz zahlreicher Untersuchungen bleibt eine zentrale Frage offen, die sich auf die Wirksamkeit und Vorhersagegenauigkeit der gängigen Methoden - das Assessment Center, die Selbstauskunftsfragebögen und KI-basierten Ansätze - bezieht. Die vorliegende Masterarbeit setzt genau hier an, um einen Beitrag zur Klärung dieser Forschungslücke zu leisten und zugleich einen Beitrag zur Optimierung der Praxis der Personaleignungsdiagnostik zu liefern. Nachfolgend wird die Problemstellung detailliert dargelegt, um den genauen Forschungskontext und die damit verbundenen Fragestellungen zu umreißen.

1.1 Problemstellung

Die Personaleignungsdiagnostik bei Führungskräften ist ein komplexer und aufwändiger Prozess, der die fachliche Passung, den culture fit¹ und das Potenzial der Kandidaten berücksichtigt (Sarges, 2013; Schuler, 2014). Die Effektivität des Führungshandelns ergibt sich aus einer Kombination verschiedener Variablen wie organisationsbezogenen, situativen und kulturellen Faktoren sowie der Persönlichkeit und Haltung der betreffenden Person. Mit steigendem Karrierelevel nehmen die Anforderungen an die Gespräche zu und verschiedene Entscheider sind in der Regel in den Prozess der Personaleignungsdiagnostik involviert. Eine Fehlentscheidung bei der Auswahl einer Führungskraft kann erhebliche finanzielle Folgen haben, von mehr als 50.000 Euro oder im Management sogar bis zum Dreifachen des Jahresgehalts (Dahmen, 2016). In einer Metaanalyse untersuchten Kristof-Brown, Zimmerman und Johnson (2005) die Auswirkungen von individueller Passung in verschiedenen Bereichen, einschließlich Person-Job Fit, Person-Organisation Fit, Person-Gruppe Fit und Person-Vorgesetzter Fit. Die Ergebnisse zeigten, dass eine erhöhte Übereinstimmung in diesen Dimensionen

¹ Der Cultural Fit Test untersucht die Übereinstimmung der individuellen Wertvorstellungen von Bewerbern mit denen des Unternehmens (Haufe, 2019). Neben fachlichen Kompetenzen und Softskills sind kulturelle Übereinstimmungen für eine erfolgreiche Zusammenarbeit relevant. Der Test berücksichtigt sowohl explizite Aspekte der Unternehmenskultur wie Vision und Strategie, als auch implizite Aspekte, wie Entscheidungsfindung und Traditionen.

mit positiven Outcomes, wie gesteigerter Arbeitszufriedenheit, verbesserter Arbeitsleistung und reduzierter Fluktuation, korreliert ist. Um Fehlentscheidungen zu minimieren und die geeigneten Kandidaten zu identifizieren, werden häufig Assessment Center eingesetzt, die als Verfahren zur Eignungs- oder Potenzialbeurteilung dienen (Obermann, 2018). Zusätzlich werden Selbstauskunftsfragebögen in Kombination mit Assessment Centern verwendet, um eine möglichst umfassende Einschätzung der Kandidaten zu erhalten. In jüngster Zeit ist auch der Einsatz von künstlicher Intelligenz (KI) in der Personaleignungsdiagnostik zunehmend relevant geworden, was laut Kanning in den kommenden Jahren eines der intensivsten Themen in der Eignungsdiagnostik sein wird (Tirrel, et al., 2021, S. 17). KI-Technologien bieten zweifellos Potenzial, wobei die zentrale Frage ist, ob sie tatsächlichen Nutzen bringen. Die Objektivität mag in vielen Fällen gegeben sein, doch bereits bei der Reliabilität beginnen die Schwierigkeiten, insbesondere wenn Anbieter keine Informationen dazu liefern. Es existieren keine verlässlichen Beweise für eine nennenswerte prognostische Validität von KI-basierten Verfahren. Derzeit sind diese Methoden weit davon entfernt, bessere Ergebnisse zu liefern als eine konsequente Anwendung wissenschaftlich fundierter Eignungsdiagnostik. Sollten sie jedoch prognostische Validität besitzen, könnten KI-Verfahren höchstens das bestehende Methodenspektrum erweitern, aber die Personalauswahl nicht grundlegend revolutionieren - zumindest nicht, wenn sich Anwender von der tatsächlichen Leistungsfähigkeit der Verfahren leiten lassen und nicht von aktuellen Trends. Zukünftig muss sich zeigen, welche KI-Verfahren tatsächlich sinnvoll sind oder wie sie gestaltet werden müssen, um Nutzen zu bringen. Dazu wird vor allem unabhängige Forschung benötigt. Wenn die Ergebnisse unabhängiger Forschung auf die Expertise kompetenter Kunden treffen, könnte sich möglicherweise ein realer Nutzen entwickeln (Kanning, 2021, S. 26f). In der aktuellen Forschungslandschaft bleibt eine wesentliche Frage offen: Welcher der Ansätze - Assessment Center, Selbstauskunftsfragebögen oder KI-basierte Methoden - liefert die größte Varianzaufklärung bei der Vorhersage der Eignung für eine Führungsrolle? Obwohl frühere Studien darauf hinweisen, dass Assessment Center und Selbstauskunftsfragebögen als effektive Instrumente dienen können, mangelt es an umfassender Forschung, die diese Ansätze in direktem Vergleich mit KI-basierten Verfahren setzt. Diese Forschungslücke stellt eine erhebliche Herausforderung für die Praxis der Personaleignungsdiagnostik dar und beeinträchtigt die Optimierung von Auswahlprozessen in Organisationen.

1.2 Zielsetzung der Arbeit und Leitfragen

Angeichts der offenen Fragen im Bereich der Personaleignungsdiagnostik und der Identifizierung potenzieller Führungskräfte konzentriert sich diese Masterarbeit darauf,

die Varianzaufklärung verschiedener Methoden, darunter Assessment Center, Selbstauskunftsfragebogen und ein KI-basiertes Verfahren, zu untersuchen und zu vergleichen. Die daraus resultierenden Erkenntnisse sollen einen Beitrag zum Verständnis dieser Prozesse leisten und eine solide Grundlage für effizientere Auswahlverfahren in Organisationen schaffen.

Die zentrale Forschungsfrage, die diese Arbeit zu beantworten versucht, lautet daher: "Wie unterscheiden sich Assessment Center, Selbstauskunftsfragebögen und eine KI-basierte Methode in ihrer prädiktiven Validität bei der Ermittlung der Eignung von Führungskräften, und inwiefern kann eine kombinierte Anwendung dieser Verfahren die Vorhersagegenauigkeit optimieren?"

Zur Beantwortung dieser Frage werden mehrere Hypothesen aufgestellt und überprüft, darunter die Annahme, dass das LeadIM Profil (Selbstbeschreibung) mit der Fremdeinschätzung im Assessment Center korreliert, und dass die Einbeziehung von Daten aus dem LeadIM Profil in eine multiple Regression die Varianzaufklärung hinsichtlich der Eignung für eine Teamleiterrolle signifikant erhöht. Darüber hinaus wird angenommen, dass die KI-basierte Einschätzung der Eignung als Führungskraft mit dem qualitativen Assessment übereinstimmt. Im Rahmen dieser Arbeit wird ein qualitativer Forschungsansatz angewandt, um diese Hypothesen zu testen. Hierbei werden Daten von N = 10 Kandidaten, die an einem 3-stündigen Assessment Center teilnehmen, erhoben und analysiert. Der Fokus liegt auf der Erfassung von vier Komponenten der Führung gemäß dem LeadIM Fragebogen: aktivierende Führung, selbstbewusste / reflektierende Führung, strukturierende / organisierende Führung und empathische Führung. Die Ergebnisse dieser Studie werden voraussichtlich wichtige Implikationen für die Personaleignungsdiagnostik und die Führungskräfteentwicklung in der Praxis haben. Sie werden aufzeigen, ob die Kombination aus Selbstauskunftsfragebogen und KI-gestützten Verfahren eine effektive und kosteneffiziente Alternative zu zeitaufwendigen und teuren Assessment Centern darstellen kann.

1.3 Übersicht über die Arbeit

Die vorliegende Arbeit konzentriert sich auf die Untersuchung und den Vergleich der Varianzaufklärung verschiedener Methoden, einschließlich Assessment-Centern, Selbstauskunftsfragebögen und KI-basierten Verfahren. Sie setzt sich aus den Hauptabschnitten Einleitung, theoretischer Teil, Methodik und Vorgehen, Ergebnisse und Diskussion zusammen.

Im einführenden Teil wird zunächst die Problemstellung detailliert dargelegt, gefolgt von der Festlegung der Zielsetzung dieser Arbeit und der Formulierung der Leitfragen, die in diesem Kontext beantwortet werden sollen. Die Einleitung endet mit einer umfassenden Übersicht über die gesamte Arbeit.

Der theoretische Teil, welcher den größten Teil der Arbeit ausmacht, führt den Leser durch eine Vielzahl von Konzepten und Methoden im Bereich der Eignungsdiagnostik. Hierbei beginnt die Arbeit mit einer generellen Einführung in die Eignungsdiagnostik und den Einsatz von KI in diesem Bereich. Darüber hinaus behandelt der Theorieteil Themen wie Assessment-Center, Entwicklung und Auswahl von Aufgaben, Durchführung von Präsentationen, Interviews und Rollenspielen sowie Methoden zur Beobachtung. Weiterhin wird das Konzept der Selbstauskunftsfragebögen und die Auswertung der Ergebnisse diskutiert. Dieser Abschnitt schließt mit der Präzisierung der Zielsetzung und den spezifischen Fragestellungen, die im Rahmen dieser Arbeit untersucht werden.

Im Abschnitt "Methode und Vorgehen" wird das gewählte Forschungsdesign sowie die angewandte Methodik detailliert erläutert. Hierbei werden die Durchführung von qualitativen Experteninterviews und die Konzeption von Rollenspielen behandelt. Die Abfolge eines Assessment-Centers wird beschrieben.

Im Ergebnisteil werden die deskriptiven Statistiken der erhobenen Variablen präsentiert. Ein wichtiger Aspekt dieses Abschnitts ist die Replikation der Ergebnisse des Forschungsprojekts und die Beantwortung der formulierten Leitfragen. In diesem Kontext werden die Ergebnisse in Bezug auf die vorgelegten Hypothesen diskutiert.

Der abschließende Abschnitt "Diskussion" bietet eine Interpretation der Ergebnisse und bewertet die Qualität des eigenen methodischen Vorgehens hinsichtlich Objektivität, Reliabilität und Validität. Hier werden die gewonnenen Erkenntnisse zusammengefasst und ein umfassender Überblick über die gesamte Arbeit gegeben, einschließlich der Einschränkungen und Perspektiven für zukünftige Forschungen.

Insgesamt handelt es sich bei der vorliegenden Arbeit um eine umfassende Untersuchung und Vergleich der Varianzaufklärung verschiedener Methoden, einschließlich Assessment-Centern, Selbstauskunftsfragebögen und KI-basierten Verfahren.

2. Theoretischer Teil

Im theoretischen Teil wird der theoretische Hintergrund der vorliegenden Arbeit dargelegt, der die Basis für die empirischen Untersuchungen darstellt. Dabei wird auf die relevanten Theorien und Modelle eingegangen, die das Fundament der untersuchten Thematik bilden. Sie bieten die notwendigen konzeptuellen Rahmenbedingungen, um die erhobenen empirischen Daten adäquat zu interpretieren. Darüber hinaus wird auf die Forschungsliteratur zu den betreffenden Themen eingegangen. Schließlich werden die spezifischen Forschungsfragen und Hypothesen, aus dem theoretischen Hintergrund heraus entwickelt und präsentiert.

2.1 Eignungsdiagnostik

Im Bereich der Eignungsdiagnostik sind die Konzepte "Potenzial" und "Kompetenz" von zentraler Bedeutung. "Potenzial" bezeichnet hier das zukünftige Leistungsniveau eines Mitarbeiters und dessen Fähigkeit, seine Leistung auf neue Aufgaben und Herausforderungen zu übertragen. "Kompetenz" hingegen steht für die allgemeinen Fähigkeiten oder Fertigkeiten einer Person, die sie in der aktuellen Situation aufweist (Obermann & Solga, 2018, S. 137).

Im Rahmen der Eignungsdiagnostik geht es darum, zu ermitteln, welche individuellen Eigenschaften mit beruflichem Erfolg korrelieren, um darauf basierend die Auswahl und Entwicklung von Personal zu unterstützen (Cyrus, 2015, S. 59). Besonders in Bezug auf Führungspotenzial ist es wichtig, Persönlichkeitsfaktoren zu identifizieren und zu messen, die eine hohe prognostische Aussagekraft für die zukünftige berufliche Entwicklung haben (Obermann & Solga, 2018, S. 137 ff).

Die Qualität in der Eignungsdiagnostik ist sowohl für die Organisation als auch für den arbeitenden Menschen von fundamentaler Bedeutung (Diagnostik- u. Testkuratorium, 2017). Um diese Qualität sicherzustellen, dient die DIN 33430 als maßgebliche Norm im Bereich der Eignungsdiagnostik. Sie basiert auf systematischen Auswertungen nationaler und internationaler Erfahrungen und identifiziert "besonders bewährte Vorgehensweisen der Eignungsdiagnostik" (Diagnostik- u. Testkuratorium, 2017), die als Grundlage für diese Norm dienen. Abbildung 1 zeigt die Architektur der DIN 33430.

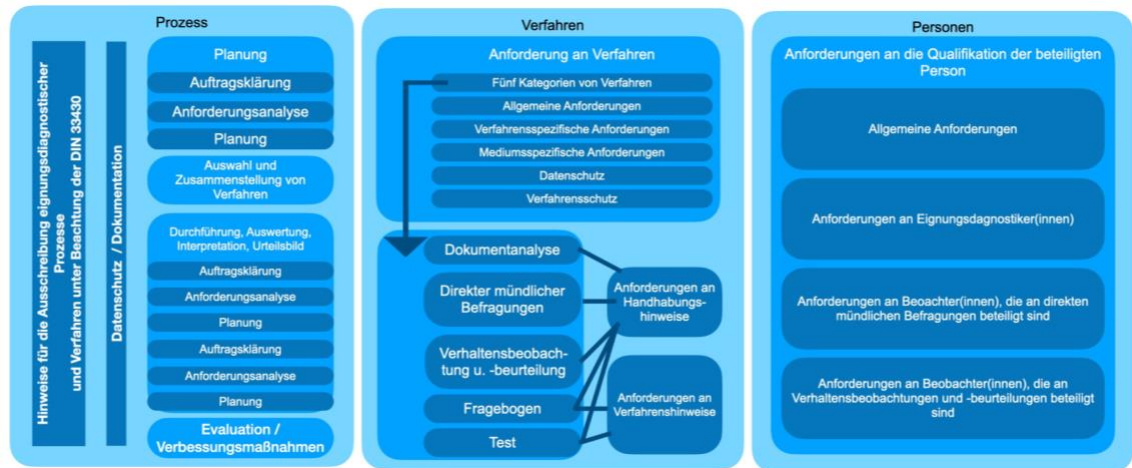


Abbildung 1 Architektur der DIN 33430 (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 7)

Der Prozess der Eignungsdiagnostik nach DIN 33430 umfasst verschiedene Stufen: Die Planung, die Auswahl und Zusammenstellung der Verfahren, die Durchführung, die Auswertung, die Interpretation und Urteilsbildung, die Evaluation und Implementierung von Verbesserungsmaßnahmen sowie Aspekte des Datenschutzes und der Dokumentation und Hinweise für die Ausschreibung eignungsdiagnostischer Prozesse.

Im Kontext der Verfahren werden fünf Kategorien in Betracht gezogen: allgemeine, verfahrens- und mediumsspezifische Anforderungen sowie daten- und verfahrensschutzspezifische Aspekte.

Bezüglich der beteiligten Personen wird die Qualifikation in den Fokus gerückt. Neben allgemeinen Anforderungen werden spezifische Anforderungen an Eignungsdiagnostiker und Beobachter, die an mündlichen Befragungen oder Verhaltensbeobachtungen beteiligt sind, gestellt. Das Anforderungsprofil ist spezifisch auf das Aufgabenfeld ausgerichtet, beispielsweise "für eine Ausbildung, ein Studium, einen Beruf" (Diagnostik- u. Testkuratorium, 2017, S. 28). Es charakterisiert die erwartete Leistung anhand von Eignungsmerkmalen und deren Ausprägungen.

Eignungsmerkmale umfassen "Qualifikationen, Kompetenzen und Potenziale sowie berufsbezogene Interessen, Bedürfnisse, Werthaltungen, Motive und andere Merkmale einer Person" (ebd.), die Voraussetzungen für die geforderte berufliche Leistung und die berufliche Zufriedenheit darstellen.

Das Eignungsprofil hingegen spiegelt die Eignungsmerkmale der Bewerberinnen und Bewerber wider (ebd.). Es wird mittels eines IST-SOLL-Vergleichs mit den anforderungsspezifischen Eignungsmerkmalen, auch als Konstrukte oder latente

Variablen bezeichnet, abgeglichen. Für eine solche Abgleichung ist es unerlässlich, die Eignungsmerkmale mithilfe von Verhaltensankerungen² zu präzisieren (ebd.).

Die Herausforderung der Eignungsdiagnostik liegt somit in der Auswahl adäquater Verfahren zur Erfassung dieser Eigenschaften und in der fachlichen Expertise ihrer Anwendung. Im Lichte dieser theoretischen Überlegungen und des anhaltenden technologischen Fortschritts wird im folgenden Abschnitt ein neuer Ansatz in der Eignungsdiagnostik beleuchtet: Der Einsatz von Künstlicher Intelligenz (KI) und dessen Auswirkungen und Potenziale für die Identifizierung und Beurteilung von Führungspotenzial.

2.1.1 Der Einsatz von Künstlicher Intelligenz in der Eignungsdiagnostik

Im Bereich der Eignungsdiagnostik bietet die Integration von Künstlicher Intelligenz (KI) vielseitige Anwendungsmöglichkeiten, etwa bei der gezielten Ansprache von Bewerbern, bei der Vorselektion auf Basis von Online-Daten oder bei der Bewertung von Bewerbungsschreiben (Kanning, 2021, S. 17ff). Zudem besteht das Potential, KI-Algorithmen in Bewerbungsgespräche einzubinden, die eventuell sogar von Computern geleitet werden könnten. Derartige Verfahren können nicht nur für externe Kandidaten verwendet werden, sondern auch bei internen Stellenbesetzungen Verwendung finden. Dabei könnten interne Daten wie Bildungshintergrund, Weiterbildung und bisherige Karrierestationen einbezogen werden. Allerdings sollte beachtet werden, dass trotz der Existenz von Anbietern und Nutzern von KI-Lösungen in diesen Bereichen, die Verbreitung eher von Marktinteressen als von Forschungsergebnissen oder tatsächlichen Kundenanforderungen angetrieben wird. Daher ist es dringend erforderlich, die tatsächlichen Fortschritte von den weniger überzeugenden Angeboten zu unterscheiden (Kanning, 2021, S. 17ff).

In diesem Kontext lassen sich KI-basierte Systeme in drei grundlegende Kategorien einteilen: Manuell programmierte, selbstlernende und kontinuierlich lernende Systeme. Während manuell programmierte Systeme explizite, menschlich definierte Entscheidungsregeln erfordern, erlernen selbstlernende Systeme eigenständig Zusammenhänge und Entscheidungskriterien aus Daten, entwickeln sich jedoch nach ihrer Implementierung nicht weiter. Kontinuierlich lernende Systeme hingegen adaptieren sich dynamisch an veränderte Datenlagen (Langer & Lazar, 2023, S. 315).

² Verhaltensanker stellen konkrete Beobachtungen dar, wie zum Beispiel, ob eine Person den Blickkontakt zu einer anderen Person halten kann. Dies kann als ein Indikator für die nicht direkt beobachtbare soziale Kompetenz dienen (Diagnostik- u. Testkuratorium, 2017, S. 28).

Meist liegt der Fokus auf einer Unterstützung des Menschen, statt einer vollständigen Automatisierung der Prozesse. Diese Unterstützung kann deskriptiv, prädiktiv oder präskriptiv sein. Sie verbessert das Verständnis der vorhandenen Daten, ermöglicht eine Verbesserung der Vorhersagegenauigkeit oder liefert konkrete Handlungsempfehlungen (Langer & Lazar, 2023, S. 315).

Insbesondere hinsichtlich der Objektivität, Reliabilität und prognostischen Validität identifiziert Kanning (2019) Potenziale durch den Einsatz von Künstlicher Intelligenz. Bei der Implementierung von KI in Eignungsdiagnostikprozessen wird sie mit dem Ziel eingesetzt, Kandidaten effizienter zu bewerten und deren Passung für bestimmte Positionen zu ermitteln. Algorithmen und maschinelles Lernen werden verwendet, um große Datenmengen zu analysieren und relevante Muster zu identifizieren, wodurch subjektive Vorurteile minimiert und objektivere Bewertungen ermöglicht werden.

Es ist jedoch zu beachten, dass KI die menschliche Urteilsfindung nicht ersetzen sollte, und ihre Ergebnisse stets in Kombination mit weiteren Bewertungsmethoden betrachtet werden sollten. Trotz der jährlichen Veröffentlichung von etwa 800 wissenschaftlichen Publikationen zum Thema Personalauswahl (Kanning, 2021, S. 18), gelingt es vielen Unternehmen nicht, diese Erkenntnisse in ihre Prozesse zu integrieren. So setzen Unternehmen häufig weiterhin unstrukturierte Interviews ein, obwohl strukturierte Interviews signifikant bessere Prognosen der beruflichen Leistung ermöglichen (Kanning, 2016, S. 18ff).

Die Objektivität in der Eignungsdiagnostik bezieht sich auf die Standardisierung des Auswahlprozesses, und eine höhere Objektivität ist gegeben, wenn das Ergebnis einer Messung weniger von dem diagnostischen Personal beeinflusst wird (Kanning, 2019, S. 196). Die Durchführungsobjektivität berücksichtigt Aspekte wie Begrüßung, räumliche Gestaltung, Fragestellungen, Hilfestellungen und Freundlichkeit (Kanning, 2019, S. 197). Die Auswertungsobjektivität bezieht sich auf die Auswertung der Untersuchung, beispielsweise die Berechnung von Punktwertsummen oder Einschätzungen auf einer Skala. Die Interpretationsobjektivität erfordert den Vergleich der Ergebnisse mit Anforderungsprofilen, die inhaltliche Interpretation oder die Anwendung statistischer Normen. In verschiedenen Forschungsmethoden lässt sich die Variation der Objektivität ohne den Einsatz von Künstlicher Intelligenz beobachten. Bei der Dokumentenanalyse ist die Durchführungsobjektivität besonders hoch, wohingegen die Auswertungs- und Interpretationsobjektivität weniger ausgeprägt sind.

Bei standardisierten Fragebögen und Testverfahren kann hingegen eine durchweg hohe Objektivität beobachtet werden. Auch das hochstrukturierte Interview sowie das Assessment-Center weisen im Vergleich zu unstrukturierten Interviews eine höhere Objektivität auf. Darüber hinaus ist es wichtig, mögliche Beurteilungsfehler und Wahrnehmungsverzerrungen zu berücksichtigen, um die Validität und Zuverlässigkeit der Eignungsdiagnostik zu erhöhen. Abbildung 2 zeigt eine grobe Einschätzung der Objektivität der unterschiedlichen Methoden. Die Integration von KI in den diagnostischen Prozess kann dazu beitragen, solche Fehler und Verzerrungen zu reduzieren. Dennoch sollte der Einsatz von KI stets in Kombination mit anderen Methoden erfolgen und unter Berücksichtigung menschlicher Expertise betrachtet werden.

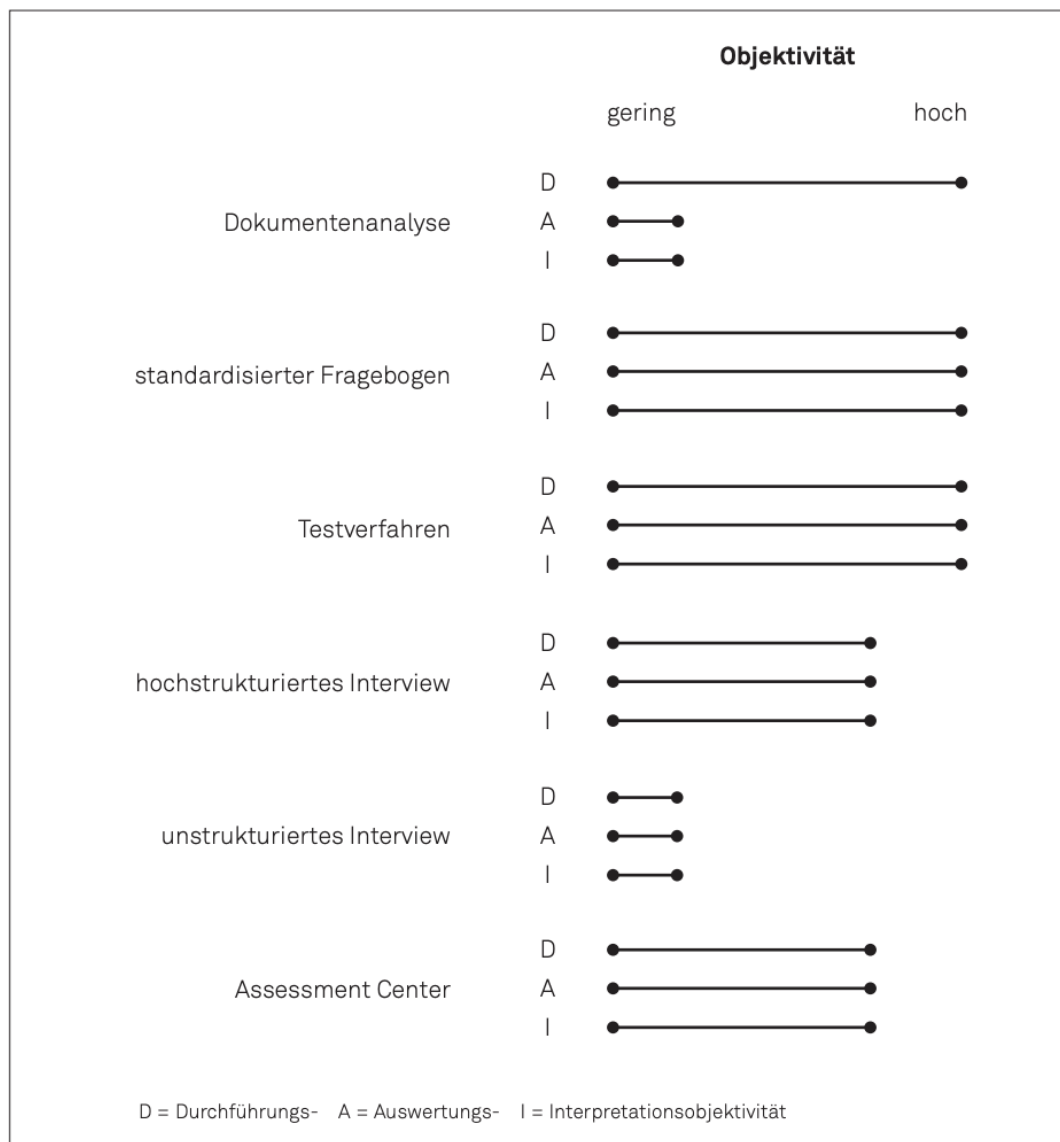


Abbildung 2 Grobe Einschätzung der Objektivität verschiedener Methoden (Quelle: Kanning, 2019, S. 201)

Nachdem der Einsatz von Künstlicher Intelligenz in der Eignungsdiagnostik beleuchtet wurde, soll nun ein spezielles Instrument der Personalauswahl in den Fokus gerückt werden, welches in der Praxis eine zentrale Rolle spielt: Das Assessment Center.

2.2 Assessment-Center

Das Assessment Center (AC) stellt ein zentrales Verfahren in der Personalauswahl und Potentialermittlung dar, welches durch die Kombination verschiedener Methoden und die Einbeziehung mehrerer Beobachter gekennzeichnet ist (Obermann, 2018, S. 1). Diese Methodenvielfalt umfasst unter anderem Beobachtungen on the job, Verhaltenssimulationen, Persönlichkeitsfragebögen und -tests sowie die Analyse der beruflichen Biografie im Rahmen von Interviews (Obermann, 2018). In diesem Sinne verbindet das Assessment Center drei wesentliche Verfahrenskategorien: Verhaltenssimulationen, psychologische Tests und Fragebögen sowie Interviews (Obermann, 2018, S. 6).

Im Detail sieht die Umsetzung der verschiedenen Verfahrenstypen folgendermaßen aus: Im Falle von Beobachtungen steht die Leistung des Kandidaten in einem realen oder simulierten Arbeitskontext im Vordergrund. Die Bandbreite der Umsetzung reicht dabei von konkreten Arbeitsproben bis hin zu Probearbeitstagen oder Praktika.

Verhaltenssimulationen zielen darauf ab, die Leistungsfähigkeit der Kandidaten durch spezifische Situationen wie Rollenspiele, Gruppendiskussionen, Präsentationen, Fallstudien oder Postkorb-Übungen zu ermitteln. Obermann (2018, S. 3) betont hierbei, dass die spezifischen Simulationen in direktem Zusammenhang mit den Anforderungen der jeweiligen Zielposition stehen. Daher werden je nach Fokus der Rolle unterschiedliche Simulationen angewendet: Beispielsweise wird für eine Vorgesetztenrolle ein Mitarbeiter-Rollenspiel durchgeführt, während für eine Führungsposition eine Gruppendiskussion angemessen ist und bei einer Entscheidungsrolle eine Postkorb-Übung implementiert wird (vgl. Obermann, 2018, S. 3).

Darüber hinaus integrieren Assessment Center psychologische Tests und Fragebögen, um die Ausprägung spezifischer Persönlichkeitsdimensionen der Kandidaten zu ermitteln. In diesem Kontext sind Interviews ein weiteres zentrales Instrument, in welchen die Teilnehmer über ihre berufliche Biografie berichten oder darlegen, wie sie zukünftige Herausforderungen in ihrer angestrebten Rolle bewältigen würden (Obermann, 2018, S. 5). Abbildung 3 visualisiert die einzelnen Verfahrenstypen.

