Die Validierung von eignungsdiagnostischen Verfahren: der Persönlichkeitsfragebogen und die virtuelle Dialogübung

Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Pia Sophie Wedemeyer

Kiel

09.01.2023

Erstgutachter: Prof. Dr. Udo Konradt
Zweitgutachter: PD Dr. Daniela Renger
Tag der mündlichen Prüfung: 19.06.2023
Durch den Prodekan für Studium und Lehre Prof. Dr. Jörg Kilian,
zum Druck genehmigt am: 26.06.2023

Danksagung

Ich bin sehr dankbar, dass ich die Möglichkeit erhalten habe, meine Promotion neben meiner ersten Berufserfahrung schreiben zu dürfen und von vielen Personen unterstützt worden zu sein.

Der erste Dank gilt meinem Betreuer und Doktorvater Prof. Dr. Udo Konradt für die ständige Erreichbarkeit sowie schnelle produktive Rückmeldungen und vor allem für die Möglichkeit, als externe Promovendin an seinem Lehrstuhl schreiben zu können. Ich danke ihm für die vielen fachlichen Ratschläge und das Vertrauen in mich und meine Arbeit. Darüber hinaus möchte ich mich auch bei dem gesamten Lehrstuhl für die Zusammenarbeit bedanken. Bei auftretenden Fragen wurde ich stets tatkräftig unterstützt. PD Dr. Jürgen Golz möchte ich dafür danken, dass er sich als Zweitgutachter meiner Arbeit zur Verfügung gestellt hat.

Ein großer Dank gilt auch meinem Arbeitgeber Moldzio & Partner, der es mir als wissenschaftlichen Mitarbeiterin erst ermöglicht hat, neben der Arbeit in einer Unternehmensberatung auch die Promotion voranzutreiben. Vielen Dank Dr. Martina Böge für die ständige Beratung zu inhaltlichen Themen sowie die aufmunternden Worte in schwierigen Situationen. Ein Dank gilt auch Dr. Thomas Moldzio und Dr. Martina Böge für das inhaltliche Korrekturlesen meiner Arbeit und meinen Kolleg:innen im Büro, welche die formalen Korrekturen durchgeführt haben und mir in stressigen Situationen den Rücken für meine Dissertation freigehalten haben.

Ich bedanke mich sehr herzlich bei meinen Eltern und der restlichen Familie. Ihr habt mich zu jeder Zeit ermutigt, die Arbeit fertig zu schreiben, hattet immer ein offenes Ohr für mich und habt mir vor allem in den letzten Schreibphasen eine ruhige und erholsame Atmosphäre ermöglicht. Ein besonderer Dank gilt meiner Mutter, die in ihrer Freizeit meine für sie fachfremde Arbeit auf Rechtschreibung, Grammatik und Verständnis überprüft hat. Darüber hinaus bedanke ich mich auch bei meinen Freunden, die mich immer wieder mit verschiedenen Aktivitäten aus stressigen Phasen herausgeholt und abgelenkt haben. Zudem waren sie immer direkt zur Stelle, wenn Probleme bei der Formatierung auftraten.

Inhaltsverzeichnis

Danksagung ... 3
Inhaltsverzeichnis ... 4
Tabellenverzeichnis ... 7
Abbildungsverzeichnis ... 9
Abkürzungsverzeichnis ... 10
Zusammenfassung ... 12
Abstract ... 13
Einleitung ... 14
Theorieteil ... 15
Berufliche Eignungsdiagnostik ... 16
Methoden der Eignungsdiagnostik ... 17
Digitalisierung in der Eignungsdiagnostik ... 20
Media Richness Theory ... 22
Media Synchronicity Theory ... 23
Verfahrens-Mediums-Matrix ... 24
Chancen und Einschränkungen ... 26
Wandel der Eignungsdiagnostik ... 28
Persönlichkeit in der Eignungsdiagnostik ... 33
Persönlichkeitsmodelle ... 33
Big Five Dimensionen ... 35
Big Five Aspekte ... 39
Persönlichkeitserfassung in der Eignungsdiagnostik ... 43
Haupt- und Nebenkriterien eines Persönlichkeitsfragebogens ... 44
Berufserfolg in der Eignungsdiagnostik ... 47
Job Performance Model ... 47
Berufserfolg und Persönlichkeit ... 50
Weitere Berufskriterien und der Zusammenhang mit Persönlichkeitsfaktoren ... 54
Gerechtigkeitswahrnehmung in der beruflichen Eignungsdiagnostik ... 56
Gerechtigkeitswahrnehmung von Auswahlprozessen ... 58
Gerechtigkeitswahrnehmung einzelner Auswahlverfahren ... 59
Modell der Bewerbendenreaktionen in Auswahlprozessen ... 63
Zusammenhänge zwischen der Gerechtigkeitswahrnehmung und der Persönlichkeit ... 69
Entwicklung und Validierung eines Fragebogens zu den Aspekten der Extraversion und Offenheit für Erfahrung ... 71
Gegenstand der Fragestellung ... 71
Kurzabriss der theoretischen und empirischen Grundlagen ... 72

Ableitung der Hypothesen 74
Methoden 78
Stichprobe 78
Vorstudie 79
Auszubildende 80
Kaufmännische Auszubildende und duale Studierende 80
Technische Auszubildende und duale Studierende 80
Expert:innen ohne Führungsverantwortung 81
Führungskräfte 81
Linienführungskräfte 82
Hochrangige Führungskräfte 82
Operationalisierung der Variablen 82
Operationalisierung der Prädiktoren 82
Persönlichkeit 82
Schlussfolgerndes Denken 83
Operationalisierung der Kriterien 84
Operationalisierung der konvergenten Validität 84
Operationalisierung der divergenten Validität 86
Operationalisierung der Kriteriumsvalidität 89
Untersuchungsdurchführung 93
Datenanalyse 95
Vorstudie 95
Hauptstudie 96
Weitere Analysen 102
Ergebnisse 103
Vorstudie 103
Hauptstudie 105
Konstruktvalidität 109
Kriteriumsvalidität 114
Weitere Analysen 119
Diskussion 123
Konstruktvalidität 123
Kriteriumsvalidität 127
Limitationen 132
Implikation und weitere Forschung 135
Die Validierung von virtuellen Dialogübungen im eignungsdiagnostischen Kontext 136
Gegenstand der Fragestellung 136

Kurzabriss der theoretischen und empirischen Grundlagen 136
Ableitung der Hypothesen 138
Weitere Analysen 142
Methoden 143
Stichprobe 143
Operationalisierung der Variablen 144
Operationalisierung der unabhängigen Variablen 144
virtuell versus nicht virtuell 144
Persönlichkeit 145
Operationalisierung der abhängigen Variablen 146
Berufserfolg 146
Gerechtigkeitswahrnehmung 151
Durchführung 154
Vor dem Workshop 155
Während des Workshops 155
Nach dem Workshop 157
Datenanalyse 157
Weitere Analysen 161
Ergebnisse 162
Diskussion 170
Weitere Analysen 173
Limitationen 175
Implikation und weitere Forschung 177
Generelle Diskussion 178
Literaturverzeichnis 181
Anhang 194
Dialogübung 194
Einschätzung der Beobachtenden 195
Profilbogen 197

Tabellenverzeichnis

Tabelle 1 Hypothesen zur Überprüfung der Gütekriterien sowie der Konstruktvalidität..... 76
Tabelle 2 Hypothesen zur Kriteriumsvalidität..... 77
Tabelle 3 Hypothesen zur inkrementellen Validität und dem Vergleich der prädiktiven Validität zwischen den Aspekten und der Dimension..... 78
Tabelle 4 Darstellungen der verschiedenen Stichprobengruppen..... 79
Tabelle 5 Zuordnung der Fragebogenskalen zur konvergenten und divergenten Validität.... 85
Tabelle 6 Darstellung und Beschreibung der Abstufungen der Kriteriumsvariablen..... 92
Tabelle 7 Erklärung des Vorgehens zur Betrachtung der inkrementellen Validität 102
Tabelle 8 Skalenkennwerte für die Skala Enthusiasmus 106
Tabelle 9 Skalenkennwerte für die Skala Durchsetzungsfähigkeit..... 107
Tabelle 10 Skalenkennwerte für die Skala Offenheit 107
Tabelle 11 Skalenkennwerte für die Skala Intellekt 109
Tabelle 12 Ergebnisse konfirmatorische Faktorenanalyse 110
Tabelle 13 Korrelationstabelle der Business Big 5..... 113
Tabelle 14 Übersicht der Hypothesenprüfung 3e der Business Big 5 114
Tabelle 15 Kennwerte der Korrelationen zwischen der Durchschnittsnote und den Aspekten 115
Tabelle 16 Kennwerte der Korrelationen zwischen dem schlussfolgernden Denken und den Aspekten 116
Tabelle 17 Kennwerte der Korrelationen zwischen der Eingruppierung und den Aspekten 116
Tabelle 18 Kennwerte der Korrelationen zwischen dem Potenzial und den Aspekten 117
Tabelle 19 Ergebnisse Strukturgleichungsmodelle zur inkrementellen Validität mit dem Kriterium Eingruppierung 117
Tabelle 20 Ergebnisse Strukturgleichungsmodelle zur inkrementellen Validität mit dem Kriterium Potenzial..... 119
Tabelle 21 Kennwerte der Korrelationen zwischen der Durchschnittsnote und den Aspekten 119
Tabelle 22 Kennwerte der Korrelationen zwischen dem schlussfolgernden Denken und den Aspekten 120
Tabelle 23 Kennwerte der Korrelationen zwischen der Eingruppierung und den Aspekten 122
Tabelle 24 Kennwerte der Korrelationen zwischen dem Potenzial und den Aspekten 123
Tabelle 25 Übersicht zur Verteilung der höchsten Bildungsgrade in den beiden Darbietungsarten (in Präsenz vs. virtuell) 144

Tabelle 26 Mittelwertvergleich der demografischen Variablen vor und nach dem Matching 163
Tabelle 27 Ergebnisse der Überprüfung der Messinvarianz 165

Abbildungsverzeichnis

Abbildung 1 Verfahrens-Mediums-Matrix von Kersting und Ziegler (2020)..... 25
Abbildung 2 Darstellung der fünf Big 5 Dimensionen und der jeweils sechs zugehörigen Facetten nach Borkenau und Ostendorf (2008)..... 38
Abbildung 3 Die Struktur des Fünf-Faktoren Modells mit Hinzunahme der jeweiligen Aspekte nach DeYoung et al. (2007)..... 42
Abbildung 4 Modell des Mechanismus zwischen Persönlichkeitsmerkmalen und Berufserfolg nach Tett und Burnett (2003) 49
Abbildung 5 Die Regeln der Bewerbendenreaktionen nach Gilliland (1993), Abbildung und Übersetzung aus Böge (2016)..... 65
Abbildung 6 Modell der Bewerbendenreaktionen nach Ryan und Ployhart (2000) (eigene Darstellung). 66
Abbildung 7 Screeplot der ersten explorativen Faktorenanalyse mit 119 Items..... 104
Abbildung 8 Screeplot der explorativen Faktorenanalyse mit 53 Items 105
Abbildung 9 Darstellung der Hypothesen nach Tett und Burnett (2003) sowie Gilliland (1993) differenziert nach der virtuellen Darbietungsart und der Darbietungsart in Präsenz. 140
Abbildung 10 Darstellung des MotivSORTS..... 152
Abbildung 11 Aufbau der Legekarten des GerechtigkeitSORT in der Onlineumfrage 154
Abbildung 12 Vereinfachte Darstellung des Strukturgleichungsmodells zur Hypothesentestung..... 160
Abbildung 13 Jitter Plot des Propensity Score Matchings 163
Abbildung 14 Histogramme des Propensity Score Matchings..... 164
Abbildung 15 Verteilungen der Antworten zu der Frage: Was ist Ihnen in Auswahlverfahren wichtig? 168

Abkürzungsverzeichnis

α	Cronbachs Alpha
β	Regressionskoeffizient
λ	Ladung
χ^2	Chi-Quadrat-Wert
ABGS	arbeitsbezogenen Belastbarkeits- und Gewissenhaftigkeitsskalen
AEOS	arbeitsbezogenen Extraversions- und Offenheitsskalen
AIC	Akaike Information Criterion
AVS	arbeitsbezogenen Verträglichkeitsskalen
BIP	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung
BIP-6F	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – 6 Faktoren
bspw.	beispielsweise
bzw.	beziehungsweise
CFA	Konfirmatorische Faktorenanalyse
CFI	Comparative-Fit-Index
CI	Konfidenzintervall
d	Effektstärke
df	Freiheitsgrad
EFA	Explorative Faktorenanalyse
et al.	et alii (lateinisch für: und andere)
F	F-Statistik
ggf.	gegebenenfalls
H	Hypothese
IST-Screening	Intelligenz-Struktur-Test Screening
M	Mittelwert
MGKA	Mehrgruppenkausalanalyse
MI	Modification Indices
MLR	Maximum-Likelihood Robust
M_{rtt}	Mittelwert der Retest-Reliabilität
N	Anzahl
n	Anzahl Teilstichprobe
NEO-FFI	NEO-Fünf-Faktoren Inventar
NEO-PI-R	NEO Personality Inventory
n.s.	Nicht signifikant
OPQ32i	Occupational Personality Questionnaire 32
p	Signifikanzniveau mit folgenden Klassifikationen: $p < .05$*; $p < .01$**; $p < .001$***
P_i	Itemschwierigkeit
r	Regressionskoeffizient
R^2	Determinationskoeffizient (aufgeklärte Varianz)
r_{it}	Trennschärfe
r_{tt}	Retest-Reliabilität
RMSEA	Root-Mean-Square-Error-of-Approximation
s.	siehe
s^2	Fehlervarianz
SD	Standardabweichung
SPJS	Selection Procedure Justice Scale
SRMR	Standardized Root Mean Square Residual
t	t-Wert

TLI	Tucker-Lewis Index
z.B.	zum Beispiel

Zusammenfassung

Die Digitalisierung nimmt immer weiter Einzug in die Arbeitswelt und hat zur Folge, dass einzelne Kompetenzen (zum Beispiel Offenheit für Erfahrung) im Berufsleben immer wichtiger werden. Auch in der Eignungsdiagnostik nehmen digitale Methoden mehr Raum ein (Fellner, 2019; Schermuly et al., 2019). Aus diesem Grund prüfte diese Studie zum einen die Gütekriterien des neu entwickelten berufsbezogenen Fragebogens AEOS (Wedemeyer & Moldzio, in Vorbereitung) zu den vier Aspekten der Dimensionen Offenheit für Erfahrung sowie Extraversion nach DeYoung et al. (2007). Zum anderen fokussierte die Studie die Methode der virtuellen Dialogübung auf Grundlage der prognostischen Validität anhand des Job Performance Modell (Tett & Burnett, 2003) und der Gerechtigkeitswahrnehmung im Vergleich zur Darbietungsart vor Ort. In Studie 1 wurde davon ausgegangen, dass der Fragebogen reliabel und valide ist. Dies wurde mit $N = 1550$ Versuchspersonen in fünf beruflichen Stichproben untersucht. Die Studie zeigte, dass der Fragebogen interne Konsistenz zwischen $\alpha = .81$ und $\alpha = .86$ besitzt und reliabel ist. Die konfirmatorische Faktorenanalyse konnte die Struktur nach DeYoung et al. (2007) replizieren. Die Kriteriumsvalidität wurde teilweise bestätigt und es bestand inkrementelle Validität über andere, wissenschaftlich basierte Verfahren hinaus. Die Studie 2 wurde in einem experimentellen Workshop anhand von zwei Stichprobengruppen (virtuell vs. vor Ort) mit $n = 50$ Personen untersucht. Es wurde davon ausgegangen, dass Unterschiede zwischen den Gruppen in der prognostischen Validität und in dem Zusammenhang der Persönlichkeit (AEOS) und der Gerechtigkeitswahrnehmung existieren. Die Ergebnisse zeigten, dass sich die prognostische Validität in den beiden Darbietungsarten unterschied, jedoch war die Gerechtigkeitswahrnehmung vergleichbar.