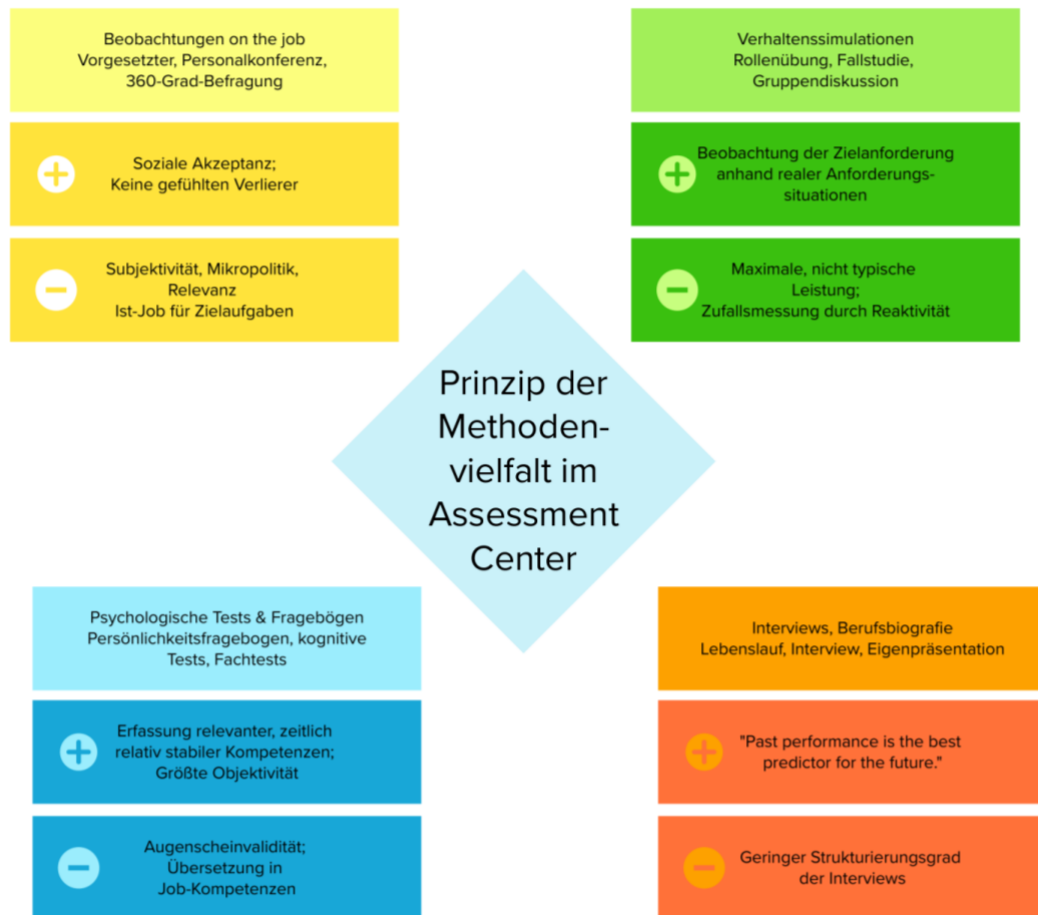


Abbildung 3 Verfahrenstypen (Quelle: In Anlehnung an Obermann, 2018, S.6)

Obermann unterscheidet zwischen dem Auswahl-AC, Potenzialanalyse-AC und dem Bildungsbedarfs-AC, wobei er den Nutzen für das Unternehmen dem Nutzen der Teilnehmer gegenüberstellt. Beispielsweise dient das Auswahl-AC dem Unternehmen, um treffsicherer Personalentscheidungen insbesondere bei überfachlichen Eigenschaften zu fällen. Beurteilungsbereiche wie „etwa Führungsverhalten, Arbeitsorganisation und analytisches Denken“ werden sichtbar. Blender und Langsamstarter lassen sich besser differenzieren. Allein durch die Anzahl der Beurteiler im Assessment werden mehr Aspekte offensichtlich. Die Fluktuationskosten reduzieren sich, da die Teilnehmer besser auch entsprechend des Culture Fits betrachtet werden können.

Das Firmenimage wird dadurch gefördert, das durch maximale Transparenz, auch gegenüber internen Konkurrenten, die Atmosphäre und Rückmeldungen auch abgelehnte Bewerber zufriedener sind. Die Beurteiler werden geschult bezüglich Beurteilungs- und Beobachtungsfähigkeiten im Alltag und bei klassischen Interviews und erhalten viele Denkanstöße. Und letztlich verbessert sich die Zusammenarbeit der Personalabteilungslinie auch durch Hinweise für gezielte Einarbeitung der Teilnehmer.

Demgegenüber gehen die Teilnehmer ein geringeres Risiko ein, die falsche Position anzutreten, und erhalten konkrete Rückmeldung über ihre Stärken und Schwächen. Darüber hinaus können sie ihre „erworbenen Fähigkeiten in berufsnahen Situationen“ demonstrieren. Angenehm ist auch, dass die Teilnehmer direkt im Anschluss an das AC ihr Ergebnis erhalten und Vergleichsmöglichkeiten zu Mitbewerbern haben.

Der Nutzen des Potenzialanalyse-AC liegt für Unternehmen im Erkennen von bisher ungenutzter Potenzialen, welche sich außerhalb von AC nur zeit- und kostenaufwendig erheben lassen würden. Auch hier ist durch die „(h)ohe Transparenz über die Defizitstruktur eines Mitarbeiters (im direkten Vergleich mit anderen Mitarbeitern“ ein direkter Vergleich für die Mitarbeiter möglich und ein direkter Vergleich mehrerer Mitarbeiter in identischen Arbeitssituationen. Auch hier liegt der Nutzen neben der Transparenz und Vergleichbarkeit in den Rückmeldungen über Stärken und Schwächen in Bezug auf die konkreten Anforderungen, in der Möglichkeit Alltagssituationen zu erleben und mit den eigenen Vorstellungen abzugleichen und der „(o)bjektiven Beurteilung der Potenziale für unterschiedliche Positionen/Laufbahnen“.

Durch diese Vorgehensweise werde Beförderungen für die Mitarbeiter auch gerechter erlebt, da mehrere Beurteiler zugegen waren, die Kriterien transparent kommuniziert werden und entsprechend Praxis-Situationen festgelegt werden.

Der letzte von Obermann skizzierte Bereich betrifft das Bildungsbedarf-AC. Hierbei erhalten die Beobachter einen guten Überblick über das vorhandene Potenzial der Teilnehmer und können somit den „Gießkanneneffekt“ bei Bildungsinvestitionen vermeiden und individuelle Karrierepfade entwickeln, bei denen der Mitarbeiter und seine Individualität in den Mittelpunkt gerückt werden und gezielte Entwicklungsmaßnahmen ausgewählt werden können. Auch ist die „Effizienz bisheriger Bildungsmaßnahmen“ überprüfbar. Für die Teilnehmer des Bildungsbedarfs-AC „werden Maßnahmen und Wege aufgezeigt, wie vorhandene Schwächen abgebaut und Stärken besser genutzt werden können“, wodurch die anschließenden Trainings- und Entwicklungsmaßnahmen erfolgreiche sind, „da der direkte Bezug zu den eigenen Schwächen erkennbar wird“ (Obermann, 2018, S. 49ff).

Für das AC ist jeweils „zu entscheiden, worin die damit verbundene eignungsdiagnostische Zielsetzung besteht: inter- oder intraindividuelle Unterschiede beschreiben, beruflichen Erfolg oder Potenzial vorhersagen oder Verbesserungen im z.B. Führungsverhalten bewirken“ (Obermann, 2018, S. 300).

2.2.1 Entwicklung eines Assessment Centers

Der Prozess der Entwicklung eines eignungsdiagnostischen Verfahrens umfasst mehrere Schritte. Zunächst werden in den Vorarbeiten (Schritt 0) die Eignungsmerkmale und Verfahren festgelegt. Anschließend wird ein Szenario entworfen (Schritt 1), das eignungsrelevantes Verhalten provozieren soll. In Schritt 2 wird ein Verfahrensprototyp konstruiert, der die Handhabungshinweise, Materialien und Instruktionen für Teilnehmer und Verfahrensbeteiligte, die Ziele und Eignungsmerkmale sowie die Verhaltensoperationalisierungen in Beobachtungs- und Beurteilungsbogen umfasst. Nach der Konstruktion des Prototyps erfolgt in Schritt 3 dessen Erprobung mit Testpersonen. Dabei liegt der Fokus auf der zielgruppenadäquaten Schwierigkeit, der erzielten Realitätsnähe, der Beobachtbarkeit der Eignungsmerkmale und der Plausibilität des Prozessablaufs. Basierend auf den Ergebnissen der Testläufe wird der Verfahrensprototyp in Schritt 4 überarbeitet und gegebenenfalls erneut getestet. Sobald der Prototyp erfolgreich überarbeitet wurde, erfolgt die Finalisierung und Konstruktion von Verfahrensvarianten (Schritt 5), bei der Handhabungshinweise sowie Beobachtungs- und Beurteilungsbogen fertiggestellt werden. Der nächste Schritt (Schritt 6) besteht in der Einschleusung des Verfahrens in den Realbetrieb, wobei die Eignungsdiagnostiker und Beobachter ihre Aufgaben anhand realistischer Beispiele einüben. Schließlich wird im laufenden Verfahren eine Qualitätssicherung (Schritt 7) durchgeführt, um die Prozesse kontinuierlich zu begleiten und zu überwachen (Diagnostik- u. Testkuratorium, 2017, S. 41).

2.2.2 Entwicklung und Auswahl der Aufgaben

Die Beobachtungsstrategien können in unterschiedlichen Kontexten Anwendung finden, sei es in einer natürlichen Umgebung wie etwa in Freizeit- oder Arbeitsumgebungen oder in einer künstlich konstruierten Situation unter kontrollierten Bedingungen. Ferner kann eine Unterscheidung zwischen direkter und indirekter Beobachtung getroffen werden. Die direkte Beobachtung findet in der jeweiligen Situation statt, während die indirekte Beobachtung auf der retrospektiven Auswertung basiert, beispielsweise durch die Analyse von Videomaterial (Diagnostik- u. Testkuratorium, 2017, S. 34). Darüber hinaus können in der Rolle des Beobachters Variationen festgestellt werden. Bei der partizipativen Beobachtung übernimmt der Beobachter eine aktive Rolle im beobachteten Geschehen, während er bei der nicht-partizipativen Beobachtung lediglich passiv anwesend ist. Diese differenzierten Rollen und Methoden tragen zur Anpassungsfähigkeit und Flexibilität der Beobachtungsstrategien bei und ermöglichen eine effiziente Arbeit in sowohl natürlichen als auch kontrollierten Umgebungen. Die systematische Beobachtung stellt insgesamt ein äußerst vielseitiges Instrument in der diagnostischen Praxis dar. Durch den flexiblen Einsatz verschiedener Beobachtungs-

und Dokumentationsmethoden können spezifische Verhaltensmerkmale identifiziert und analysiert werden. Dies erlaubt eine detaillierte und individualisierte Einschätzung des beobachteten Individuums und leistet einen wesentlichen Beitrag zur Verbesserung der diagnostischen Präzision. Tabelle 1 zeigt die unterschiedlichen Verfahrenskategorien entsprechend der Kompetenzen auf. (Obermann, 2018, S. 112). Im Durchschnitt werden aktuell 5,6 unterschiedliche Verfahrenselemente eingesetzt (ebd).

Kompetenz	Verfahrens-Kategorie	Verfahrens-Simulation						Tests- und Fragebögen		Biographie und weitere Methoden			
		Interview	Präsentation	Rollenübung	Fallstudie	Fact-Finding	Postkorb	Kognitive Tests	Persönlichkeitsfragebögen	Interview	Arbeitsprobe	Kollegenurteil	Selbsteinschätzung
Unternehmerische Kompetenz	Unternehmerisches Denken / Handeln				X	X	X	X	X	X	X	X	X
	Analysefähigkeit				X	X	X	X	X	X	X	X	X
	Handlungs-Orientierung				X		X			X	X	X	X
	Kunden-orientierung				X		X			X	X	X	X
	Arbeitsorganisation		X		X		X			X	X	X	X
Führungs-Kompetenz	Mitarbeiter-motivation		X	X						X	X	X	X
	Ergebnis-orientierung	X	X	X						X	X	X	X
Soziale Kompetenz	Konfliktfähigkeit	X	X	X					X	X	X	X	X
	Überzeugungs-kraft	X	X	X						X	X	X	X
	Kontaktfähigkeit	X	X	X					X	X	X	X	X
	Kooperations-fähigkeit	X	X	X					X	X	X	X	X
Persönliche Kompetenz	Veränderungs-bereitschaft								X	X	X	X	X
	Leistungs-motivation								X	X	X	X	X
	Belastbarkeit								X	X	X	X	X

Tabelle 1 Matrix: Mit welchen Modulen können einzelne Kompetenzen gemessen werden? (Quelle: in Anlehnung an Obermann, 2018, S. 112)

Im weiteren Verlauf wird die Aufmerksamkeit auf die Präsentation, das Interview, das Rollenspiel sowie den Persönlichkeitsfragebogen gelenkt, da diese Methoden für die praktische Umsetzung von besonderer Relevanz sind.

2.3 Präsentationen

Präsentationen stellen eine Form von Verhaltenssimulationen dar, in denen das Verhalten der Probanden beobachtet und evaluiert wird. Im Unterschied zu Interviews, in denen hypothetisches Verhalten diskutiert wird, konzentriert sich die Beurteilung bei Präsentationen auf das „tatsächliche Vorgehen im Rahmen einer Simulation der entsprechenden Situation“ (Obermann, 2018, S. 3). Präsentationen tragen zur Visualisierung sowohl von „organisatorisch-analytischen Fähigkeiten als auch die sprachliche Sicherheit im Auftreten von Gruppen“ bei (Obermann, 2018, S. 134). Im Kontext einer Präsentation erhält der Assessee die Aufgabe, „zu einem frei gewählten oder vorgegebenen Thema oder einer spezifischen Fragestellung eine Präsentation vor Publikum (Beobachtende, Assessore und Assessorinnen) zu halten“ (Eck et al., 2016, S. 44). Es ist zu beachten, dass, obwohl die Validität der Übung Präsentation bestätigt

ist, ihre relative Validität im Vergleich zu anderen Modulen wie dem kognitiven Test und dem strukturierten Interview geringer ist (Obermann, 2018, S. 135). Abbildung 4 zeigt eine Übersicht über die unterschiedlichen Präsentationsthemen.

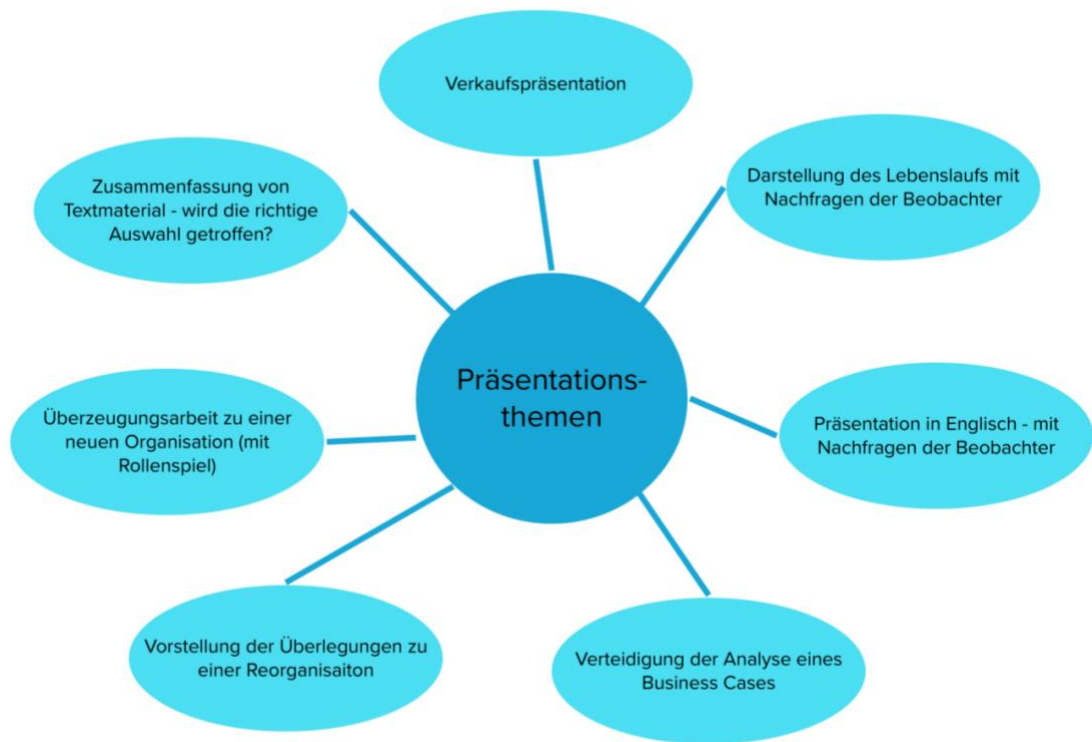


Abbildung 4 Präsentationsthemen (Quelle: modifizierte Darstellung Obermann, 2018, S. 134)

2.4 Das Interview

Das Interviewverfahren stellt einen fundamentalen Ansatz in der Datenerfassung dar. Es lassen sich diverse Typologien differenzieren: quantitative, qualitative, standardisierte, halbstandardisierte und nicht-standardisierte Interviews. Das quantitative standardisierte Interview differenziert sich von qualitativen Interviews durch seinen eng begrenzten Variationsrahmen. Im standardisierten Interview ist der Spielraum für die Befragung minimal, da die Fragen des Fragebogens mündlich übertragen werden. „Die Richtlinien beginnen beim Wortlaut der Fragen und beziehen sich auf ihre Reihenfolge, verlaufen über genau festgelegte Antwortvorgaben und legen bis ins Detail fest, wie sich der Interviewer zu verhalten hat: Was muss er vorlesen, was darf er auf keinen Fall vorlesen? Wann muss er welche Kärtchen oder Listen vorzeigen? Muss er bestimmte Fragen betonen?“ (Reinhardt et al., 2020, S. 13ff). Charakteristisch für das standardisierte Interview ist die Nutzung eines Fragebogens mit minimalen Spielräumen für die Interviewer und Befragten. Die Anzahl und Reihenfolge der Fragen sind

festgelegt, was zu einem begrenzten Gestaltungsspielraum führt. Emotionen und subjektive Wahrnehmungen werden durch das standardisierte Interview nicht erfasst (ebd.).

Ein Leitfadeninterview wird gemäß einem Leitfaden durchgeführt. „Grundlage eines halbstandardisierten Interviews ist ein mehr oder weniger flexibel aufgebautes und anzuwendendes Frageschema (Leitfaden). Der Interviewer darf unverständene Fragen erklären, es ist ihm gestattet, auf Hauptfragen Nebenfragen zu stellen, um die Antworten zu präzisieren“ (ebd.). Charakteristisch für ein halbstandardisiertes Interview ist der Einsatz eines Interviewleitfadens mit offenen Fragen, was zu einer Vergleichbarkeit der Daten und zu einer strukturierten Orientierung führt. Es handelt sich um eine offene Gesprächsführung mit Antwortspielräumen und der Integration von Schlüssel- und Evaluationsfragen. Der Interviewer hat Freiheiten wie beispielsweise das Nachfragen, um die Antworten zu präzisieren, auch dürfen Fragen umformuliert werden oder die Reihenfolge der Fragen angepasst werden.

Der Interviewer muss einen guten Überblick über das bereits Gesagte haben (ebd.). „Der Begriff Leitfadeninterview oder leitfadengestütztes Interview verkörpert einen Oberbegriff, unter dem verschiedene qualitative Interviewverfahren subsumiert werden können“ (Reinhardt et al., 2020, S. 17).

Das nicht-standardisierte Interview bietet vollständige Freiheitsgrade für den Interviewer, wobei sich die Fragen spontan im Gesprächsverlauf ergeben. Die Grundlage für das Interview bildet höchstens ein Stichwortkatalog. Diese Form des Interviews wird angewandt, wenn es wenig Wissen über das Interviewthema gibt. Daher ergeben sich folgende Charakteristika: Es wird weder ein Fragebogen noch ein Leitfaden verwendet. Die Fragen können frei formuliert und angeordnet werden. Die Kontrolle des Interviews obliegt dem Interviewer, was bedeutet, dass eine ausgeprägte Vertrautheit zum Untersuchungsgegenstand erforderlich sein muss. Das Eignungsinterview, eine wesentliche Methode der direkten mündlichen Befragung, stellt ein weit verbreitetes und effizientes Verfahren dar, welches viele Funktionen erfüllt. „Es bietet neben dem gegenseitigen Kennenlernen die Möglichkeit „mehr über die Kompetenzen und Potenziale sowie Qualifikationen zu erfahren“ (Diagnostik- u. Testkuratorium, 2017, S. 67) und gibt dem Kandidaten andererseits die Chance, die Organisation sowie das Tätigkeitsfeld kennenzulernen.

Ein halbstrukturiertes Interview besteht aus zwei Abschnitten. Im ersten Teil werden offene Fragen gestellt, um explizit verfügbare Hypothesen und Komponenten subjektiver Theorien zu identifizieren (Mey & Mruck, 20, S. 322). Theoriegeleitete und Konfrontationsfragen dienen dazu, implizite Wissensbestände zu erfassen und die sich

entwickelnden subjektiven Theorien kritisch zu prüfen (ebd.). Der zweite Teil des Interviews umfasst die Validierung der Ergebnisse mit den Befragten.

Für qualitative Befragungen geben Boltz und Göring (2015, S. 403) verschiedene Regeln zur Formulierung von Fragen an. Erstens sollten Fragen einfach und direkt in der Sprache der Zielgruppe formuliert sein, ohne doppelte Verneinungen, Fremdwörter oder Fachbegriffe. Zweitens sollten Faktenfragen auf die für die Studie relevanten Kernfragen reduziert werden. Drittens sollten Fragen so gestaltet sein, dass sie Probanden zum Erzählen und Beschreiben von Sachverhalten anregen, wofür sich offene Fragen eignen. Viertens sollten Warum-Fragen vermieden werden, da sie Befragte in eine Rechtfertigungssituation bringen können. Fünftens sind Suggestivfragen zu vermeiden, um die Antwortrichtung offen zu halten. Sechstens sollten geschlossene Fragen, falls überhaupt, nur als Filterfunktion verwendet werden, bei der auf eine Ja- oder Nein-Antwort unterschiedliche Folgefragen gestellt werden, die nur bei einem entsprechenden Antworttyp sinnvoll sind (Bolz & Döring, 2015, S. 403).

Ein Interviewleitfaden dient als Werkzeug für halbstrukturierte Interviews und basiert auf einem Katalog offener Fragen, zu denen die Befragten in eigenen Worten Stellung nehmen sollen (Boltz & Döring, S. 359). Die Reihenfolge der Fragen ist nicht vorgegeben, und Fragen können ausgelassen werden, wenn sie durch vorherige Fragen bereits beantwortet wurden.

Um einen Interviewleitfaden zu erstellen, sollten folgende Schritte durchgeführt werden (Helfferich, 2019, S. 677-678):

- 1) Sammeln von Fragen, wobei zunächst die Quantität vor Qualität im Vordergrund steht.
- 2) Überprüfen der Fragen hinsichtlich ihrer Relevanz für das Forschungsvorhaben.
- 3) Sortieren der Fragen nach Indikatoren und nach Erzählaufforderung, Aufrechterhaltungsfragen und konkreten Nachfragen innerhalb der Dimension.
- 4) Subsummieren von Fragen, die möglicherweise bereits in einem Oberbegriff oder einer Erzählaufforderung enthalten sind.

Insgesamt demonstrieren die dargelegten Aspekte die Vielfältigkeit und Komplexität der Interviewführung als Forschungsmethode. Die korrekte Auswahl und Anwendung der verschiedenen Interviewformate und -techniken ist wesentlich für die Qualität der erzielten Forschungsergebnisse und erfordert ein hohes Maß an Fachkenntnis und Erfahrung. Darüber hinaus sollte stets eine sorgfältige Planung und Vorbereitung der Interviews erfolgen, um die Relevanz und Genauigkeit der gesammelten Daten zu

gewährleisten. Hierzu zählen insbesondere die Entwicklung geeigneter Fragebögen oder Leitfäden, die Anpassung des Fragensdesigns an die spezifischen Bedürfnisse und Eigenschaften der Zielgruppe sowie die effektive Verwaltung und Analyse der erhobenen Daten.

2.5 Das Rollenspiel

Die Ursprünge der Rollenspielanwendung lassen sich auf den klinisch-psychotherapeutischen Bereich zurückverfolgen und sind eng mit der Arbeit des Psychiaters Jacob Levy Moreno (1890-1973) verknüpft, dem Pionier des Psychodramas, der Soziometrie und der Gruppenpsychotherapie (Bregas, 2022, S. 5). Im Kontext des Rollenspiels liegt der Schwerpunkt auf dem Training von zwischenmenschlichen und potenziell zukünftigen Situationen, dessen Bedeutung sich über das psychotherapeutische Feld hinaus seit den 1970er Jahren auch in der Managementausbildung verstärkt hat (ebd.). Das Hauptziel eines effektiven Rollenspiels besteht darin, zwischen geeigneten und weniger geeigneten Kandidaten im Hinblick auf zu beobachtende und zu bewertende Kompetenzen zu differenzieren (Bregas, 2022, S. 17). Ein wesentliches Merkmal des Rollenspiels ist, dass sowohl die Hauptakteure als auch die Beobachter durch Reflexion lernen, wodurch die gesamte Gruppe in den Prozess integriert wird.

Im eignungsdiagnostischen Kontext stellen Rollenspiele simulierte Gesprächssituationen dar, in denen ein Teilnehmer mit mindestens einem Rollenspieler zu einem vorher definierten Thema interagiert (Bregas, 2022, S. 7). Bei der Implementierung von Rollenspielen in Assessment Centern sind bestimmte Vorannahmen zu berücksichtigen, die bei bekannter Kenntnis in der Auswertungsphase berücksichtigt werden können. In der wissenschaftlichen Diskussion über die Bewertung von Rollenspielen im Assessment-Center-Kontext werden unterschiedliche Annahmen und Gegenargumente erörtert (Bregas, 2022, S. 8). Eine dieser Annahmen ist, dass ein erfolgreiches Rollenspiel auf Führungskompetenzen in der Praxis hindeutet. Das Gegenargument dazu betont jedoch, dass Führung mehr ist als die Summe der Mitarbeitergespräche und dass Gesprächskompetenz nicht zwangsläufig mit Führungskompetenz gleichzusetzen ist. Eine weitere Annahme postuliert, dass das erfolgreiche Navigieren durch eine kritische Situation im Rollenspiel, die für eine Führungsposition typisch ist, auch für andere typische Situationen dieser Führungsposition anwendbar ist. Das Gegenargument dazu verweist jedoch darauf, dass die Wahrscheinlichkeit, auch andere Situationen zu beherrschen, von der Ähnlichkeit der Situationen abhängt und nicht zwangsläufig verallgemeinert werden

kann. Die dritte Annahme besagt, dass ein Kandidat, der im Rollenspiel erfolgreich bzw. nicht erfolgreich ein Mitarbeitergespräch führen kann, dies auch später in der Praxis entsprechend (nicht) kann. Das Gegenargument dazu hinterfragt jedoch die Konstanz des wahrgenommenen Verhaltens und betont, dass das Potenzial zur Kompetenzerwerbung nicht übersehen werden sollte. Die vierte Annahme verallgemeinert die Situation und behauptet, dass das im Rollenspiel gezeigte Verhalten auf die Praxis übertragbar ist. Das Gegenargument dazu weist jedoch darauf hin, dass situative Faktoren des Assessment Centers das Rollenspiel und dessen Ergebnisse beeinflussen können und daher berücksichtigt werden sollten. Die fünfte Annahme verallgemeinert die Personen und geht davon aus, dass ein Kandidat, der in der Lage ist, einen Gesprächspartner im Rollenspiel abzuholen und zu überzeugen bzw. eine gemeinsame Lösung zu erarbeiten, dies auch bei anderen Gesprächspartnern erreichen kann. Das Gegenargument dazu weist darauf hin, dass diese Vereinfachung der Vielfältigkeit der menschlichen Natur nicht gerecht wird. Die letzte Annahme postuliert, dass das Verhalten, das ein Kandidat im Assessment Center im Rollenspiel zeigt, seinem "natürlichen" Verhalten entspricht, das er/sie auch im Arbeitsalltag zeigen würde. Das Gegenargument dazu betont den Aufforderungscharakter des Assessment Centers, bei dem der Kandidat aufgefordert wird, sich in eine bestimmte Situation zu begeben und sich auf bestimmte Weise zu verhalten. Das Befolgen dieser Anweisungen wird als positiv bewertet, während das Nichtbefolgen der Persönlichkeit des Kandidaten angelastet wird (Bregas, 2022, S. 8).

In der folgenden SWOT-Analyse werden die Stärken, Schwächen, Chancen und Risiken von Rollenspielen im Assessment-Center (AC) im Vergleich zu anderen Übungstypen dargestellt (Bregas, 2022, S. 9). Die Stärken der Rollenspiele liegen in ihrer hohen scheinbaren Validität bzw. Nähe zur Realität, der direkten Beobachtbarkeit sozialer Kompetenzen, der verbesserten Standardisierbarkeit im Vergleich zu beispielsweise Gruppendiskussionen, der hohen Akzeptanz bei allen Teilnehmergruppen und der ökonomischen Durchführbarkeit. Allerdings haben Rollenspiele auch Schwächen, wie die eher punktuelle Erfassung sozialer Kompetenzen, die starke Abhängigkeit von der Qualität des Rollenspielers, die erhöhte Subjektivität in der Bewertung durch die hohe Kontextabhängigkeit und die Laborsituation, in der kein reales Verhalten beobachtet wird, sondern nur ein "als-ob-Charakter" besteht. Die Chancen von Rollenspielen liegen in der Vielfalt der Gestaltungsmöglichkeiten und Schwerpunktsetzung sowohl für die Konzipierenden als auch für die Teilnehmenden sowie in dem relativ geringen Aufwand bei der Konzeption. Es sind verschiedene Parallelversionen möglich. Zu den Risiken gehört, dass die Teilnehmenden des Assessment Centers möglicherweise noch nicht

über den erforderlichen Erfahrungshintergrund verfügen, um situationsgerecht in die Rolle zu schlüpfen. Zudem besteht die Gefahr des sozial erwünschten Verhaltens, bei dem die Teilnehmenden bewusst schauspielern.

2.5.1 Rollenspielarten

Die Durchführung von Mitarbeitergesprächen stellt eine der am häufigsten vorkommenden Rollenspielformen dar, bei der die Interaktion zwischen Führungskraft und Mitarbeiter im Zentrum steht. Dennoch eröffnen sich weitere Möglichkeiten für Rollenspiele, die Kundengespräche, Beratungsgespräche, Gespräche unter Kollegen oder sogar Einkaufs-, Verkaufs- oder Verhandlungsgespräche simulieren. Die Vielfalt der Gesprächsanlässe, die durch Rollenspiele in Assessmentcentern repräsentiert werden können, ist groß. Einige Anlässe eignen sich dabei besser zur Darstellung mittels Rollenspiele als andere. So bietet beispielsweise ein Delegationsgespräch eine geeignetere Grundlage für ein Rollenspiel als ein Unterweisungsgespräch. Ebenso eignen sich Ermahnungsgespräche, Klärungsgespräche oder kritische Feedbackgespräche bezüglich Leistung oder Verhalten besser als positive Feedbackgespräche oder reine Lobäußerungen (Bregas, 2022, S. 13).

Im Kontext akuter Handlungsnotwendigkeiten sind Interventionsgespräche aufgrund mangelnder Motivation oder Konflikten im Team besonders gut geeignet. Ebenso können Kommunikationsprozesse während Veränderungsphasen gut in Rollenspielen dargestellt werden. Hingegen eignen sich Interventionsgespräche aufgrund von Alkoholmissbrauch oder mangelnder Hygiene weniger gut. Im Hinblick auf Leistungsbeurteilungen zeigen sich Zielvereinbarungsgespräche oder Beurteilungsgespräche als besonders geeignete Szenarien für Rollenspiele, wohingegen Gehalts- oder Jahresgespräche weniger geeignet sind. Im Bereich der Personalentwicklung kann ein fachlich-persönliches Entwicklungsgespräch gut in einem Rollenspiel abgebildet werden, während Laufbahnplanungs- oder Mentoring-Gespräche³ sich weniger gut eignen (ebd.).

2.5.2 Rollenspielkonzeption

Bei der Ausgestaltung von Rollenspielen ist essenziell, dass die zugrundeliegende Situation sowohl repräsentativ als auch relevant für die angestrebte Führungsposition ist (Jung, 2003). Es ist von Bedeutung, die Realität so zu simulieren, dass die Teilnehmenden sich selbst in der Situation darstellen können (ebd.), was voraussetzt, dass die Rollenbeschreibung ausreichend Freiheiten gewährt, sodass die

³ Eine Übersicht über verschiedene Kontexte von Mitarbeitergesprächen findet sich bei Bregas, 2022, S. 14.

Teilnehmenden authentische Handlungsmuster anwenden und situationsangemessen sowie spontan agieren können (ebd.).

Bregas formuliert folgende Empfehlungen zur Konzeption: Er unterscheidet zwischen der Anlassebene und der grundlegenden Ebene. Der Anlass ist meist eine konkrete Situation, während die grundlegende Ebene den eigentlichen Auslöser für das Vorkommnis repräsentiert. Beide Ebenen sollten berücksichtigt werden, um zukünftige Vorkommnisse dieser Art zu vermeiden. In eignungsdiagnostischen Rollenspielen sollte keine Vermischung von Themen stattfinden, obwohl dies in der Realität oft der Fall ist. Zudem sollte die Situation nicht dazu genutzt werden, mehrere Themen gleichzeitig anzusprechen. Jung betont, dass ein Rollenspiel umso mehr Akzeptanz und Gültigkeit erlangt, je realitätsnäher es gestaltet ist (Bregas, 2022, S. 20).

Im Anschluss an ein Rollenspiel sollte stets eine Selbstreflexion integriert werden. Bregas schlägt einige Fragen zur Selbstreflexion vor (Bregas, 2022, S. 21). Es ist wichtig, zwischen Kompetenz und Potenzial zu differenzieren. Bei der Kompetenz geht es darum, ob der Kandidat in vergleichbaren Situationen ähnlich agiert, was in der Situation angemessen ist. Das Potenzial hingegen bezieht sich auf die Frage, ob dem Kandidaten zugetraut wird, sich die Kompetenz in der Zukunft anzueignen. Hinsichtlich der Vorbereitungszeit und der Dauer des Rollengesprächs haben sich 10-20 Minuten für die Vorbereitung und 15-20 Minuten für die Durchführung als optimal erwiesen (Bregas, 2022, S. 24). Bei der Anweisung der Teilnehmer weisen Bolte und Jung (1995, S. 79-80) darauf hin, dass die Texte nicht nur leicht verständlich sein sollen, sondern die Kandidaten zur Identifikation anregen und motivieren und Spaß machen sollen. Eine "überlegte Dramaturgie" sollte jeder Übung zugrunde liegen, um den Spannungsbogen bis zum geplanten Ende aufrechtzuerhalten und nicht vorzeitig zu entlassen. Dabei sollten Argumente bei den Beteiligten zahlreich und vielschichtig sein. Interessenkonflikte sollten deutlich angelegt sein. Es ist wichtig, den Teilnehmern zu ermöglichen, variabel zu handeln und nicht das Erreichen einer bestimmten Lösung in den Mittelpunkt zu stellen (Bolte & Jung, 1995, S. 79-80).

In diesem Sinne ist eine akkurate und sorgfältige Konzeption und Durchführung von Rollenspielen im Assessmentcenter von entscheidender Bedeutung. Sie ermöglicht nicht nur eine genaue Beurteilung der Kompetenzen und Potenziale der Kandidaten, sondern stellt auch sicher, dass diese Beurteilung auf einer soliden und realitätsnahen Grundlage stattfindet. Die vorgeschlagenen Reflexionsfragen von Bregas (2022, S. 21) dienen dabei der Selbstbewertung der Teilnehmer, dem Abgleich der tatsächlich

erreichten Ziele mit den ursprünglich intendierten, der Identifikation alternativer Handlungsmöglichkeiten sowie der Einschätzung der eigenen Performance aus der Perspektive des Gesprächspartners. Zudem ermutigen sie zur Übertragung des im Rollenspiel gezeigten Verhaltens auf vergleichbare Situationen in der Realität.

Abschließend sei nochmals hervorgehoben, dass Rollenspiele ein wertvolles Instrument in der Kompetenz- und Potenzialbewertung darstellen können, sofern sie auf realitätsnahen und bedeutsamen Situationen basieren, eine ausreichende Handlungsfreiheit für die Teilnehmer bieten und mit adäquaten Reflexionsprozessen verbunden sind. Jedoch sollten sie mit Bedacht eingesetzt und kritische Themen, wie mangelnde Hygiene, Alkoholmissbrauch und sexuelle Belästigung, ausgespart bleiben. Indem solche Best Practices befolgt werden, kann sichergestellt werden, dass Rollenspiele sowohl für die Beurteilenden als auch für die Teilnehmer eine bereichernde und aufschlussreiche Erfahrung darstellen.

2.6 Beobachtung

In der Praxis der Diagnostik ist die systematische Beobachtung von Verhaltensmerkmalen ein fundamentales Element. Dies umfasst die Identifizierung und Dokumentation spezifischer Beobachtungseinheiten, welche als charakteristische Verhaltensindikatoren mit deutlichem Beginn und Ende definiert werden. Diese Beobachtungseinheiten tragen zur präzisen Beschreibung abstrakter Eignungsmerkmale bei (Diagnostik- u. Testkuratorium, 2017, S. 45ff). Abbildung 5 illustriert den Prozess der Schlussfolgerung von beobachtbarem Verhalten auf ein Eignungsmerkmal.

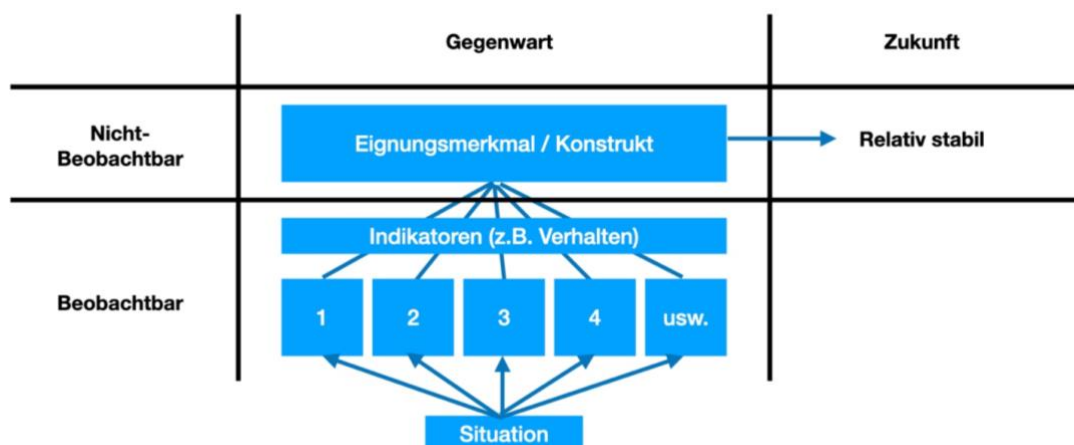


Abbildung 5 Schlussfolgerung von beobachtbarem Verhalten auf ein Eignungsmerkmal (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 31)

Verschiedene Strategien, darunter der Ereignisstichprobenansatz und der Zeitstichprobenansatz, werden zur Durchführung dieser Beobachtungen verwendet. Die Beobachtungsmodalitäten können sich je nach Kontext und Zielsetzung unterscheiden. Hierbei können Beobachterrollen variieren, basierend darauf, ob der Beobachter direkt an der Situation teilnimmt oder eine rein beobachtende Rolle einnimmt. Des Weiteren erfolgt eine Unterscheidung zwischen unsystematisch-freier und systematisch-gebundener Beobachtung, wobei die unsystematisch-freie Beobachtung keine vorgegebenen Kriterien aufweist und ergebnisoffen ist, während die systematisch-gebundene Beobachtung auf vordefinierten Regeln und Kriterien basiert (Diagnostik- u. Testkuratorium, 2017, S. 34).

Eine weitere Unterscheidung kann hinsichtlich der Art der Beobachtung getroffen werden, wobei zwischen Selbst- und Fremdbeobachtung unterschieden wird. Selbstbeobachtung beinhaltet eine Reflexion des eigenen Verhaltens, beispielsweise in Form von Tagebuchprotokollen, während Fremdbeobachtungen von Dritten, wie Bekannten oder geschulten Experten, durchgeführt werden. Ein weiterer relevanter Aspekt ist die Unterscheidung zwischen Feld- und Laborsituationen. Feldbeobachtungen finden in natürlichen Umgebungen, wie Freizeit oder Arbeitsumfeld, statt, während Laborsituationen kontrollierte, künstlich geschaffene Szenarien repräsentieren. Die Beobachtung kann direkt oder indirekt erfolgen. Bei der direkten Beobachtung wird die Situation unmittelbar wahrgenommen, während bei der indirekten Beobachtung die Auswertung im Nachhinein auf Basis von Videoanalysen stattfindet. Abbildung 6 beschreibt die wesentlichen Arbeitsschritte bei der Verhaltensbeobachtung und -beurteilung.



Abbildung 6 Wesentliche Arbeitsschritte bei der Verhaltensbeobachtung und -beurteilung (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 35)

Innerhalb des Beobachtersystems wird festgelegt, wann und wo die Beobachtung stattfindet, welche Verhaltensweisen beobachtet werden und wie die Ergebnisse dokumentiert werden (Diagnostik- u. Testkuratorium, 2017, S. 37).

Beispielsweise erfolgt die Konkretisierung des Eignungsmerkmals "Kooperation" durch verhaltensnahe Operationalisierung. Positive Verhaltensindikatoren könnten dabei etwa Unterstützung anderer Teilnehmer, Berücksichtigung der Belange anderer Teilnehmer, Demonstration von Interesse für die Situation anderer Teilnehmer und die Fähigkeit zur Entwicklung von Kompromissen zwischen widerstrebenden Interessen sein. Negative Verhaltensindikatoren könnten hingegen das ausschließliche Verfolgen eigener Interessen, das Übergehen der Belange und Wünsche anderer Teilnehmer, das Ignorieren der Situation anderer Teilnehmer oder das sture Beharren auf dem eigenen Standpunkt umfassen. Diese konkretisierten Eignungsmerkmale ermöglichen spezifische Verhaltensbeobachtungen, die auf bestimmte Verhaltensweisen hinweisen. Zum Beispiel könnte im Kontext der Operationalisierung "Andere Teilnehmer/-innen unterstützen" beobachtet werden, ob ein Teilnehmer einem anderen einen unklaren Punkt erläutert, das Argument eines anderen Teilnehmers bekräftigt oder in den Vorschlag eines anderen Teilnehmers einwilligt.

Diese Beispiele verdeutlichen, wie die Operationalisierung auf konkrete Verhaltensweisen angewendet werden kann, um das Eignungsmerkmal "Kooperation" zu erfassen und zu bewerten.

Die Urteilsbildung variiert entsprechend ihren Beobachtungstechniken und der eignungsdiagnostischen Praxis. „Daten, die mit dem Ereignisstichproben- oder dem Zeitstichprobenansatz gewonnen wurden, enthalten Informationen zur Häufigkeit und Dauer“, die für das Merkmal relevant sind und mittels Klassifizierung eingestuft werden, sodass der Gesamtwert durch eine „gewichtete Aggregation der Einzelinformationen“ entsteht (ebd.). Bei der eignungsdiagnostischen Praxis werden Beurteilungsskalen eingesetzt. Beim Verhaltensbeobachtungsansatz werden erst im Nachhinein die passenden Eignungsmerkmale zugewiesen, während beim Bezugsrahmenansatz die Notizen im Nachhinein ausschließlich hinsichtlich der Zuweisung überprüft werden. Die Beurteilung erfolgt auf Basis der subjektiven Einschätzung bezüglich Häufigkeit, Dauer und Intensität. Abbildung 7 präsentiert Beispiele für numerische Urteile (Operationalisierungsansatz und verhaltensverankerte Beurteilungsskalen) (Diagnostik- u. Testkuratorium, 2017, S. 49).

(a) Operationalisierungsansatz	(b) Verhaltensverankerte Beurteilungsskala										
<table border="1"> <tr> <th>Kandidat/-in:</th> <th>Beobachter/-in:</th> </tr> <tr> <td colspan="2"> Kooperation: <ul style="list-style-type: none"> <input type="checkbox"/> + Andere Teilnehmer (TN) unterstützt <input type="checkbox"/> + Belange anderer TN berücksichtigt <input type="checkbox"/> + Interesse für die Situation anderer TN gezeigt <input type="checkbox"/> + Kompromisse zwischen widerstrebenden Interessen entwickelt <input type="checkbox"/> - nur eigene Interessen verfolgt <input type="checkbox"/> - Belange und Wünsche anderer TN übergangen <input type="checkbox"/> - Situation anderer TN ignoriert <input type="checkbox"/> - Auf eigenem Standpunkt beharrt </td> </tr> <tr> <td colspan="2"> Eignungsmerkmalsbezogene Bewertung: <div style="text-align: center;"> gar nicht erfüllt vollständig erfüllt ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ </div> </td> </tr> </table>	Kandidat/-in:	Beobachter/-in:	Kooperation: <ul style="list-style-type: none"> <input type="checkbox"/> + Andere Teilnehmer (TN) unterstützt <input type="checkbox"/> + Belange anderer TN berücksichtigt <input type="checkbox"/> + Interesse für die Situation anderer TN gezeigt <input type="checkbox"/> + Kompromisse zwischen widerstrebenden Interessen entwickelt <input type="checkbox"/> - nur eigene Interessen verfolgt <input type="checkbox"/> - Belange und Wünsche anderer TN übergangen <input type="checkbox"/> - Situation anderer TN ignoriert <input type="checkbox"/> - Auf eigenem Standpunkt beharrt 		Eignungsmerkmalsbezogene Bewertung: <div style="text-align: center;"> gar nicht erfüllt vollständig erfüllt ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ </div>		<table border="1"> <tr> <th>Kandidat/-in:</th> <th>Beobachter/-in:</th> </tr> <tr> <td colspan="2"> Kooperation: <ul style="list-style-type: none"> ① Unterstützt andere TN und berücksichtigt Belange anderer TN, zeigt Interesse für die Situation anderer TN und entwickelt Kompromisse zwischen widerstrebenden Interessen ② ③ Arbeitet unauffällig nach Ansprache mit anderen TN zusammen, zeigt keine besondere Engagement zur Integration anderer ④ ⑤ Verfolgt nur eigene Interessen und übergeht Belange und Wünsche anderer TN, ignoriert Situation anderer TN und beharrt auf eigenem Standpunkt </td> </tr> </table>	Kandidat/-in:	Beobachter/-in:	Kooperation: <ul style="list-style-type: none"> ① Unterstützt andere TN und berücksichtigt Belange anderer TN, zeigt Interesse für die Situation anderer TN und entwickelt Kompromisse zwischen widerstrebenden Interessen ② ③ Arbeitet unauffällig nach Ansprache mit anderen TN zusammen, zeigt keine besondere Engagement zur Integration anderer ④ ⑤ Verfolgt nur eigene Interessen und übergeht Belange und Wünsche anderer TN, ignoriert Situation anderer TN und beharrt auf eigenem Standpunkt 	
Kandidat/-in:	Beobachter/-in:										
Kooperation: <ul style="list-style-type: none"> <input type="checkbox"/> + Andere Teilnehmer (TN) unterstützt <input type="checkbox"/> + Belange anderer TN berücksichtigt <input type="checkbox"/> + Interesse für die Situation anderer TN gezeigt <input type="checkbox"/> + Kompromisse zwischen widerstrebenden Interessen entwickelt <input type="checkbox"/> - nur eigene Interessen verfolgt <input type="checkbox"/> - Belange und Wünsche anderer TN übergangen <input type="checkbox"/> - Situation anderer TN ignoriert <input type="checkbox"/> - Auf eigenem Standpunkt beharrt 											
Eignungsmerkmalsbezogene Bewertung: <div style="text-align: center;"> gar nicht erfüllt vollständig erfüllt ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ </div>											
Kandidat/-in:	Beobachter/-in:										
Kooperation: <ul style="list-style-type: none"> ① Unterstützt andere TN und berücksichtigt Belange anderer TN, zeigt Interesse für die Situation anderer TN und entwickelt Kompromisse zwischen widerstrebenden Interessen ② ③ Arbeitet unauffällig nach Ansprache mit anderen TN zusammen, zeigt keine besondere Engagement zur Integration anderer ④ ⑤ Verfolgt nur eigene Interessen und übergeht Belange und Wünsche anderer TN, ignoriert Situation anderer TN und beharrt auf eigenem Standpunkt 											

Abbildung 7 Beispiele für numerische Urteile (Operationalisierungsansatz und verhaltensverankerte Beurteilungsskalen) (Quelle: Diagnostik- u. Testkuratorium, 2017, S. 49)

Beim Operationalisierungsansatz werden verhaltensnahe Teilaspekte beurteilt, während bei der verhaltensverankerten Beurteilungsskala die Abstufungen über Verhaltensbeispiele beschrieben werden.

2.6.1 Beobachtung von Rollenspielen

Im Kontext der diagnostischen Praxis gewinnt neben der Untersuchung individueller Verhaltensweisen auch die Beobachtung von Interaktionen zwischen Teilnehmern an Bedeutung. Um diesen Prozess zu unterstützen, werden spezifische Kriterien etabliert. Diese berücksichtigen nicht nur die Sach- und Appell-Ebene, sondern auch die Selbstoffenbarungs- und Beziehungsebene. Crisand (1998) weist auf drei verschiedene Arten von Informationen hin, die im Rahmen der Beobachtung erfasst werden können: Verbale Informationen, die sich im sprachlichen Inhalt manifestieren, paraverbale Informationen, die durch Merkmale wie Tempo, Rhythmus, Melodie und Stimmlage gekennzeichnet sind, und nonverbale Informationen, die sich in der Mimik, Gestik und dem äußeren Erscheinungsbild niederschlagen (Crisand, 1998). Des Weiteren identifiziert Bregras (2022) spezifische Facetten von Kompetenzen, die sich eindeutig aus dem Verhalten ableiten lassen und somit zur Beobachtung geeignet sind. Beispiele hierfür sind "Gesprächsführung", "Rhetorik", "Dialektik und Argumentation", "Durchsetzungsfähigkeit", und "Konfliktmanagement" (Bregras, 2022, S. 37). Schließlich sollte bei der Auswertung von Rollenspielen der Kontext der existierenden Kultur des jeweiligen Unternehmens oder der Organisation berücksichtigt werden. Die Interpretation und Bewertung von Verhaltensweisen und Interaktionen muss in Bezug auf die spezifischen Normen, Werte und Erwartungen der gegebenen organisatorischen Umgebung erfolgen.

2.6.2 Ablauf der Verhaltensbeobachtung

Im initialen Stadium des Prozesses ist die Erstellung von Handlungsleitfäden und Prozessplänen unerlässlich, einschließlich Beobachtungs- und Bewertungsformularen sowie erforderlichen Untersuchungsmaterialien. Während der Beobachtungsphase sollte das demonstrierte Verhalten adäquat dokumentiert und im weiteren Verlauf oder posthum kategorisiert werden. Dies könnte in Form einer fortlaufenden Beobachtungsprotokollierung realisiert werden, oder die Informationen könnten mental archiviert werden. Nach der Beobachtungsperiode werden die gesammelten Daten hinsichtlich der zu diagnostizierenden Eignungsmerkmale analysiert und bewertet. Schließlich wird eine Gesamtbeurteilung formuliert, die die verschiedenen Aspekte integriert und zusammenfasst ("Diagnostik- u. Testkuratorium", 2017).

Es ist wichtig, auf potenzielle Fehlerquellen in den einzelnen Phasen des Prozesses zu achten. Die erste Fehlerquelle manifestiert sich vor der eigentlichen Beobachtungsphase. Wenn die Eignungsmerkmale zu stark überlappen oder inhomogen sind, kann es für Beobachter schwierig sein, Verhaltensindikatoren eindeutig zuzuordnen. Eine unzureichende Operationalisierung könnte dazu führen, dass Beobachter gezwungen sind, Verhaltensindikatoren eigenständig zu definieren. Fehlplanungen und suboptimale Gestaltung von Beobachtungs- und Bewertungsinstrumenten könnten den Prozess unnötig komplizieren und die Aufmerksamkeit fehlleiten ("Diagnostik- u. Testkuratorium", 2017, S. 53). Eine umfassende Konstruktion, Vorprüfung und korrektive Maßnahmen bei identifizierten Defiziten können dazu beitragen, diese Fehler zu minimieren. Das Beobachtungsumfeld kann ebenfalls erhebliche Fehler verursachen. Dazu gehören ungünstige Blickwinkel und Lichtverhältnisse, Störungen im Prozessablauf wie Unterbrechungen oder auch technische Probleme bei Video- und Tonaufnahmen. Gegenmaßnahmen können eine sorgfältige Vorbereitung, Vorkehrungen für Notfallszenarien und Gelassenheit umfassen. Die dritte Fehlerquelle sind situationsbedingte Fehler, die aus suboptimalen Beobachtungsverfahren resultieren, etwa wenn zu wenige Beobachtungsmöglichkeiten existieren oder übermäßige Dynamik die Beobachtung erschwert. Hier können sorgfältige Konstruktion und Vorab-Prüfung sowie kontrollierte Interventionen zur Reduzierung der Dynamik helfen. Eine weitere Fehlerquelle betrifft die Kandidaten. Eine Person könnte sich anders verhalten als erwartet, was durch eine Fehlinterpretation der Situation und daraus resultierende intensive Selbstdarstellungstechniken verursacht werden kann. Hier ist die Vorprüfung sehr wichtig, und im Nachhinein kann ein Interview mit dem Teilnehmer Aufschluss geben. Beobachterbedingte Fehler können Verzerrungen bereits bei der Informationsaufnahme, bei der Erinnerung oder bei der Beurteilung selbst verursachen. Der Primacy-Effekt, Rosenthal-/Pygmalion-Effekt oder Sympathie und Antipathie gegenüber den Kandidaten können zu Informationsaufnahme Fehlern führen. Erinnerungsverzerrungen können durch den Primacy- und Recency-Effekt, Rückschaufehler und Häufigkeits-Validitäts-Fehler hervorgerufen werden. Es ist zu beachten, dass der Primacy-Effekt zeitstabil ist, während der Recency-Effekt nur bei einem unmittelbaren Abruf der Informationen nach der Beobachtung wirksam ist (ebd.). Während des Beurteilungsprozesses können diverse Effekte zu Fehlern führen. Dazu zählen der Anker-/Kontrasteffekt, der Halo-Effekt, Beobachter mit einer Tendenz zur sozialen Erwünschtheit, logische und theoretische Fehler, die Beobachterdrift, kulturabhängige Beurteilungen, und Antworttendenzen bei der Verwendung von Skalen.

Zusätzlich können Fehler bei der Kommunikation und Integration der Einzelbeurteilungen auftreten. Beispielsweise kann ein Beobachter seine geäußerte Meinung aufgrund eines wahrgenommenen Konformitätsdrucks an die Meinung einer Referenzperson anpassen. Die Erkenntnis und das Bewusstsein dieser potenziellen Fehlerquellen sind entscheidend, um angemessene Strategien zur Minimierung dieser Risiken zu entwickeln und damit die Qualität und Zuverlässigkeit des Beobachtungs- und Bewertungsprozesses zu gewährleisten. Darüber hinaus ist es wichtig, die identifizierten Fehlerquellen nicht isoliert zu betrachten, sondern sie in einem ganzheitlichen Kontext zu sehen. Fehler in einem Bereich können weitreichende Auswirkungen auf andere Bereiche und somit auf das Gesamtergebnis der Beobachtung und Bewertung haben. Eine sorgfältige Planung und Vorbereitung der Beobachtungs- und Bewertungsprozesse können dazu beitragen, mögliche Fehlerquellen zu minimieren. Die frühzeitige Erkennung und Korrektur von Mängeln in der Gestaltung von Beobachtungs- und Beurteilungsinstrumenten können helfen, den Prozess zu optimieren und die Genauigkeit der gesammelten Daten zu verbessern. Es ist auch notwendig, die Beobachter sorgfältig zu schulen und sie auf mögliche Verzerrungen und Fehler hinzuweisen. Eine fundierte Schulung kann dazu beitragen, das Bewusstsein für potenzielle Fehlerquellen zu schärfen und Strategien zur Minimierung dieser Risiken zu erarbeiten. Die Berücksichtigung der physischen Bedingungen der Beobachtungsumgebung, einschließlich des Blickwinkels, der Lichtverhältnisse und möglicher Störungen, ist ebenfalls ein wichtiger Aspekt bei der Reduzierung von Beobachtungsfehlern. Auch technische Aspekte, wie die Qualität von Video- und Tonaufzeichnungen, sollten nicht vernachlässigt werden.

Die Einbeziehung der Beobachteten in den Prozess kann ebenfalls zur Reduzierung von Fehlern beitragen. Eine genaue Vorstellung der zu erwartenden Situation und des Verhaltens kann dazu beitragen, die Wahrscheinlichkeit von Missverständnissen und unerwartetem Verhalten zu reduzieren. Abschließend ist zu betonen, dass trotz aller Bemühungen zur Minimierung von Fehlern immer noch ein gewisses Maß an Unsicherheit besteht. Daher ist es wichtig, dass die Ergebnisse der Beobachtung und Bewertung mit einer angemessenen Vorsicht interpretiert werden. Dabei sollte stets die Möglichkeit in Betracht gezogen werden, dass einzelne Ergebnisse durch unbeabsichtigte Fehler beeinflusst worden sein könnten. Abbildung 8 visualisiert den Detailablauf einer Verhaltensbeobachtung und -beurteilung mit der Markierung der möglichen Fehlerquellen.



Abbildung 8 Detailablauf einer Verhaltensbeobachtung und -beurteilung mit einer Markierung möglicher Fehlerquellen (Quelle: modifizierte Darstellung aus Diagnostik- u. Testkuratorium, 2017, S. 52)

2.7 Durchführung eines Assessment Centers

In Bezug auf Organisation und Durchführung sind diverse Aspekte zu berücksichtigen, wie beispielsweise das Management von Einladungen der Teilnehmenden, die Festlegung von Zeitplänen, die Zuweisung von Rollen, die Erstellung einer detaillierten Agenda und die Bereitstellung ausgewählter Informationen für die Teilnehmenden im Vorfeld. Zudem kann die Zielsetzung des Assessment Centers den Teilnehmenden vorab kommuniziert werden. Des Weiteren ist es ratsam, Informationen über die Art der Übungen, beteiligte Personen, Verwendung und Speicherung der AC-Daten sowie Tipps zur Vorbereitung und allgemein relevante Verhaltensweisen bereitzustellen. Abbildung 9 verdeutlicht, welche Informationen den Teilnehmenden vor dem Assessment Center zur Verfügung gestellt werden sollten.

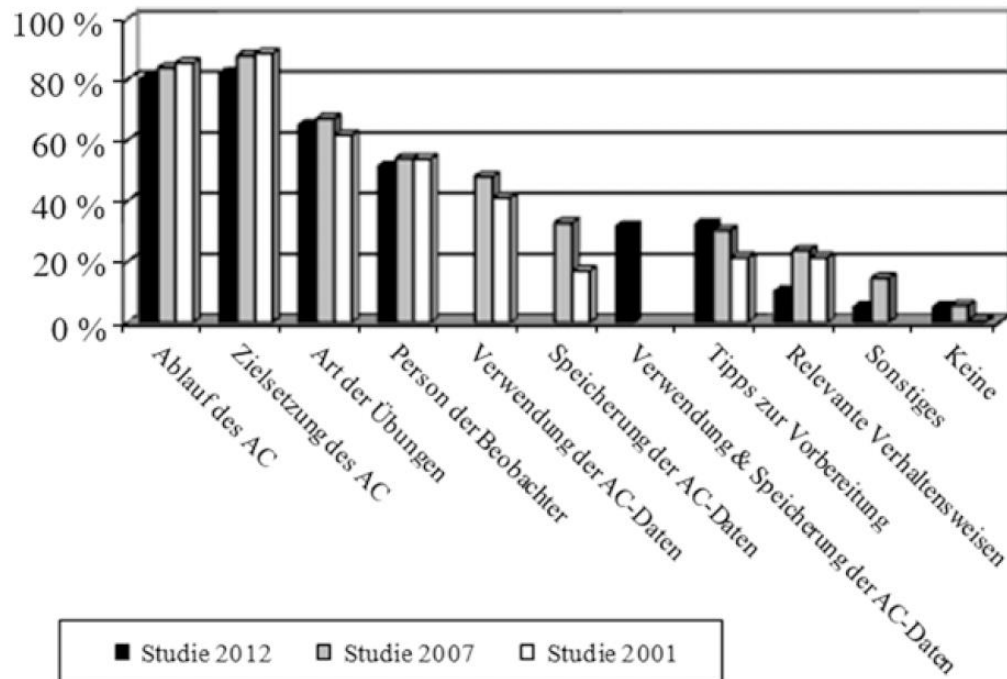


Abbildung 9 Welche Informationen erhält der Teilnehmer vor dem AC? (Quelle: Obermann et al. 2012, zitiert nach Obermann, 2018, S. 247)

2.8 Gesamturteil

Im Kontext von Assessment Centern stellt die Gesamtbeurteilung einen integralen Bestandteil dar, durch den das gesammelte Verhaltensprofil der Teilnehmenden zusammengefasst und interpretiert wird. Diese Beurteilung bildet die Grundlage für Entscheidungen über Personalentwicklung und -auswahl und berücksichtigt die Vielfalt der während des Assessment Centers gesammelten Informationen.