Stichwörter: Validität, Eignungsdiagnostik, Big Five

Abstract

Digitalisation is part of our personal and professional life. As a result, individual competences (especially openness to experience) are becoming increasingly important in professional life. In aptitude testing, digital methods are also used more often (Fellner, 2019; Schermuly et al., 2019). For this reason, this study tested the criteria of good quality (i.e. validity, reliability) of the newly developed job-related questionnaire AEOS (Wedemeyer & Moldzio, in preparation) on the four aspects of the personality dimensions openness to experience and extraversion according to DeYoung et al. (2007). Furthermore, the study focused on the method of the dialogue exercise in a virtual setting. Using the Job Performance Model (Tett & Burnett, 2003) the prognostic validity and perception of justice were tested in comparison to an on-site situation. In Study 1, it was assumed that the questionnaire was reliable and valid. This was investigated with $N = 1550$ participants in five occupational samples. The study showed that the questionnaire had internal consistency between $\alpha = .81$ and $\alpha = .86$ and was reliable. The confirmatory factor analysis was able to replicate the structure of DeYoung et al. (2007). Criterion validity was partially confirmed. Likewise, incremental validity beyond other inventories was evidenced. In study 2, an experimental setting (workshop) with two samples (virtual vs. on-site group, $n = 50$ participants each) was used. It was assumed that there are differences between the samples regarding prognostic validity, regarding personality (AEOS) and justice perceptions. The results showed that prognostic validity differed in the two samples, but justice perceptions were comparable.

Keywords: Validity, Aptitude diagnostics, Big Five

Einleitung

Die ganze Arbeitswelt ist im Wandel. Digitalisierung und Agilität halten immer weiter Einzug und haben zur Folge, dass einzelne Kompetenzen im Berufsleben immer wichtiger werden (Schermuly et al., 2019). Schermuly et al. (2019) stellten im Zuge dessen mithilfe ihrer Forschung fest, dass die Big Five Dimension Offenheit für Erfahrung in der Arbeitswelt immer bedeutsamer wird. Diesen Wandel muss auch die Personalauswahl berücksichtigen. Es sollte überdacht werden, ob die aktuell erfragten Kompetenzen relevant für den zukünftigen beruflichen Erfolg der Bewerbenden sind und ob andere Kompetenzen in Auswahlverfahren überprüft werden sollten. Ein weiterer wichtiger Punkt ist, dass sich die Tätigkeiten zum Beispiel durch Projektarbeiten stetig ändern. Und so muss nicht nur ein:e Mitarbeiter:in anhand eines Anforderungsprofils für ein Projekt gefunden werden, sondern es muss regelmäßig überprüft werden, welches Projekt folgt und welche Kompetenzen bzw. eventuelle Fördermaßnahmen anschließend notwendig sind (Kersting, 2021). Aus diesem Grund gestaltet sich die Personalauswahl immer komplexer.

Neben der Digitalisierung und deren neuen Möglichkeiten in der Personalauswahl bzw. der Eignungsdiagnostik scheint eine immer größere Diskrepanz zwischen der Wirtschaft und der Wissenschaft zu bestehen (Kanning, 2022). So werden zunehmend nicht wissenschaftlich fundierte Methoden verwendet. Beispielsweise werden Persönlichkeitsfragebögen, welche eine gute Vorhersagekraft für den beruflichen Erfolg besitzen (Barrick & Mount, 1991), selten verwendet, da sie meist nicht berufsbezogen formuliert sind (Schuler et al., 2007). Um die Diskrepanz zwischen der Wissenschaft und Wirtschaft zu verringern sowie Persönlichkeitsfragebögen für die Wirtschaft attraktiver zu gestalten, wurde in der aufgeführten Arbeit ein berufsbezogener Persönlichkeitsfragebogen zur Offenheit für Erfahrung sowie Extraversion entwickelt und die Gütekriterien betrachtet. Wie oben beschrieben ist die Offenheit für Erfahrung durch die Digitalisierung zu einer wichtigen Kompetenz herangereift (Schermuly et al., 2019). Die Extraversion sagt vor allem im Vertrieb und bei Führungskräften den Berufserfolg voraus und sollte daher bei der Personalauswahl berücksichtigt werden (Barrick & Mount, 1991).

Ein weiterer Aspekt, der durch die Corona-Pandemie in den Jahren 2020 und 2021 in den Vordergrund gerückt ist, ist das Arbeiten im Homeoffice. In dieser Zeit bestand teilweise eine Homeoffice-Pflicht für Arbeitnehmende, die nicht zwangsläufig im Unternehmen vor Ort arbeiten mussten (Bundesregierung, 2022). Dies waren zum Beispiel Arbeitnehmende mit einer Bürotätigkeit, jedoch kaum Mitarbeitende aus der Produktion oder aus dem Gesundheitssystem. Diese Homeoffice-Pflicht verbunden mit strengen Hygienekonzepten

betrafen auch einen großen Teil der Mitarbeitende der Personalauswahl und Potenzialerkennung. Denn es mussten auch zu Pandemiezeiten Stellen besetzt oder Talente gefördert werden. Aus diesem Grund wurden diese Verfahren immer mehr virtuell durchgeführt. Es wurden nicht nur Vorstellungsgespräche per Videokonferenz geführt, sondern auch getestet, in wie weit Assessment Center digital genauso praktikabel sind, wie die ursprünglich vor Ort durchgeführten. Es existierten keinerlei Forschungsarbeiten zu der Fragestellung, ob virtuelle Assessment Center genauso valide sind wie die traditionell verwendeten Verfahren vor Ort (bspw. das Interview). Kersting (2021) beschrieb jedoch auch unabhängig von der Pandemie, dass virtuelle Simulationsaufgaben (zum Beispiel Präsentationen und Dialogübungen) im Rahmen der Digitalisierung öfter eingesetzt werden sollten. Simulationsaufgaben leben von der Ähnlichkeit der Übung zu der eigentlichen Tätigkeit. Die Tätigkeit findet durch die Digitalisierung zunehmend virtuell statt, beispielsweise Gespräche mit Kund:innen. Daher kann es sogar realistischer sein, dies anhand einer virtuellen Dialogübung per Videokonferenz darzustellen.

Diese Arbeit möchte die Frage nach der Validität von virtuellen Assessment Centern anhand von Simulationsübungen, konkret sogenannter Dialogübungen, beantworten. Dies ist nicht nur für den weiteren Verlauf der Pandemie bedeutend, sondern auch hinsichtlich des Fortschrittes der Globalisierung sowie des Klimaschutzes (zum Beispiel könnten Reisen immer teurer oder auch grundsätzlich in Frage gestellt werden). Nach der Abschaffung der Homeoffice-Pflicht fragen sich die Unternehmen, ob es überhaupt notwendig ist, in den „alten Modus“ zurückzukehren, oder ob das Homeoffice nicht sogar positive Faktoren bzw. Vorteile bietet. Diese Frage gilt auch für Verfahren der beruflichen Eignungsdiagnostik. Sollten die virtuellen Dialogübungen genauso valide sein, wie die vor Ort durchgeführten, dann könnten hohe Reisekosten sowie hoher Zeitaufwand gespart werden, indem nicht alle Beobachtende und Bewerbende an einen gemeinsamen Ort reisen müssten. Auf diese Art könnten vor allem wichtige Führungskräfte global agierender Unternehmen aus anderen Ländern oder Standorten den Auswahl- oder Förderprozess mit begleiten. Darüber hinaus würden die Unternehmen einen weiteren Beitrag zur Klimaneutralität leisten. Ein Vorteil der virtuellen Verfahren wäre zudem, dass die Stellenbesetzungen durch den geringeren Zeitaufwand beschleunigt werden kann (Fellner, 2019).

Theorieteil

Im nachfolgenden Abschnitt werden wichtige Begrifflichkeiten definiert sowie der aktuelle Forschungsstand erklärt. Ausgehend von der Klärung des Begriffs „*Eignungsdiagnostik*“ werden relevante Verfahren, Modelle und Kompetenzen

ausdifferenziert. Abschließend wird die Gerechtigkeitswahrnehmung der Bewerbenden in einzelnen eignungsdiagnostischen Verfahren betrachtet, da diese unter anderem maßgeblich die Entscheidungsfindung beeinflusst, ob ein:e Bewerber:in eine Stelle annimmt oder nicht (Gilliland, 1993).

Berufliche Eignungsdiagnostik

In diesem Abschnitt wird die berufliche Eignungsdiagnostik definiert sowie eine Klärung der Bestandteile beruflicher Eignungsdiagnostik herbeigeführt. Die berufliche Eignungsdiagnostik ist die Grundlage für die Personalauswahl sowie Potenzialerkennung und steht somit im Vordergrund dieser Arbeit (Ziegler & Bühner, 2012). Um die berufliche Eignungsdiagnostik definieren zu können, ist es zentral die Personalauswahl zu erklären, da die Eignungsdiagnostik ein Hilfsmittel für diese darstellt. Blickle (2014) beschreibt die Personalauswahl als Lösung für ein Zuordnungsproblem zwischen Personen und Arbeitsplätzen. Darunter ist zu verstehen, dass sich mehrere Personen auf eine vakante Stelle bewerben und nun die geeignetste Person für die vakante Stelle gefunden werden soll. In diesem Fall wird von der Passung der Person für die Stelle gesprochen. Dem gegenüber muss auch die Stelle zu der Person passen, damit nicht nach kürzester Zeit die Stelle erneut ausgeschrieben werden muss, da die zuvor ausgewählte Person aus verschiedenen Gründen (zum Beispiel durch falsche Vorstellungen der Tätigkeit) gekündigt hat (Blickle, 2014).

Es stellt sich daher die Frage, wie man die beste Passung feststellen kann. An dieser Stelle ist die psychologische Diagnostik bedeutsam, die Ziegler und Bühner (2012) als einen essentiellen Bestandteil der praktischen Tätigkeit von Psycholog:innen beschreiben. Diese Diagnostik wird zielgerichtet eingesetzt, um bestimmte Fragestellungen zu beantworten. In der Personalauswahl muss zum Beispiel die Frage nach der Eignung der bestpassendsten Person auf eine vakante Stelle beantwortet werden (Ziegler & Bühner, 2012). Zur Beantwortung dieser Fragestellungen verwendet die Diagnostik passende Methoden, wie zum Beispiel Tests, Fragebögen, Interviews und Verhaltensbeobachtungen.

Mittels der Eignungsdiagnostik sollen somit die Zusammenhänge zwischen Merkmalen einer Person und beruflichen Erfolgskriterien anhand von diagnostischen Methoden betrachtet werden (Schuler, 2000). Dabei können die Merkmale beispielsweise Fähigkeiten und Verhaltensweisen einer Person sein. Als Erfolgskriterium wird zum Beispiel der berufliche Erfolg anhand von Gehalt oder Beförderungen gemessen. Wichtig ist, dass die ausgewählte Diagnostik von den Bewerbenden als gerecht wahrgenommen wird, um Ablehnungen der angebotenen Stellen (Selbstselektion; Moldzio, 2014) oder gar Klagen

seitens der Bewerbenden gegen die Unternehmen zu vermeiden. Daher haben Westhoff et al. (2010) die DIN 33430 entwickelt, mithilfe derer man einen diagnostischen Prozess gerecht und standardisiert entwickeln und durchführen kann. Diese Norm beinhaltet Richtwerte für diagnostische Verfahren zur Objektivität, Reliabilität, Validität, Eichung und Verfälschbarkeit.

Neben den Prozessen der Personalauswahl bedienen sich auch die Potenzialerkennungsverfahren eignungsdiagnostischer Prozesse (Rehrl et al., 2006). Unternehmen nutzen die sogenannte Potenzialanalyse, um mittels der Diagnostik Stärken und Schwächen der Mitarbeitenden herauszufiltern und anschließende Entwicklungsmaßnahmen zu beschließen (Rehrl et al., 2006). Anhand der Analyse wird für die Unternehmen sichtbar, welche Mitarbeitenden sogenannte Potenzialträger:innen sind bzw. sein könnten und explizit auf beispielsweise Führungsrollen vorbereitet werden sollten. Die Eignungsdiagnostik in der Personalauswahl sowie in der Potenzialerkennung bedient sich der gleichen Methoden, welche nachfolgend erklärt werden.

Methoden der Eignungsdiagnostik

In diesem Abschnitt werden die wichtigsten diagnostischen Methoden der Eignungsdiagnostik erklärt. Diese können nach zwei bekannten Modellen gegliedert werden. Dies sind zum einen der Trimodale Ansatz von Schuler (2006) und zum anderen der CUBE-Ansatz von Kersting (2006). Diese Modelle verdeutlichen, dass die Passung der Personen zur vakanten Stelle aus verschiedenen Perspektiven betrachtet werden kann und die unterschiedlichen Perspektiven miteinander kombiniert werden sollten (Schuler, 2006). Im Nachgang werden die Möglichkeiten der Digitalisierung in der Eignungsdiagnostik beschrieben.

Der Trimodale Ansatz nach Schuler 2006 besagt, dass die diagnostischen Methoden in drei Gruppen geclustert werden können. Die erste der drei Gruppen ist die konstruktorientierte Ebene, welche Eigenschaften oder Kompetenzen der Bewerbenden erfasst. Es werden zum Beispiel mittels Intelligenztests oder Persönlichkeitsfragebögen stabile Merkmale der Bewerbenden erfasst, anhand derer man Schlüsse über das spätere Arbeitsverhalten ziehen kann. Die zweite Gruppe beinhaltet die sogenannten simulationsorientierten Verfahren, bei denen die Bewerbenden Aufgaben, welche der späteren Tätigkeit ähneln, bearbeiten müssen. Dabei wird von den im Verfahren gezeigten Verhaltensweisen auf zukünftiges Verhalten am Arbeitsplatz geschlossen. Zu diesen Aufgaben zählen zum Beispiel Konfliktgespräche/ Dialogübungen, Präsentationsübungen,

Postkorbübungen, Gruppendiskussionen und Fallstudien, welche oft Bestandteile von Assessment Centern sind. Die letzte Gruppe des Trimodalen Ansatzes sind die biografieorientierten Verfahren. Hier wird versucht, aufgrund von vergangenem Verhalten, zukünftiges Verhalten vorherzusagen. Dieser Teil beinhaltet zum Beispiel das Sichten der Bewerbungsunterlagen oder ein biografiegeleitetes Interview. Wenn Bewerbende sich beispielsweise ehrenamtlich engagieren, kann darauf geschlossen werden, dass sich diese Personen auch am Arbeitsplatz in höherem Maße einsetzen (Schuler, 2006). Diese drei Gruppen sollten im Sinne des Trimodalen Ansatzes in einem Auswahlverfahren kombiniert verwendet werden (Schuler, 2006).