In der Praxis ist es oft unerlässlich, eine systematische Herangehensweise zur Konsolidierung der verschiedenen Perspektiven und Beobachtungen einzusetzen, die während eines Assessment Centers anfallen. Eine bewährte Methode, um diese Herausforderung zu bewältigen, ist die Anwendung von Beobachterkonferenzen⁴. Diese ermöglichen eine gezielte Diskussion und Integration der unterschiedlichen Beobachtungen und Eindrücke der Beobachterinnen und Beobachter.

Die Anwendung von Beobachterkonferenzen innerhalb des Assessment Center (AC)-Kontexts ist potenziell anfällig für Manipulation und subjektive Entscheidungsfindung,

⁴ Eine Beobachterkonferenz bezeichnet eine Gruppendiskussion, die typischerweise am Ende eines Assessment Centers (AC) abgehalten wird. In einer solchen Konferenz kommen die Beobachter, die während des gesamten AC-Prozesses die Kandidaten bewertet haben, zusammen, um ihre Beobachtungen und Einschätzungen auszutauschen und eine gemeinsame Entscheidung über die Eignung der Kandidaten für die jeweilige Position oder Aufgabe zu treffen.

insbesondere im Rahmen des qualitativen, klinischen Ansatzes, der sowohl im deutschsprachigen Raum als auch international bevorzugt wird (Kuncel et al., 2013). Trotz der Präferenz für diesen qualitativen Ansatz legen Studien nahe, dass das quantitative Verfahren in außeruniversitären Kontexten eine überlegene prognostische Gültigkeit in Bezug auf den späteren Berufserfolg aufweist und weniger fehlerhafte Urteile enthält (Kuncel et al., 2013). Der qualitative Ansatz birgt Vorteile wie Konsistenz und kollektive Entscheidungsfindung unter den Beobachtern, eine erhöhte Akzeptanz und Effizienz bei den involvierten Hierarchieebenen, stärkere Unterstützung für unerwünschte Ergebnisse und die Berücksichtigung individueller Aspekte.

Im Gegensatz dazu zeichnet sich das quantitative Verfahren durch eine höhere Entscheidungsqualität, eine präzisere Prognose des Berufserfolgs, Unabhängigkeit von Statusdifferenzen und Gruppendruck, sowie Zeit- und Kostenersparnis und eine breitere Datenabdeckung aus. Um die Vorzüge beider Methoden zu nutzen, können qualitative und quantitative Verfahren miteinander kombiniert werden. Beispielsweise kann das rechnerisch ermittelte Ergebnis als Grundlage für Diskussionen dienen, wobei jede Abweichung begründet und dokumentiert werden muss. Eine alternative Herangehensweise könnte darin bestehen, Diskussionen auf Grenzfälle oder größere Abweichungen zwischen den Beurteilern zu konzentrieren.

Eine weitere Option wäre die Nutzung der Vielfalt der Beobachter durch die Erhebung des Durchschnitts der anonym erhobenen qualitativen Urteile der einzelnen Beobachter.

Schließlich ist zu berücksichtigen, dass AC-Dimensionen im Durchschnitt höhere Validitäten als einzelne AC-Aufgaben aufweisen. Daher sind Diskussionen auf der Ebene einzelner Aufgaben weniger zielführend, da sie einen übermäßigen Zufallsanteil enthalten (Obermann, 2018, S. 260ff).

Zusammenfassend lässt sich festhalten, dass die Anwendung von Assessment Centern und insbesondere die Durchführung von Beobachterkonferenzen ein komplexer Prozess ist, der sowohl quantitative als auch qualitative Ansätze erfordert, um eine fundierte und faire Beurteilung der Kandidaten sicherzustellen. Indem die Stärken beider Ansätze genutzt und potenzielle Fehlerquellen berücksichtigt werden, kann die Qualität der Entscheidungsfindung verbessert werden.

2.8.1 Rückmeldung an die Teilnehmer

Im Einklang mit den Qualitätsstandards des deutschsprachigen Arbeitskreises Assessment Center hat jeder Teilnehmer eines Assessment Centers das Recht auf

personalisiertes Feedback. Dies fördert nicht nur das Verständnis der Teilnehmer für die Ergebnisse, sondern bietet ihnen auch die Möglichkeit, aus der Erfahrung zu lernen.

Obermann (2018, S. 266) weist darauf hin, dass dieses Feedback in verschiedenen Formaten bereitgestellt werden kann, einschließlich schriftlicher, mündlicher oder kombinierter Formen. Die Effektivität des Feedbacks kann durch eine Vielzahl von Faktoren beeinflusst werden. Auf Seiten des Feedbackgebers sind die wahrgenommene Kompetenz, die zugeschriebenen Motive und die grundsätzliche Einstellung gegenüber dem Feedbackempfänger sowie die Wahl der eingesetzten Instrumente von entscheidender Bedeutung. Hinsichtlich des Feedbackempfängers beeinflussen Faktoren wie der Selbstwertschutz, das Ausmaß der Diskrepanz zwischen Selbst- und Fremdbild und das Ausmaß der Selbstwirksamkeit die Wirksamkeit des Feedbacks. Bei der Gestaltung des Feedbacks spielen die Reihenfolge der Feedbackpunkte, die Anzahl der kommunizierten Entwicklungsfelder, die Realitätsnähe der Übungen, die Spezifität des Feedbacks, der Bezugspunkt des Feedbacks, die Form (schriftlich oder mündlich), die Verbindung des Feedbacks mit den gesetzten Zielen, die Verständlichkeit, der Zeitpunkt der Entscheidungsfindung und allgemein das Ergebnis des Assessment Centers eine entscheidende Rolle für die Wirksamkeit des Feedbacks beim Empfänger. Darüber hinaus können Kontextfaktoren wie zeitliche Aspekte und die wahrgenommene Unterstützung sowie der wahrgenommene Nutzen für die Entwicklung die Wirksamkeit des Feedbacks beeinflussen (Arbeitskreis AC 2008, zitiert nach Obermann, 2018, S. 270).

2.8.2 Assessment Center versus Development Center

Ein Development Center wird charakterisiert durch die Integration von diagnostischen Verfahren, wie sie in Assessment Centern Anwendung finden, mit pädagogischen Interventionen zur Lernförderung und Coaching-Ansätzen zur Unterstützung von Lern- und Karrierezielen (Eck et al., 2016, S. 44). In Analogie zum Assessment Center ist es initial notwendig, spezifische Fragestellungen bezüglich der Kandidaten zu definieren, während gleichzeitig eine geeignete Auswahl an Methoden berücksichtigt werden muss, um eine valide Beantwortung dieser Fragestellungen sicherzustellen.

In der Konzeption und Implementierung eines Development Centers ist es essenziell, verstärkt Feedback in Form von Lernzyklen zu integrieren, einen erweiterten Fokus auf bisher ungenutzte Fähigkeiten oder Talente jenseits der eng definierten Fragestellung zu legen, und eine tiefgreifende Berichterstattung zu garantieren. Diese sollte ein umfassendes Feedback und Beratung zu folgenden Aspekten umfassen:

- Fähigkeiten (Skills, Kompetenzen)
- Limitationen (Grenzen)
- Offenkundige Potentiale sowie
- Strategien zur Optimierung bestehender Talente oder zur Modifikation ungeeigneter Lern- bzw. Problemlösungsstrategien (Eck et al., 2016, S. 44).

Gemäß Obermann (2018, S. 247) weist das Resultat eines Development Centers einen qualitativen Charakter auf. Neben Stärken werden auch Entwicklungsfelder identifiziert und adressiert. folgende Abbildung 10 zeigt ein Beispiel einer qualitativen Ergebniszusammenfassung.

Stärken	
Hohe Kompetenz, Mitarbeiter zu motivieren:	Vermittelt Wertschätzung und Optimismus; berücksichtigt die Interessen seiner Mitarbeiter und bemüht sich um ihre Arbeitszufriedenheit; ist hilfsbereit.
Ausgeprägte Kundenorientierung:	Reagiert gut auf Markterfordernisse und Bedarfssignale von Kunden; realisiert Akquisitionschancen, präsentiert gute Initiativen zur Steigerung der Service- und Kundenorientierung sowie des Vertriebs
Gutes unternehmerisches Denken:	Denkt und agiert kosten- sowie ertragsorientiert, geht gut mit Unternehmenszahlen um.
Sehr gutes analytisches Denken:	Sehr gute Abstraktionsfähigkeit, gute Analyse von Zahlen- und Tabellenmaterial sowie von Texten.
Entwicklungsfelder	
Geringe Überzeugungskraft:	Zu wenig tiefergehende Argumente und Argumentationsketten, könnte stärker mit dem Nutzen für den Gesprächspartner argumentieren sowie eigene Begeisterung vermitteln.
Geringe Prozessorientierung:	Fokussiert Einzelfalllösungen, könnte stärker auf schlecht laufende Prozesse eingehen.

Abbildung 10 Beispiel qualitative Ergebniszusammenfassung (Quelle: modifizierte Darstellung Obermann, 2018, S. 258)

2.9 Selbstauskunftsfragebögen

Fragebögen dienen typischerweise der Selbsteinschätzung von Individuen bezüglich ihrer Persönlichkeitsmerkmale, Einstellungen, Interessen oder Werte. Sie können ebenso zum Erheben eines Fremdbildes einer Person verwendet werden, indem etwa Vorgesetzte, Kollegen oder Mitarbeiter ihre Wahrnehmungen teilen. In diesem Zusammenhang liegt der Schwerpunkt eher auf Personalentwicklung als auf Personalauswahl. In der Personalauswahl deutscher Unternehmen weisen Fragebögen jedoch eine eher marginale Präsenz auf, wobei lediglich 20% der befragten Unternehmen die Nutzung von Persönlichkeitsfragebögen bestätigen. In einer Stichprobe von 999 Bewerbern berichteten etwa 25% von der Konfrontation mit Fragebögen. Trotz eines ansteigenden Trends in der Anwendung von Fragebögen über die Jahre, existiert bei Arbeitgebern eine gewisse Skepsis oder Unwissenheit bezüglich ihrer Nutzung. Eine mögliche Quelle dieser Skepsis könnte die Vulnerabilität von Fragebögen für sozial erwünschte Selbstdarstellung sein. Die oft unzureichende fachliche Kompetenz von Entscheidungsträgern resultiert häufig in der Anwendung von aus wissenschaftlicher Perspektive fragwürdigen Fragebögen. Diese weisen entweder unklare oder ungenügende Gütekriterien und Normierungsstichproben auf oder basieren auf typologischen Verfahren, die in der wissenschaftlichen Forschung, seit Jahrzehnten keine signifikante Rolle mehr spielen.

Selbsteinschätzungsbögen erfreuen sich bei Bewerbern einer moderaten Akzeptanz und schneiden im Vergleich zu Leistungstests weniger positiv ab. Die Akzeptanz erhöht sich mit der Erfahrung der Bewerber, wobei eine häufigere Anwendung von Fragebögen zu einer gesteigerten Akzeptanz führt. Ein wesentliches Problem bei der Anwendung von Fragebögen ist ihre Verfälschbarkeit, die im Vergleich zu Leistungstests größer ist. Bewerber könnten versucht sein, ihre Antworten den mutmaßlichen Erwartungen des Arbeitgebers anzupassen. Laut einer Metaanalyse von Viswesvaran und Ones (1999) liegt die Verfälschbarkeit von Persönlichkeitsfragebögen zwischen einer halben und fast einer Standardabweichung. In der Praxis ist die tatsächliche Verfälschung jedoch geringer, da die meisten Bewerber das Risiko einer maximalen Verfälschung nicht eingehen wollen oder diese als unfair empfinden. Ob durch Verfälschung tatsächlich ein Vorteil erlangt werden kann, hängt von verschiedenen Faktoren ab, einschließlich der Erkennung sozial erwünschter Antworten und der potenziellen negativen Bewertung von übermäßigen Merkmalsausprägungen. Empfehlungen, stets die Mitte der Skala zu markieren, sind häufig nicht zielführend. Ein weiteres Problem von Fragebögen ist ihre Beschränkung auf das Selbstbild des Respondenten. Individuen mit verzerrtem Selbstbild können unbeabsichtigt falsche Angaben machen. Während sich solche

mittleren Verzerrungen durch Normierung tendenziell eliminieren lassen, sind extreme Verzerrungen nicht auf diese Weise neutralisierbar. In solchen Fällen wird eine Kombination von Fragebögen mit anderen diagnostischen Methoden, wie etwa einem Assessment-Center, empfohlen (Kanning, 2019, S. 517ff). Die vorliegenden Betrachtungen verdeutlichen, dass Fragebögen trotz ihrer vielfältigen Einsatzmöglichkeiten und ihres Potenzials in der Personalauswahl und -entwicklung sowohl Vorzüge als auch Grenzen aufweisen. Eine sorgfältige Auswertung und Interpretation der gewonnenen Daten sowie die Kombination mit weiteren diagnostischen Instrumenten können dazu beitragen, die Validität der Ergebnisse zu verbessern und den bestmöglichen Nutzen aus diesen Werkzeugen zu ziehen.

2.9.1 LeadIM Fragenbogen

In den jüngsten wissenschaftlichen Arbeiten zur Führungsdiagnostik hat sich der LeadIM-Selbstauskunftsfragebogen als bedeutend erwiesen, der von Christoph Janssen und seinen Kollegen entwickelt wurde (Janssen et al., 2022). Sein Hauptziel besteht darin, motivationsorientierte und selbstreflektierende Aspekte des Führungsverhaltens zu vereinen, indem er verschiedene Führungskompetenzen wie Überzeugungskraft und bestimmte Einstellungen wie das Streben nach Gestaltung berücksichtigt. Der LeadIM baut auf der Tradition berufsbezogener Persönlichkeitsinventare auf, wie beispielsweise das Berufliche Persönlichkeitsinventar (BIP) (Hossiep und Paschen, 2019) und der Fragebogen zur integrativen Führung (FIF) (Rowold und Poethke, 2017). Sein einzigartiger Mehrwert liegt jedoch in der Berücksichtigung von motivatorischen und selbstschutzbezogenen Facetten der Führungspersönlichkeit. In einer Studie, die die konvergente und diskriminante Validität des LeadIM-Fragebogens im Vergleich zum deutschen "Multifactor Leadership Questionnaire" (MLQ) untersuchte, wurden die innovativen Aspekte des LeadIM hervorgehoben. Es wurde erwartet, dass insbesondere das Führungs-Mindset (Leadership Mindset) und das Führungsverhalten (Leadership Strategies) im LeadIM mit den entsprechenden Skalen des MLQ korrelieren würden. Die zusätzlichen im LeadIM erfassten Motivatoren (Boost Factors) und persönlichen Schutzfaktoren (Protectors) sollten sich jedoch von den MLQ-Skalen unterscheiden. Die empirischen Ergebnisse bestätigten eine stabile Sechsfaktorenstruktur, die über 70% der Varianz erklärte, was etwa 10% mehr als die Varianzerklärung des MLQ-Fragebogens ausmachte. Insbesondere die "Protectors" und "Boost Factors" im LeadIM stellten eigenständige, führungsbezogene Persönlichkeitsdimensionen dar, die über die durch MLQ erfassten Facetten der transformationalen und transaktionalen Führung hinausgingen. Zusammenfassend lässt sich sagen, dass der LeadIM-Fragebogen eine wertvolle Ergänzung zum MLQ darstellt und dabei hilft, zusätzliche Persönlichkeitsfacetten im Führungskontext zu messen, was die kontinuierliche

Bedeutung der Diagnose von Führungskompetenzen und -potenzialen unterstreicht (Janssen et al., 2022, S. 6ff).

2.10 Ergebnisse Forschungsprojekt

Im Rahmen eines Forschungsprojekts wurde das Ausmaß der Konkordanz zwischen Personalentscheidungsträgern und einem algorithmischen Modell in Bezug auf die Beurteilung der Passgenauigkeit von Bewerberprofilen für eine Teamleiterposition in einem Unternehmenskontext analysiert. Dazu wurden 120 Kandidatenprofile mittels eines Selbsteinschätzungsfragebogens (LeadIM) evaluiert, der auf die Identifizierung von Führungspotenzialen, Haltungen und Verhaltensweisen abzielt. Diese Bewertungen wurden von zwölf erfahrenen Personalentscheidern durchgeführt. Der LeadIM-Fragebogen (Janssen et al., under review) kartiert vier Bereiche des Führungshandelns, basierend auf 128 Items und 5-stufigen Likert-Skalen. Die Ergebnisse werden auf einer normierten STEN-Skala dargestellt. Die Entscheidungsfindung wurde anhand eines vorgegebenen Anforderungsprofils durchgeführt und betraf die Eignung (1) oder Nichteignung (0) der Kandidatenprofile. Gleichzeitig wurde eine binär-logistische Regression auf Basis der Profildaten implementiert, die Wahrscheinlichkeitswerte für Nichteignung ($p < 0.5$) und Eignung ($p > 0.5$) berechnete und entsprechende Zuordnungen vornahm ($p < 0.5 \rightarrow '0'$; $p > 0.5 \rightarrow '1'$). Die binär-logistische Regression, ein mathematisches Modell für dichotome abhängige Variablen (Fahrmeir, Kneib & Lang, 2009), ermittelt standardisierte β -Parameter, die Wahrscheinlichkeitswerte berechnen. Die Analyse der 120 Testdatensätze ergab, dass alle vier Führungskomponenten positive β -Parameter aufweisen, was impliziert, dass höhere Werte eine höhere Wahrscheinlichkeit für die Eignung als Teamleiter bedeuten. Der Wert $\text{Exp}(B)$ repräsentiert die "Odds Ratio", die den Einfluss der Führungskomponenten auf die Wahrscheinlichkeit der Gruppenzuordnung verdeutlicht. Die Konfidenzintervalle für alle vier Komponenten sind signifikant positiv, was auf einen signifikanten Einfluss auf die Klassifikation der Probandenprofile in die Gruppen 1 (geeignet) oder 0 (nicht geeignet) hinweist. Die Gesamtmodellgüte wurde als ausreichend für weitere statistische Analysen bewertet (Nagelkerkes R-Quadrat = 0,819; Omnibus-Test der Modellkoeffizienten n.s., Hosmer-Lemeshow-Test n.s.). Ein Panel aus 12 Experten in der Personalauswahl für Führungspositionen bewertete die 120 Profile unabhängig von der mathematischen Berechnung binär (1: geeignet, 0: nicht geeignet). Diese Bewertung erfolgte ausschließlich auf Basis des jeweiligen LeadIM-Profiles und unter Berücksichtigung einer Stellenbeschreibung als Referenz. Persönliche Daten wie Alter, Geschlecht oder Berufserfahrung waren den Entscheidern nicht zugänglich. Um Reihenfolgeeffekte zu vermeiden, wurden die Profile in zufälliger Reihenfolge präsentiert.

Die Ergebnisse offenbarten Unterschiede in der Bewertungsstrenge. Durch die Eliminierung der beiden extremen Bewertungen erhöhte sich Fleiss Kappa von 0,358 auf 0,431 (moderate Übereinstimmung) (Landis und Koch, 1977), was auf eine moderate Inter-Rater-Reliabilität hindeutet. Die Bewertungen der Personalentscheidungsträger wurden dann mit den Wahrscheinlichkeitsberechnungen und Gruppenzuordnungen der binär-logistischen Regression verglichen. Eine Übereinstimmungsrate von mindestens 70% war erforderlich, d.h., mindestens 7 von 10 Entscheidern mussten eine '1' für das jeweilige Profil vergeben. Für das Regressionsmodell war eine Mindestwahrscheinlichkeit von $p=0,5$ als Kriterium für die Gruppenzuordnung '1' festgelegt. Insgesamt stimmten die Personalentscheider und das Regressionsmodell bei 13 der 120 Profile überein, dass diese für die Teamleiterrolle geeignet waren. In 2 Fällen war das Modell kritischer als die Personalentscheider, und in 2 Fällen war es umgekehrt. Vergleiche der Mittelwerte zeigten signifikante Unterschiede auf den LeadIM-Skalen zwischen den als geeignet eingestuften Profilen, sowohl im Regressionsmodell als auch bei den Personalentscheidern. Es gab jedoch keine signifikanten Unterschiede zwischen den durchschnittlichen Profilbewertungen und den Streuungen in den Skalenwerten zwischen menschlichen Beurteilern und dem Regressionsmodell. Die Durchschnittswerte der als geeignet eingestuften Profile lagen in beiden Fällen zwischen 6,7 und 9,1, während die als ungeeignet eingestuften Profile Skalenwerte zwischen 4,9 und 6,1 aufwiesen. Die Studie offenbart eine hohe Übereinstimmung von über 96% zwischen der Einschätzung menschlicher Personalentscheider und dem binär-logistischen Regressionsmodell. Beide Systeme zeigten eine strenge Beurteilungsweise und stellten hohe Anforderungen, um ein Profil als geeignet für eine Teamleiterrolle zu bewerten. Das Regressionsmodell erwies sich als sehr robust gegenüber der Auswahl der Stichprobe und zeigte eine klare Diskriminierungsfähigkeit bei der Gruppenzuordnung. Die Ergebnisse der Studie verdeutlichen, dass in 116 von 120 Fällen die Einschätzungen der Personalentscheider mit den Resultaten der binär-logistischen Regression übereinstimmten (13 x "geeignet", 103 x "nicht geeignet"). In zwei Fällen bewerteten die menschlichen Entscheider ein Profil als geeignet, während der Algorithmus dies nicht tat, und in weiteren zwei Fällen war das Gegenteil der Fall. Zusammenfassend weisen die Befunde dieser Untersuchung einen hohen Grad an Konkordanz zwischen der menschlichen Beurteilung und den Ergebnissen der binär-logistischen Regression auf.

2.11 Eigene Fragestellungen

Die Wirtschaftspsychologie bietet eine Vielzahl von Methoden zur Ermittlung der Eignung von Individuen für Führungsrollen, darunter Assessment Center, Selbstauskunftsfragebögen und neuerdings KI-basierte Verfahren. Frühere Studien deuten darauf hin, dass Assessment Center und Selbstauskunftsfragebögen effektive Instrumente für die Auswahl und Bewertung von Führungskräften sein können. Jedoch besteht eine erhebliche Forschungslücke in Bezug auf den direkten Vergleich dieser traditionellen Methoden mit KI-basierten Verfahren. Angesichts der beschleunigten Digitalisierung und Technologisierung in Organisationen ist es von entscheidender Bedeutung, diesen Vergleich zu erforschen. Daher wird in dieser Masterarbeit der inkrementelle Nutzen eines Persönlichkeitsfragebogens (LeadIM Profil) in der Managementdiagnostik untersucht. Das vorliegende Forschungsdesign stellt das LeadIM Profil, einen Selbstauskunftsfragebogen, in den Mittelpunkt der Datenerhebung. Dieser Ansatz wird mit einer Bewertung im Assessment Center sowie einer künstlichen Intelligenz-gestützten Einschätzung der Eignungswahrscheinlichkeit kontrastiert. Die zugrundeliegende Annahme ist, dass die Integration dieser verschiedenen Methoden zu einer umfangreicheren und präziseren Prognose der Eignung für Führungspositionen führen wird.

Im Zuge dessen wurden folgende Hypothesen aufgestellt:

Hypothese 1: Das LeadIM Profil (Selbstbeschreibung) korreliert hoch mit der Fremdeinschätzung im Assessment Center.

Hypothese 2: Die Hinzunahme der Datenpunkte aus dem LeadIM Profil im Rahmen einer multiplen Regression erhöht in signifikanter Weise die Varianzaufklärung im Kriterium „Eignung für eine Teamleiterrolle“.

Hypothese 3: Die KI basierte Wahrscheinlichkeitseinschätzung der Eignung als Führungskraft stimmt mit dem qualitativen Assessment überein.

2.11.1 Präzisierung der Zielsetzung

Das Hauptziel dieser Masterarbeit ist es, zu untersuchen, welcher der Ansätze - Assessment Center, Selbstauskunftsfragebögen (im Speziellen das LeadIM Profil) oder KI-basierte Methoden - die größte Varianzaufklärung bei der Vorhersage der Eignung für eine Führungsrolle liefert. Dabei soll der inkrementelle Nutzen des LeadIM Profils in

der Managementdiagnostik ermittelt und die Effektivität der KI-basierten Methoden im Vergleich zu den traditionellen Ansätzen beurteilt werden.

3. Methode und Vorgehen

Das folgende Kapitel erläutert die methodologische Vorgehensweise und die angewandten Untersuchungstechniken innerhalb der Masterarbeit. Dabei liegt der Schwerpunkt auf der systematischen Datenerfassung, der Auswertungsprozedur und der Integration der Resultate in den theoretischen Kontext der Eignungsdiagnostik. Die Datenerfassung umfasst mehrere Phasen, die darauf abzielen, ein ganzheitliches Bild der Führungskompetenzen der untersuchten Individuen zu erlangen. Anfänglich werden die Studienteilnehmer aufgefordert, den LeadIM-Selbstauskunftsfragebogen auszufüllen, der auf vier fundamentalen Führungskomponenten basiert. Zusätzlich absolvieren die Teilnehmer ein dreistündiges Assessment Center, in welchem sie verschiedenen Übungen und Situationen ausgesetzt werden, um ihre Führungskompetenzen zu evaluieren. Wie in Kapitel 2.2.2 bezüglich der unterschiedlichen Methoden aufgeführt, wurden verschiedene Beobachtungsmethoden integriert, um eine detaillierte und individualisierte Einschätzung des beobachteten Individuums zu ermöglichen und die diagnostische Präzision zu verbessern. Dazu gehören ein strukturiertes Interview, ein simuliertes Mitarbeitergespräch und eine Teamansprache. Parallel zur Datenerfassung kommt die binär-logistische Regression mit Sigmoid-Funktion zum Einsatz. Diese Funktion modelliert den Zusammenhang zwischen den unabhängigen Variablen (wie Führungskompetenzen, Persönlichkeitsmerkmale, Erfahrungen) und der abhängigen Zielvariable (Eignung für eine Führungsrolle). Durch dieses statistische Verfahren lässt sich die Wahrscheinlichkeit der Eignung für eine Führungsposition anhand einer oder mehrerer unabhängiger Variablen vorhersagen. Nach der Datenerfassung schließt sich die Auswertungsphase an. Die qualitativen Daten aus den Interviews und simulierten Gesprächen werden hinsichtlich der vier Führungskomponenten des LeadIM Fragebogens analysiert, während die metrischen Daten aus dem Selbstauskunftsfragebogen und der KI-basierten Methode statistisch ausgewertet werden. Die gewonnenen Daten werden mittels Triangulation verarbeitet, wodurch die Ergebnisse des Selbsteinschätzungsfragebogens, des Assessment Centers und der KI-gestützten Methode gemeinsam betrachtet und analysiert werden. Diese Methode ermöglicht die Identifikation möglicher Übereinstimmungen und Diskrepanzen zwischen den Resultaten der verschiedenen Methoden und trägt zur Aufdeckung potenzieller Stärken und Schwächen jedes Verfahrens bei. Auf diese Weise kann die Validität und Reliabilität der gewonnenen Erkenntnisse erhöht werden. Die gesammelten Befunde

werden im Kontext der bestehenden Literatur zu Eignungsdiagnostik und Assessment Center interpretiert und hinsichtlich ihrer Implikationen für die Praxis und zukünftige Forschung diskutiert. Dabei wird besonderes Augenmerk auf die Einhaltung von Qualitätsstandards und Normen, wie die DIN 33430, gelegt.

3.1 Forschungsdesign

Um die beschriebene Fragestellung zu beantworten, wird ein qualitatives Forschungsdesign gewählt, das die theoretisch abgeleiteten Forschungshypothesen überprüft. Die Studie verfolgt einen mixed-method Ansatz, der qualitative und quantitative Daten miteinander verknüpft. Im Zentrum steht die Untersuchung der Führungskompetenzen von Führungskräften unter Verwendung dreier unterschiedlicher Verfahren: (1) eines Assessment Centers, (2) eines Selbsteinschätzungsfragebogens (LeadIM) und (3) einer KI-basierten Methode. Die zentrale Fragestellung lautet: Inwieweit kann ein Assessment Center im Vergleich zu einem Selbsteinschätzungsfragebogen und einer KI-basierten Methode zur Bewertung der Führungskompetenzen beitragen? Im Rahmen des Assessment Centers werden die Teilnehmenden einem Experteninterview, simulierten Mitarbeitergesprächen und Teamansprachen unterzogen. Die Gespräche und Ansprachen werden durch zwei Personen unabhängig beobachtet und bewertet. Die Bewertungen erfolgen, wie in Kapitel 2.6. aufgeführt, systematisch-gebunden anhand einer zuvor entwickelten Kriterienliste, die auf den vier Führungskomponenten des LeadIM Fragebogens basiert. Folgende werden die unterschiedlichen Methoden, die im Assessment Center Anwendung finden detailliert vorgestellt.

3.2 Qualitatives Experteninterview

Das gewählte qualitative Forschungsdesign ermöglicht die Erfassung subjektiver Perspektiven der interviewten Personen und verwendet ihre Darstellungsmuster als primäre Datenquelle für die Interpretation im Analyseprozess. Die Auswertung der Daten erfolgt durch einen Beobachter und den Interviewer. Durch die Durchführung von zehn Interviews wird eine hinreichende Varianz an Meinungen und Urteilen erfasst, um psychologische Repräsentativität bzw. exemplarische Generalisierbarkeit zu gewährleisten. Um die Nachvollziehbarkeit der gewonnenen Erkenntnisse zu sichern, wird regelgeleitet verfahren. Der hier verwendete Interviewleitfaden, dessen Gestaltung in Kapitel 2.4 als Basis für ein halbstrukturiertes Interview erläutert wurde, ermöglicht eine systematische Auswertung durch den Beobachter und den Interviewer in jedem durchgeführten Interview. Im Folgenden wird das zugrundeliegende Forschungsdesign detaillierter dargestellt.