Der zweite Ansatz ist der CUBE-Ansatz von Kersting (2006), der seinen Namen dem dreidimensionalen Aufbau zu verdanken hat. Die erste Seite des CUBEs besteht aus fünf Komponenten, welche die Datenquelle darstellen. Die erste Komponente besteht aus den biografischen Daten, welche wie bei Schuler (2001) anhand von Bewerbungsunterlagen erfasst werden können. Als zweites werden die Selbst- und Fremdberichte benannt, die anhand von Interviews und beispielsweise Führungskräfteeinschätzungen erfasst werden können. Die dritte und vierte Komponente bilden Fragebogendaten (zum Beispiel Persönlichkeitsfragebögen) und die Leistungs(test)daten (zum Beispiel Intelligenztests). Als letztes führt Kersting (2006) die Simulationsdaten auf, welche wie bei Schuler (2001) mithilfe von beispielsweise Konfliktgesprächen/ Dialogübungen, Gruppendiskussionen oder Präsentationsübungen erfasst werden können. Die zweite Seite des CUBEs differenziert, ob die Datenquellen verhaltensorientiert oder eigenschaftsorientiert erhoben wurden. So kann man beispielsweise, um Verhaltensweisen zu beurteilen, im Interview erfragen, wie die Bewerbenden eine gewisse Situation meistern würden. Andersherum können Bewerbende in einem Persönlichkeitsfragebogen Eigenschaften wie zum Beispiel die Gewissenhaftigkeit bewerten. Die letzte Ebene des CUBEs repräsentiert die Zeitdimension. Diese Zeitdimension teilte Kersting (2006) in die drei Zeitfenster Vergangenheit, Gegenwart und Zukunft auf, welche alle Bestandteile des Interviews sein können. Auf diese Art kann nach vergangenen Verhaltensmustern gefragt werden, jedoch auch danach, wie eine Person in bestimmten Situationen in der Zukunft handeln würde, falls sie die Stelle angeboten bekommt.

Nachdem die einzelnen diagnostischen Verfahren aufgeführt wurden, stellt sich die Frage, welche Methoden am geeignetsten für die prädiktive Validität in der Personalauswahl bzw. für die Vorhersage von der Arbeitsleistung sind. Diese Frage beantworteten Schmidt und Hunter im Jahr 1998, in dem sie eine Metaanalyse mit 85 Studien erstellten, in der sie die prädiktive Validität von 19 verschiedenen Auswahlmethoden in Bezug auf Arbeits- und

Ausbildungsleistung betrachteten. Diese metaanalytische Herangehensweise war wichtig, da zuvor durch unterschiedlich große Stichproben sowie diverse Kriterien viele verschiedene Ergebnisse in Bezug auf die Validität existierten. Die Metaanalyse schaffte durch die Zusammenfassung aller Stichproben und Ergebnisse einen genaueren Überblick über die Zusammenhänge und die Gültigkeit der Auswahlverfahren. Die Ergebnisse zeigen, dass der Intelligenztest mit einem Wert von $r = .51$ hinter der Arbeitsprobe ($r = .54$) neben dem strukturierten Einstellungsinterview ($r = .51$) eine der besten prädiktiven Validitäten in Bezug auf die Arbeitsleistung bietet. Der Vorteil der Intelligenztests scheint jedoch im Vergleich zu anderen Verfahren darin zu bestehen, dass sie ein ökonomisches und geringe Kosten verursachendes Verfahren sind. Persönlichkeitsfragebögen (hier Integritäts-Gewissenhaftigkeitstests) werden mit einem $r = .31$ im hinteren Mittelfeld gelistet, wohingegen das Alter, Graphologie, Interessen sowie Berufserfahrung kaum einen Prädiktor für die Arbeitsleistung zu sein scheinen. Zusätzlich zu der prädiktiven Validität betrachtete diese Studie die inkrementelle Validität (Schmidt & Hunter, 1998). Diese thematisierte die Frage, inwiefern die prädiktive Validität erhöht wird, wenn Auswahlverfahren kombiniert werden, in diesem Fall in Kombination mit Intelligenztests als dem validesten Verfahren. Dabei wird ersichtlich, dass die Persönlichkeitsfragebögen zwar keine große prädiktive Validität aufweisen, jedoch in Kombination mit den Intelligenztests die Vorhersagekraft für die Arbeitsleistung um 18% erhöhen und somit die viert beste Kombination aufweisen. Arbeitsproben, Integritätstests sowie strukturierte Interviews erhöhen die prädiktive Validität um bis zu 27%, sind jedoch auch zeit- und kostenintensiver. Bei der Betrachtung der Validität in Bezug auf die Arbeitsleistung in Trainingsprogrammen erhöht der Persönlichkeitsfragebogen die Validität sogar um 16% und nimmt damit hinter dem Integritätstest den zweiten Platz ein. Die Ergebnisse der Studie von Schmidt und Hunter (1998) verdeutlichen, dass neben der prädiktiven Validität auch die inkrementelle Validität betrachtet werden sollte. Darüber hinaus muss sich immer die Frage gestellt werden, ob der Kosten- und Zeitaufwand im Verhältnis zur Validitätssteigerung steht (Schmidt und Hunter, 1998). Da diese Ergebnisse mittlerweile über 20 Jahre alt sind und sich die Personalauswahl durch die Digitalisierung verändert hat, sollten diese Ergebnisse in einer neuen Forschungsarbeit aktuell betrachtet und neu bewertet werden. Einige neuere Studien betrachteten einzelne Bestandteile beispielsweise die Persönlichkeit in Bezug auf die Arbeitsleistung (z.B. Judge et al., 2013), aber eine erneuerte Übersichtsarbeit über alle eignungsdiagnostischen Verfahren wäre interessant. Dieses Problems nahmen sich Sackett et al. im Jahr 2022 an und erstellten eine Metaanalyse angelehnt an die Schmidt und Hunter

Studie (1998). Die Ergebnisse verdeutlichten, dass die prädiktive Validität in allen Verfahren im Vergleich zu Schmidt und Hunter (1998) deutlich sank. So besaß beispielsweise die Arbeitsprobe zum Kriterium nur noch eine Korrelation von r = .33 anstelle von r = .54. Dieser Abfall der Korrelationen ist zum einen durch eine konservativere Vorgehensweise mit anderen statistischen Korrekturen zu erklären. Des Weiteren betrachtete diese Studie über die ebenfalls von Schmidt und Hunter (1998) verwendeten Studien hinaus neuere Studien. Aufgrund der konservativen Vorgehensweise von Sackett et al. (2022) sollten die Ergebnisse von Schmidt und Hunter (1998) nicht komplett vernachlässigt werden. Auch in der Reihenfolge der Verfahren auf Basis der Validität änderte sich einiges. Die Verfahren mit der größten Validität waren in der Studie von Sackett et al. (2022) die strukturierten Interviews (r = .42) und Arbeitswissenstests (r = .40). Assessment Center (r = .29) und kognitive Fähigkeitstests (r = .31) lagen in Bezug auf die Validität im mittleren Bereich. Die Gewissenhaftigkeitsfragebögen lagen mit einer Korrelation von r = .19 im hinteren Mittelfeld. Schmidt und Hunter (1998) boten einen großen Mehrwert für weiterführende Forschungsarbeiten durch die Erfassung der inkrementellen Validität. Sackett et al. (2022) überprüften jedoch die inkrementelle Validität in ihrer Studie nicht. Dies sollte nachfolgend von Forschungsarbeiten untersucht werden, um der Praxis Hinweise liefern zu können, welche Kombination aus Verfahren am validesten ist.

Digitalisierung in der Eignungsdiagnostik. Im gesamten gesellschaftlichen, politischen wie wirtschaftlichen Leben hält die Digitalisierung Einzug. Dieser Abschnitt hat zum Ziel, die Möglichkeiten, welche durch die Digitalisierung geschaffen wurden, aufzuzeigen, jedoch auch Grenzen zu berichten. So kann beispielsweise die Dauer des Auswahlprozesses reduziert werden, jedoch müssen auch vermehrt Datenschutzrichtlinien beachtet werden (Fellner, 2019). Auch in der Personalauswahl bzw. Eignungsdiagnostik haben sich die Prozesse immer mehr digitalisiert und vereinfacht, beispielsweise durch Apps (Kersting & Ziegler, 2020). Darüber hinaus hat die Corona-Pandemie in den Jahren 2020 und 2021 der Forschung digitaler Diagnostikmöglichkeiten einen weiteren Aufschwung bereitet (z.B. Basch & Melchers, 2020), da die Personalauswahl trotz Distanzgebot weitergeführt werden musste. Lange Zeit fand die Personalauswahl persönlich, mittels paper-pencil Tests und höchstens per Telefon statt (Schuler et al., 2007). Nun gibt es seit einigen Jahren die Möglichkeit, diese Test- und Fragebogenverfahren auf Computern oder mobilen Endgeräten zu präsentieren (Geister & Rastetter, 2009). Dadurch werden nicht nur vorhandene Fragebögen oder Testverfahren internetbasiert präsentiert, sondern auch neue Verfahren, wie

beispielsweise das Self-Assessment, immer beliebter (Geister & Rastetter, 2009). Unter Self-Assessments versteht man Online-Tests, welche Aufgaben beinhalten, die Anforderungen eines gewissen Unternehmens entsprechen (Geister & Rastetter, 2009). Diese Selbsttests können potenzielle Bewerbende bearbeiten und herausfinden, ob sie für eine Bewerbung geeignet wären und das Unternehmen ihnen potentiell zusagen würde. Das Unternehmen kann auf diese Art die Anzahl von ungeeigneten Bewerbungen minimieren und so auch Zeit und Aufwand einsparen. Neben dem Self-Assessment bieten die herkömmlichen Test- und Fragebogenverfahren nach wie vor eine große Chance in der Eignungsdiagnostik (Fellner, 2019) und sollten nicht nur webbasiert in Form eines sogenannten Online-Assessments (Konradt & Sarges, 2003) in der Vorauswahl durchgeführt werden. So können beispielsweise Antwortzeiten Aufschluss darüber geben, ob eventuell bei der Bearbeitung von Rechenaufgaben ein Taschenrechner als Hilfsmittel verwendet wurde und die Rechenfähigkeit im nächsten Prozessschritt vor Ort wiederholt abgefragt werden sollte (Fellner, 2019; Hertel et al., 2003). Zudem können auch Stimm- und Sprachanalysen erstellt werden. Diese sind allerdings noch nicht weitreichend erforscht und bieten große datenschutzrechtliche Herausforderungen (Fellner, 2019). Neben der Möglichkeit, Antwortzeiten sowie Stimm- und Sprachanalysen zu verwenden, stellt die Nutzung von Online-Assessments weitere Vorteile dar. So kann nicht nur diese Art von Vorauswahl zeit- und ortsunabhängig stattfinden und somit die Flexibilität der Bewerbenden sowie des Unternehmens erhöhen. Durch die standardisierte onlinebasierte Auswertung, welche nicht mehr einzeln per Hand durchgeführt wird, kann der Prozess beschleunigt werden (Konradt & Sarges, 2003).

Nicht nur der direkte Auswahlprozess kann onlinebasiert stattfinden, sondern auch die Recrutierung, welche E-Recruitment genannt wird (Konradt & Sarges, 2003). So verlagerte sich schon die Stellensuche von beispielsweise Annoncen in der Zeitung auf das Internet oder das Intranet, welches zumeist die internen Homepages eines Unternehmens sind (Konradt & Sarges, 2003). Diese Art des E-Recruitments stellt nach Konradt & Sarges (2003) Vorteile für die Bewerbenden und die Unternehmen dar. So können die Bewerbenden zu jeder Zeit den Status der Bewerbung abrufen und bleiben auf dem Laufenden. Darüber hinaus können diese zum Beispiel durch Chatfunktionen vorab mit dem Unternehmen in den Kontakt treten. Die Vorteile des Unternehmens liegen darin, einen innovativen Eindruck bei den Bewerbenden zu hinterlassen und es bietet die Möglichkeit, die Informationstiefe durch beispielsweise aufgezeigte Entwicklungschancen oder das Arbeitsklima zu vergrößern. So könnte sich das Unternehmen ggf. ein besseres Image erarbeiten.

Doch die Digitalisierung hat nicht nur Auswirkungen auf die Bewerbenden, sondern bietet auch den Beobachtenden in einem Assessment Center Vereinfachung. Es gibt viele Apps, welche die Mitschriften und Bewertungen der Beobachtenden abspeichern und für die Konferenz der Beobachtenden am Ende des Verfahrens automatisch grafisch aufbereiten (Kersting, 2021).

Bei der Benutzung von virtuellen Verfahren kann die Reichhaltigkeit und Synchronität der verwendeten Medien für den Erfolg der Eignungsdiagnostik entscheidend sein und daher werden verschiedene Theorien in den nächsten Abschnitten erläutert (Kersting & Ziegler, 2020). So können beispielsweise aufgezeichnete Interviews durch die Beobachtenden mehrfach betrachtet werden, jedoch können auch Verfahren (zum Beispiel die Gruppendiskussion) für die virtuelle Durchführung zu komplex sein (Kersting & Ziegler, 2020).

Media Richness Theory. Nicht alle Medien sind für jede eignungsdiagnostische Methode geeignet, da gewisse Methoden zu reichhaltig sind, um diese beispielsweise durch eine Eingabe in einem Chatfeld zu erfassen (Kersting & Ziegler, 2020). Daher muss, bevor auf die Möglichkeiten von Personalauswahlverfahren eingegangen wird, zu Beginn erst einmal die Media Richness Theory von Daft und Lengel (1986) erklärt werden, da es verschiedene Medien mit unterschiedlicher Reichhaltigkeit gibt, welche in der Eignungsdiagnostik berücksichtigt werden sollten. Daft und Lengel (1986) gehen davon aus, dass eine bestimmte Aufgabe nur erfüllt werden kann, wenn das passende Medium dafür ausgewählt wird, da die Medien unterschiedlich reichhaltig sind und deshalb die Informationen unterschiedlich vielfältig übermittelt werden können. Aus diesem Grund empfehlen die Forscher vor der Nutzung eines Mediums die Anforderungen der Situation mit den Eigenschaften des Kommunikationsmediums abzugleichen, um die Situation bestmöglich zu meistern (Kersting & Ziegler, 2020). Überträgt man dies auf die Eignungsdiagnostik, sollte für die Simulation eines Konfliktgespräches nicht ein Medium verwendet werden, welches ausschließlich Textnachrichten übermitteln kann. So würden viele Informationen, wie zum Beispiel Emotionen und non-verbale Kommunikation, verloren gehen.

Die Reichhaltigkeit eines Mediums wird anhand von vier Kriterien definiert (Daft & Lengel, 1986). Das erste Kriterium ist die Vielfalt der Sprache, welche beispielsweise angibt, ob ein Medium per Audio, Video oder Text Informationen austauschen kann. Darüber hinaus ist die Anzahl von Möglichkeiten (Kriterium 2), in welcher Art Informationen übermittelt werden können, ein entscheidender Aspekt. Aus diesem Grund kann es bedeutend für das Verständnis sein, ob der Tonfall verändert werden und somit beispielsweise Ironie erkannt

werden kann oder nicht. Die letzten beiden Kriterien bilden der Grad der Personalisierung sowie die Schnelligkeit der Rückmeldung. Daraus folgend kann eine Person nicht über jedes Medium eine persönliche Note mit einfließen lassen und es kann zur Verzögerung des Informationsaustausches kommen, beispielsweise bei einem Telefonat ohne ausreichend gute Verbindung.

Kersting und Ziegler (2020) haben diese Theorie auf die Mediumsmöglichkeiten der Eignungsdiagnostik übertragen und ein sogenanntes Ranking der Reichhaltigkeit erstellt. Als „armes“ Medium bezeichnen sie eine non-verbale Eingabe, welche bei der Test- und Fragebogenbearbeitung dargeboten wird. In diesem Fall wird nur anhand von gesetzten Kreuzen oder anderen Antwortmöglichkeiten kommuniziert. Da alle Bewerbenden den gleichen Fragebogen bearbeiten müssen, kann keine persönliche Note mit eingebracht werden. Das „zweitärmste“ Medium ist der Chat. Dieser ist reichhaltiger als die non-verbalen Eingaben, da dort auch eine Audiofunktion besteht sowie ein Freitext geschrieben werden kann und somit die Kommunikation etwas persönlicher wird. Zudem findet in einem Chat eine Interaktion mit einer anderen Person oder einem Computer statt. Das Telefon ist das nächst reichhaltige Medium, welches oft in der Vorauswahl für ein Telefoninterview verwendet wird. Bei diesem Medium wird schneller eine Rückmeldung erhalten als bei einem Chat, jedoch ist dort auch nur die Audiofunktion verfügbar. Als die beiden reichhaltigsten Medien bezeichnen Kersting und Ziegler (2020) das Video sowie die Face-to-Face Interaktion. Da im Video meist nur ein Ausschnitt des Körpers sichtbar ist und die Personen sich nicht gegenüberstehen, können vereinzelt Informationen unbemerkt/ unbeachtet bleiben. Daher ist die Face-to-Face Interaktion das reichhaltigste Medium in der Eignungsdiagnostik. Die aktuelle Eignungsdiagnostik zeigt (Kersting & Ziegler, 2020), dass „ärmere“ Medien für die Test- und Fragebogenverfahren ausreichend sind, jedoch ein guter Mix aus allen Medien in einem Personalauswahlprozess vorhanden sein sollte.

***Media Synchronicity Theory*.** Im vorherigen Abschnitt wurde sichtbar, dass die Eignungsdiagnostik mit verschieden reichhaltigen Medien arbeiten kann. Ein Kriterium für den Grad der Reichhaltigkeit war die Schnelligkeit der Rückmeldung. Aufbauend auf diesem Aspekt entwickelten Dennis und Valacich (1999) die Media Synchonicity Theory, welche die Synchronität der Medien beleuchtet. Diese Theorie zeigt auf, welche Vorteile die Digitalisierung bieten kann, wenn die Kommunikation nicht synchron vor Ort stattfinden muss und stellt die Grundlage für die Verwendung von digitalen Auswahlverfahren dar (Kersting, 2021). Die Synchronität ist das Ausmaß, in dem Individuen zur selben Zeit an der

gleichen Aktivität arbeiten und einen gemeinsamen Fokus auf bestimmte Tätigkeiten legen (Dennis und Valacich, 1999). Die Synchronität bezieht sich in der Eignungsdiagnostik auf die Schnelligkeit der Rückmeldung sowie auf die Frage, inwieweit die Bewerbenden und die Beobachtenden gleichzeitig agieren. Die einzelnen diagnostischen Verfahren können synchron, das bedeutet gleichzeitig, oder asynchron (zeitversetzt) ablaufen. Die Sichtung der Bewerbungsunterlagen findet ausschließlich ohne die Bewerbenden statt und ist somit asynchron. Die Recruiter:innen betrachten die Unterlagen orts- und zeitunabhängig und geben den Bewerbenden nachfolgend ein Feedback. Test- und Fragebogenverfahren können sowohl synchron als auch asynchron stattfinden. In einigen Unternehmen werden Test- und Fragebogenverfahren vor Ort mit Testleitenden durchgeführt und finden somit synchron statt. Die Test- und Fragebögen können aber ebenfalls asynchron von den Bewerbenden zuhause online bearbeitet werden. Telefoninterviews und Vorstellungsgespräche laufen fast immer synchron ab (Kersting, 2021). Videointerviews können sowohl synchron als auch asynchron stattfinden. Zum einen kann ein Interview via Videokonferenz geführt werden und somit synchron ablaufen. Dank der Digitalisierung ist es ebenfalls möglich, dass die Bewerbenden ein Video, in dem sie gewisse Fragen beantworten sowie sich selber vorstellen, aufnehmen und auf eine Plattform hochladen. Diese Videos können Beurteilende des Unternehmens zu einer beliebigen Zeit bewerten und daher ist diese Art des Videointerviews asynchron.

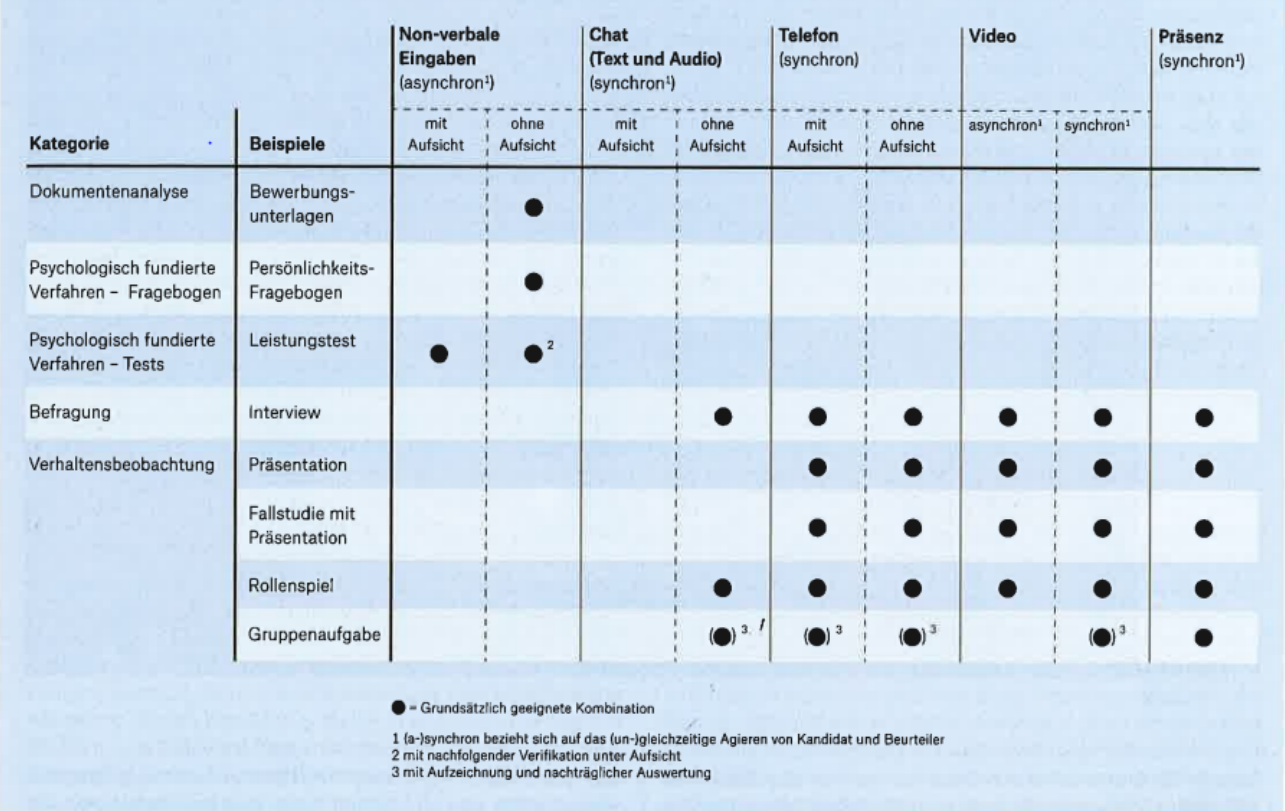
Synchrone Verfahren, beispielsweise Interviews, können auch gleichzeitig asynchron sein, wenn diese aufgezeichnet werden. Dies bringt den Vorteil mit sich, dass die Beobachtenden im Nachgang das Interview wiederholt betrachten und somit Beobachtungsfehler minimiert werden können (Kersting & Ziegler, 2020). Diese Variante birgt jedoch datenschutztechnische Schwierigkeiten, da die Einwilligung an ein Stellenangebot gekoppelt ist und daher die Bewerbenden keine wirkliche Chance haben, dieser Aufzeichnung zu widersprechen (Kersting & Ziegler, 2020).

Verfahrens-Mediums-Matrix. Angelehnt an die beiden Theorien zur Reichhaltigkeit und Synchronizität haben Kersting und Ziegler (2020) eine Verfahrens-Mediums-Matrix erstellt (s. Abbildung 1), welche die Eignung der Verfahrenskonstellationen durchleuchtet und die beiden Theorien kombiniert. Diese Theorie zeigte nicht nur die Möglichkeiten der Verwendung von verschiedenen reichhaltigen und synchronen Verfahren, sondern vor allem auch Grenzen in der Praxis auf. So zeigte die Theorie, dass die Verwendung einer Gruppenübung aufgrund ihrer Komplexität nur in Präsenz stattfinden sollte.

In dieser Theorie wurde zusätzlich der Aspekt Sicherheit mit einbezogen, welcher das Mitschneiden von Inhalten und die Verfälschbarkeit von beispielsweise Testverfahren berücksichtigt. Dies ist entscheidend, da durch den Mitschnitt Assessmentaufgaben eines Unternehmens veröffentlicht werden könnten und somit spätere Bewerbende durch Trainings die Bearbeitung verfälschen könnten (Kersting & Ziegler, 2020).

Abbildung 1

Verfahrens-Mediums-Matrix von Kersting und Ziegler (2020)



Kategorie	Beispiele	Non-verbale Eingaben (asynchron[1]) mit Aufsicht	Non-verbale Eingaben (asynchron[1]) ohne Aufsicht	Chat (Text und Audio) (synchron[1]) mit Aufsicht	Chat (Text und Audio) (synchron[1]) ohne Aufsicht	Telefon (synchron) mit Aufsicht	Telefon (synchron) ohne Aufsicht	Video asynchron[1]	Video synchron[1]	Präsenz (synchron[1])
Dokumentenanalyse	Bewerbungs-unterlagen		●							
Psychologisch fundierte Verfahren – Fragebogen	Persönlichkeits-Fragebogen		●							
Psychologisch fundierte Verfahren – Tests	Leistungstest	●	●[2]							
Befragung	Interview				●	●	●	●	●	●
Verhaltensbeobachtung	Präsentation					●	●	●	●	●
	Fallstudie mit Präsentation					●	●	●	●	●
	Rollenspiel				●	●	●	●	●	●
	Gruppenaufgabe				(●)[3]	(●)[3]	(●)[3]		(●)[3]	●

● = Grundsätzlich geeignete Kombination

1 (a-)synchron bezieht sich auf das (un-)gleichzeitige Agieren von Kandidat und Beurteiler
2 mit nachfolgender Verifikation unter Aufsicht
3 mit Aufzeichnung und nachträglicher Auswertung

Die Matrix der Theorie ist so aufgebaut, dass die Verfahren in fünf Kategorien gegliedert werden (Kersting & Ziegler, 2020). Zudem wird die Reichhaltigkeit in die oben genannten fünf Kategorien (non-verbale Eingaben, Chat, Telefon, Video und Präsenz) aufgeteilt und zusätzlich in mit und ohne Aufsicht sowie synchron und asynchron unterteilt. Kersting und Ziegler (2020) empfehlen, die Dokumentenanalyse bzw. die Sichtung der Bewerbungsunterlagen sowie Persönlichkeitsfragebögen, welche zu der Kategorie psychologisch fundierte Fragebögen gezählt werden, asynchron und ohne Aufsicht zu erheben. Psychologisch fundierte Tests (zum Beispiel Leistungstests) können sowohl mit als auch ohne Aufsicht durchgeführt werden (s. Abbildung 1). Bei einer Durchführung ohne Aufsicht ist es bei auffälligen Ergebnissen ratsam, diese in einem weiteren Gespräch zu hinterfragen oder den Bewerbenden einen weiteren ähnlichen Test zu präsentieren (Kersting

& Ziegler, 2020). Alle drei Kategorien (Dokumentenanalyse, Persönlichkeitsfragebögen und Leistungstests) werden im Bereich der Reichhaltigkeit zu den non-verbalen Eingaben gezählt. Befragungen, wie beispielsweise ein Interview, können per Chat, Telefon, Video und in Präsenz durchgeführt werden. Der Chat und das Telefon können ohne Aufsicht und das Video sowohl synchron als auch asynchron stattfinden. Verhaltensübungen wie Präsentationen, Fallstudien oder Rollenspiele können per Telefon, Video und in Präsenz unter sämtlichen Bedingungen stattfinden. Lediglich das Rollenspiel ist auch in einem Chat ohne Aufsicht möglich. Gruppenübungen zählen zu der Kategorie der Verhaltensbeobachtungen. Sie müssen jedoch gesondert betrachtet werden, da sie durch die Anzahl an Personen sehr komplex sind. Lediglich in Präsenz können diese ohne Einschränkungen durchgeführt werden. Weiterhin sind Rollenspiele im Chat ohne Aufsicht, im Telefonat unter allen Bedingungen und in einem synchronen Video möglich. Dies muss dann jedoch aufgezeichnet und nachträglich ausgewertet werden, da die Beobachtungen sehr aufwändig sind.

Chancen und Einschränkungen. In den vorherigen Abschnitten wurden anhand von einzelnen Theorien verschiedene Möglichkeiten der Digitalisierung in der Eignungsdiagnostik präsentiert, jedoch auch Grenzen aufgezeigt. So müssen beispielsweise in einer Gruppenübung einige Bedingungen abgeklärt werden, damit diese virtuell gut funktionieren kann (Kersting & Ziegler, 2020). Doch unabhängig von der Machbarkeit der Digitalisierung in der Eignungsdiagnostik stellt sich die Frage, welchen Mehrwert die digitalen Methoden bieten und welche Risiken beachtet werden müssen, damit eine virtuelle Personalauswahl funktioniert. Aus diesem Grund werden in diesem Abschnitt die Vor- und Nachteile der digitalen Eignungsdiagnostik nebeneinandergestellt, um Handlungsempfehlungen herauszuarbeiten.

Viele Autor:innen sind der Meinung, dass die Digitalisierung in der Personalauswahl große Chancen bietet, jedoch auch einige Risiken birgt, wenn die Diagnostik nicht gut durchgeführt wird (z.B. Fellner, 2019; Geister & Rastetter, 2009; Ott et al., 2017). Beispielsweise sehen Geister und Rastetter (2009) einen großen Vorteil in der zeitlichen und räumlichen Flexibilität. Die Bewerbenden bekommen zum Beispiel einen Link zu einem Testverfahren per Mail zugeschickt und können diesen zu ihrer gewünschten Zeit ortsunabhängig bearbeiten. Darüber hinaus können Kosten und Aufwand bei einem Assessment Center gespart werden, wenn das Assessment Center virtuell per Videokonferenz stattfindet und nicht alle Bewerbenden an einen gemeinsamen Ort reisen müssen (Fellner, 2019). Da die Auswertung direkt automatisch in dem dargebotenen Programm ablaufen kann,

könnte zusätzlich Zeit eingespart, der Bewerbungsprozess verkürzt und durch die standardisierte Auswertung die Objektivität erhöht werden (Hertel et al., 2003). Kirbach et al. (2004) gehen sogar davon aus, dass durch diese Standardisierung ein kompletter Bewerbungsprozess auf drei Wochen zeitlich reduziert werden kann und somit das Risiko, dass geeignete Bewerberenden zwischenzeitlich eine andere Stelle antreten, minimiert wird. Fellner (2019) zählt auch die reduzierte Fehleranfälligkeit und die erhöhte Effizienz der Standardisierung zu den großen Vorteilen, warnt jedoch davor, die Erfolgsfaktoren (zum Beispiel die Objektivität und Validität) nicht aus den Augen zu verlieren. Ebenso sollte bei wachsender Anzahl an virtuellen Angeboten (zum Beispiel für Online-Assessments) von verschiedensten Anbietenden jedes Verfahren vor der Nutzung kritisch beleuchtet werden.