Das leitfadengestützte Interview als Erhebungsinstrument

Im Kontext des qualitativen Forschungsdesigns der vorliegenden Studie dient das subjektive Verständnis der Befragten und deren Darstellungsmuster als essenzielle Datenquelle für den Interpretationsprozess. Die Methode des leitfadengestützten Interviews nimmt in diesem Kontext eine zentrale Rolle ein. Dieses teilstandardisierte Interviewverfahren stellt eine der am häufigsten genutzten Befragungsformen in der qualitativen Forschung dar. Die Teilstandardisierung resultiert aus den Leitfragen, die in jedem Interview zur Beantwortung stehen. Trotz der Anwendung eines Leitfadens bleibt die Formulierung und Reihenfolge der Fragen flexibel, um den natürlichen Charakter des Gesprächsverlaufs zu gewährleisten (Gläser & Laudel, 2006, S. 39; Bortz & Döring, 1995, S. 290). Zur Erfassung der Expertenmeinungen wird das Online-Interview als Untersuchungsinstrument eingesetzt. Mit einer geplanten Dauer von etwa 45 Minuten ist dieses Instrument darauf ausgerichtet, die zentralen Aspekte adäquat zu erfassen. Die Konzeption des Leitfadens erfolgt auf der Grundlage einer theoriegeleiteten Vorgehensweise (Gläser & Laudel, 2006, S. 108). Hierbei werden aus der Theorie abgeleitete Leitfragen in konkrete Interviewfragen transformiert. Der Leitfaden (s. Anhang D Interviewleitfaden) folgt einer logisch-thematischen Struktur. Jedes Interview startet mit einer allgemeinen Einführung in das Thema sowie der Vorstellung der Interviewer. Zudem wird der Ablauf des Interviews dargelegt. Das Interview gliedert sich in die Bereiche Begrüßung und Einleitung, persönliche Datenabfrage (Name, Beruf, Alter, Geschlecht, Führungsposition ja/nein, Branche, Datum, Beginn und Ende des Interviews) sowie den Hauptteil. Dieser besteht aus Fragen zu vier Kategorien: aktivierende Führung, selbstbewusste und reflektierende Führung, strukturierende und prinzipienorientierte Führung, empathische und integrierende Führung. Die Interviewten werden aufgefordert, ihre Erfahrungen aus diesen Bereichen mithilfe von Beispielsituationen aus ihrem (Führungs)alltag zu illustrieren. Dies beinhaltet, dass sie dazu aufgefordert werden, Szenarien aus ihrem (Führungs)alltag zu schildern, die ihre Fähigkeiten und Verhaltensweisen in den jeweiligen Führungskategorien aufzeigen. In der Kategorie "aktivierende Führung" beispielsweise werden die Teilnehmenden dazu ermutigt, Situationen zu beschreiben, in denen sie ihre Initiative zeigen mussten. Im Kontext der "selbstbewussten und reflektierenden Führung" sollen die Befragten Szenarien aufzeigen, in denen sie ihre Kolleginnen und Kollegen überzeugen und dabei Sinn stiften mussten. In der Kategorie "strukturierende und prinzipienorientierte Führung" werden die Interviewten gebeten, eine Situation zu illustrieren, in der ihre Entschlossenheit gefordert war. Schließlich, im Rahmen der "empathischen und integrierenden Führung", sollen die Teilnehmenden eine Situation darlegen, in der sie ein hohes Maß an Empathie und Integrationsfähigkeit aufzeigen mussten. Die

anschließende Bewertung der Antworten erfolgt durch einen unabhängigen Beobachter und den Interviewer selbst, basierend auf einer 5-Punkte-Skala. Dieses Vorgehen ermöglicht eine Beurteilung der individuellen Führungsfähigkeiten und -stile, die sich aus den geschilderten Situationen ergeben. Der vollständige Interviewleitfaden, ist im Anhang D Interviewleitfaden zu finden.

3.3 Konzeption des Rollenspiels

Die Evaluation führungsbezogener Kompetenzen durch Rollenspiele ist eine etablierte Methode in Assessment-Centern (ACs). In der folgenden Übung führt der AC-Teilnehmer als Teamleiter der Brikolage GmbH ein Gespräch mit einem seiner Mitarbeiter. Das Team, bestehend aus sieben Personen mit unterschiedlichen Teilaufgaben, ist stark voneinander abhängig, um gemeinsame Ziele zu erreichen. Daher hat die rechtzeitige und korrekte Erfüllung individueller Aufgaben einen erheblichen Einfluss auf die Gesamtleistung des Teams. Ziel dieses Rollenspiels ist es, eine Leistungsrückmeldung zu geben und auf private Probleme des Mitarbeiters zu reagieren, um bestehende Unzufriedenheit im Team zu lösen. Der Mitarbeiter, der im Rollenspiel vertreten wird, ist seit mehreren Jahren in der Abteilung tätig und war bisher immer zuverlässig und leistungsfähig. In den letzten Monaten hat sich dies jedoch geändert, wobei der Mitarbeiter Fehler macht und regelmäßig am Freitag und Montag krankheitsbedingt abwesend ist. Dies hat zu Unzufriedenheit und Überlastung im Team geführt, und es liegt an der Führungskraft, diese Situation zu bewältigen. In der Rolle des Mitarbeiters ist der Teilnehmer mit einer Krebserkrankung konfrontiert, die regelmäßige Behandlungen und daraus resultierende Fehlzeiten erfordert. Der Mitarbeiter möchte seine Krankheit nicht mit dem Team teilen und hat Angst, seinen Job zu verlieren, wenn seine Erkrankung bekannt wird. Dieser Kontext führt zu einem komplexen und emotional belastenden Szenario für die Führungskraft im Rollenspiel. Die Beurteilung der Leistung der Führungskraft im Rollenspiel basiert auf verschiedenen Führungskompetenzen:

Selbstbewusste und reflektierende Führung: Hierbei werden das Selbstvertrauen der Führungskraft, ihre Resilienz und ihre Fähigkeit zur Selbstreflexion bewertet. Außerdem wird beurteilt, wie gut sie ihre Lösungsstrategien überzeugend und sinnvoll vermitteln kann.

Strukturierte und ambitionierte Führung: Hierbei wird die Fähigkeit der Führungskraft bewertet, eine Lösung zu strukturieren und zu planen und ihre Willensstärke, an ihren Zielen trotz Widerständen festzuhalten.

Empathische und integrierende Führung: Hierbei wird die Fähigkeit der Führungskraft bewertet, empathisch auf den Mitarbeiter einzugehen, seine Situation zu verstehen und ihn zu motivieren, seine Perspektive zu ändern und seine Verhaltensweisen anzupassen.

Dieses Rollenspiel ermöglicht eine umfassende Bewertung der Führungsqualitäten des Teilnehmers in einem herausfordernden Kontext und bietet wertvolle Erkenntnisse für die Auswahl und Entwicklung von Führungskräften.

Im Detail werden folgende Kriterien zur Evaluierung der Führungsqualitäten herangezogen:

Selbstbewusste und reflektierende Führung:

- Selbstvertrauen: Die Führungskraft sollte sicher und selbstbewusst auftreten, auch in schwierigen Situationen. Sie sollte in der Lage sein, Probleme anzusprechen und klare Entscheidungen zu treffen.
- Resilienz: Die Führungskraft sollte zeigen, dass sie belastende Situationen bewältigen kann und bei Rückschlägen standhaft bleibt. Ihre emotionale Widerstandsfähigkeit ist wichtig, um das Team effektiv zu führen.
- Selbstreflexion: Nach dem Gespräch sollte die Führungskraft ihre Leistung und das Ergebnis des Gesprächs reflektieren und Bereiche für mögliche Verbesserungen identifizieren.
- Überzeugen und Sinn vermitteln: Es ist entscheidend, dass die Führungskraft eine klare Vision und Lösungsstrategie vermitteln kann, die sowohl für den betroffenen Mitarbeiter als auch für das Team sinnvoll ist.

Strukturierte und ambitionierte Führung:

- Strukturieren und Planen: Nach der Klärung der Situation sollte die Führungskraft einen klaren Plan für die Zukunft erstellen, der sowohl den Bedürfnissen des betroffenen Mitarbeiters als auch des Teams gerecht wird.
- Willensstärke: Die Führungskraft sollte in der Lage sein, an ihren Zielen festzuhalten, auch wenn sie auf Widerstand stößt. Dies erfordert Mut und Entschlossenheit.

Empathische und integrierende Führung:

- Empathie: Die Führungskraft sollte die Fähigkeit zeigen, sich in den betroffenen Mitarbeiter hineinzuversetzen, um sein Verhalten besser zu verstehen und angemessen darauf zu reagieren.
- Motivieren und aktivieren: Es ist wichtig, dass die Führungskraft den Mitarbeiter dazu motivieren kann, sein Verhalten zu ändern und seine Perspektive zu überdenken. Dies kann dazu beitragen, das Teamklima zu verbessern und die Leistung des Mitarbeiters zu steigern.

Das Rollenspiel bietet eine wertvolle Möglichkeit, die Führungsqualitäten des Teilnehmers zu evaluieren und zu entwickeln. Es stellt eine realistische und herausfordernde Situation dar, in der sowohl die Fachkompetenz als auch die soziale und emotionale Kompetenz der Führungskraft gefordert sind. Durch die Bewertung der Leistung der Führungskraft anhand dieser Kriterien können wertvolle Erkenntnisse für die Personalentwicklung und -auswahl gewonnen werden.

3.4 Teamansprache

Bei dieser Übung werden die Führungsqualitäten des Teilnehmers im Rahmen einer Teamansprache zur Einführung einer neuen Zeiterfassungssoftware bewertet. Die Übung umfasst vier zentrale Aspekte der Führungsqualität, die auf der Basis von vier Kategorien erhoben werden: Aktivierende und energetisierende Führung, Selbstbewusste Führung, Strukturierende und ambitionierte Führung sowie Empathische und integrierende Führung. Im Rahmen der Kategorie Aktivierende und energetisierende Führung wird die Fähigkeit des Teilnehmers beurteilt, Initiative zu ergreifen und das Team zu inspirieren. Es wird bewertet, ob der Teilnehmer den Mut hat, Neues zu wagen und das Team dazu ermutigen kann, sich auf die Einführung der neuen Software einzulassen. In der Kategorie Selbstbewusste Führung wird das Selbstvertrauen des Teilnehmers bewertet. Es wird untersucht, ob er resiliente Eigenschaften zeigt und wie gut er in der Lage ist, das Team von der Notwendigkeit der neuen Software zu überzeugen und den Sinn dieser Veränderung zu vermitteln.

In der Kategorie Strukturierende und ambitionierte Führung wird das Selbstvertrauen des Teilnehmers sowie seine Ambitionen bewertet. Es wird analysiert, wie er die Einführung der neuen Software plant und strukturiert und ob er ambitionierte Ziele für das Team setzt. In der Kategorie Empathische und integrierende Führung wird die Fähigkeit des Teilnehmers bewertet, Empathie zu zeigen und das Team zu verbinden und zu integrieren. Es wird beurteilt, wie gut der Teilnehmer das Team motivieren und aktivieren kann, sich auf die Einführung der neuen Software einzulassen und die damit

verbundenen Veränderungen zu akzeptieren. Die Bewertung basiert auf der Beobachtung des Teilnehmers während der Teamansprache und auf der Analyse seiner Kommunikations- und Interaktionsfähigkeiten. Es wird beurteilt, wie gut er das Team informiert, wie er auf Fragen und Bedenken des Teams reagiert und wie effektiv er das Team motiviert und aktiviert. Dabei wird sowohl die verbale als auch die nonverbale Kommunikation des Teilnehmers berücksichtigt. Die Bewertung findet durch eine 5-stufige Skala statt.

3.4.1 Stichprobe

In der vorliegenden Untersuchung wurden insgesamt zehn Individuen einbezogen. Das Alter der Teilnehmenden variierte zwischen 23 und 60 Jahren. Die Geschlechterverteilung in der Stichprobe zeigte eine Dominanz männlicher Teilnehmer (60 Prozent), während weibliche Teilnehmerinnen 40 Prozent repräsentierten. Die berufliche Herkunft der Teilnehmenden zeigte eine breite Streuung. Ein Großteil der Probanden (60 Prozent) stammte aus dem Sozialsektor, während 10 Prozent im Gesundheitsamt tätig waren. Darüber hinaus waren 20 Prozent der Teilnehmenden im Wirtschaftssektor aktiv. Vor Beginn der Untersuchung wurden die Teilnehmer über die geplante Agenda und die angewendeten Methoden im Rahmen der Development Center informiert. Diese zehn geplanten Development Center fanden zwischen Anfang Januar und Mitte März 2023 statt. Direkt im Anschluss an die Durchführung begann die Phase der Datenanalyse.

Die Rekrutierung der Probanden basierte auf den folgenden Kriterien:

- Aktuelle oder frühere Inhaber einer Führungsposition
- Personen, die eine Führungsrolle anstreben
- Zugehörigkeit zu unterschiedlichen Unternehmensformen, darunter der Wirtschaftssektor, der soziale Bereich und der öffentliche Dienst
- Eine Streuung der Mitarbeiterzahlen unter den Teilnehmenden
- Ein explizites Interesse am Forschungsfeld der Eignungsdiagnostik, Assessment Center, und der Anwendung von Künstlicher Intelligenz (KI) zur Unterstützung der Personalauswahl

Die Probandenrekrutierung wurde innerhalb des persönlichen Berufs- und Freundesnetzwerks durchgeführt, wobei eine professionelle Distanz zu den Probanden gewahrt wurde (vgl. Gläser & Laudel, 2006, S. 114).

3.5 Assessment Center Ablauf (2 Seiten)

Die Teilnehmenden erhalten im Zusammenhang mit der Durchführung des Online Assessments eine Teamseinladung und eine Erläuterung bezüglich der im Assessment Center angedachten Methoden. Darüber hinaus erhalten alle Teilnehmenden den Link zum Selbstauskunftsfragebogen (LeadIM) mit der bitte den Selbstauskunftsfragebogen im Vorfeld auszufüllen und die Ergebnisse direkt im Anschluss an das Assessment Center zur Verfügung zu stellen. Die Agenda des remote Assessment Centers ist wie folgt:

- 5 Minuten Einführung und Vorstellungsrunde
- 45 Minuten Interview
- 30 Minuten Vorbereitungszeit für die Teamansprache, parallel dazu erfolgt die Beurteilung bezüglich des Interviews
- 20 Minuten Teamansprache
- 30 Minuten Vorbereitungszeit für das Mitarbeitergespräch, parallel dazu erfolgt die Beurteilung bezüglich der Teamansprache
- 20 Minuten Mitarbeitergespräch
- 20 Minuten Beurteilung der Teamansprache
- 10 Minuten Auswertungsgespräch für die Teilnehmer

Das Interview wird durch einen Interviewer geführt und von der Beobachterin dokumentiert. Im Anschluss an das Interview erfolgt ein Abgleich der Bewertungen zwischen dem Interviewer und der Beobachterin. Die Bewertungen zwischen Interviewer und Beobachterin stimmen jeweils überein. Zur Vorbereitung der Teamansprache und des Mitarbeitergesprächs erhält der Teilnehmer eine Instruktion in der die Situation des Rollenspiels aufgeführt ist. Die Instruktion befindet sich im Anhang B Teamansprache und Anhang C Mitarbeitergespräch. Bevor das Rollenspiel startet hat der Teilnehmende die Möglichkeit Fragen zu stellen. Im Anschluss an die Teamansprache und das Mitarbeitergespräch werden die Teilnehmenden befragt ob sie so, wie sie in dem Rollenspiel und der Teamansprache agiert haben, auch in einer nicht gestellten Situation agieren würden, vlg. Kapitel 2.5.2 bezüglich Reflexionsfragen im Rollenspiel. Anschließend erfolgt eine Bewertung durch die Beurteiler. Abbildung 11 zeigt exemplarisch eine Ergebnisübersicht des durchgeführten Einzel-Assessment-Centers (AC). Es enthält die Bewertungen der Teilnehmerin in verschiedenen Kategorien, die auf ihre Führungskompetenzen abzielen. Die Straffierten Bereiche wurden in den jeweiligen Settings nicht erhoben. Die Kategorien sind in vier Hauptgruppen unterteilt: Aktivierende und energetisierende Führung, selbstbewusste Führung, strukturierende und ambitionierte Führung sowie empathische und integrierende Führung. In jeder

Hauptgruppe werden verschiedene Aspekte der Führungskompetenzen bewertet, und die Ergebnisse werden in Form von Punkten dargestellt. Die Punkte reichen von 1 (niedrig) bis 5 (hoch) und zeigen das Ausmaß, in dem die Teilnehmerin in den jeweiligen Aspekten erfolgreich war. Für die Kategorie "Aktivierende und energetisierende Führung" werden die Teilbereiche Lernmotivation, Initiative, Gestalten und Vorgehen sowie Inspirieren und Neues wagen bewertet. In der Kategorie "Selbstbewusste Führung" werden Selbstvertrauen, Resilienz, Selbstreflexion und Überzeugen und Sinn vermitteln beurteilt. In der Kategorie "Strukturierende und ambitionierte Führung" werden Strukturieren und Planen, Willensstärke, Selbstvertrauen und Ambition bewertet. Schließlich umfasst die Kategorie "Empathische und integrierende Führung" die Teilbereiche Empathie, Verbinden und Integrieren, Entwickeln und Fördern sowie Motivieren und Aktivieren. Die Ergebnisse aus dem LeadIM wurden rekodiert, d.h. LeadIM 1 und 2 werden auf 1 rekodiert, LeadIM 3 und 4 werden auf 2 rekodiert, LeadIM 5 und 6 werden auf 3 rekodiert, LeadIM 7 und 8 werden auf 4 rekodiert und LeadIM 9 und 10 werden auf 5 rekodiert. In den letzten beiden Spalten sind die Durchschnitte einmal mit Betrachtung des LeadIM und einmal ohne Betrachtung der Bewertungen des LeadIM Fragebogens aufgeführt. Die Bewertungen erfolgen anhand einer zuvor entwickelten Kriterienliste, die auf den vier Führungskomponenten des LeadIM Fragebogens basiert.

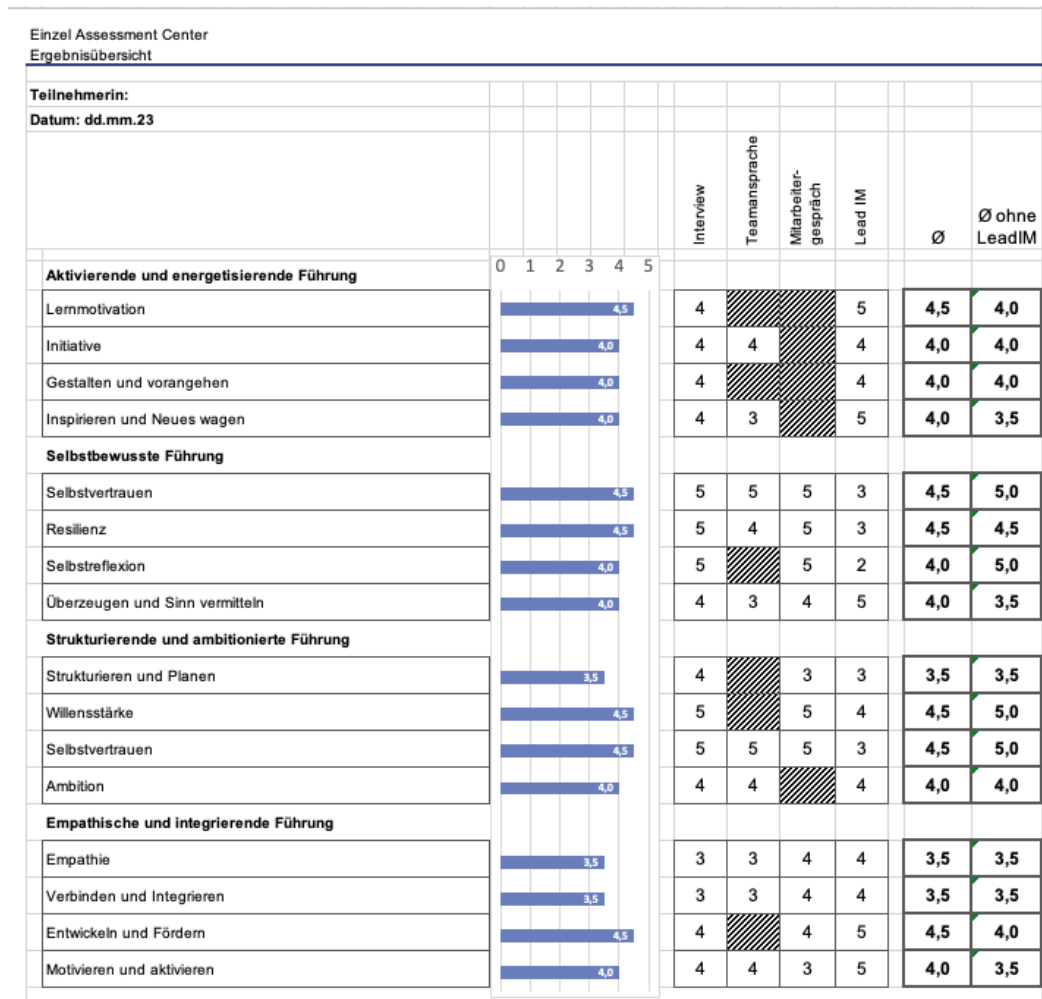


Abbildung 11 Exemplarische Darstellung der Ergebnisse eines Teilnehmers, die Straffierten Felder wurden in den jeweiligen Settings nicht erhoben (Quelle: eigene Darstellung)

4. Ergebnisse

Im Folgenden Kapitel werden die Ergebnisse der durchgeführten Untersuchungen präsentiert, welche die empirische Grundlage dieser Arbeit darstellen. Zu Beginn werden zunächst die deskriptiven Statistiken der erhobenen Daten präsentiert. Diese erlauben einen ersten Überblick über die zentralen Tendenzen der erhobenen Variablen, bevor weiterführende Analysen, wie beispielsweise die Hauptkomponentenanalyse und die logistische Regression angewandt werden.

4.1 Deskriptive Statistik zu den erhobenen Variablen

Die Gesamtstichprobe setzt sich aus einem Datensatz von 130 Personen zusammen, wobei 10 Personen an dem Assessment teilnehmen und 120 Datensätze aus einem größeren Datensatz (N=1411) eines Managementaudits zufällig ausgewählt wurden. Die nachfolgende Grafik stellt die Altersverteilung der in Betracht gezogenen Gesamtstichprobe dar.

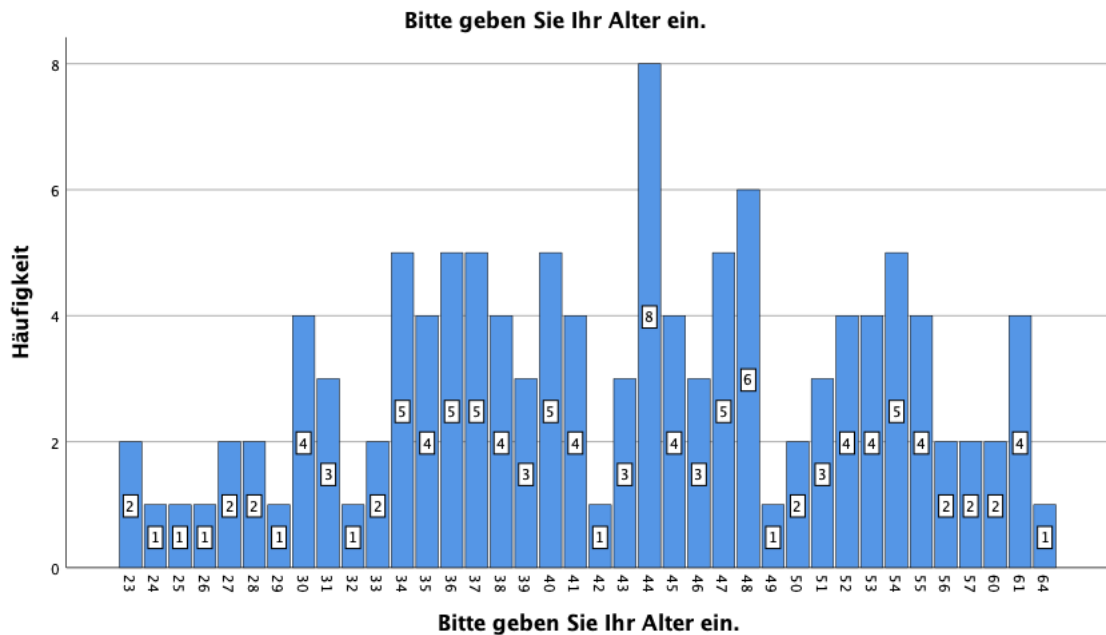


Abbildung 12 Alter der Teilnehmer (Quelle: Ausgabe SPSS)

In der Gesamtstichprobe sind das weibliche Geschlecht 34-mal, das männliche Geschlecht 91-mal vertreten. Bei 5 Datensätzen fehlt die Angabe. Bei der Datenanalyse zu Bildungsabschlüssen ergibt sich ein differenziertes Bild, 17 der Teilnehmer verfügen über einen Dokortitel. Ein Großteil der Stichprobe, konkret 62 Personen, besitzt einen Hochschulabschluss, darunter Masterabschlüsse und Diplome. Zudem haben 17 Teilnehmer einen Bachelorabschluss. Im Bereich der beruflichen Bildung gibt es neun Teilnehmer, die eine duale Ausbildung absolviert haben, während sechs Teilnehmer eine Berufsausbildung abgeschlossen haben. Darüber hinaus besitzen 13 Teilnehmer einen sekundären Bildungsabschluss, einschließlich Gymnasium, Realschule oder Hauptschule. Für sechs Teilnehmer liegen keine spezifischen Informationen zum höchsten Bildungsabschluss vor, s. Abbildung 13.

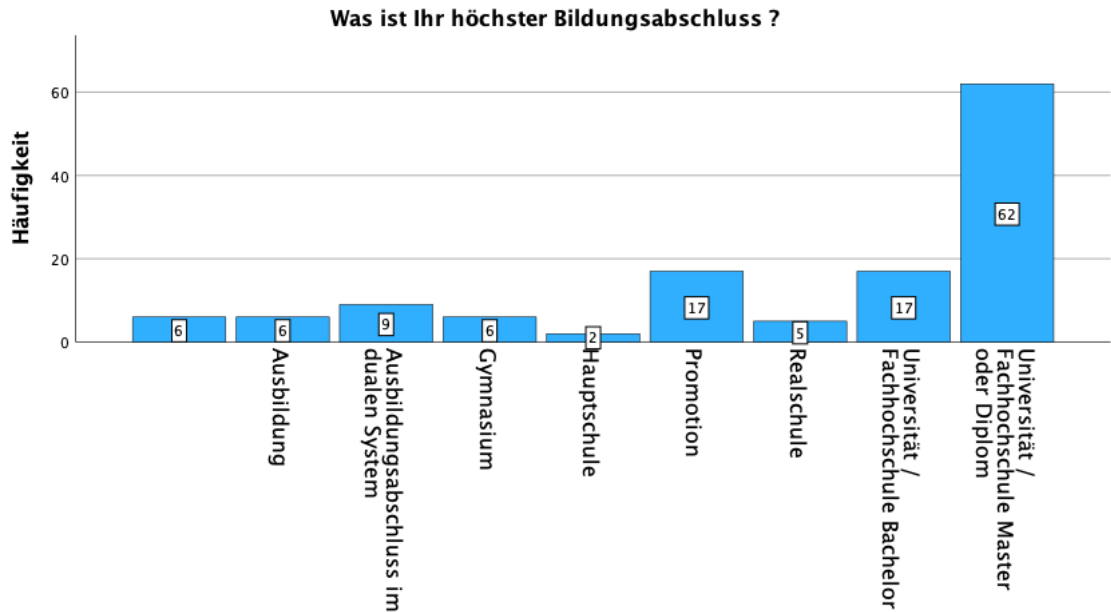


Abbildung 13 höchster Bildungsabschluss (Quelle: Ausgabe SPSS)

Die Analyse des beruflichen Status der Teilnehmer ergibt ein klares Bild: Die Mehrheit, nämlich 111 Teilnehmer, ist in Vollzeit beschäftigt. Eine kleinere Gruppe von sieben Teilnehmern ist in Teilzeit tätig. Drei Teilnehmer sind derzeit ohne Beschäftigung. Weitere vier Teilnehmer befinden sich in Ausbildung oder Studium. Bei fünf Teilnehmern liegen keine ausreichenden Informationen zum aktuellen beruflichen Status vor.

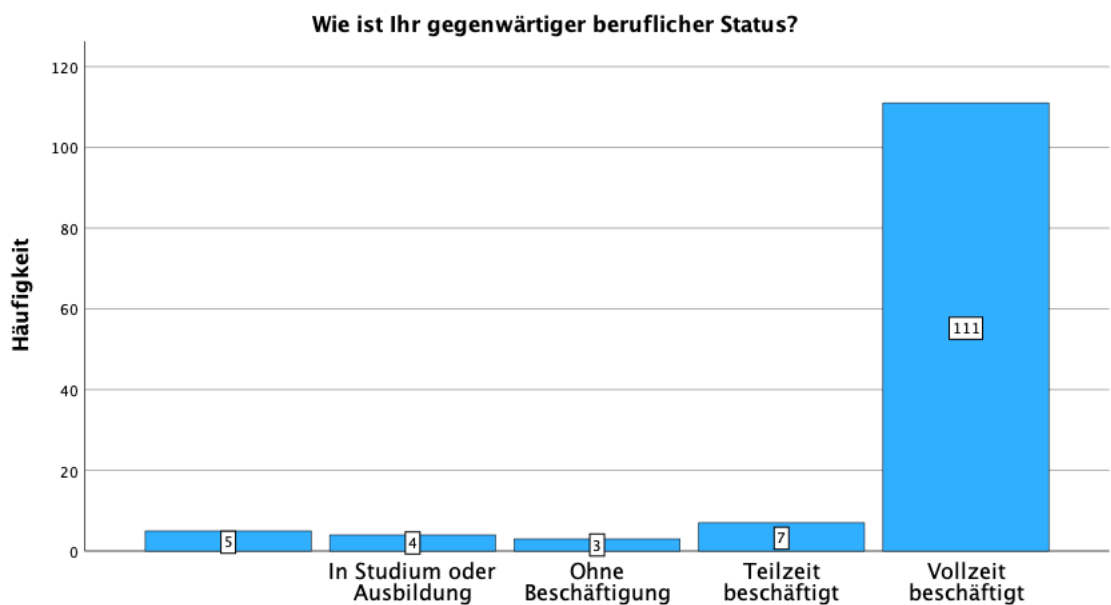


Abbildung 14 Beruflicher Status (Quelle: Ausgabe SPSS)

Die Analyse der gegenwärtigen Rollen innerhalb der Stichprobe offenbart folgende Verteilung: Es lassen sich vier Mitglieder der Geschäftsleitung identifizieren, 19 Personen bekleiden die Position eines Abteilungsleiters und 10 Teilnehmer fungieren als Bereichsleiter. Die Rolle des Teamleiters wird von 56 Teilnehmern eingenommen,

während 24 Personen als Spezialisten tätig sind. Vier Personen agieren als Sachbearbeiter und ein Teilnehmer befindet sich in der Ausbildung. Für 12 Datensätze lassen sich keine Informationen hinsichtlich der gegenwärtigen Rolle erheben.