Geister und Rastetter (2009) sehen neben den Vorteilen von onlinebasierten diagnostischen Methoden auch große Nachteile, die zuvor abgeklärt werden müssen. Beispielsweise muss sichergestellt werden, dass das verwendete Programm keine speziellen Anforderungen an die Soft- oder Hardware aufweist, damit die Bewerbenden an einem üblichen Gerät die Testungen bearbeiten können. Darüber hinaus sollte zu Beginn der Umstellung von der analogen in die onlinebasierte Darbietungsart darauf geachtet werden, wie kostenintensiv die Programmierung ist. Da die Bewerbenden nicht zwangsläufig die Testungen unter Aufsicht bearbeiten, sollte das Unternehmen überlegen, ob es zum Beispiel in einem weiteren Bewerbungsprozess die kognitiven Kompetenzen nochmals überprüft, um sicher zu stellen, dass die Person die Testung selbstständig durchgeführt hat. Der gravierendste Nachteil ist die Reduktion des persönlichen Kontaktes, zum Beispiel vor dem Hintergrund des Personalmarketings. Durch die onlinebasierte Vorauswahl findet der persönliche Kontakt erst zu einem späteren Zeitpunkt oder lediglich virtuell statt.

Hertel et al. (2003) zeigten auf, dass der Einsatz verschiedener computergestützter Verfahren möglich wäre, jedoch die Validität beispielsweise bei virtuellen Postkorbaufgaben noch nicht ausreichend erforscht ist. Sie zeigten des Weiteren auf, dass Leistungstests online schneller bearbeitet werden können und somit die Normierung der paper-pencil Verfahren nicht verwendet werden kann, da Leistungstests zumeist Speedtests sind.

Damit die Auswahlverfahren digital funktionieren, müssen die Unternehmen einige Gegebenheiten beachten (Hertel et al., 2003). So sollte die Zugänglichkeit eines Computers für jede:n Bewerber:in gleich sein und somit könnte eine Lösung darin bestehen, dass sich die Bewerbenden in einer regionalen Niederlassung einfinden und dort digitale Endgeräte zur Verfügung gestellt werden. Die Verwendung von regionalen Niederlassungen würde auch verhindern, dass sich die Bewerbenden bei der Bearbeitung Hilfe holen und somit die

Ergebnisse gezielt verfälschen. Es wird empfohlen, dass allen Bewerbenden Übungsmöglichkeiten gegeben werden, damit ein Trainingseffekt nicht die Ergebnisse verzerrt, weil vereinzelt Bewerbenden das Verfahren schon einmal durchlaufen haben (Hertel et al., 2003). Hertel et al. (2003) führten mit der Selektivität einen weiteren Problempunkt der digitalen Verfahren an. Es könnte sein, dass gewisse Bewerbenden bestimmte Medien der Verfahren bevorzugen und somit ggf. aus dem Bewerbungsprozess ausscheiden. So können beispielsweise Bewerbenden ihre Bewerbung zurückziehen, da sie sich bei asynchronen Interviews nicht wohlfühlen.

Dieser Abschnitt hat gezeigt, dass bei der Auswahl der geeigneten Darbietungsart von eignungsdiagnostischen Methoden viele Faktoren zu berücksichtigen sind, damit die Personalauswahl oder Potenzialerkennung allen Bewerbenden die gleichen Chancen liefert und nicht zur Selektivität aufgrund einer geringeren Akzeptanz bzw. Medienpräferenzen führt (Hertel et al., 2003). In den nachfolgenden Abschnitten werden die einzelnen Methoden differenzierter erklärt. Darüber hinaus wird auf die Einsatzhäufigkeiten der Methoden eingegangen, auch im Kontext zu der Darbietungsart.

Wandel der Eignungsdiagnostik

Die vorherigen Abschnitte haben gezeigt, dass die Digitalisierung viele Risiken und Chancen in der Eignungsdiagnostik darstellt. Es stellt sich die Frage, wie diese onlinebasierten Methoden in der Praxis angenommen werden und was einzelne Forschungsprojekte beispielsweise über die Reaktionen der Bewerbenden zu den neuartigen Methoden und den verschiedenen Darbietungsarten aussagen. Bevor die neuen Methoden betrachtet werden, steht zunächst der Wandel der Eignungsdiagnostik in den letzten 30 Jahre im Fokus, um zu zeigen, welche Methoden in der Wirtschaft bisher präferiert wurden. Dies könnte auch Aufschlüsse über die zukünftige Nutzung von onlinebasierten Methoden geben und Hinweise liefern, welche Methoden von den Unternehmen eher kritisch gesehen werden könnten. Dies könnte der Wissenschaft Hinweise liefern, welche Forschungsarbeiten praxisrelevant sein könnten. So zeigten Hossiep et al. (2015) auf, dass nicht valide Persönlichkeitsfragebögen den wissenschaftlich stark erforschten Fragebögen von Unternehmen vorgezogen werden. Dies könnte für die Wissenschaft ein Zeichen sein, dass die wissenschaftlich stark erforschten Fragebögen nicht die praxisrelevanten Merkmale erfassen, oder die Wissenschaft den Mehrwert dieser Fragebögen den Unternehmen näherbringen muss.

Schuler et al. haben 2007 zum dritten Mal 125 deutsche Unternehmen danach befragt, welche Personalauswahlverfahren sie mit welcher Häufigkeit in der Führungskräfteauswahl verwenden. Die anderen beiden Erhebungszeitpunkte waren 1985 (Schulz et al., 1985) und 1993 (Schuler et al., 1993). Die Umfragen beschränkten sich nicht nur auf große Unternehmen einer gewissen Branche, sondern die Forschung schaffte einen Weitblick über viele Branchen und Firmengrößen hinweg. Sowohl im Jahr 1993 als auch im Jahr 2007 wurden am häufigsten Bewerbungsunterlagen (bis zu 99%) analysiert und Einstellungsinterviews durchgeführt (bis zu 81%). Jedoch änderte sich die Art der Interviews. Sowohl die Fachabteilungen als auch die Personalabteilungen führten 2007 vermehrt strukturierte Interviews und nicht wie zuvor unstrukturierte Interviews durch. Im Vergleich zum Jahr 1993 ist das strukturierte Telefon-Interview bei 32% der Befragten zum festen Bestandteil geworden und hat einen kleinen Beitrag zur Digitalisierung geleistet. Assessment Center haben einen starken Aufwind erlebt. Ihr Anteil ist um 18,6% auf 57,6% innerhalb von 14 Jahren gestiegen. Dabei ist auffällig, dass Gruppendiskussionen als eignungsdiagnostische Methode in einem Assessment Center um 8,6% abgenommen haben (von 51,0% auf 42,4%). Wenn sich dieser Trend fortführen sollte, wäre dies für den Einsatz des onlinebasierten Assessment Centers hilfreich, da Kersting und Ziegler (2020) Schwierigkeiten in der Übertragung von Gruppenübungen in die virtuelle Darbietungsart sehen. Darüber hinaus ist auffällig, dass Persönlichkeitsfragebögen sowie Testverfahren in 20% bis 40,8% der Fälle eingesetzt werden. Vor allem beim Betrachten der Digitalisierung erkennt man, dass lediglich 1,6% der Fragebogenverfahren und tatsächlich keinerlei Testverfahren online stattfanden. Der geringe Einsatz von Persönlichkeitsfragebögen sowie Testverfahren zeigt einen Widerspruch zu der Studie von Schmidt und Hunter (1998), welche eine höhere prädiktive Validität von Arbeitsleistung im Vergleich zu demografischen Daten fanden, die in Form von Bewerbungsunterlagen bei fast allen Unternehmen betrachtet wurden.

Hossiep et al. (2015) griffen die Untersuchungen von Schuler et al. (2007) auf und betrachteten den Einsatz von Persönlichkeitsfragebögen differenzierter. Sie erkannten bei der Befragung von 120 großen Unternehmen in Deutschland, dass zwei Drittel der Befragten Persönlichkeitsfragebögen einsetzen. Zumeist wurden allerdings Typentests eingesetzt, welche den Bewerbenden die Möglichkeit geben, aus nur zwei Antworten zu wählen. Dadurch kann beispielsweise eine Wahl zwischen „Ich bin introvertiert“ und „Ich bin extravertiert“ getroffen werden. Diese Verfahren werden in Bezug auf die Aussagekraft in der Wissenschaft als fraglich eingeordnet (Hossiep et al., 2015). Demgegenüber wurde beispielsweise der wissenschaftlich fundierte und stark erforschte NEO-FFI (Borkenau &

Ostendorf, 2008) bei lediglich etwa 5% der Verfahren verwendet. Dies zeigt eine große Diskrepanz zwischen Wissenschaft und Praxis in Bezug auf die Verwendung von standardisierten und validierten Persönlichkeitsfragebögen auf.

Armoneit (2019) führte die Trendstudie von Schuler et al. (2007) weiter und befragte wiederholt deutsche Unternehmen nach der Einsatzhäufigkeit von Personalauswahlverfahren. Die Analyse zeigte, dass online-basierte Verfahren zunehmend genutzt wurden. So war beispielsweise die Zahl der händischen Analysen von Bewerbungsunterlagen in Papierform rückläufig, während die Online-Analysen von Bewerbungsunterlagen stark zunahmen und bei 75% lagen. Zudem vollzog sich ein Wandel in den Interviewformen. Die Einsatzhäufigkeit des Telefoninterviews nahm weiter zu und die neue Variante eines Videointerviews wurde zwischen 8,6% (unstrukturiert) und 12,1% (strukturiert) eingesetzt. Armoneit (2019) erkannte genauso wie Hossiep et al. (2015) einen Anstieg der Einsatzhäufigkeiten von Persönlichkeitsfragebögen und Testverfahren. Auch in diesen beiden Verfahren ist ein klarer digitaler Trend zu erkennen. So wurden beispielsweise laut Armoneit (2019) nur noch in 19,3% der Verfahren paper-pencil Persönlichkeitsfragebögen verwendet, dahingegen 23,6 % online gestützte Fragebogenverfahren. In der Studie wird deutlich, dass weitestgehend die üblichen Verfahren in eine online-basierte Form übersetzt und weniger neue Verfahren onlinebasiert entwickelt wurden. So wurden zum Beispiel Online-Self-Assessments im Jahr 2018 in nur 2,9 % der Personalauswahlverfahren eingesetzt. Dies bedeutet im Vergleich zur Studie von Schuler et al. (2007) sogar einen Negativtrend von 1,1%.

In diesem Abschnitt wurde deutlich, dass die Unternehmen im Einsatz digitaler Verfahren in der Eignungsdiagnostik zurückhaltend sind und Online-Assessments sowie Videointerviews sehr selten (maximal 12% der Unternehmen) einsetzen. Wenn sie diese jedoch einsetzen, dann bevorzugen sie häufig die einfache Übersetzung der Eignungsdiagnostik in onlinebasierte Varianten wie beispielsweise bei Persönlichkeitsfragebögen.

Verschiedene Forschungsarbeiten haben sich der Frage gewidmet, ob man diese Verfahren bedenkenlos in die virtuelle Darbietungsart verlagern kann (z.B. Brenner et al., 2016; Konradt & Sarges, 2003; Petri et al., 2019).

Chuah et al. (2006) widmeten sich der Frage, ob Fragebögen, die am Computer oder im Internet bearbeitet werden, auch die gleiche Aussagekraft und Validität aufweisen, wie die traditionellen paper-pencil Fragebögen. Dies haben sie anhand eines Fragebogens zu den Big Five Persönlichkeitsmerkmalen untersucht. Die Studie zeigte, dass es vereinzelte signifikante Unterschiede in den Messeigenschaften zwischen den Darbietungsarten gab. Dabei lag die

Effektstärke zwischen d = -.20 und d = .19, was nach Cohen (1998) als gering zu bewerten ist. Diese Studie bestärkt das Vorgehen der Wirtschaft, dass immer mehr Persönlichkeitsfragebögen online eingesetzt werden, da die Messeigenschaften der verschiedenen Darbietungsarten vergleichbar zu sein scheinen. Diese Erkenntnis wird auch durch die Forschung von Meade et al. (2007) gestützt, welche ebenso herausfanden, dass die Qualität online-basierter Persönlichkeitsfragebögen mit der von paper-pencil Fragebögen vergleichbar ist. So war beispielsweise die Reliabilität in der Gewissenhaftigkeitsskala in allen Darbietungsarten nicht signifikant verschieden (α_{online} = .84 & $\alpha_{paper\text{-}pencil}$ = .83). Jedoch empfehlen die Autor:innen, dass vor der ersten Durchführung die Vergleichbarkeit kontrolliert werden sollte. Joubert und Kriek testeten 2009 diese Fragestellung anhand eines anderen Persönlichkeitsfragebogens (Occupational Personality Questionnaire 32; OPQ32i) in zwei verschiedenen Studien und kamen nach der Analyse des Strukturgleichungsmodells auch zu der Erkenntnis, dass die Persönlichkeitsfragebögen unabhängig von der Darbietungsform valide sind. Die Effektstärke der Skalenunterschiede in der ersten Studie lagen zwischen d = -.02 und d = .57 im sehr geringen bis mittleren Bereich, jedoch scheinen die größeren Effektstärken durch Stichprobenunterschiede begründet zu sein. In der zweiten Studie war die Effektstärke mit d = .01 bis d = .41 geringer. In beiden Studien waren die Reliabilitäten der Skalen in beiden Darbietungsarten vergleichbar (Studie 1: α_{online} = .74 & $\alpha_{paper\text{-}pencil}$ = .72; Studie 2: α_{online} = .76 & $\alpha_{paper\text{-}pencil}$ = .75). Anhand dieser drei Studien ist sichtbar, dass bei Persönlichkeitsfragebögen von einer Replizierbarkeit der Ergebnisse von paper-pencil auf eine online-basierte Version auszugehen ist. Aus diesem Grund wird nachfolgend bzw. in der ersten Studie nicht explizit auf die Darbietungsart eingegangen.

Die Übersichtsarbeit von Hertel et al. (2003) zeigte dem gegenüber, dass die Leistungstests, welche Speedtests sind, nicht einfach in die onlinebasierte Version übersetzt werden können. Es stellte sich heraus, dass die Bearbeitung online zumeist schneller geht und somit die Normwerte der paper-pencil Version nicht verwendet werden können.

Petri et al. (2019) untersuchten die Frage, ob unbeaufsichtigte Onlinetestungen in gleichem Maße von den Bewerbenden akzeptiert werden wie beaufsichtigte Eignungstests. Gesichtspunkte dieser Studie waren die Kontrollierbarkeit, die Belastungsfreiheit (zum Beispiel fühlen sich die Bewerbenden überfordert), die Augenscheinvalidität sowie die Messqualität. Die Testung war Bestandteil eines Bewerbungsverfahrens für Ausbildungsberufe. Diese Studie hat gezeigt, dass die Auszubildendenstichprobe die Kontrollierbarkeit und Belastungsfreiheit der unbeaufsichtigten Testung positiver bewertet (Kontrollierbarkeit: $M_{beaufsichtigt}$ = 5.17 & $M_{unbeaufsichtigt}$ = 5.31; Belastungsfreiheit: $M_{beaufsichtigt}$

= 4.46 & $M_{unbeaufsichtigt}$ = 5.16). Darüber hinaus fiel die Gesamtbewertung der Akzeptanz bei unbeaufsichtigten Testungen positiver aus ($M_{beaufsichtigt}$ = 4.57 & $M_{unbeaufsichtigt}$ = 4.79). Lawrence et al. (2009) untersuchten in ihrer Studie die Wahrnehmung von Bewerbenden in Abhängigkeit von dem Ort, an dem sie die Testung durchgeführt haben. Sie fanden heraus, dass die Personen, die eine Störung während der Testung hatten, beispielsweise durch eine schlechte Internetverbindung, Lärm oder andere anwesende Personen, den Test negativer wahrnahmen als Personen, welche die Testung störungsfrei und in Ruhe Zuhause bearbeiteten (F = 5.134, p < .01). Diese beiden Studien zeigen, dass Onlinetestungen positiv wahrgenommen werden, jedoch ein Hinweis zur ruhigen Raumatmosphäre und eine gesicherte Internetverbindung für die Wahrnehmung der Bewerbenden hilfreich wäre, damit diese Faktoren die Wahrnehmung der Bewerbenden nicht beeinflussen (Lawrence et al., 2009; Petri et al., 2019).