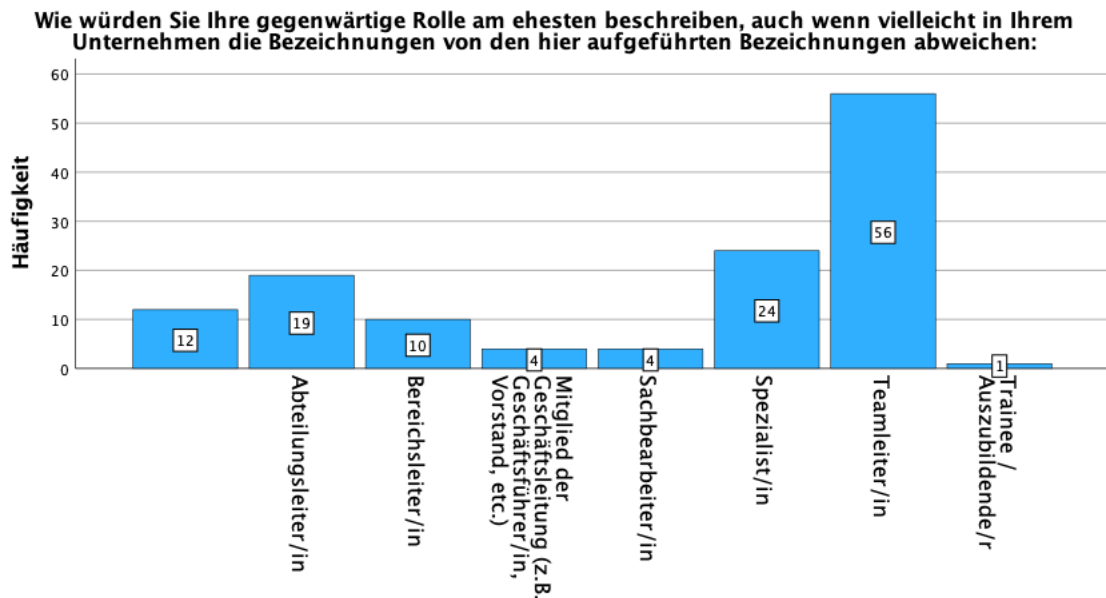


Abbildung 15 Rolle (Quelle: Ausgabe SPSS)

Im folgenden Kapitel vollzieht sich die Replikation der Befunde des Forschungsprojekts. Die Replikation der Forschungsergebnisse dient zur Überprüfung und Bestätigung der Zuverlässigkeit und Validität der ursprünglichen Erkenntnisse. Daher konzentriert sich dieses Kapitel auf die kritische Wiederholung und Überprüfung der vorangegangenen Untersuchungsergebnisse. Das Ziel ist, eine Bestätigung oder möglicherweise eine Revision der ursprünglichen Befunde und Schlussfolgerungen zu erreichen.

4.2 Replikation der Ergebnisse des Forschungsprojektes

Die Hauptkomponentenanalyse (PCA) erlaubt die wesentlichen Variablen zu identifizieren, die die größte Varianz in einem gegebenen Datensatz repräsentieren. Dies geschieht durch Auswertung der Korrelationen zwischen den verschiedenen Variablen. Es können bereits in der initialen Korrelationsmatrix Cluster identifiziert werden, die im späteren Verlauf eine eigene Komponente bilden. Das Ziel der Durchführung der Faktorenanalyse für jede der beiden Datenmengen ist die Bestätigung der Stabilität und Reproduzierbarkeit der Faktorenstruktur über verschiedene Stichproben hinweg. In diesem Kontext bezeichnet die Faktorenstruktur die Zuordnung von Items zu bestimmten Faktoren, die aufgrund ihrer hohen Korrelation in Erscheinung treten. Es zeigt sich eine konsistente Darstellung der gleichen vier Komponenten sowohl im ursprünglichen Datensatz von 120 Fällen als auch im erweiterten Datensatz von 130 Fällen, was eine stabile und zuverlässige Faktorenstruktur impliziert. Allerdings ist bei der Interpretation dieses Ergebnisses Vorsicht geboten. Zwar ist die Konsistenz der

Faktorenstruktur in beiden Datensätzen ein positiver Indikator, jedoch kann die Hinzufügung von 10 weiteren Probanden zum Datensatz, insbesondere im Vergleich zum ursprünglichen Umfang von 120 Beobachtungen, geringfügige Auswirkungen haben.

Erklärte Gesamtvarianz

Abbildung 16 präsentiert das Ergebnis der Hauptkomponentenanalyse. Die Eigenwerte, die die Varianz der Komponenten darstellen, geben Auskunft darüber, welche der Komponenten relevant sind. Das Kaiser-Guttman-Kriterium besagt, dass Eigenwerte größer als 1 bedeutungsvolle Faktoren repräsentieren (Brandt, 2020, S. 590). Die Division der Eigenwerte durch die Anzahl der Komponenten gibt Aufschluss über den Anteil der Varianz, den diese Komponente erklärt. Die Ergebnisse deuten darauf hin, dass es vier bedeutungsvolle Faktoren gibt. Die erste Komponente erklärt 43% der Varianz, die zweite Komponente 11% der Varianz, die dritte Komponente 9% der Varianz, die vierte Komponente 6% der Varianz. Somit erklären die vier Komponenten zusammen ungefähr 69% der Gesamtvarianz der Daten.

Komponente	Erklärte Gesamtvarianz								
	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	6.958	43.489	43.489	6.958	43.489	43.489	3.277	20.482	20.482
2	1.646	10.288	53.776	1.646	10.288	53.776	3.224	20.150	40.633
3	1.458	9.110	62.886	1.458	9.110	62.886	2.398	14.987	55.619
4	1.076	6.723	69.609	1.076	6.723	69.609	2.238	13.990	69.609
5	.871	5.442	75.051						
6	.633	3.955	79.007						
7	.574	3.586	82.593						
8	.497	3.106	85.699						
9	.461	2.881	88.580						
10	.394	2.463	91.044						
11	.332	2.072	93.116						
12	.301	1.883	94.999						
13	.248	1.548	96.547						
14	.211	1.318	97.865						
15	.185	1.155	99.019						
16	.157	.981	100.000						

Extraktionsmethode: Hauptkomponentenanalyse.

Gesamterklärte Varianz durch die vier Komponenten

Abbildung 16 Erklärte Gesamtvarianz (Quelle: Ausgabe SPSS)

Die Rotationsmethode wird angewandt, um die Ladungen der Variablen auf eine geringere Anzahl von Faktoren zu maximieren und Cross-Ladungen (Variablen mit hohen Ladungen auf mehrere Faktoren) zu minimieren. Dadurch wird die Varianz, die von jedem Faktor erklärt wird, neu verteilt. Vor der Rotation erklärt die erste Komponente 43% der Varianz, nach der Rotation 20%, wobei die kumulierte Varianz unverändert bleibt, wie Tabelle 2 verdeutlicht.

Komponente	Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	6.958	43.489	43.489	3.277	20.482	20.482
2	1.646	10.288	53.776	3.224	20.150	40.633
3	1.458	9.110	62.886	2.398	14.987	55.619
4	1.076	6.723	69.609	2.238	13.990	69.609

N=120. Extraktionsmethode: Hauptkomponentenanalyse. Kaiser-Meyer-Olkin-Kriterium

Tabelle 2 Komponenten und erklärte Varianz aus 16 LeadIM Variablen (Quelle: eigene Darstellung)

Im Fall der Hauptkomponentenanalyse (PCA), in der alle Variablen ursprünglich als Eingabe dienen, entsprechen die anfänglichen Gemeinsamkeiten in der Regel dem Wert 1, was darauf hindeutet, dass jede Variable 100% ihrer eigenen Varianz erklärt. Mithilfe von Extraktionskommunalitäten wird der Anteil der Gesamtvarianz jeder Variable bestimmt, der durch die nach der PCA extrahierten Faktoren erklärt wird (Brandt, 2020, S. 583). Dies zeigt die einzigartige Varianz in jeder Variable, die von der zugrundeliegenden Variable erfasst wird. "Boost Factor 01" weist mit 0,539 darauf hin, dass ungefähr 53,9% der Varianz in dieser Variable durch die vier extrahierten Faktoren erklärt werden. Alle Variablen zeigen hohe Gemeinsamkeiten, was darauf hindeutet, dass die ausgewählten Faktoren einen erheblichen Teil der Varianz in den Variablen erfassen. Dies weist auf eine zuverlässige und valide Faktorstruktur hin, wie Abbildung 17 zeigt.

Kommunalitäten		
	Anfänglich	Extraktion
Boost Factor 01	1.000	.539
Boost Factor 02 (10er Skala)	1.000	.745
Boost Factor 03 (10er Skala)	1.000	.670
Boost Factor 04 (10er Skala)	1.000	.671
Protector 01 (10er Skala)	1.000	.789
Protector 02 (10er Skala)	1.000	.718
Protector 03 (10er Skala)	1.000	.652
Protector 04 (10er Skala)	1.000	.791
Leadership Mindset 01 (10er Skala)	1.000	.759
Leadership Mindset 02 (10er Skala)	1.000	.624
Leadership Mindset 03 (10er Skala)	1.000	.639
Leadership Mindset 04 (10er Skala)	1.000	.596
Leadership Strategy 01 (10er Skala)	1.000	.733
Leadership Strategy 02 (10er Skala)	1.000	.793
Leadership Strtegy 03 (10er Skala)	1.000	.674
Leadership Strategy 04 (10er Skala)	1.000	.745

Extraktionsmethode: Hauptkomponentenanalyse.

Abbildung 17 Kommunalitäten (Quelle: Ausgabe SPSS)

Die Faktorladungen repräsentieren die Stärke und Richtung der Beziehung zwischen jedem Item und dem zugrundeliegenden Faktor und können Werte zwischen -1 bis 1 annehmen. Durch die rotierte Komponentenmatrix lässt sich eine klare Unterscheidung zwischen den Variablen erreichen, um darzustellen, wie stark die einzelnen Variablen auf einen bestimmten Faktor laden (Eckstein, 2016, S. 338). Die Komponentenmatrix, dargestellt in Abbildung 18, zeigt aktuell, dass BF03, P02, P03, LS01 und LS03 am stärksten positiv auf die erste Komponente laden, BF01, BF04, LM01 und LM03 am stärksten positiv auf die zweite Komponente laden, P01, LM02, LS02 und LS04 auf die dritte Komponente und BF02, BF03, P04 und LM04 am stärksten auf die vierte Komponente laden.

	Komponente				Kommunalität
	1	2	3	4	
P02	0,787	0,175	-0,036	0,257	0,718
LS03	0,703	0,406	-0,071	0,096	0,674
LS01	0,672	0,454	0,266	0,068	0,733
P03	0,664	-0,164	0,264	0,339	0,652
BF03	0,588	0,203	0,164	0,506	0,67
LM03	0,186	0,775	0,045	-0,045	0,639
BF04	0,143	0,747	0,023	0,304	0,671
LM01	0,394	0,687	0,316	0,178	0,759
BF01	0,026	0,635	0,313	0,193	0,539
P01	-0,106	0,043	0,852	0,221	0,789
LM02	0,163	0,155	0,757	0,023	0,624
LS04	0,424	0,485	0,575	-0,004	0,745
LS02	0,490	0,448	0,549	0,226	0,793
LM04	0,210	0,035	0,047	0,741	0,596
BF02	0,076	0,484	0,099	0,703	0,745
P04	0,454	0,154	0,283	0,693	0,791

N=120. Hauptkomponentenanalyse. Varimax Rotationsmethode mit Kaiser-Normalisierung.
Die Rotation ist in 8 Iterationen konvergiert.

Abbildung 18 Komponentenmatrix, nach Koeffizientengröße sortiert (Quelle: Modifizierte Darstellung Ausgabe SPSS)

Bei der Betrachtung des erweiterten Datensatzes mit 130 Daten zeigt sich, dass die Ergebnisse sehr ähnlich ausfallen. Tabelle 3 stellt die vier Komponenten dar, die momentan 69,4% der Gesamtvarianz in der Stichprobe erklären.

Komponente	Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	6.789	42.430	42.430	3.112	19.451	19.451
2	1.764	11.028	53.458	2.961	18.505	37.957
3	1.451	9.069	62.527	2.559	15.991	53.948
4	1.109	6.933	69.460	2.482	15.512	69.460

N=130. Extraktionsmethode: Hauptkomponentenanalyse. Kaiser-Guttman-Kriterium

Tabelle 3 Komponenten und erklärte Varianz aus 16 LeadIM Variablen (Quelle: Modifizierte Darstellung SPSS)

Die Reihenfolge der Variablen ändert sich im Vergleich zur Komponentenmatrix der ersten Stichprobe (N=120). Die Gruppe mit P01, LM02, LS02 und LS04 bleibt identisch

in beiden Stichproben, s. Tabelle 4 Sie lädt stark in die vierte Komponente in der erweiterten Stichprobe und in die dritte Komponente. Ebenso bleibt die Gruppe, die stark in die zweite Komponente lädt, mit BF04, LM03, BF01, BF02 und LM01 bestehen. Der einzige Unterschied besteht darin, dass BF02 tendenziell stärker mit der zweiten Komponente in der 130-Stichprobe (0,63) im Vergleich zur 120-Stichprobe korreliert. Die Gruppe von LS03, LS01 und P02 der dritten Komponente würde nicht zusammen mit P03 und BF03 in dieselbe Kategorie fallen, wie es in der ersten Stichprobe der Fall ist.

	Komponente				Kommunalität
	1	2	3	4	
P04	0.828	0.240	0.160	0.214	0.692
BF03	0.729	0.252	0.312	0.138	0.715
P03	0.680	-0.159	0.284	0.236	0.737
LM04	0.641	0.211	0.054	-0.051	0.624
BF04	0.218	0.740	0.291	-0.010	0.712
LM03	-0.030	0.736	0.331	0.083	0.660
BF01	0.214	0.649	-0.035	0.357	0.680
BF02	0.590	0.630	-0.008	0.052	0.754
LM01	0.230	0.600	0.497	0.307	0.595
LS03	0.219	0.234	0.782	-0.036	0.783
LS01	0.199	0.370	0.691	0.289	0.655
P02	0.552	0.032	0.621	0.018	0.695
P01	0.115	0.093	-0.216	0.845	0.790
LM02	0.016	0.026	0.199	0.784	0.461
LS02	0.320	0.379	0.399	0.620	0.747
LS04	0.141	0.397	0.435	0.573	0.814

N=130. Hauptkomponentenanalyse. Varimax Rotationsmethode mit Kaiser-Normalisierung.
Die Rotation ist in 8 Iterationen konvergiert.

Tabelle 4 Rotierte Komponentenmatrix, N=130 (Quelle: Modifizierte Darstellung Ausgabe SPSS)

4.3 Beantwortung der Leitfragen

In diesem Kapitel werden die Ergebnisse in Bezug auf die vorgestellten Hypothesen analysiert. Ziel dieses Kapitels ist es, die zugrundeliegenden Leitfragen dieser Arbeit durch systematische Auswertung zu beantworten.

Zunächst wird Hypothese 1 betrachtet, die eine hohe Korrelation zwischen der Selbstbewertung im Rahmen des LeadIM-Profiles und der Fremdeinschätzung im Assessment Center postuliert. Der Vergleich zwischen der subjektiven Selbstbewertung

und der Fremdeinschätzung erlaubt uns Rückschlüsse auf die Präzision der Selbsteinschätzung sowie auf die Validität und Zuverlässigkeit des Assessment Centers.

Folgend wird Hypothese 2 betrachtet. Hypothese 2 hat die Annahme, dass die Inklusion von Datenpunkten aus dem LeadIM-Profil eine signifikante Verbesserung der erklärten Varianz in Bezug auf das Kriterium "Eignung für eine Teamleiterposition" hervorbringt. Hierbei wird untersucht, inwiefern das LeadIM-Profil einen zusätzlichen explanatorischen Wert zur Vorhersage der Eignung für eine Teamleiterposition aufweist.

Abschließend wird Hypothese 3 betrachtet, die vermutet, dass die KI-gestützte Einschätzung der Wahrscheinlichkeit der Eignung für eine Führungsposition mit der qualitativen Bewertung im Assessment Center übereinstimmt. Die Übereinstimmung zwischen KI-Prognosen und menschlicher Bewertung liefert wichtige Informationen über die Validität und Zuverlässigkeit von KI-gestützten Einschätzungen im Kontext von Führungskompetenzen.

Insgesamt hat dieses Kapitel das Ziel, einen umfassenden Einblick in die Beantwortung der formulierten Hypothesen und damit verbundenen Leitfragen zu bieten.

4.3.1 Ergebnisse zu Hypothese 1

Im Mittelpunkt der ersten Forschungshypothese steht die Frage, ob das LeadIM Profil (Selbstbeschreibung) eine signifikante Korrelation mit der Fremdeinschätzung im Assessment Center (AC) aufweist. Eine hohe und statistisch signifikante Korrelation zwischen diesen beiden Maßen würde darauf hindeuten, dass das LeadIM Profil und die AC-Einschätzung ähnliche Aspekte der Eignung für eine Führungsrolle erfassen.

Im ersten Schritt wird eine entsprechende Korrelationsanalyse durchgeführt, um den Grad der Assoziation zwischen den LeadIM-Variablen und der AC-Einschätzung zu ermitteln. Im Falle einer starken Korrelation könnten die beiden Maße als nahezu austauschbar betrachtet werden. Es wird erwartet, dass eine Ersetzung der LeadIM-Variablen durch AC-Variablen in einem Regressionsmodell, das die Eignung für Führungsrollen vorhersagt, ähnlich gute Schätzungen liefert.

Es wird eine Korrelationsmatrix erstellt, um alle möglichen Korrelationen zwischen den LeadIM-Variablen und den AC-Einschätzungen darzustellen. Tabelle 5 zeigt das Ergebnis. Die Signifikanzniveaus der Korrelationen sind zur leichteren Identifikation mit Sternen gekennzeichnet.

Bei genauerer Betrachtung der spezifischen Beziehungen, wie z.B. zwischen "Aktivierende Lernmotivation" und BF01, zeigt sich, dass keine der untersuchten Beziehungen statistisch signifikant sind. Daher sollte das Ergebnis mit Vorsicht interpretiert werden.

	Aktivierende				Selbstbewusste				Strukturierende				Empathische			
	Lern	Init	Gest	InsN	Sver	Res	Sref	ÜSv	StP	Will	Sver	Amb	Emp	VerI	EntF	MotA
BF01	0,23	0,11	0,13	-0,01	0,20	-0,36	0,29	-0,18	0,32	0,16	0,19	0,03	0,05	-0,21	-0,31	-0,15
BF02	.660*	0,06	0,40	0,00	0,34	-0,19	-0,10	-0,03	0,38	0,25	0,53	0,33	-0,16	-0,37	0,02	-0,37
BF03	0,49	-0,26	0,22	-0,21	0,22	-0,49	-0,09	-0,28	0,10	-0,01	0,24	0,01	-0,19	-0,35	-0,35	-0,58
BF04	0,60	0,18	0,50	0,31	.695*	0,39	0,26	0,34	0,29	0,37	.712*	0,60	-0,44	-0,51	0,00	-0,38
P01	-0,51	0,19	-0,09	0,03	-0,41	-0,13	-0,07	-0,41	-0,52	-0,37	-0,49	-0,56	0,01	0,08	-0,36	0,19
P02	.673*	0,44	.703*	0,37	0,26	0,02	-0,13	0,19	0,33	0,35	0,39	0,36	-0,17	-0,16	0,21	0,00
P03	0,20	-0,19	0,16	0,00	-0,04	-0,08	0,47	-0,12	-0,07	-0,08	-0,05	0,08	0,18	0,26	-0,18	-0,12
P04	0,55	-0,08	0,18	-0,14	0,35	-0,22	0,34	0,00	0,42	0,34	0,54	0,34	0,30	0,02	0,12	-0,19
LM01	0,58	0,27	0,57	0,23	.843**	0,28	0,24	0,32	0,16	0,54	.713*	0,43	-0,39	-0,32	0,02	-0,24
LM02	-0,04	.692*	0,44	0,45	0,09	0,44	-0,10	0,20	-0,08	0,27	0,00	0,05	-0,25	0,11	0,23	0,52
LM03	.738*	0,36	.666*	0,39	.842**	0,44	0,44	0,46	0,40	0,61	.854**	.714*	-0,29	-0,32	0,16	-0,19
LM04	0,54	-0,19	0,11	-0,07	0,61	-0,16	-0,03	0,19	0,63	0,32	0,52	0,44	-0,46	.660*	-0,17	-0,58
LS01	0,27	0,25	0,44	0,25	0,55	0,45	-0,21	0,16	-0,13	0,16	0,45	0,25	.804**	-0,63	-0,12	-0,35
LS02	0,44	0,35	0,42	0,24	0,56	0,19	.705*	0,25	0,37	0,58	0,53	0,44	0,18	0,17	0,13	0,17
LS03	-0,08	0,25	0,14	0,27	-0,22	0,32	-0,16	0,06	-0,29	-0,20	0,02	0,00	0,02	-0,06	0,18	0,10
LS04	.648*	0,39	0,58	0,34	0,40	0,49	0,41	0,36	0,37	0,51	.707*	.712*	0,18	0,16	0,55	0,18

** Korrelationskoeffizienten sind auf dem 0,01-Niveau (zweiseitig) signifikant. N=10

* Korrelationskoeffizienten sind auf dem 0,05-Niveau (zweiseitig) signifikant.

BF: Boost Factors, PO: Protectors, LM: Leadership Mindest, LS: Leadership Strategies

Tabelle 5 Spearman Korrelation (Quelle: Modifizierte Darstellung Ausgabe SPSS)

Bei den Korrelationen zwischen den AC-Einschätzungen und den Selbsteinschätzungen zum Leadership Mindset und den "Boost Factors" (BF) zeigen sich insgesamt höhere Werte als bei den Leadership Strategies (LS). Es zeigen sich hohe Korrelationen zwischen LM01 und "Gestalten und Vorangehen" (0,57), LM04 und "Strukturieren und Planen" (0,63), sowie LS04 und "Entwickeln und Fördern" (0,55).

Die "Boost Factors" weisen moderate Korrelationen mit "Ambition", "Selbstvertrauen" und "Lernmotivation" auf, während die "Leadership Strategies" nur geringe Korrelationen mit "Überzeugen und Sinn Vermitteln" und "Motivieren und Aktivieren" aufweisen. Beachtenswert ist jedoch, dass LS04 eine moderate Korrelation mit "Entwickeln und Fördern" zeigt.

Die Korrelation für "Empathie" (P01) und "Resilienz" (P02) ist fast Null. Dies könnte darauf hindeuten, dass diese Faktoren im Rahmen des Assessment Centers

möglicherweise schwieriger zu erfassen sind, oder dass Selbst- und Fremdeinschätzung in diesen Bereichen nicht übereinstimmen.

Die Nichtexistenz von signifikanten Korrelationen in diesen Bereichen könnte bei kleineren Stichproben auch auf eine mangelnde statistische Power zurückzuführen sein, um tatsächlich bestehende Beziehungen zwischen den Variablen zu erkennen. Darüber hinaus zeigt LS01 eine negative signifikante Verbindung zur "Empathie", was weiter untersucht werden sollte.

Insgesamt deuten die Ergebnisse darauf hin, dass das LeadIM Profil und die AC-Einschätzung in einigen Bereichen ähnliche Aspekte der Eignung für eine Führungsrolle erfassen. Angesichts der kleinen Stichprobengröße und der damit verbundenen Einschränkungen sollten die Ergebnisse jedoch mit Vorsicht interpretiert werden.

Ergänzend wird die Berechnung der Spearman-Rangkorrelationen zwischen den Selbsteinschätzungen (dargestellt durch das LeadIM Profil) und den Schätzergebnissen zu den vier Führungskomponenten (1: Selbstreflektierte Führung, 2: Initiative und gestalterische Führung, 3: Empathische und integrierende Führung, 4: Rationale, willensstarke, planende und strukturierende Führung) durchgeführt.

Dabei werden die Werte der LeadIM-Variablen standardisiert, gemessen auf einer 5-Punkte-Likert-Skala, um die Kompatibilität mit den Assessment Center-Maßen zu gewährleisten, welche ebenfalls standardisiert und in Form von z-Werten vorliegen.

Es ist ersichtlich, dass zumindest zwei statistisch signifikante Korrelationen aus den hypothetischen Beziehungen zwischen den Variablen resultieren. Insbesondere zeigen Leadership Mindset 1 und 3 eine starke Korrelation mit der Komponente "Initiative und gestalterische Führung". Beispielsweise wies "Gestalten und vorangehen" (LM01) vorher keine signifikante Korrelation mit den Elementen in der "Initiative und gestalterische Führung"-Komponente auf. Interessanterweise korreliert Leadership Mindset 3 (LM03) hoch und signifikant mit drei der Führungskomponenten (s. Tabelle 6). Bezüglich der "Empathischen und integrierenden Führung" ist es weiterhin schwierig, signifikante Korrelationen in der Stichprobe zu identifizieren, wobei viele davon negativ sind. Die Korrelation mit "Motivieren und aktivieren" (LS02) nimmt jedoch an Stärke zu.

Zusammenfassend lässt sich sagen, dass die Ergebnisse gemischte Unterstützung für die Hypothese bieten, dass das LeadIM-Profil und die AC-Einschätzung ähnliche

Aspekte der Führungsqualifikation erfassen. Während einige der Korrelationen die Hypothese stützen, gibt es auch Bereiche, in denen die Hypothese nicht unterstützt wird. Angesichts der beschränkten Stichprobengröße und der damit verbundenen statistischen Einschränkungen sollten die Ergebnisse mit Vorsicht interpretiert werden.

	Aktivierende	Selbstbewusste	Strukturierende	Empathische
BF01	0,17	0,03	0,25	0,05
BF02	,276	,029	,398	-,166
BF03	0,11	-0,14	0,13	-0,20
BF04	0,47	0,56	0,45	-0,26
P01	-0,13	-0,35	-0,43	-0,07
P02	0,56	0,14	0,44	-0,13
P03	0,07	0,14	0,04	0,23
P04	0,19	0,16	0,46	0,35
LM01	.638*	0,58	0,49	-0,06
LM02	0,43	0,20	0,14	-0,07
LM03	.687*	.728*	.660*	-0,01
LM04	0,16	0,27	0,43	-0,39
LS01	0,37	0,32	0,14	-0,59
LS02	0,54	0,59	0,58	0,45
LS04	0,49	0,51	0,61	0,33

** Signifikant auf dem 0,01-Niveau, * Signifikant auf dem 0,05-Niveau. N=10

BF: Boost Factors, PO: Protectors, LM: Leadership Mindest, LS: Leadership Strategies.

Tabelle 6 Spearman (Quelle: Modifizierte Darstellung Ausgabe SPSS)

4.3.2 Ergebnisse zu Hypothese 2

Zur Beantwortung der zweiten Forschungsfrage, mit der Annahme, dass durch die Hinzunahme der Datenpunkte aus dem LeadIM Profil sich die Varianzaufklärung im Kriterium „Eignung als Teamleiterrolle“ signifikant erhöht, wird die Auswirkung des Einsatzes der AC-Daten anstelle der LeadIM-Faktorwerte verglichen. Angesichts der begrenzten Stichprobe, die sowohl AC-Daten als auch über LeadIM-Daten verfügen, werden für die 10 Personen die LeadIM Faktorwerte durch ihre AC-Ergebnisse ersetzt. Die verbleibenden 120 Personen verwenden weiterhin die Daten aus dem LeadIM-PCA abgeleiteten Faktorwerte.

Die AC-Schätzungen sind auf z-Werte standardisiert, mit einem Mittelwert von 3 und einer Standardabweichung von 1,5.

Tabelle 7 zeigt die Assessment Center Schätzungen der vier Komponenten als z-Werte.

Leadership-Komponente	121	122	123	124	125	126	127	128	129	130
Aktivierende	0.08	0.00	-0.42	0.00	0.17	0.33	-0.42	0.08	0.58	0.92
Selbstbewusste	0.08	0.08	-0.08	-0.25	0.42	0.50	0.17	0.50	1.00	0.83
Strukturierende	-0.25	0.17	-0.50	-0.17	0.58	0.25	0.25	0.08	0.92	0.67
Empathische	-0.08	-0.33	0.08	-0.24	-0.17	-0.25	0.58	-0.33	0.42	0.92

Tabelle 7 AC-Schätzungen der vier Komponenten (als z-Werte) (Quelle: eigene Darstellung)

Bei der Überprüfung der Modellgüte zeigt sich, dass die Signifikanz < 0.05 ist, d.h. das Modell ist besser als das 0 Modell, das bedeutet, dass das Modell die abhängige Variable signifikant erklären kann, s. Abbildung 19. Bei der Kombination mit den 10 Assessment Center Teilnehmerergebnissen, können 59% der Varianz mit dem Modell erklärt werden, wohingegen das Modell mit ausschließlicher Betrachtung der LeadIM Profile 61% der Varianz mit dem Modell erklärt werden können. Es deuten die Ergebnisse darauf hin, dass das Pseudo R-Square etwas schlechter wird, wenn einige der Beobachtungen durch die AC-Werte ersetzt werden.

120 mit LeadIM, 10 mit AC

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	49.953	4	<.001
	Block	49.953	4	<.001
	Model	49.953	4	<.001

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	50.887 ^a	.319	.591

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

130 alle mit LeadIM

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	51.855	4	<.001
	Block	51.855	4	<.001
	Model	51.855	4	<.001

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	48.985 ^a	.329	.610

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Abbildung 19 Modell Zusammenfassung AC versus LeadIM (Quelle: Ausgabe SPSS)

Die Regressionskoeffizienten haben ausschließlich positive Werte, was bedeutet, dass alle Komponenten einen positiven Beitrag zur Wahrscheinlichkeit für eine Führungsposition geeignet zu sein beitragen, s. Tabelle 8 und Tabelle 9. Die Regressionskoeffizienten der ersten 3 Führungskomponenten sind bei beiden signifikant von 0 verschieden (Wald Test). Alle „Odd Ratios“ sind größer 1 und das 95% Konfidenzintervall liegt außerhalb von 1, was einen positiven Beitrag zur Einschätzung als Führungskraft geeignet bedeutet (Held, 2010). In beiden Fällen hat die Führungskomponente Empathie den größten Einfluss. Dennoch ist der Koeffizient mit einem hohen Maß an Unsicherheit behaftet, wie am breiten Konfidenzintervall zu erkennen ist.

Komponente	Komponenten bezeichnung	Koeffizient β	Standardfehler	Wald	Sig	Exp(β)	95% Konfidenzintervall für Exp(β)	
							Unterer Wert	Oberer Wert
1	Strukturierende Führung	1,58	0,569	7,688	0,006	4,843	1,588	14,772
2	Aktivierende Führung	1,419	0,531	7,156	0,007	4,135	1,462	11,698
3	Selbstbewusste Führung	1,137	0,462	6,040	0,014	3,116	1,259	7,714
4	Empathische Führung	1,830	0,523	12,236	0,000	6,233	2,236	17,376
	Konstante	-3,940	0,747	27,801	0,000	0,019		

Abhängige Variable: geeignet (1), nicht geeignet (0). Zehn Beobachtungen enthalten Informationen aus dem AC anstelle des LeadIM.