Brenner et al. betrachteten 2016 asynchrone Videointerviews und befassten sich mit den Reaktionen von Bewerbenden auf diese neue Technologie. Die Versuchspersonen beantworteten neben dem asynchronen Videointerview auch einen Fragebogen zur Selbstwirksamkeit sowie einen Big Five Persönlichkeitsfragebogen, damit Zusammenhänge zwischen der Akzeptanz und der Persönlichkeit untersucht werden konnten. Im Anschluss an das Interview wurden noch einmal Fragebögen zur Benutzerfreundlichkeit, Nützlichkeit, wahrgenommenen eigenen Leistung, Ernsthaftigkeit des Gesprächs sowie der Akzeptanz bzw. Verfahrensgerechtigkeit beantwortet. Die Studie zeigt, dass die Nützlichkeit (β = .62, p < .001) und die Benutzerfreundlichkeit (β = .22, p < .05) einen signifikanten Einfluss auf die Einstellung gegenüber der neuen Technologie aufweisen und somit auf eine einfache Handhabung der neuen Technologie geachtet werden sollte (R^2 = .59, $F_{(12,93)}$ = 11.01, p < .001; Steigerung der Varianzaufklärung um 34% durch Nützlichkeit und Benutzerfreundlichkeit). Ein weiterer Aspekt war, dass die Big Five Dimension Offenheit für Erfahrung einen moderierenden Effekt zwischen der wahrgenommenen Nützlichkeit und der Einstellung zu dem Videointerview hat (β = .15, p < .005). So konnte gezeigt werden, dass gewisse Persönlichkeitsmerkmale bei der Akzeptanz von neuen Technologien eine Rolle spielen könnten. In einem weiteren Artikel führten Basch und Melchers (2020) die aktuelle Forschung zu online-basierten Einstellungsinterviews fort. Im Vordergrund standen Telefon- und Videointerviews. Auffällig schien die schwächere Bewertung der Bewerbenden in einer online-basierten Form im Vergleich zu Face-to-Face zu sein. Der Grund für dieses Ungleichgewicht ist jedoch noch nicht erforscht. Darüber hinaus scheinen die Bewerbenden den Telefon- und Videointerviews kritischer gegenüber zu stehen als Interviews vor Ort.

Diese Studien legen nahe, dass es notwendig ist, die Forschung zu digitalen Personalauswahlverfahren zu vertiefen sowie die Verfahren weiterzuentwickeln. Darüber hinaus wird deutlich, dass bisher primär onlinebasierte Test- und Fragebogenverfahren und Einstellungsinterviews in der Forschung berücksichtigt wurden. Simulationsaufgaben, bei denen die Interaktion zwischen zwei Menschen im Vordergrund steht, wurden dagegen nicht betrachtet. Diese Lücke gilt es zu schließen, da Simulationsaufgaben ein wichtiger Bestandteil von Assessment Centern sind, welche das Unternehmen sehr viel Aufwand und Geld kosten. Somit können bei einer validen Simulationsaufgabe Reiseaufwand und Kosten für Räumlichkeiten vor Ort eingespart werden (Fellner, 2019). Ein weiterer Aspekt ist der Zusammenhang zwischen der neuen Technologie und Persönlichkeitsmerkmalen, wie die Forschung von Brenner et al. (2016) sichtbar machte. Weitere Forschung kann helfen, Zusammenhänge zu erklären und Implikationen für die Praxis abzuleiten.

Persönlichkeit in der Eignungsdiagnostik

Die Persönlichkeit ist neben der fachlichen Eignung ein wichtiger Bestandteil der Eignungsdiagnostik (Barrick & Mount, 1991). Die vorherigen Kapitel haben nicht nur gezeigt, dass die Persönlichkeitsfragebögen in Unternehmen immer mehr eingesetzt werden, sondern auch, dass sich die relevanten Persönlichkeitsausprägungen in Zeiten der Digitalisierung und der Agilität ändern (Schermuly et al. 2019). Darüber hinaus wurde ersichtlich, dass vor allem in der Verwendung valider Persönlichkeitsfragebögen eine Diskrepanz zwischen Wissenschaft und Wirtschaft besteht (Hossiep et al., 2015). Aus den angeführten Gründen legt diese Arbeit ihren Schwerpunkt auf die Entwicklung und Validierung eines neuen Persönlichkeitsfragebogens. Um jedoch genauer auf den neuen Fragebogen eingehen zu können, müssen zunächst die Persönlichkeit bzw. gewisse Persönlichkeitsmodelle erklärt werden, auf die diese Arbeit aufbaut. Dies ist das Ziel der nächsten Abschnitte. Hermann (1976) definierte die Persönlichkeit als ein relativ stabiles Verhaltenskorrelat, welches bei jedem Menschen einzigartig ausgeprägt ist. Das bedeutet, dass jeder Mensch eine spezielle Kombination aus verschiedenen Merkmalen besitzt, die das Verhalten der Person beeinflussen.

Persönlichkeitsmodelle

Hermann (1976) definierte die Persönlichkeit als eine Kombination aus verschiedenen Merkmalen. Andere Forschungsarbeiten betrachteten in unterschiedlichsten Persönlichkeitsmodellen, auf welche Art und Weise diese Merkmale ggf. zusammengefasst werden können und somit detailliertere Beschreibungen der Persönlichkeit zu erhalten (z.B.

Cattell, 1946; Costa & McCrae, 1992; DeYoung et al., 2007). Aus diesem Grund hat dieser Abschnitt zum Ziel, die Entstehung und Weiterentwicklung dieser Modelle aufzuzeigen und somit die Persönlichkeit besser sowie detaillierter zu verstehen.

Wenn Persönlichkeitsmodelle betrachtet werden, dann kann festgestellt werden, dass Persönlichkeit aus verschieden Blickwinkeln angesehen wird (z.B. Schmithüsen & Krampen, 2015). So betrachtet beispielsweise der sozial-kognitive und handlungstheoretische Ansatz die Persönlichkeitsmerkmale im Zusammenhang mit Lern- und Handlungsmodellen (Schmithüsen & Krampen, 2015). Ein weiterer Ansatz, welcher bei dieser Forschung im Vordergrund steht, ist der „*Eigenschaftstheoretische Ansatz der Persönlichkeit*“, welcher sich mit der umfassenden Persönlichkeitsbeschreibung des Menschen befasst (Schmithüsen & Krampen, 2015). 1936 beschäftigten sich Allport zusammen mit Odbert erstmals mit diesem Ansatz, indem sie 18.000 Wörter sammelten, welche verschiedene Persönlichkeitsausprägungen beschrieben. Dies wird auch der lexikalische Ansatz genannt und bildete die Grundlage für die weitere Forschung. Allport entwickelte diesen Ansatz, da er glaubte, dass Persönlichkeitsneigungen einer Person in verschiedenen Situationen dadurch zu erklären sind, dass jede Person gewisse Persönlichkeitswesenszüge besitzt, die mit dem Nervensystem verknüpft sind (Allport & Allport, 1921). Cattell (1946) hat diesen lexikalischen Ansatz aufgegriffen und weitergeführt. Er betrachtete diese Wörter anhand einer Clusteranalyse und fand 35 verschiedene Variablen, welche er Surface traits („*Oberflächeneigenschaften*“) nannte. Im Nachgang untersuchte er diese 35 Cluster anhand einer Faktoranalyse und fand 16 Persönlichkeitsmerkmale. Darüber hinaus bestimmte er fünf Faktoren zweiter Ordnung.

Eysenck (1967) führte die Forschung von Cattell (1946) fort und betrachtete ebenfalls mithilfe einer Faktorenanalyse die Persönlichkeitsmerkmale. Er fand zunächst zwei übergeordnete Faktoren (Extraversion und Neurotizismus) und fügte im Anschluss noch Psychotizismus hinzu. Dieser Ansatz wurde als PEN-Modell bezeichnet. Darüber hinaus wurde deutlich, dass viele Persönlichkeitsbeschreibungen statistische Zusammenhänge aufweisen, somit zusammen auftreten und gemeinsam betrachtet werden sollten. Demgegenüber gibt es auch viele Wortpaare, welche sich gegenüberstehen, wie beispielsweise optimistisch und pessimistisch. Eysenck war der einzige Forscher, welcher den eigenschaftsorientierten Ansatz mit einem biopsychologischen Ansatz verknüpfte. Er vermutete, dass die Eigenschaft Extraversion mit dem aufsteigenden retikulären Aktivierungssystem (ARAS) des Nervensystems zusammenhängt, da die entgegengesetzte Eigenschaft Introversion im ARAS empfindlicher reagieren soll. Diese Behauptung über die

Verknüpfung der Persönlichkeit mit dem Nervensystem würde dafürsprechen, dass die Persönlichkeit angeboren ist, jedoch von der Umwelt geformt wird.

Big Five Dimensionen. Nach jahrelanger Forschung und aufbauend auf den obengenannten Persönlichkeitsmodellen hat sich das Fünf-Faktoren Modell, auch Big Five genannt, in der Wissenschaft und Wirtschaft etabliert. Diese Big Five Dimensionen wurden in vielen Studien hinsichtlich verschiedener Kriterien (zum Beispiel Arbeitsleistung oder Gerechtigkeitswahrnehmung) untersucht und sind somit einer der am stärksten erforschte Persönlichkeitsansätze (z.B. Barrick & Mount, 1991; Hausknecht et al. 2014). Das Big Five Modell entwickelten Costa und McCrae (1992) im englischsprachigen Raum. Sie erfassten in ihrem NEO Personality Inventory (NEO-PI-R) insgesamt 30 Submerkmale dieser fünf Dimensionen (s. Abbildung 2). Borkenau und Ostendorf (2008), welche mit dem NEO-Fünf-Faktoren Inventar (kurz NEO-FFI) eine Kurzform des NEO-PI-R entwickelten, etablierten dieses Modell im deutschsprachigen Raum. Dieses gekürzte Inventar mit 60 anstatt 240 Items schafft es, einen groben aber vollständigen Persönlichkeitsunterschied zwischen Personen zu erfassen. Das Modell besagt, dass fünf Persönlichkeitsdimensionen existieren, welche jeweils sechs untergeordnete Facetten besitzen, anhand derer jede Persönlichkeit beschrieben werden kann. Die Big Five Dimensionen lauten Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit. Wie jedes andere weit verbreitete Modell ist auch das Big Five Modell umstritten und wird kritisiert. In der Kritik geht es zumeist um die Faktorenstruktur. So wurden zum Beispiel im HEXACO-Modell sechs Faktoren nachgewiesen (z.B. Ashton et al., 2002). Jedoch ist festzuhalten, dass in der Entstehungsgeschichte des Modells diverse Faktorenanalysen zum gleichen Ergebnis dieser fünf Dimensionen kamen, auch wenn die Bezeichnungen variierten (Costa & McCrae, 1992; Goldberg, 1990).

Die Dimension Neurotizismus beschreibt die Fähigkeit, emotional stabil zu handeln und nicht bei Stress aus dem Gleichgewicht zu geraten (Ostendorf & Angleitner, 2004). Menschen mit einer hohen Ausprägung lassen sich beispielsweise als hilflos, unsicher, launisch oder nervös beschreiben. Der positiv umgepolte Begriff zu Neurotizismus ist Belastbarkeit oder Ausgeglichenheit (Moldzio et al., 2019) bzw. emotionale Stabilität (DeYoung et al., 2007). Wenn im Folgenden die Begriffe Belastbarkeit bzw. Ausgeglichenheit verwendet werden, dann ist immer die positive Bezeichnung von der Dimension Neurotizismus gemeint. Personen, die eine niedrige Ausprägung im Neurotizismus aufweisen, sind demnach gefestigt, ruhig und robust (Ostendorf & Angleitner,

2004). Die erste Facette beschreibt die Ängstlichkeit, welche betrachtet, inwiefern eine mit vielen negativen Gedanken behaftete Person dadurch unruhig agiert (Ostendorf & Angleitner, 2004). Diese Facette hinterfragt keine spezifischen Ängste und soll auch keine Angststörung thematisieren. Die Facette Reizbarkeit misst den Grad, inwieweit eine Person Ärger erleben und ausleben kann, beispielsweise, wenn diese Person frustriert ist (Ostendorf & Angleitner, 2004). Eine Person mit einer hohen Ausprägung in dieser Facette wird als empfindlich, missmutig und hitzig beschrieben. Die dritte Facette ist die Depression, welche die Gegensätze von Bedrücktheit und Niedergeschlagenheit bis hin zum Optimismus sowie zur Unbekümmertheit erfasst (Ostendorf & Angleitner, 2004). Eine hohe Ausprägung in der Facette Befangenheit besagt, dass eine Person beispielsweise schüchtern oder verlegen auftritt, wohingegen eine Person mit einer niedrigen Ausprägung selbstsicher und unbefangen agieren soll (Ostendorf & Angleitner, 2004). Die fünfte und vorletzte Facette von Neurotizismus ist die Impulsivität, welche die Fähigkeit beschreibt, Verlangen zu kontrollieren und beispielsweise dem Essenswunsch zu widerstehen (Ostendorf & Angleitner, 2004). Die Facette Verletzlichkeit beschreibt die Fähigkeit Stress zu bewältigen. Personen mit einer hohen Ausprägung werden eher als sensibel, stressanfällig sowie verletzlich beschrieben (Ostendorf & Angleitner, 2004).

Die Dimension Extraversion beschreibt den Grad der Geselligkeit und Gesprächigkeit einer Person (Ostendorf & Angleitner, 2004). Extravertierte Personen bewegen sich selbstbewusst in großen Gruppen und neigen eher zum Optimismus. Introvertierte Personen werden eher als ruhig und zurückhaltend beschrieben. Die erste Facette der Extraversion ist die Herzlichkeit. Diese Facette beschreibt die zwischenmenschlichen Beziehungen, das heißt inwieweit eine Person eher förmlich und reserviert oder eher freundlich und umgänglich agiert (Ostendorf & Angleitner, 2004). Personen mit einer hohen Ausprägung in der Facette Geselligkeit werden als gesprächig und kontaktfreudig beschrieben. Personen mit einer niedrigen Ausprägung sind eher Einzelgänger, die nicht gerne unter viele Menschen gehen. Die dritte Facette der Extraversion ist die Durchsetzungsfähigkeit, welche den Grad der Dominanz und sozialen Überlegenheit misst (Ostendorf & Angleitner, 2004). Aktivität ist eine weitere Facette der Extraversion und beschreibt das Bedürfnis, beschäftigt zu sein. Personen mit einer niedrigen Ausprägung werden als passiv und gemütlich bezeichnet. Abenteuerfreude und Risikobereitschaft sind Eigenschaften, welche in der Facette Erlebnissuche abgefragt werden. Personen mit einer niedrigen Ausprägung werden als bedächtig und vorsichtig handelnd beschrieben. Positive Energie ist die letzte Facette der

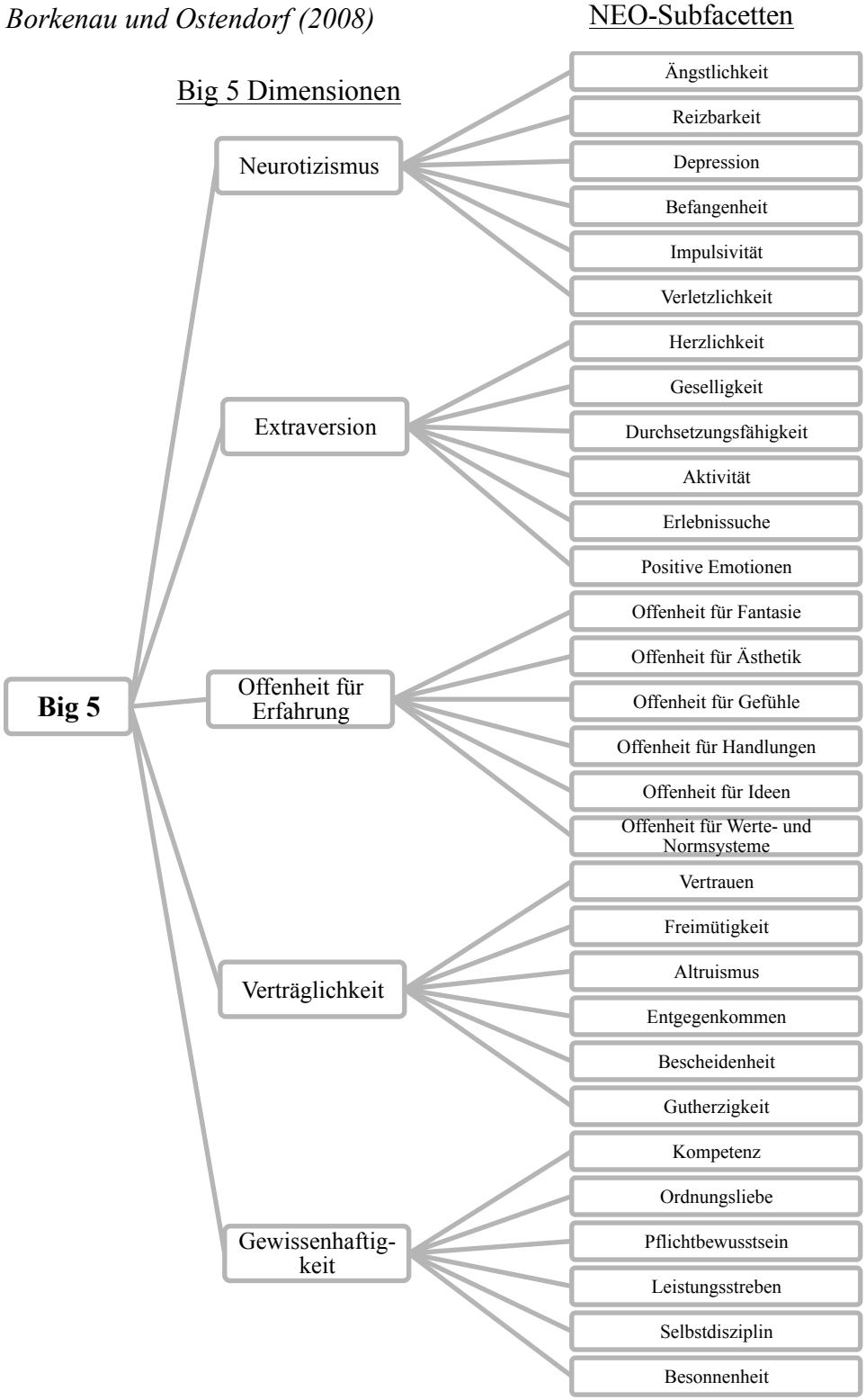
Extraversion und beschreibt das Ausmaß der Begeisterungsfähigkeit und Ausgelassenheit (Ostendorf & Angleitner, 2004).

Als dritte Dimension wird die Offenheit für Erfahrung genannt, welche Lust auf neue Erfahrungen, Erlebnisse sowie Eindrücke schildert (Ostendorf & Angleitner, 2004). Die Offenheit bezieht sich in den einzelnen Facetten auf unterschiedliche Aspekte. So beschreibt die Offenheit für Fantasien das lebhafte Vorstellungsvermögen oder die Tagträumerei. In der zweiten Facette wird die Offenheit gegenüber der Ästhetik, beispielsweise der Musik, Poesie oder Kunst hinterfragt (Ostendorf & Angleitner, 2004). Inwiefern eine Person empfänglich für die eigenen Gefühle ist und diese zulassen kann, wird in der Facette Offenheit für Gefühle gemessen. Personen mit einer niedrigen Ausprägung in der Facette Offenheit für Handlungen fällt es schwer, neue Aktivitäten oder Speisen auszuprobieren und sie ziehen Altbewährtes diesen vor. Die Facette Offenheit für Ideen beschreibt das Ausmaß der Wissbegierde und Lernbereitschaft von Personen (Ostendorf & Angleitner, 2004). Die Offenheit für Normen- und Wertesysteme beschreibt die Fähigkeit, jegliche Art von Werten, egal ob soziale, religiöse oder politische Werte, zu hinterfragen, jedoch auch Autoritäten zu respektieren.

Die vorletzte Dimension wird Verträglichkeit genannt und thematisiert interpersonelle soziale Beziehungen (Ostendorf & Angleitner, 2004). Sie misst die Frage, inwieweit eine Person hilfsbereit und gutmütig handelt, jedoch auch egoistisch sein kann. Die Facette Vertrauen beschreibt die Grundannahme, dass das Handeln von Personen durch gute Absichten bestimmt ist (Ostendorf & Angleitner, 2004). Daher werden Personen mit einer hohen Ausprägung als vertrauensvoll und gutgläubig beschrieben. Personen, die anderen schmeicheln und diese durch Täuschungen beeinflussen, besitzen eher eine niedrige Ausprägung in der Facette Freimütigkeit. Personen mit einer hohen Ausprägung werden auch als gradlinig und offenherzig beschrieben. Die Facette Altruismus beschreibt die Fähigkeit, sich um andere Personen zu sorgen und Rücksicht auf andere zu nehmen (Ostendorf & Angleitner, 2004). Die Frage, wie Personen in zwischenmenschlichen Konflikten umgehen, beantwortet die Facette Entgegenkommen. Personen mit einer hohen Ausprägung geben in einem Konflikt eher nach und vergeben anderen Personen schneller. Die Facette Bescheidenheit beschreibt die Zurückhaltung einer Person, welche jedoch nicht auf einem geringen Selbstwertgefühl oder Selbstbewusstsein beruht (Ostendorf & Angleitner, 2004). Die letzte Facette der Verträglichkeit ist Gutherzigkeit. Personen mit einer hohen Ausprägung werden als gutmütig und warmherzig beschrieben.

Abbildung 2

Darstellung der fünf Big 5 Dimensionen und der jeweils sechs zugehörigen Facetten nach Borkenau und Ostendorf (2008)



Die letzte Dimension ist die Gewissenhaftigkeit, welche die Zielstrebigkeit und Entschlossenheit einer Person thematisiert (Ostendorf & Angleitner, 2004). Die erste Facette Kompetenz beschreibt die Endscheidungsfähigkeit und Leistungsfähigkeit einer Person,

jedoch auch die Fähigkeiten, die eigenen Fertigkeiten einschätzen zu können (Ostendorf & Angleitner, 2004). Personen mit einer niedrigen Ausprägung in der Facette Ordnungsliebe werden als unsystematisch und unorganisiert arbeitend beschrieben. Als Pflichtbewusstsein wird die Facette beschrieben, gewissenhaft, zuverlässig und sorgfältig zu handeln. Als arbeitsscheu, faul sowie ziellos können Personen mit einer niedrigen Ausprägung in der Facette Leistungsstreben beschrieben werden. Ausdauer und Beharrlichkeit werden in der Facette Selbstdisziplin gemessen (Ostendorf & Angleitner, 2004). Personen mit einer niedrigen Ausprägung schreiben sich die Eigenschaften Sprunghaftigkeit und Undiszipliniertheit zu. Die Facette Besonnenheit thematisiert die Reflexion der eigenen Handlungen, jedoch auch die Art der Entscheidungsfindung, ob wohlüberlegt oder auch spontan.

Big Five Aspekte. Die Forschung von Costa und McCrae im Jahr 1992 zeigt, dass das Fünf-Faktoren Modell aus zwei Ebenen besteht. Übergeordnet stehen fünf Faktoren, die jedoch jeweils sechs untergeordnete Facetten besitzen. Nachfolgend haben viele Wissenschaftler:innen die hierarchische Struktur der Big Five hinterfragt. Jang et al. (2002) untersuchten Kovarianzen zwischen den einzelnen Facetten und fanden anhand von Zwillingsstudien in Deutschland und Kanada heraus, dass diese Kovarianzen aufgrund von genetischen und umweltbedingten Faktoren zustande kommen. Diese Studie nahmen DeYoung et al. (2007) zum Anlass, um zu überprüfen, ob zwischen den Dimensionen und den Facetten eine weitere Ebene existiert. Sie fanden pro Dimension zwei Aspekte, welchen die Facetten eine Hierarchieebene zugeordnet werden können. Judge et al. (2013) betrachteten anhand einer Metaanalyse die Vorhersagbarkeit der Aspekte und Dimensionen von Berufserfolg gemessen an 410 verschiedenen Stichproben, bei denen die Korrelationen zwischen den Big Five Dimensionen und Arbeitsleistungen (Generelle Arbeitsleistung, Aufgabenleistung, kontextuelle Leistung) untersucht wurden. Es zeigen sich anhand der konfirmatorischen Faktoranalysen, dass die jeweils sechs Facetten signifikant auf den jeweiligen zwei Aspekten nach DeYoung et al. (2007) luden und die Gesamtstärke der Ladungen betrug $\lambda = .65$. Darüber hinaus luden die zwei Aspekte jeweils auf der zugehörigen Dimension. Die Betrachtung der Vorhersage von Berufserfolg zeigte, dass in 13 von 15 Fällen (3 Kriterien x 5 Dimensionen) die Facetten die meiste Varianz aufklärten, die Aspekte die zweitmeiste (ausgenommen bei zwei Kriterien die meiste) und die Dimensionen die geringste Varianz aufklärten (Facetten: $R^2 = .24 - .03$; Aspekte: $R^2 = .10 - .01$; Dimensionen: $R^2 = .10 - .001$). Diese Studie stützte somit die Erkenntnisse von DeYoung et al. (2007) und

wies darauf hin, dass die Aspekte und Facetten wichtig für die Vorhersage von Berufserfolg sind und nicht nur die Dimension im Allgemeinen betrachtet werden sollen. Die Struktur der zwischengeschalteten Aspekte zeigt Abbildung 3.

DeYoung et al. (2007) nannten die Aspekte der Dimension Neurotizismus „*volatility*“ und „*withdrawal*“. „*Volatility*“ kann als Unbeständigkeit ins Deutsche übersetzt werden und hinterfragt, inwiefern negative Emotionen in Form von Wut und Impulsivität ausgedrückt werden. Angelehnt an diese Beschreibung sind die beiden Facetten Reizbarkeit und Impulsivität von Costa und McCrae (1992) diesem Aspekt zugeordnet. „*Withdrawal*“ (auf Deutsch Sensitivität) beinhaltet die restlichen vier Facetten Ängstlichkeit, Depression, Befangenheit und Verletzlichkeit. Dieser Aspekt erfasst die Empfänglichkeit und Zulässigkeit von negativen Emotionen. Moldzio et al. (2019) nutzten in der Entwicklung eines berufsbezogenen Persönlichkeitsfragebogen zu den Aspekten der Gewissenhaftigkeit sowie des Neurotizismus zwei andere Begriffe für diese Aspekte. Da sie das positive Pendant Belastbarkeit betrachteten, entstanden die Begriffe Soziale Belastbarkeit und Dauerbelastbarkeit. Soziale Belastbarkeit beinhaltet den Aspekt, wie belastbar eine Person in sozial anspruchsvollen Situationen, wie beispielsweise bei einer Präsentation ist, und ob sie eventuell nervös wird. Demgegenüber beschreibt der Aspekt Dauerbelastbarkeit die Fähigkeit, bei langandauernden Belastungen, wie zum Beispiel bei vielen Überstunden oder langer körperlich anspruchsvoller Tätigkeit, leistungsfähig zu bleiben. Zu beachten ist, dass durch den Berufskontext ein gewisser inhaltlicher Unterschied zu der Definition der Aspekte nach DeYoung et al. (2007) vorliegt. In diesem Kontext werden aufgrund der geringeren Relevanz die Facetten „*depression*“ und „*angry hostility*“ teilweise ausgeklammert bzw. vernachlässigt (Moldzio et al., 2021). Nachfolgend wird in dieser Forschungsarbeit die Definition der Aspekte nach Moldzio et al. (2021) als Grundlage dienen, da die Arbeit an diese Studie anknüpft.

Die zweite Dimension ist die Extraversion, welche in zwei Aspekte, die DeYoung et al. (2007) „*enthusiasm*“ und „*assertiveness*“ benannten, unterteilt werden kann. „*Enthusiasm*“, welches mit Enthusiasmus ins Deutsche übersetzt werden kann, beschreibt die affektive Komponente der Dimension, beispielsweise die Erwartungshaltung an positive Emotionen. Die Facetten Herzlichkeit, Geselligkeit und Positive Emotionen werden diesem Aspekt zugeordnet. Eine Besonderheit in Bezug auf die Aspekte der Extraversion ist, dass die Facette Erlebnissuche in einer Faktoranalyse auf beide Aspekten (Enthusiasmus und Durchsetzungsstärke) lädt (DeYoung et al., 2007). Der zweite Aspekt dieser Dimension ist

die Durchsetzungsfähigkeit, welche die gleichnamige Facette sowie Aktivität und Erlebnissuche beinhaltet.