Tabelle 8 Regressionskoeffizienten, logistische Regression, LeadIM, AC (Quelle: modifizierte Darstellung Ausgabe SPSS)

Komponente	Komponentenbezeichnung	Koeffizient β	Standardfehler	Wald	Sig.	Exp(β)	95% Konfidenzintervall für Exp(β)	
							Unterer Wert	Oberer Wert
1	Strukturierende Führung	1,59	0,607	6,846	0,009	4,889	1,489	16,052
2	Aktivierende Führung	1,049	0,456	5,284	0,022	2,856	1,167	6,987
3	Selbstbewusste Führung	0,719	0,380	3,578	0,059	2,052	0,974	4,323
4	Empathische Führung	2,824	0,747	14,300	0,000	16,842	3,897	72,782
	Konstante	-4,401	0,942	21,813	0,000	0,012		

Abhängige Variable: geeignet (1), nicht geeignet (0). N=130

Tabelle 9 Regressionskoeffizienten, logistische Regression, LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)

Die Prognosegüte des Modells im Vergleich mit den LeadIM beziehungsweise AC ist unverändert, s. Tabelle 10 und Tabelle 11, das bedeutet, dass die Ergebnisse identisch bezüglich der Eignung für eine Teamleiterrolle sind. Hierbei ist schwer zu beurteilen, ob sie gleichbleiben, weil die Informationen aus dem AC identisch sind, oder ob der Effekt nicht aufzuzeigen ist, weil nur 10 der Beobachtungen ersetzt werden.

		Vorhergesagt		
		Nicht geeignet	Geeignet	% Korrekt
Beobachtet	Nicht geeignet (n = 113)	111	2	98,2%
	Geeignet (n = 17)	4	13	76,5%
	Gesamtprozentsatz			95,4%

Logistische Regression,
N=130; Der Schwellenwert
beträgt 0,5.

Zehn Beobachtungen enthalten Informationen aus dem AC anstelle des LeadIM.

Tabelle 10 Klassifikationstabelle der Eignung für eine Teamleiterrolle AC + LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)

		Vorhergesagt		
		Nicht geeignet	Geeignet	% Korrekt
Beobachtet	Nicht geeignet (n = 113)	111	2	98,2%
	Geeignet (n = 17)	4	13	76,5%
	Gesamtprozentsatz			95,4%

Logistische Regression,
N=130; Der Schwellenwert
beträgt 0,5.

Zehn Beobachtungen enthalten Informationen aus dem AC anstelle des LeadIM.

Tabelle 11 Klassifikationstabelle der Eignung für eine Teamleiterrolle LeadIM (Quelle: modifizierte Darstellung Ausgabe SPSS)

4.3.3 Ergebnisse zu Hypothese 3

Bei Hypothese 3 wird angenommen, dass die von künstlicher Intelligenz unterstützte Prognose bezüglich der Eignung für eine Führungsposition mit der qualitativen Beurteilung im Assessment Center konsistent ist. Mit Hilfe einer binären logistischen Regression wird untersucht, ob die von Menschen vorgenommenen Einschätzungen der Eignung für eine Führungsrolle durch ein algorithmisches Modell bestätigt werden kann (Baltes-Görtz, 2012). Als Eingabevariablen in der logistischen Regression werden die Factor Scores der PCA-Analyse verwendet. Diese fassen die Informationen der 16 Items in den identifizierten 4 Faktoren zusammen, was die Regressionsanalyse und Interpretation der Ergebnisse vereinfachen. Es wird getestet, wie gut das Modell (KI-Methode) die Personen anhand der ausschließlich aus dem LeadIM-Fragebogen extrahierten Informationen in „geeignet“ oder „nicht geeignet“ klassifiziert. Dafür wurden die 10 Assessmentteilnehmer allein mit den AC-Daten klassifiziert, wohingegen die Klassifizierung für die 120 Daten anhand der LeadIM-Profile vorgenommen wurde. Das bedeutet, das geprüft wird, ob die Einschätzung der Beobachter des Assessment Centers bezüglich der Teilnehmer als geeignet oder nicht geeignet, durch das Regressionsmodell nur auf Grundlage der LeadIM Profildaten bestätigt werden kann. Der Chi-Quadrat Test wird verwendet, um anzuzeigen, ob die unabhängigen Variablen insgesamt mit der abhängigen Variable zusammenhängen. In Statistischer Hinsicht handelt es sich um einen gemeinsamen Signifikanztest der erklärenden Variable. Er vergleicht das Modell mit den vier Prädiktoren mit einem Modell ohne Prädiktoren (nur einer Konstante). Hier ist der Chi-Quadrat-Wert von 51,855 (s. Abbildung 20) signifikant und deutet darauf hin, dass die Koeffizienten des Modells kollektiv mit der Ergebnisvariable zusammenhängen. Genauer gesagt, ist mindestens eine der Prädiktorvariablen im Modell signifikant.

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	51.855	4	<.001
	Block	51.855	4	<.001
	Modell	51.855	4	<.001

Abbildung 20 Omnibus-Test (Quelle: Ausgabe SPSS)

Die Tabelle „Modellzusammenfassung“ (s. Abbildung 21) in SPSS liefert weitere Maße zur Model Fit. Die angegebenen R-Quadrat-Werte werden als Pseudo-R-Quadrat bezeichnet. Sie sind ähnlich wie das R-Quadrat in linearen Regressionsmodellen, aber nicht vollständig, daher ist bei der Interpretation Vorsicht geboten. Das R-Quadrat erfasst, wie gut die Varianz in der abhängigen Variablen durch die Variablen im Modell erklärt wird. Ein Wert von eins würde eine perfekte Anpassung bedeuten (Backhaus et al., 2018, S. 77). Bei Pseudo R-Quadrat gibt es jedoch viele verschiedene Formeln, und es ist nicht klar, ob es wie ein normales R-Quadrat interpretieren sollte. Das Problem mit Cox & Snell's ist, dass es niemals den Wert eins erreichen kann und tatsächlich sehr weit von eins entfernt sein kann. Stattdessen ist das Nagelkerke's eine skalierte Version, bei der der maximale Wert eins beträgt, der als bestmögliches Modell interpretiert werden kann (ebd).

Modellzusammenfassung

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	48.985 ^a	.329	.610

a. Schätzung beendet bei Iteration Nummer 8, weil die Parameterschätzer sich um weniger als .001 änderten.

Abbildung 21 Modellzusammenfassung (Quelle: Ausgabe SPSS)

Bei der Interpretation der Regressionskoeffizienten ergibt sich folgendes Bild, s. auch Abbildung 22. Der Effekt von Faktor score 1 auf die Wahrscheinlichkeit für eine Teamleiterrolle geeignet zu sein, ist auf dem 0.01 Niveau signifikant. Die Koeffizientschätzungen für die Variablen kann nur insofern ausgewertet werden, dass daran erkennbar ist, ob die Variablen positiv oder negativ mit der Wahrscheinlichkeit der Eignung zusammenhängen. Die Spalte Exp(B) gibt die Veränderung der Wahrscheinlichkeit an, dass ein Ereignis eintritt, oder das „Odds Ratio“, wenn $\exp(b)$ größer als 1 ist. Dann kann gesagt werden, dass die Odds $\exp(b)$ mal höher sind. Für den Faktor 1 ist eine Erhöhung des Factor Score um eine Einheit mit einer etwa 4,9-

fachen Erhöhung der Chancen verbunden, für die Rolle geeignet zu sein, wobei die anderen Faktoren konstant bleiben.

		Variablen in der Gleichung						95% Konfidenzintervall für EXP (B)	
		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	Unterer Wert	Oberer Wert
Schritt 1 ^a	REGR factor score 1 for analysis 1	1.587	.607	6.846	1	.009	4.889	1.489	16.052
	REGR factor score 2 for analysis 1	1.049	.456	5.284	1	.022	2.856	1.167	6.987
	REGR factor score 3 for analysis 1	.719	.380	3.578	1	.059	2.052	.974	4.323
	REGR factor score 4 for analysis 1	2.824	.747	14.300	1	<.001	16.842	3.897	72.782
	Konstante	-4.401	.942	21.813	1	<.001	.012		

a. In Schritt 1 eingegebene Variablen: REGR factor score 1 for analysis 1, REGR factor score 2 for analysis 1, REGR factor score 3 for analysis 1, REGR factor score 4 for analysis 1.

Abbildung 22 Regressionskoeffizienten (Quelle: Ausgabe SPSS)

Die relativen Größen der Koeffizienten helfen die Auswirkung oder Bedeutung jedes Faktors bei der Beeinflussung der Wahrscheinlichkeit, für eine Teamleiterrolle geeignet zu sein, zu verstehen. Alle Komponenten wirken sich positiv auf die Wahrscheinlichkeit aus, für eine Teamleiterstelle geeignet zu sein. Bei der Betrachtung der ersten drei Komponenten scheint die 'Strukturierende' Komponente wichtiger zu sein, um die Chancen zu erhöhen, gefolgt von der 'Aktivierende' und der 'Selbstbewusste'.

Die 'Empathiekomponente' hat einen Koeffizienten, der einen sehr großen Einfluss auf die Wahrscheinlichkeit anzeigt. Gemäß den Ergebnissen würde diese Komponente daher als ein sehr starker Prädiktor für die Wahrscheinlichkeit angesehen werden. Gleichzeitig weist sie jedoch ein sehr breites Konfidenzintervall auf (zwischen 3,9 und 72,8). Dies deutet darauf hin, dass mit der Schätzung des Koeffizienten eine erhebliche Unsicherheit verbunden ist, denn „je größer das Konfidenzintervall ist, desto unsicherer ist die Schätzung“ (Backhaus et al., 2018, S, 89).

Zum Vergleich wurde die gleiche Regression durchgeführt, bei der jedoch die vierte Komponente ausgelassen wurde. Das Modell hat alle "nicht geeignet" richtig vorhergesagt, konnte jedoch keine einzige "geeignet" richtig Vorhersagen.

Insgesamt verliert das Modell also deutlich an Performance, wenn diese Komponente nicht berücksichtigt wird, s. Tabelle 12.

Komponente	Komponentenbezeichnung	Koeffizient β	Standard- fehler	Wald	Sig	Exp(β)	95% Konfidenzintervall für Exp(β)	
							Unterer Wert	Oberer Wert
1	Strukturierende Führung	1.59	0.607	6.846	0.009	4.889	1.489	16.052
2	Aktivierende Führung	1.049	0.456	5.284	0.022	2.856	1.167	6.987
3	Selbstbewusste Führung	0.719	0.380	3.578	0.059	2.052	0.974	4.323
4	Empathische Führung	2.824	0.747	14.300	0.000	16.842	3.897	72.782
	Konstante	-4.401	0.942	21.813	0.000	0.012		

Abhängige Variable: geeignet (1), nicht geeignet (0). N=130

Tabelle 12 Regressionskoeffizienten, logistische Regression, LeadIM

Die Klassifikationstabelle, s. Tabelle 13, gibt Auskunft darüber zu beurteilen, wie gut der Algorithmus die beobachtete Klassifizierung vorhersagt und damit auch die Forschungsfrage: Ob die qualitativen Bewertungen im Assessment Center für eine Teamleiterrolle durch das Regressionsmodell nur auf Grundlage der LeadIM-Profilaten bestätigt werden.

		Vorhergesagt		% Korrekt
		Nicht geeignet	<u>Geeignet</u>	
Beobachtet	Nicht geeignet (n = 113)	111	2	98.2%
	Geeignet (n = 17)	4	13	76.5%
	Gesamtprozentsatz			95.4%

Logistische Regression, N=130; Der Schwellenwert beträgt 0,5.

Tabelle 13 Klassifikationstabelle Eignung für eine Teamleiterrolle

Standardmäßig liegt der Klassifizierungsgrenzwert bei 0,5. Wenn die vorhergesagte Wahrscheinlichkeit $< 0,5$ ist, wird der Fall der Kategorie 0 zugeordnet, wenn die Wahrscheinlichkeit $\geq 0,5$ ist, wird der Fall der Kategorie 1 zugeordnet. Von der Gesamtzahl der Personen, die nach Ansicht der Beobachter nicht für eine Führungsrolle geeignet sind, wurden 111 vom Algorithmus richtig eingestuft und 2 wurden der falschen Kategorie zugeordnet. Das bedeutet, dass 98% der Personen in der Kategorie "nicht geeignet" auch vom Modell als "nicht geeignet" eingestuft wurden. Von den insgesamt siebzehn Personen, die als geeignet für eine Leitungsrolle eingestuft wurden, hat das Modell 13 davon richtig klassifiziert. Dies ergibt eine Gesamtübereinstimmung von 76,5%. Das Modell ist bei der

Vorhersage von 0's effektiver als bei 1 's. Das ist keine Überraschung, denn es ist schwieriger, Fälle richtig vorherzusagen, für die es in der Stichprobe viel weniger "Beispiele" gibt. Insgesamt werden 95,4 korrekt klassifiziert.

5. Diskussion

Diese Master-Thesis zielt darauf ab, die Varianzaufklärung verschiedener methodischer Ansätze - das Assessment-Center, der Selbstauskunftsfragebogen und die künstliche Intelligenz (KI) gestützte Methode - in Bezug auf die Eignungseinschätzung für Führungspositionen zu vergleichen und zu bewerten. Im nachfolgenden Diskussionsteil wird dieses Thema in seinen Facetten beleuchtet, die Befunde in einen umfassenderen Kontext gestellt und Implikationen sowie zukünftige Forschungsrichtungen abgeleitet.

5.1 Interpretation Ergebnisse

Die Diskussion der vorgelegten Befunde hat sowohl theoretische als auch praktische Implikationen für die Managementdiagnostik. Im Folgenden wird auf beide Aspekte eingegangen.

Die Ergebnisse liefern Anhaltspunkte für die Annahme, dass das Verhältnis zwischen Selbsteinschätzung und Fremdeinschätzung in der Managementdiagnostik komplex ist und von verschiedenen Faktoren abhängen kann. Einerseits scheinen einige Aspekte des Leadership Mindsets und die sogenannten "Boost Factors" gut mit den Fremdeinschätzungen im Assessment Center zu korrelieren. Dies könnte darauf hindeuten, dass Führungskräfte in der Lage sind, ihre eigenen Fähigkeiten in diesen Bereichen einzuschätzen, und dass diese Einschätzungen von anderen wahrgenommen und anerkannt werden.

Andererseits weisen die Leadership Strategies weniger starke Korrelationen auf, was darauf hindeuten könnte, dass Führungskräfte Schwierigkeiten haben, ihre eigenen Strategien genau zu bewerten oder dass diese Strategien von anderen weniger gut wahrgenommen werden. Dies könnte darauf zurückzuführen sein, dass diese Strategien komplexer oder weniger offensichtlich sind oder dass sie stärker von Kontextfaktoren abhängen.

Die fehlende signifikante Korrelation bei "Empathie" und "Resilienz" könnte darauf hindeuten, dass diese Aspekte schwerer zu bewerten sind, möglicherweise weil sie weniger direkt beobachtbar sind oder weil sie stärker von subjektiven Interpretationen abhängen. Dies legt nahe, dass Assessment Center diese Aspekte möglicherweise nicht

optimal erfassen können oder dass zusätzliche Informationsquellen benötigt werden, um eine genaue Einschätzung zu ermöglichen.

Die negative Korrelation zwischen "Empathie" und der Leadership Strategy "LS01" wirft weitere Fragen auf und erfordert weitere Untersuchungen. Es könnte sein, dass hohe Werte in "Empathie" bestimmte Strategien weniger wahrscheinlich machen oder dass bestimmte Strategien die Wahrnehmung von Empathie verringern.

Praktisch gesehen könnten diese Erkenntnisse dazu beitragen, die Managementdiagnostik zu verbessern, indem sie dazu anregen zusätzliche Methoden zur Bewertung bestimmter Aspekte einzuführen. Sie könnten auch dazu beitragen, das Verständnis und die Entwicklung von Führungskräften zu fördern, indem sie das Bewusstsein für mögliche Diskrepanzen zwischen Selbst- und Fremdeinschätzung schärfen.

Die Diskrepanz zwischen den Selbsteinschätzungen und den Fremdeinschätzungen in den Bereichen "Empathie" und "Resilienz" könnte auf mehrere Faktoren zurückzuführen sein. Erstens könnten die Kriterien, die im Assessment Center verwendet werden, um diese Aspekte zu bewerten, nicht ausreichend sein. Zweitens könnten die Kandidaten diese Fähigkeiten nicht effektiv demonstrieren, was zu einer niedrigeren Bewertung führt. Schließlich könnte die fehlende Signifikanz in diesen Bereichen aufgrund der kleinen Stichprobengröße auch auf eine mangelnde statistische Power zurückzuführen sein.

Es ist wichtig, bei der Interpretation dieser Ergebnisse zu berücksichtigen, dass "Empathie" und "Resilienz" entscheidende Führungsmerkmale sind. Laut einer Quelle von Forbes, sind Empathie und Resilienz wichtige Merkmale effektiver Führung, insbesondere in Zeiten der Krise. Sie sind ansteckend und können Stabilität, Optimismus und Positivität fördern. Darüber hinaus sind sie eng mit Vertrauen und Integrität verbunden, die als grundlegende Merkmale starker Beziehungen betrachtet werden (Kennedy, 2020).

In Bezug auf die Verwendung von Selbsteinschätzungen und 360-Grad-Bewertungen in der Managementdiagnostik weisen einige Quellen darauf hin, dass beide Methoden ihre Vorzüge haben. Selbsteinschätzungen sind nützlich für Personen, die mehr über Führung lernen möchten und können dazu beitragen, das Bewusstsein für individuelle Führungsstärken und Verbesserungsbereiche zu fördern (Sears, 2019). Auf der anderen

Seite können 360-Grad-Bewertungen ein vollständiges Bild davon liefern, wie ein Führer sich selbst im Vergleich zu dem Feedback von anderen sieht, was bei der Verbesserung von Beziehungen am Arbeitsplatz hilfreich sein kann (ebd).

Die vorliegende Untersuchung bewertete die Rolle zusätzlicher Datenpunkte aus dem LeadIM-Profil, um festzustellen, ob sie die Erklärungskraft für die "Eignung für eine Teamleiterrolle" signifikant erhöht. Die Studie analysierte die Auswirkungen des Austauschs der LeadIM-Faktorwerte durch Assessment Center (AC) Daten.

Ergebnisse zeigten eine signifikante Anpassung des Modells mit AC-Daten ($p < 0.05$), was auf die Eignung beider Modelle hinweist. Interessanterweise führte die Kombination von AC- und LeadIM-Daten zu einer Varianzerklärung von 59%, während das auf LeadIM-Daten basierende Modell 61% der Varianz erklärte. Das legt nahe, dass die Integration von AC-Daten nicht notwendigerweise zu einer signifikanten Verbesserung der Varianzerklärung beiträgt.

Die Regressionskoeffizientenanalyse zeigte, dass alle Führungsdimensionen einen positiven Einfluss auf die Wahrscheinlichkeit haben, als geeignet für eine Führungsrolle bewertet zu werden. Interessanterweise zeigte die "empathische" Führungsdimension den stärksten Einfluss, obwohl dieser Effekt mit erheblicher Unsicherheit behaftet ist, wie das breite Konfidenzintervall zeigt. Forschung unterstreicht die Wichtigkeit von Empathie in der Führung, sie trägt zu positiven Ergebnissen bei, einschließlich Innovation, Engagement, Mitarbeiterbindung, Inklusivität und Work-Life-Balance (Brower, 2021).

Trotz der ähnlichen Vorhersagegenauigkeit der beiden Modelle für die Eignung als Teamleiter, führten die zusätzlichen AC-Daten nicht zu verbesserten Vorhersagen. Ob dies auf eine Übereinstimmung der Informationen aus dem AC und dem LeadIM zurückzuführen ist oder auf den Umstand, dass nur ein kleiner Teil der Beobachtungen (10 von 130) durch AC-Daten ersetzt wurde, bleibt unklar.

Zusammenfassend lässt sich sagen, dass diese Ergebnisse die Hypothese nicht eindeutig unterstützen. Während die Integration von AC-Daten in die LeadIM-Daten nicht zu einer signifikanten Verbesserung der Modellanpassung oder Vorhersagegenauigkeit führte, könnte der geringe Anteil der durch AC-Daten ersetzten Beobachtungen einen Einfluss auf dieses Ergebnis gehabt haben. Weitere Forschungen mit einem größeren Anteil von AC-Daten könnten dazu beitragen, diese Frage zu klären.

Die Ergebnisse der Analyse zeigen eine teilweise Bestätigung der dritten Hypothese, dass die künstliche Intelligenzgestützte Einschätzung der Wahrscheinlichkeit der Eignung für eine Führungsposition mit der qualitativen Bewertung im Assessment Center übereinstimmt. Diese Feststellung hat insofern Auswirkungen auf die Managementdiagnostik, insbesondere in Bezug auf die Verwendung von KI-Technologien in der Personalbewertung und -auswahl.

Erstens zeigt die Fähigkeit der KI, die Eignung von Individuen für eine Führungsposition zu beurteilen, dass KI-Technologien zu einem nützlichen Instrument für die Bewertung des Führungspotenzials werden könnten. Es wurde festgestellt, dass generative KI-Modelle in Unternehmen vielfältige Anwendungsmöglichkeiten haben, von der Beantwortung von Kundenfragen bis zur Unterstützung von Ingenieuren bei der Entwicklung aktualisierter Produktversionen. Dies deutet darauf hin, dass KI-Modelle auch in der Lage sein könnten, komplexe menschliche Fähigkeiten wie Führungsqualitäten zu bewerten.

Zweitens ist jedoch auch Vorsicht geboten, da KI-Modelle nicht ohne Risiken sind. Sie können anfällig für "Halluzinationen" sein, d.h. sie können Fragen mit plausiblen, aber unwahren Behauptungen beantworten. Darüber hinaus sind die zugrunde liegenden Gründe oder Quellen für eine Antwort nicht immer gegeben. Dies bedeutet, dass Unternehmen bei der Integration von generativer KI ohne menschliche Aufsicht in Anwendungen, in denen Fehler Schaden anrichten können oder wo Erklärbarkeit benötigt wird, vorsichtig sein sollten (Chui et al., 2023).

Allgemein ist zu beachten, dass KI-Modelle, insbesondere jene, die auf maschinellem Lernen basieren, auf den Daten trainiert werden, die ihnen zur Verfügung gestellt werden. Wenn es in der Stichprobe weniger Beispiele für "geeignete" Kandidaten gibt, könnte dies die Fähigkeit des Modells beeinträchtigen, solche Kandidaten korrekt zu identifizieren.

In Anbetracht dieser Faktoren sollte die Verwendung von KI in der Managementdiagnostik als Ergänzung zu bestehenden Bewertungsmethoden und nicht als Ersatz betrachtet werden. Es ist auch wichtig, dass Führungskräfte, die KI in ihren Bewertungsprozessen einsetzen, eine gewisse Vertrautheit mit den Technologien aufweisen und klare Geschäftsziele für deren Einsatz festlegen.

5.2 Qualität des eigenen Vorgehens (Objektivität, Reliabilität, Validität)

Im Kontext dieser Studie werden die Prinzipien der Objektivität, Reliabilität und Validität von diagnostischen Verfahren berücksichtigt, um die Qualität der gewonnenen Daten zu gewährleisten. Diese Prinzipien stellen eine zuverlässige Grundlage für die Bewertung und Interpretation der Ergebnisse dar und sind essenziell für die Methodik jeder wissenschaftlichen Untersuchung.

Die Objektivität des Verfahrens wird durch die Anwendung standardisierter Verfahren und Instrumente sichergestellt. Dabei erfolgen alle Bewertungen unabhängig von individuellen Vorlieben oder Meinungen der Forschenden, um eine objektive Einschätzung der Führungskompetenzen zu gewährleisten (Obermann, 2018). Darüber hinaus werden Expertenbewertungen von mindestens zwei unabhängigen Personen durchgeführt, um die Objektivität der Studie weiter zu erhöhen.

Hinsichtlich der Reliabilität wird durch die Verwendung von konsistenten Methoden und Instrumenten eine zuverlässige Messung der Führungskompetenzen gewährleistet. In diesem Zusammenhang wird ein strukturierter Interviewleitfaden in allen Interviews verwendet, und die Bewertungen erfolgen auf Grundlage einer vorab definierten Kriterienliste (Obermann, 2018). Diese Maßnahmen zielen darauf ab, eine möglichst zuverlässige Bewertung der Führungskompetenzen zu gewährleisten und eine hohe Interrater-Reliabilität und Retest-Reliabilität zu erreichen.

Um die Validität der Studie zu gewährleisten, werden die Daten aus verschiedenen Methoden trianguliert. Die Verwendung von sowohl qualitativen als auch quantitativen Daten sowie unterschiedlichen Methoden zur Erfassung der Führungskompetenzen ermöglicht eine umfassende und genaue Einschätzung der untersuchten Konstrukte (Obermann, 2018). Darüber hinaus wird die Übereinstimmung der Ergebnisse zwischen den verschiedenen Verfahren analysiert, um die Validität der verwendeten Instrumente zu überprüfen.

Trotz dieser umfassenden Maßnahmen zur Sicherung von Objektivität, Reliabilität und Validität stellen diese Aspekte fortwährende Herausforderungen dar, insbesondere im Kontext von Assessment-Centern und diagnostischen Verfahren. Daher ist eine fortlaufende Evaluation dieser Aspekte und eine ständige Verbesserung der verwendeten Methoden und Instrumente von entscheidender Bedeutung, um die Qualität der gewonnenen Daten zu gewährleisten und aussagekräftige Schlussfolgerungen zu ermöglichen (Obermann, 2018).

Trotz der Berücksichtigung von Qualitätssicherungsprinzipien wie Objektivität, Reliabilität und Validität in dieser Studie sind einige kritische Aspekte in Bezug auf unser methodisches Vorgehen zu beachten. Erstens kann die Wahl der Klassifizierungsgrenze, die für das Modell gewählt wurde, zu variierenden Ergebnissen führen, was die Vorhersagegenauigkeit in Bezug auf die Eignung der Kandidaten beeinflusst. Ein Wert von 70% könnte beispielsweise zu weniger genauen Vorhersagen führen. Hier könnte eine ROC-Analyse helfen, ein besseres Verständnis für das Gleichgewicht zwischen wahren positiven und falsch positiven Raten zu erlangen und somit zur Optimierung der Klassifizierungsgrenze beitragen.

Zweitens beruht unsere Methode der Vorhersagegenauigkeit auf den gleichen Daten, die zur Schätzung der Koeffizienten (Modelltraining) verwendet wurden. Eine stärkere Methode wäre die Verwendung eines Train-Test-Split-Ansatzes, bei dem ein Teil der Daten zur Modellbildung und ein anderer Teil zur Evaluierung der Vorhersagegenauigkeit verwendet wird. In unserem aktuellen Ansatz könnte kritisiert werden, dass unser Modell möglicherweise überoptimiert ist, da es auf den gleichen Daten getestet wurde, auf denen es trainiert wurde. Ohne einen Test auf neuen, unbekannten Daten fehlt uns die Sicherheit, wie unser Modell in der Praxis abschneiden würde.

Drittens könnte eine umfassendere Untersuchung anderer Modelle wie Decision Trees, Random Forests oder Support Vector Machines eine umfangreichere und robustere Analyse ermöglichen. Jedes dieser Modelle hat seine eigenen Stärken und Schwächen und kann unterschiedliche Aspekte der Daten beleuchten. Es ist daher wichtig, künftige Forschung in diese Richtung zu erweitern, um eine umfassendere Perspektive auf unsere Vorhersagegenauigkeit zu erhalten.

Viertens, die Qualität des eigenen methodischen Vorgehens beim Assessment Center steht in direktem Zusammenhang mit der Repräsentativität der Stichprobenauswahl, die jedoch in diesem Fall in Frage gestellt werden kann. Die gewählte Herangehensweise, Stichproben basierend auf persönlicher Unterstützung auszuwählen, statt diejenigen auszuwählen, die ein breites und repräsentatives Spektrum von Fähigkeiten, Erfahrungen und Eigenschaften repräsentieren, kann erhebliche Auswirkungen auf die Gültigkeit und Zuverlässigkeit der daraus resultierenden Bewertungen haben. Ein solches voreingenommenes Vorgehen könnte zu einer sogenannten Bestätigungsverzerrung führen, bei der die Ergebnisse tendenziell die bestehenden

Ansichten und Erwartungen des Beurteilenden widerspiegeln, anstatt objektive Indikatoren der Leistung oder des Potenzials darzustellen. Dies weist auf die Notwendigkeit hin, das Auswahlverfahren für Stichproben zu überdenken und zu optimieren, um eine größere Objektivität und Gerechtigkeit zu gewährleisten.

5.3 Ausblick und weiteres Vorgehen

Zukünftige Forschungen könnten sich auf mehrere kritische Bereiche konzentrieren, um die Genauigkeit und Anwendbarkeit der in dieser Studie angewandten Methoden zu verbessern. Zunächst sollte die optimale Klassifizierungsgrenze, die in diesem Modell verwendet wird, weiter untersucht werden. Die Leistung des Modells kann durch die gewählte Klassifizierungsgrenze beeinflusst werden, wobei unterschiedliche Grenzwerte unterschiedliche Vorhersageergebnisse liefern können. Die Anwendung einer ROC-Kurvenanalyse könnte dabei helfen, einen optimalen Kompromiss zwischen der wahren positiven Rate und der falsch positiven Rate zu finden und somit die Klassifizierungsgenauigkeit zu verbessern.

Zweitens sollten zukünftige Untersuchungen auch eine robustere Evaluierung der Modellleistung in Betracht ziehen, indem sie auf die Technik des Train-Test-Splits zurückgreifen. Bei diesem Ansatz wird ein Teil der Daten für das Modelltraining und ein anderer Teil für die Überprüfung der Vorhersagegenauigkeit verwendet. Durch die Anwendung dieser Methode könnte verhindert werden, dass das Modell zu stark auf die Trainingsdaten optimiert wird und somit möglicherweise an Generalisierungsfähigkeit verliert. Es wäre von Vorteil, wenn das Modell in der Lage wäre, genaue Vorhersagen für neue, bisher nicht beobachtete Daten zu treffen.

Drittens könnte ein umfassender Vergleich unterschiedlicher Modelle, wie Decision Trees, Random Forests und Support Vector Machines, wertvolle Erkenntnisse liefern. Diese Modelle könnten unterschiedliche Aspekte der Daten hervorheben und dabei helfen, die Vorhersagegenauigkeit weiter zu verbessern. Darüber hinaus könnten Untersuchungen in dieser Richtung auch dazu beitragen, die Stärken und Schwächen verschiedener Vorhersagemodelle besser zu verstehen und geeignete Modelle für spezifische Kontexte auszuwählen.

Darüber hinaus könnte die Integration weiterer relevanter Daten, etwa aus umfangreicheren oder differenzierteren Assessmentverfahren, eine Verbesserung der Vorhersagegenauigkeit ermöglichen. Zudem könnte es sinnvoll sein, unterschiedliche Modelle oder Verfahren in einer komplementären Weise zu kombinieren, um eine umfassendere und robustere Einschätzung der Eignung der Kandidaten zu ermöglichen.

Schließlich wäre es auch wertvoll, die Praktikabilität und die Auswirkungen der Implementierung dieser Modelle und Verfahren in realen Einstellungen weiter zu erforschen. Dies könnte beispielsweise durch Feldstudien oder Experimente in Zusammenarbeit mit Organisationen und Unternehmen erfolgen, die an innovativen Verfahren zur Kandidatenauswahl interessiert sind.

Literaturverzeichnis

Achouri, C. (2015). Human Resources Management. Eine praxisbasierte Einführung. 2. Auflage. Springer Gabler.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R., (2018). Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 15. Auflage. Springer Gabler.

Baltes-Götz, B. (2012). Logistische Regressionsanalyse mit SPSS. Universität Trier.

Borts, J. & Döring, N. (19995). Forschungsmethoden und Evaluation. (2. Überarbeitete Auflage). Berlin: Springer.

Brandt, H. (2020). Exploratorische Faktorenanalyse (EFA). In Testtheorie und Fragebogenkonstruktion. Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2020.

Bregas, J. (2022). Rollenspiel im Assessment Center. Aufbau, Konzeption und Durchführung für Eignungsdiagnostiker. Springer-Verlag GmbH.

Brower, T. (2021). Empathy Is The Most Important Leadership Skill According To Research. Abgerufen am 20. Mai von <https://www.forbes.com/sites/tracybrower/2021/09/19/empathy-is-the-most-important-leadership-skill-according-to-research/?sh=6571900d3dc5>

Bundeszentrale für politische Bildung. (o.J.). Vorurteile und Stereotypen. Abgerufen am 16. April 2023 von <https://m.bpb.de/lernen/grafstat/fussball-und-nationalbewusstsein/130843/m-01-06-vorurteile-und-stereotypen>

Chui, M., et al. (2023). What every CEO should know about generative AI. McKinsey Digital. Abgerufen am 30. Mai 2023 von <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/what-every-ceo-should-know-about-generative-ai#/>

Clance, P. R., & Imes, S. A. (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice*, 15(3), 241-247. doi:10.1037/h0086006

DAHMEN (2016). Kosten und Auswirkungen von nicht- und Fehlbesetzungen. Veröffentlicht in *Arbeitnehmerüberlastung, Fachkräftemangel, Zeitarbeit*. Abgerufen von <https://dahmen-personal.de/kosten-nichtbesetzung-fehlbesetzungen/> am 05.08.2022.

Diagnostik- u. Testkuratorium (Hrsg.), Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430, Springer 2017.

Dorsch – Lexikon der Psychologie. (o.J.). Attributionsfehler. Abgerufen am 16. April 2023 von <https://dorsch.hogrefe.com/stichwort/attributionsfehler>

Dorsch – Lexikon der Psychologie. (o.J.). Logischer Fehler. Abgerufen am 16. April 2023 von <https://dorsch.hogrefe.com/stichwort/logischer-fehler>

Dorsch – Lexikon der Psychologie. (o.J.). Projektion. Abgerufen am 16. April 2023 von <https://dorsch.hogrefe.com/stichwort/projektion>

Döring, N. & Bortz, J. (2016). Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (5.). Berlin: Springer. Doi:10.1007/978-3-642-41089-5

Eck, C., D., Jöri, H., Vogt, M., (2016). Assessment Center. Entwicklung und Anwendung – mit 57 AC-Übungen und Checklisten. 3. Auflage. Springer.

Eckstein, P., P., (2016). Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler. 8., überarbeitete und erweiterte Auflage. Springer Gabler Fachmedien Wiesbaden.

Enders, T., Hieronimus, S., Meyer-Guckel, V., Schünemann, I. (2018). *Wie Future Skills die Personalarbeit verändern. Stifterverband für die Deutsche Wissenschaft e.V. in Kooperation mit McKinsey&Company.*

Kennedy, Y., (2020). The Importance Of Resiliency, Integrity And Leading With Empathy. *Forbes*. Abgerufen am 5 Mai 2023 von

<https://www.forbes.com/sites/forbeshumanresourcescouncil/2020/12/15/the-importance-of-resiliency-integrity-and-leading-with-empathy/?sh=3f7e3696678f>

Gläser, J. & Laudel, G. (2006). Experteninterviews und qualitative Inhaltsanalyse (2. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.

HAUFE (2019). Cultural Fit: Was ist das eigentlich? Zugegriffen am 24.02.2022.
<https://www.myonboarding.de/magazin/cultural-fit-was-bedeutet-das-eigentlich?akttyp=organische%20suche&med=google&aktnr=84834&wnr=04393672>

Hartmann, P (2002): Indikator. In: Endruweit, Günter & Trommsdorff, Gisela (Hrsg.), Wörterbuch der Soziologie, Stuttgart: Lucius & Lucius, S. 223-224.

Held, U. (2010). Was ist eine «Odds Ratio»–und wann wird sie verwendet?. In *Swiss Medical Forum* (Vol. 10, No. 37, pp. 634-635). EMH Swiss Medical Publishers.

Helfferrich, C. (2019). Leitfaden- und Experteninterviews. Handbuch Methoden der empirischen Sozialforschung. Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature.

Hoff, E.-H. (1985). Datenerhebung als Kommunikation: Intensivbefragungen mit zwei Interviewern. In G. Jüttemann (Hrsg.), *Qualitative Forschung in der Psychologie : Grundfragen, Verfahrensweisen, Anwendungsfelder* (S. 161-186). Weinheim: Beltz.
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-5617>

Janssen, C., Mühlfelder, M., Küchler, S., Paschen, M., Karg, K., Oenning, S., (2022). Entwicklung und Evaluierung eines Fragebogens zur Selbsteinschätzung von Führungskompetenzen und -potenzialen: LeadIM. Profil M Beratung für Human Resources Management GmbH & Co. KG Wermelskirchen.

Kanning, U. P. (2021). Künstliche Intelligenz in der Eignungsdiagnostik. *Digitales Human Resource*. Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer nature.

Kanning, U. P. (2019). Standards der Personaldiagnostik (2. Aufl.) Hogrefe.

Kanning, U.P. (2016). Über die Sichtung von Bewerbungsunterlagen in der Praxis der Personalauswahl. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 60(19), 18-32. <https://doi.org/10.1026/0932/a000193>

Kratz, H. J., (2013). 20 Rollenspiele für Führungssituationen. Beltz Verlagsgruppe.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (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.

Langer, M., Lazar, V., (2023). Künstliche Intelligenz in eignungsdiagnostischen Interviews. Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2023.

Mey, G., Mruck, K., (2020). Qualitative Interviews. Handbuch Qualitative Forschung in der Psychologie. Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature.

Tirrel, H., Winnen, L., Ranwehr, R. (2021). Digitales Human Ressource Management. Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2021.

Obermann, C. (2018). Assessment Center. Entwicklung, Durchführung, Trends Mit neuen originalen AC-Übungen. 6. Auflage. Springer.

Obermann, C., & Solga, M. (2018). Jobinterviews professionell führen. Über 400 Interviewfragen für die erfolgreiche Bewerberauswahl. Springer Gabler.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston.

Reinhardt, R., Ornau, F., Tennert, F. (2020). Interviewtechnik. 3. Auflage., Studienbrief der SRH Fernhochschule. Riedlingen.

Roth, W. (2021). Die resiliente Führungskraft. Sich selbst und andere gesund führen. Springer Gabler.

Sarges, W. (2013). Management-Diagnostik. Göttingen: Hogrefe

Sears, J. (2019). How to. Choose Between a Self or 360 Leadership Assessment. Abgerufen am 4. Mai 2023 von <https://www.flashpointleadership.com/blog/how-to-choose-between-a-self-or-360-leadership-assessment>

Schuler, H. (2014). Psychologische Personalauswahl: Eignungsdiagnostik für Personalentscheidungen und Berufsberatung. Göttingen: Hogrefe

Schlüter, A. (2020). Dunning-Kruger-Effekt: Weshalb inkompetente Menschen oft das größte Selbstbewusstsein haben. Geo.de. Abgerufen am 16. April 2023 von <https://www.geo.de/wissen/23942-rtkl-psychologie-dunning-kruger-effekt-weshalb-inkompetente-menschen-oft-das-groesste>

Spektrum der Wissenschaft. (o.J.). Halo-Effekt. Abgerufen am 16. April 2023 von <https://www.spektrum.de/lexikon/psychologie/halo-effekt/6232>

Stangl, W. (2016). Florida-Effekt. Stangl-Taller – Lexikon der Psychologie. Abgerufen am 16. April 2023 von <https://lexikon.stangl.eu/10246/florida-effekt>

Stangl, W. (2020). Selektive Wahrnehmung. Stangl-Taller – Lexikon der Psychologie. Abgerufen am 16. April 2023 von <https://lexikon.stangl.eu/1708/selektive-wahrnehmung>

Stangl, W. (2012). Überlegenheitsillusion. Stangl-Taller – Lexikon der Psychologie. Abgerufen am 16. April 2023 von <https://lexikon.stangl.eu/9527/ueberlegenheitsillusion>

Stulle, K. P. (2020). Digitalisierung der Management-Diagnostik. Aktuelle Instrumente, Trends, Herausforderungen. Springer Gabler.

Wiemer, K. (2020). Kopfzerbrechen: Die Psychologie des Umgangs mit Risiko und Krise. Wissenschaftskommunikation.de. Abgerufen am 16. April 2023 von <https://www.wissenschaftskommunikation.de/kopfzerbrechen-die-psychologie-des-umgangs-mit-risiko-und-krise-24731/>

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, 9(2, Pt.2), 1-27. doi:10.1037/h0025848

Anhang A Beurteilungsfehler und Wahrnehmungsverzerrungen

Der **Halo-Effekt** beschreibt einen Beurteilungsfehler, bei dem ein bestimmtes Merkmal andere Merkmale dominiert und somit Rückschlüsse auf diese Merkmale ermöglicht. Beispielsweise werden attraktive Personen häufig als intelligenter eingeschätzt (Amelang & Bartussek, 2023). Der Primäreffekt, auch erster Eindruck genannt, bezieht sich auf das Gedächtnisphänomen, dass früher präsentierte Informationen besser behalten werden. Dies führt oft dazu, dass ein anfängliches Bauchgefühl über Personen entwickeln wird, das sich später als zutreffend erweist.

Der **Pygmalion-Effekt**, auch als Rosenthal-Effekt (Rosenthal & Jacobson, 1968) bekannt, besagt, dass die Leistung eines Mitarbeiters maßgeblich von den Erwartungen der Führungskraft abhängt, da diese die Leistung unbewusst fördert oder hemmt und somit unterschiedliche Leistungen des Mitarbeiters ermöglicht. Die selbsterfüllende Prophezeiung, auch Andorra-Effekt genannt, tritt auf, wenn eine Person das Ergebnis einer Situation übernimmt oder ihre Wahrnehmung der Erwartungen anderer das Ergebnis beeinflusst oder verursacht. Wenn der Mitarbeiter beispielsweise davon ausgeht, dass der Vorgesetzte seine Leistung nicht anerkennen wird, interpretiert er alles in diese Richtung und bestätigt seine Wahrnehmung ständig durch negative Vorurteile.

Die **Projektion** bezieht sich auf die Übertragung unerwünschter Eigenschaften auf andere Personen, wodurch etwas in andere Personen hineininterpretiert wird, was eigentlich in einem selbst liegt. Stereotypen sind wiederholte Urteile über Personen, Gruppen, Dinge oder Verhaltensweisen, die dazu beitragen, sich in Situationen zurechtzufinden und Komplexität schnell zu reduzieren (Dorsch – Lexikon der Psychologie, o.J.). Die selektive Wahrnehmung beschreibt das Phänomen, dass bewusst nur Teile unserer Umwelt wahrgenommen werden und jene Teile ausgeblendet werden, die nicht mit unseren Annahmen übereinstimmen (Stangl, 2023).

Der **Attributionsfehler** bezieht sich auf die Überschätzung von Einflüssen wie Persönlichkeit, Einstellungen und Meinungen auf das Verhalten anderer, während situative Einflüsse unterschätzt werden (Dorsch – Lexikon der Psychologie, o.J.). Die Überlegenheitsillusion, umgangssprachlich als Selbstüberschätzung bezeichnet, beschreibt die Tendenz von Menschen, ihre eigenen Fähigkeiten und Stärken systematisch und konsequent zu überschätzen, während ihre eigenen Fehler und Schwächen unterschätzt werden (Stangl, 2023).

Priming bezeichnet den Effekt, dass das Denken durch einen vorherigen Stimulus beeinflusst wird, was die Verarbeitungsgeschwindigkeit erhöht. Der Florida-Effekt zeigt, dass Priming sogar das Verhalten beeinflussen kann. Framing bezieht sich auf den Interpretationsrahmen, der Einfluss darauf hat, wie eine Situation oder Beziehung wahrgenommen wird. Je nachdem, wie der Rahmen gesetzt ist, kann eine Situation positiv oder negativ wahrgenommen werden (Stangl, 2016).

Der **Mere Exposure Effect** beschreibt die Tendenz, Dinge zu mögen, die uns vertraut erscheinen. Dies liegt daran, dass vertraute Dinge leichter verarbeitet werden, was zu einem positiven Gefühl führt (Zajonc, 1968). Der logische Fehler tritt auf, wenn persönliche Überzeugungen die Schlussfolgerung von assoziierten Merkmalen bedingen, im Gegensatz zum Halo-Effekt, bei dem es zu einer diffusen Übertragung eines Merkmals auf andere Merkmale kommt. Zum Beispiel kann aus eingeschränkten Deutschkenntnissen auf geringe kognitive Fähigkeiten geschlossen werden.

Der **Dunning-Kruger-Effekt** bezieht sich auf die kognitive Verzerrung inkompetenter Menschen, das heißt, je weniger eine Person über einen Sachverhalt weiß, desto mehr überschätzt sie sich in dieser Hinsicht. Außerdem sind sie nicht in der Lage, ihre Inkompetenz zu erkennen und unterschätzen die Fähigkeiten von Experten (Schlüter, 2020).

Das **Impostersyndrom**, auch als Hochstaplerphänomen bezeichnet, beschreibt das Phänomen, dass Personen nicht in der Lage sind, ihre Erfolge zu verinnerlichen. Obwohl es offensichtliche Beweise für ihre Erfolge und Kompetenzen gibt, führen sie diese auf Glück oder äußere Umstände zurück (Clance & Imes, 1978).

Anhang B Teamansprache

Ziel dieser Übung

In dieser Übung führt der Teilnehmer/ die Teilnehmerin des Assessment-Centers in der Rolle des Head of Service der Brikolage GmbH ein Gespräch mit seinem Team. Ziel des Gesprächs ist es das Team für eine Neueinführung einer Software zu motivieren.

Ihre Rolle

Sie sind Simon/Simone Müller, 30 Jahre alt und seit 2 Jahren bei der Brikolage GmbH tätig. Im Anschluss an Ihr Studium waren Sie zunächst in einem mittelständischen Softwarefirma angestellt, bevor Sie aufgrund Ihres Wunsches nach beruflicher Veränderung zur Brikolage GmbH wechselten. Aufgrund ihrer vielfältigen Erfahrung sind Sie ein geschätzter Know-How-Träger. Im Rahmen von internen Changeprozessen gab es eine Teamumstrukturierung, so dass für eins der Teams eine neue Führungskraft gesucht wurde. Sie haben kürzlich die Teamleitung übernommen. Das Team ist seit mehreren Jahren in der Konstellation zusammen und besteht aus 7 Experten für unterschiedliche Bereiche und Themen. Zwei von Ihren neuen Mitarbeiter:Innen im Team haben sich auf Ihren Posten der Teamleitung beworben. Darüber hinaus ist ein Teammitglied seit über 6 Monaten dauerkrank, so dass die anderen Teammitglieder seine Aufgaben weitestgehend mitübernehmen müssen und über eine hohe Anzahl an Überstunden klagen. Ihre erste Aufgabe ist es eine neue Zeiterfassungssoftware in ihrem Team einzuführen. Bis jetzt existierte keine Zeiterfassung, es wird darauf vertraut, dass die Mitarbeiter ihre Stunden absolvieren.

Sie halten gleich eine Teamansprache, in der Sie das Team bezüglich der neuen Zeiterfassungssoftware informieren.

Argumente für das Gespräch

- **Nachvollziehbarkeit der Arbeitszeiten ist seit Corona nicht mehr für die Geschäftsführung gegeben**

Im Rahmen von Corona wurde für die Mitarbeiter die Möglichkeit geschaffen zu 2/3 ihre Arbeit aus dem Homeoffice heraus zu absolvieren. Die Geschäftsführung hat Ihnen gegenüber angezweifelt, ob die Mitarbeiter wirklich ihre vertraglich vereinbarten Arbeitszeiten einhalten. Seit Corona wäre die Arbeitsleistung im Unternehmen reduziert und bis heute nicht wieder auf dem Stand von vor Corona.

- **Der Gesetzgeber fordert eine nachvollziehbare Zeiterfassung auch bezogen auf die eingehaltenen Pausenzeiten**

Zwar ist das Urteil des Europäischen Gerichtshofes zur verpflichtenden Arbeitszeiterfassung noch nicht im deutschen Recht verankert, allerdings hat das Bundesarbeitsgericht (BAG) die EU-Entscheidung bestätigt. Damit ist Arbeitszeiterfassung verpflichtend. *„Unternehmen, die keine Lösungen zur umfassenden Arbeitszeiterfassung anbieten, befinden sich nun in einem rechtswidrigen Zustand – allerdings muss der Gesetzgeber das Urteil auch noch umsetzen! Setzen Arbeitgeber diese Verpflichtung allerdings nicht zeitnah um, führt dies voraussichtlich in arbeitsrechtlichen Prozessen zu einer Beweislastumkehr zulasten des Arbeitgebers. Im Ernstfall: Behauptet der Arbeitnehmer nachvollziehbar, bestimmte Stunden erbracht zu haben, ob läge es an dem Arbeitgeber, darzulegen und zu beweisen, dass diese Stunden dennoch nicht erbracht wurden.“* Die Geschäftsführung möchte proaktiv handeln und dementsprechend das Arbeitszeitmanagement angehen.

- **Überstundenregelung & Sorgfaltspflicht der Führungskraft**

Gerade Ihr Team klagt immer wieder über Überstunden und eine hohe Auslastung. Der Arbeitgeber ist gesetzlich verpflichtet sich an das Arbeitszeitgesetz zu halten in dem beispielsweise geregelt ist, dass die werktägliche Arbeitszeit von 8 Stunden pro Tag nicht überschritten werden darf. Diese darf zwar ausnahmsweise auf 10 Stunden ausgedehnt werden, muss allerdings innerhalb der nächsten 6 Monate ausgeglichen werden. Darüber hinaus ist eine Ruhezeit von mindestens 11 ununterbrochenen Stunden zu gewährleisten.

Anhang C Mitarbeitergespräch

Der Hintergrund, der dem Teammitglied gegenüber kommuniziert wird:

Vor ein paar Wochen hat die Geschäftsführung ACTION NOW ausgerufen, was bedeutet, dass mindestens 20% der Kosten eingespart werden müssen, da das letzte Jahr coronabedingt sehr schlecht verlaufen ist. Erste Kündigungen sind ausgesprochen worden.

Zwar handelt es sich bei den Mitarbeitern um sehr junge Mitarbeiter, die noch in der Probezeit waren oder sehr alte, die kurz vor der Rente stehen.

Ihr Mitarbeiter Manfred ist seit mehreren Jahren in Ihrer Abteilung tätig. Mit seinen Leistungen waren Sie immer äußerst zufrieden, nie gab sein Verhalten zu keinen Klagen Anlass. Seit einiger Zeit bemerkten Sie, dass Ihrem Mitarbeiter Fehler unterlaufen, die früher undenkbar gewesen waren. Darüber hinaus meldet sich der Mitarbeiter seit Monaten immer mal wieder jeweils für den Freitag und Montag krank. Manfred hat jeweils eine Krankmeldung, gleichwohl keiner weiß, warum das Teammitglied jeweils krankgeschrieben wird. Das Team ist schon auf Sie zukommen, und hat sich beschwert, dass das Teammitglied sehr Monaten unzuverlässig sei und keine gute Zusammenarbeit mehr mit Manfred möglich sei. Das Team müsse viele Aufgaben von Manfred mit übernehmen, und seien so schon maximal überlastet. Häufig käme es sogar zu Fehlern, die dann dazu führen, dass noch mehr Arbeit anstände. Peter, einer ihrer Mitarbeiter war letzte Woche bei Ihnen und hat sich völlig frustriert über Manfred ausgelassen, er war äußerst wütend, da er immer häufiger Manfreds Aufgaben übernehmen und Manfreds Fehler ausbügeln müsse. Als sich ein Kunde über Manfred beschwert hat, haben auch andere Teammitglieder mit ihnen gesprochen, dass auf das Manfred kein Verlass und keine Teamarbeit möglich sei. An den Tagen, an denen Manfred arbeitet, ist zu beobachten, dass das er durch Unkonzentriertheit und Launigkeit auffällt. Das Team ist frustriert, da Manfred sich immer weiter zurückzieht und weder auf Verständnis oder nachfragen aus dem Team reagiert. Sie halten es für Ihre Pflicht, die Gesamtsituation zur Sprache zu bringen ...

Anhang D Interviewleitfaden

Interviewleitfaden zur situativen Erfragung des Führungshandelns, entsprechend der vier Komponenten der Führung (aktivierende Führung, selbstbewusste/reflektierende Führung, empathische Führung, strukturierende/organisierende Führung).

Interviewleitfaden „Führungshandeln“

Begrüßung und Einleitung

Sehr geehrte Frau - Sehr geehrter Herr

Herzlich Willkommen.

Wir begrüßen Sie zur heutigen Interviewteilnahme zum Thema „Führungshandeln“ und bedanken uns für Ihre Bereitschaft der Teilnahme. Im Voraus möchten wir uns vorstellen und den Interviewablauf sowie den Forschungszweck der heutigen Befragung erläutern. Am Ende des Interviews ist Zeit eingeplant, um aufkommende Fragen oder Anregungen zu besprechen.

Vorstellung Herr Prof. Dr. Manfred Mühlfelder

Mein Name ist Dagmar Rostek, ich studiere Wirtschaftspsychologie und im Rahmen meiner Master-Thesis zum Thema „Analyse des inkrementellen Nutzens eines Persönlichkeitsfragebogens in der Managementdiagnostik“ führen wir dieses Interview durch, mit der Fragestellung, welchen inkrementellen Nutzen ein KI-Algorithmus auf Basis der binär logistischen Regression oder kategorialer Regressionsmodelle beispielsweise einem Assessment Center liefern kann, um die Entscheidungssicherheit bezüglich geeignet oder ungeeignet für eine bestimmte Rolle beziehungsweise Funktion zu erhöhen.

Die Fragestellungen werden offen sein und lassen ausreichend Raum bei der Beantwortung. Bitte erzählen Sie alles, was Ihnen zu den Fragen einfällt. Die Gesprächsdauer beträgt ca. 45 Minuten. Es gibt keine richtigen oder falschen Antworten, ihre Äußerungen werden wertungsfrei betrachtet.

Um die nachfolgende Auswertung zu vereinfachen, bitten wir Sie das Gespräch aufzeichnen zu können. Wir benötigen hierzu eine konkrete Einwilligung sowie eine Einverständniserklärung für Ihre Teilnahme. Wir versichern Ihnen, dass Ihre Daten nach dem Datenschutzgesetz vertraulich und anonymisiert behandelt werden.

Sollte die Einverständniserklärung noch nicht vorliegen sollte, bitten diese umgehend zur Verfügung zu stellen, sowie der Genehmigung für die Tonbandaufzeichnung.

Persönlicher Teil

Name:

Beruf:

Alter:

Geschlecht:

Führungsrolle

seit:

Branche:

Datum:

Beginn:

Ende:

Hauptteil

Im folgenden Abschnitt des Interviews werden wir Ihnen Fragen zum Thema Führungshandeln stellen.

Kategorie aktivierende Führung

Lernmotivation: Beschreiben Sie die letzte Situation, in der Sie etwas Neues gelernt haben.

Initiative: In welcher beruflichen Situation war Ihre Initiative gefordert?

Gestalten und vorangehen: Beschreiben Sie eine Situation, in der Sie bedeutsames bewegt haben.

Inspizieren und Neues wagen: Wann haben Sie sich mal ins ungewisse Begeben? Steve Jobs Gen, was inspiriert dich

Aktivierende Führung	Selbstaussage					Fremdeinschätzung				
	1	2	3	4	5	1	2	3	4	5
<ul style="list-style-type: none"> • Zeigt Lernwille und Entwicklungsmotivation, verlässt die Komfortzone • Zeigt Tatendrang, geht Risiken ein, setzt die Aufgaben schnell um • Ermutigt andere, fördert Lösungskompetenz bei anderen, ist Coach und Berater für Kollegen/ Mitarbeiter • Ist neugierig, unkonventionell und einfallsreich 										
Gesamtbewertung										

Interpretation Skala
 1 stark entwicklungsbedürftig
 4 persönliche Stärke

2 entwicklungsbedürftig
 5 besondere Stärke

Neutral (erfüllt Anforderungen)

Kategorie Selbstbewusste reflektierende Führung

Selbstvertrauen: In welcher Situation war Ihr Selbstvertrauen gefragt? (Selbstauskunft versus Beobachtereinschätzung im Interview, Teammotivation und Rollenspiel)

Resilienz: Angenommen Sie sind mit einer Aufgabe überfordert, wie gehen Sie damit um? ODER In welcher Situation war ihre Resilienz gefordert?

Selbstreflexion: Gibt es in Ihrem Arbeitskontext Zeit für Reflexion? (Selbstauskunft versus Beobachtereinschätzung im Interview, Teammotivation und Rollenspiel)

Überzeugen und Sinn vermitteln: In welcher Situation mussten Sie Kollegen von etwas überzeugen und eine Sinnhaftigkeit aufzeigen? (Selbstauskunft versus Beobachtereinschätzung im Interview, Teammotivation und Rollenspiel)

Durchsetzen und Steuern: Was war das Letzte, was Sie gegen Widerstand durchsetzen mussten? (Selbstauskunft versus Beobachtereinschätzung im Interview, Teammotivation und Rollenspiel)

Selbstbewusste reflektierende Führung	Selbstaussage					Fremdeinschätzung				
	1	2	3	4	5	1	2	3	4	5
• Glaubt an die eigene Leistungsfähigkeit										
• Tritt souverän auf, ist robust und gelassen										
• Zeigt eine gute Selbsteinschätzung, hat einen guten Zugang zu sich selbst und den eigenen Gefühlen, das eigene Verhalten wird schlüssig interpretiert										
• Kann andere Sichtweisen beeinflussen, die eigenen Überzeugungen vertreten und hat eine hohe persönliche Präsenz										
• Vertritt eigene Interessen, kann Konflikte eingehen und Forderungen formulieren										
Gesamtbewertung										

Interpretation Skala
 1 stark entwicklungsbedürftig
 4 persönliche Stärke

2 entwicklungsbedürftig
 5 besondere Stärke

Neutral (erfüllt Anforderungen)

Kategorie Strukturierende und prinzipienorientierte Führung

Strukturieren und Planen: Wie strukturieren Sie ihre Arbeit entsprechend der Prioritäten?

Willensstärke: In welcher Situation war ihre Willensstärke gefordert?

Selbstvertrauen: In welcher Situation war Ihr Selbstvertrauen gefragt? (Selbstauskunft versus Beobachtereinschätzung im Interview, Teammotivation und Rollenspiel)

Ambition: Beschreiben Sie eine Situation, bei der Sie allein durch Ihr Durchhaltevermögen das angestrebte Ziel erreicht haben.

Strukturierende und prinzipienorientierte Führung	Selbstaussage					Fremdeinschätzung				
	1	2	3	4	5	1	2	3	4	5
<ul style="list-style-type: none"> Ist genau, zeigt Überlegenheit und Prinzipientreue Zeigt Selbstdisziplin, Durchhaltevermögen und Selbstkontrolle Glaubt an die eigene Leistungsfähigkeit, ist zuversichtlich und furchtlos Zeigt Zielorientierung, Einsatzbereitschaft und Leistungsmotivation 										
Gesamtbewertung										

Interpretation Skala

1 stark entwicklungsbedürftig

4 persönliche Stärke

2 entwicklungsbedürftig

5 besondere Stärke

Neutral (erfüllt Anforderungen)

Kategorie Empathische, integrierende Führung

Empathie: Beschreiben Sie eine Situation, in der Sie empathisch auf einen Ihrer Mitarbeiter eingegangen sind.

Verbinden und Integrieren: Wie würden Sie einen neuen Mitarbeiter ins Team integrieren?

Entwickeln und Fördern: Wie würden Sie Kollegen oder Mitarbeiter Fördern?

Motivieren und aktivieren: Wie würden Sie Kollegen oder Mitarbeiter aktivieren?

Empathische, integrierende Führung						Selbstaussage					Fremdeinschätzung				
						1	2	3	4	5	1	2	3	4	5
<ul style="list-style-type: none"> • Zeigt Einfühlungsvermögen, ein echtes Interesse an Anderen und Verständnis für andere Perspektiven • Zeigt Kontaktstärke, Teamorientierung Nähe und Austausch • Ermutigt andere, fördert Reflexion und Lösungskompetenz bei anderen, Tritt als Coach und Berater auf • Vermittelt einen positiven Spirit, geht geschickt auf andere Bedürfnisse und unterschiedliche Persönlichkeiten ein, hat eine energetisierende Wirkung 															
Gesamtbewertung															

Interpretation Skala
 1 stark entwicklungsbedürftig
 4 persönliche Stärke

2 entwicklungsbedürftig
 5 besondere Stärke

Neutral (erfüllt Anforderungen)

Schluss teil

Vielen Dank für Ihre Kooperation und Teilnahme am heutigen Interview. Sollten Sie Fragen oder Anregungen haben, können wir diese jetzt sehr gerne besprechen. Gibt es von Ihrer Seite Anmerkungen, die Sie uns mitteilen möchten, oder haben Sie Fragen?

Verabschiedung & Dan

Eidesstattliche Erklärung

Versicherung

Name: Dagmar Rostek
Matrikel-Nr.: 2004169
Studiengang: Wirtschaftspsychologie

Hiermit versichere ich, dass ich die vorliegende Master-Thesis mit dem Thema

„Analyse des inkrementellen Nutzens eines Persönlichkeitsfragebogens in der Managementdiagnostik.“

ohne fremde Hilfe erstellt habe. Alle verwendeten Quellen wurden angegeben. Ich versichere, dass ich bisher keine Master-Thesis mit gleichem oder ähnlichem Thema an einer anderen Hochschule eingereicht habe.

Datum: 31.05.2023

Unterschrift:

The image shows a handwritten signature in black ink. The signature is written in a cursive style, starting with a large 'D' followed by a period and the name 'Rostek'. The signature is positioned above a horizontal line.