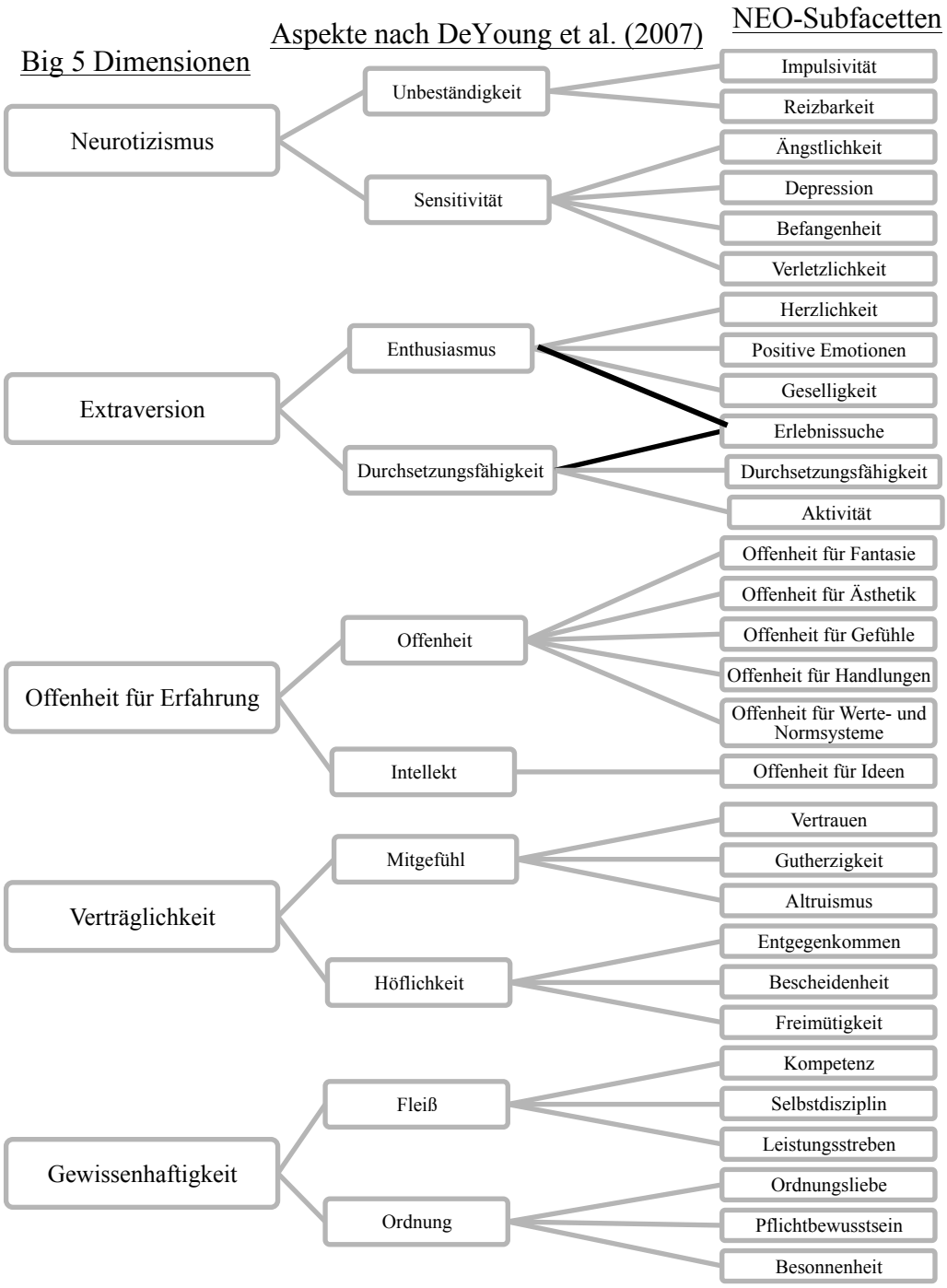
Auch in der Dimension Offenheit für Erfahrung besteht eine Besonderheit in der Zuordnung der sechs Facetten zu den beiden Aspekten „*intellect“* und „*openness“*. So werden die Facetten nicht gleichmäßig auf die Aspekte aufgeteilt, sondern DeYoung et al. (2007) fanden heraus, dass nur die Facette Offenheit für Ideen auf dem Aspekt Intellekt lädt. Diese abweichende Struktur könnte durch die Forschungsarbeit von DeYoung et al. (2005) erklärt werden. Sie besagt, dass ausschließlich die Facette Offenheit für Ideen mit fluider Intelligenz und dem Arbeitsgedächtnis, während die anderen Facetten mit kristalliner Intelligenz zu korrelieren scheinen. Die Inhalte dieser beiden Aspekte unterscheiden sich dahingehend, dass der Aspekt Intellekt die Fähigkeit der Informationsverarbeitung mit Hilfe des logischen Denkens beschreibt und sich der Aspekt Offenheit auf sensorische Verarbeitungen sowie auf Fantasien bezieht (DeYoung et al., 2013).

Der Dimension Verträglichkeit sind die beiden Aspekte „*compassion“* und „*politeness“* untergeordnet, welche mit den Begriffen Mitgefühl und Höflichkeit ins Deutsche übersetzt werden können. Moldzio et al. (in Vorbereitung) fanden jedoch im Berufskontext diese Bezeichnungen inhaltlich nicht wieder und übersetzten diese mit Einfühlungsvermögen und Bescheidenheit. Diese zwei Bezeichnungen werden nachfolgend verwendet. Bei dieser Dimension sind wiederum jeweils drei Facetten einem Aspekt gleichmäßig zugeordnet. Die Facetten Vertrauen, Gutherzigkeit und Altruismus werden dem Aspekt Einfühlungsvermögen zugeschrieben und beschreiben die Tendenz, anderen Vertrauen entgegenzubringen, die eigenen Interessen zurückzustellen sowie sich emotional in andere hineinversetzen zu können (Weisberg et al., 2011). Dem Aspekt Bescheidenheit werden die drei Facetten Entgegenkommen, Freimütigkeit und Bescheidenheit zugeschrieben. Dieser Aspekt beschreibt die Kompromissbereitschaft anderen gegenüber.

Die letzte Dimension Gewissenhaftigkeit besitzt die beiden untergeordneten Aspekte „*industriousness“* und „*orderlines“*, welche als Fleiß und Ordnung übersetzt werden (Moldzio et al., 2019). Der Aspekt Fleiß beschreibt das Streben nach einem Ziel. Ihm werden die untergeordneten Facetten Kompetenz, Selbstdisziplin sowie Leistungsstreben zugeordnet. Ordnungsliebe, Pflichtbewusstsein und Besonnenheit sind die drei Facetten, welche dem Aspekt Ordnung untergeordnet sind. Dieser Aspekt hinterfragt, inwiefern eine Person bedacht, planvoll sowie vorsichtig handelt.

Abbildung 3

Die Struktur des Fünf-Faktoren Modells mit Hinzunahme der jeweiligen Aspekte nach DeYoung et al. (2007).



Anmerkung. Die dicken schwarzen Linien stellen die Besonderheit dar, dass eine Facette zwei Aspekten zugeordnet werden kann.

Persönlichkeitserfassung in der Eignungsdiagnostik

Nachdem in den vorangegangenen Abschnitten verschiedene Persönlichkeitsmodelle erklärt und diskutiert wurden, hat der kommende Abschnitt zum Ziel, aufzuzeigen, wie diese Modelle in der Personalauswahl erfasst werden können.

Bei der Frage, wie Persönlichkeit in der Eignungsdiagnostik erfasst werden kann, hilft der Blick auf den Trimodalen Ansatz von Schuler (2006). Dieser zeigt, dass die Persönlichkeit auf verschiedenen Ebenen erfasst werden kann. So kann beispielsweise in einem strukturierten biografischen Interview der biografische Ansatz verwendet und erfragt werden, wie eine Person in einer bestimmten Situation gehandelt hat. Dabei kann von einem früher gezeigten Verhalten auf ein zukünftiges Verhalten geschlossen werden. Die Durchsetzungsfähigkeit kann ebenso anhand eines simulativen Ansatzes beobachtet werden. Beispielsweise kann eine Dialogübung mithilfe eines Konfliktgespräches durchgeführt werden, um die Durchsetzungsstärke einer Person zu bewerten. Neben dem simulativen Ansatz liegt der Fokus dieser Studie auf der Persönlichkeitserfassung mittels des Eigenschafts- oder Konstruktansatzes (Schuler, 2006). Dieser beschreibt die Nutzung von wissenschaftlich fundierten Fragebögen zur Erfassung von Eigenschaften oder eines Konstruktes. Im nächsten Abschnitt werden die wichtigsten Fragebögen für diese Studie kurz erklärt (eine ausführliche Beschreibung erfolgt in den Kapiteln Persönlichkeit, Operationalisierung der konvergenten Validität & Operationalisierung der divergenten Validität).

Ein oft verwendeter Fragebogen ist das oben genannte NEO-Persönlichkeitsinventar, kurz NEO-PIR, von Costa und McCrae (1992), welches die fünf Faktoren mit 240 Items misst. Dieser Fragebogen ist lang und eher unökonomisch, daher wird die Kurzform NEO-Fünf-Faktoren-Inventar (NEO-FFI) von Borkenau und Ostendorf (2008) mit 60 Items in der Praxis bevorzugt. Sowohl der NEO-PI-R als auch der NEO-FFI erfassen Verhaltensweisen global im Alltag und nicht in einem beruflichen Kontext.

Angelehnt an die oben erklärten zehn Aspekte nach DeYoung et al. (2007) wurden die arbeitsbezogenen Belastbarkeits- und Gewissenhaftigkeitsskalen, kurz ABGS (Moldzio et al., 2019) entwickelt, welche die vier Aspekte Soziale Belastbarkeit, Dauerbelastbarkeit, Fleiß und Ordnung der zwei Dimensionen Neurotizismus und Gewissenhaftigkeit erfassen. Dieser Fragebogen mit seinen 41 Items unterscheidet sich nicht nur durch die Erfassung der Aspekte vom NEO-FFI, sondern auch durch die berufsbezogenen formulierten Items. Angelehnt an die ABGS wurde die AVS (Moldzio, Böge & Wedemeyer, in Vorbereitung) entwickelt, welche zusammen mit der ABGS Teil der Business-Big 5 (Moldzio, Wedemeyer & Böge, in

Vorbereitung) ist und die beiden Aspekte Einfühlungsvermögen und Bescheidenheit der Verträglichkeit anhand von 19 Items erfasst. Ziel dieser Studie ist es, einen berufsbezogenen Fragebogen zu den Aspekten der beiden Dimensionen Extraversion und Offenheit für Erfahrung zu entwickeln, zu validieren und somit die Business-Big 5 zu vervollständigen.

Ein weiterer, in der Personalauswahl oft verwendeter berufsbezogener Fragebogen, ist das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) von Hossiep und Paschen (1998), welches für die Personalauswahl und Personalentwicklung konzipiert wurde. Dieser Fragebogen wurde konzipiert, da zuvor die meisten Fragebögen nicht ausschließlich für die Personalauswahl entwickelt wurden und wie bei dem NEO-FFI der Berufsbezug fehlte. Der Fragebogen ist nicht theoriebasiert entwickelt worden und basiert daher nicht auf den oben genannten Persönlichkeitsmodellen (Hossiep & Paschen, 1998), jedoch wurde der Fragebogen in diesem Abschnitt aufgezählt, um zu zeigen, welche alternativen Fragebögen der Personalauswahl zur Verfügung stehen und in der Praxis eingesetzt werden. Der Fragebogen erfasst 14 Dimensionen, die in die Überkategorien Arbeitsverhalten, berufliche Orientierung, soziale Kompetenz sowie psychische Konstitution eingeteilt werden können. Er ist mit 210 Items genauso wie der NEO-PI-R relativ lang und nicht ökonomisch. Daher haben Hossiep und Krüger (2012) einen ökonomischeren Fragebogen mit 48 Items entwickelt, welchen sie das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – 6 Faktoren, kurz BIP-6F nannten. In einer Analyse fanden sie nicht wie bei den BIP vier Überkategorien, sondern sechs, welche als Engagement, Disziplin, Dominanz, Stabilität, Kooperation und Sozialkompetenz benannt wurden und dem Fragebogen den Namen verliehen.

Haupt- und Nebenkriterien eines Persönlichkeitsfragebogens

Der vorherige Abschnitt hat gezeigt, dass in der Psychologie viele Fragebögen oder selbstkonzipierte Skalen existieren, welche gewisse Merkmale messen sollen. Um jedoch einen geeigneten Fragebogen zu erkennen, wurden einige anerkannte Kriterien entwickelt, welche nachfolgend erklärt werden (Bühner, 2021). Dies soll einen Überblick zur Fragebogengüte vermitteln.

Persönlichkeitsfragebögen gehören zu den psychometrischen Tests und haben zum Ziel, abgrenzbare Persönlichkeitsmerkmale zu erfassen sowie eine Aussage über den Grad einer individuellen Merkmalsausprägung zu ermöglichen (Lienert & Raatz, 1998). Dies bedeutet, dass ein Fragebogen so konzipiert ist, dass er beispielsweise den Grad der Extraversion in kleinen Schritten differenzieren kann und dadurch introvertierte und

extrovertierte Personen erfassen kann. Um dies zu überprüfen, existieren verschiedene Gütekriterien, die auch Inhalt der DIN 33430 sind (Westhoff et al., 2010). Die DIN 33430 ist eine Richtlinie, die dabei helfen soll, einen diagnostischen Prozess gerecht und standardisiert zu entwickeln.

Die Objektivität, Reliabilität sowie Validität zählen dabei zu den Hauptkriterien (Bühner, 2021). Die Objektivität beinhaltet die Frage, inwiefern ein Ergebnis unabhängig von Testleitenden ist (Bühner, 2021) und misst somit den Grad der Standardisierung des Verfahrens. Die Objektivität kann auf drei verschiedene Arten bewertet werden. Zum einen bewertet die Durchführungsobjektivität, inwiefern während der Testung die gleichen Rahmenbedingungen bestanden und Testleitende ggf. Verhaltensvariationen besaß (Bühner, 2021). Bei Persönlichkeitsfragebögen ist die Durchführungsobjektivität zum Beispiel bei einem Onlinefragebogen gegeben, da alle Teilnehmenden die gleiche Instruktion erhalten und keine Testleitung benötigt wird. Die Auswertungsobjektivität beschreibt die Unabhängigkeit des Ergebnisses von der Person, die beispielsweise den Fragebogen auswertet (Beauducel & Leue, 2014). Dies kann zum Beispiel durch eine programmierte Auswertung über ein System standardisiert werden. Die letzte Objektivitätsart ist die Interpretationsobjektivität, welche bewertet, ob das Interpretationsergebnis unabhängig von dem/der Auswerter:in ist (Bühner, 2021). Um die Interpretationsobjektivität gewährleisten zu können, werden große Normstichproben und geprüfte Gütekriterien benötigt (Bühner, 2021).

Das zweite Hauptkriterium ist die Reliabilität, welche den Grad der Messgenauigkeit beschreibt (Bühner, 2021). Die Messgenauigkeit ist jedoch unabhängig davon, welches Konstrukt gemessen werden soll. Es geht lediglich darum, inwiefern das Ergebnis bei einer mehrfachen Anwendung konsistent ist (Beauducel & Leue, 2014). Auch die Reliabilität kann auf drei verschiedenen Arten erfasst werden. Die am häufigsten verwendete Art ist die interne Konsistenz, welche auch Halbierungsreliabilität genannt wird (Bühner, 2021). Diese misst, ob zwei Hälften des Fragebogens hoch mit einander korrelieren und so messgenau sind. Die Retest-Reliabilität beinhaltet die Frage, inwiefern eine weitere Beantwortung zu einem späteren Zeitpunkt mit der Erstbeantwortung korreliert und die Erfassung somit zeitstabil ist (Bühner, 2021). Die letzte und komplexeste Reliabilitätsart ist die Paralleltest-Reliabilität, welche zwei Tests, die das gleiche Konstrukt mit unterschiedlichen Materialien zu zwei verschiedenen Messzeitpunkten erheben, mit einander vergleicht (Beauducel & Leue, 2014). Die Schwierigkeit dabei ist, dass die Items die gleichen psychometrischen Eigenschaften besitzen sollten (Beauducel & Leue, 2014). Die Paralleltest-Relaibilität ist beispielsweise wichtig, wenn ein Test für kognitive Fähigkeiten zwei Versionen besitzt.

Das letzte Hauptkriterium ist die Validität, welche überprüft, ob der Fragebogen das misst (zum Beispiel das Konstrukt), was er auch messen soll (Bühner, 2021). Dies wird auch als Gültigkeit bezeichnet. Teil der Validität ist die Inhaltsvalidität, auch Augenscheinvalidität genannt, welche überprüft, ob jedes Item die Merkmale beispielsweise von dem Persönlichkeitsmerkmal Extraversion erfasst (Bühner, 2021). Die Überprüfung ist komplex, da in der Praxis nicht immer bekannt ist, wie umfangreich ein Merkmal ist und ob jemals alle Facetten dieses Merkmals erfasst werden können. Die Konstruktvalidität ist demgegenüber einfacher statistisch zu erfassen und beinhaltet die Frage, ob der Fragebogen das Konstrukt misst, welches er messen soll (Bühner, 2021). Dabei kann auf der einen Seite durch eine Faktoranalyse die faktorielle Validität erfassen und betrachten, ob die Items auch auf den richtigen Faktoren laden (Bühner, 2021). Andererseits kann die Konstruktvalidität anhand des nomologischen Netzwerkes erfasst werden, welches die divergente und konvergente Validität beinhaltet (Beauducel & Leue, 2014). Dabei wird der Fragebogen anhand des Konstruktes auf zwei verschiede Arten in ein nomologisches Netz eingeordnet. Zum einen werden in der konvergenten Validität Zusammenhänge zwischen dem Fragebogen und anderen konstruktverwandten Fragebögen betrachtet (Bühner, 2021). So betrachtet man beispielsweise die Korrelation zwischen zwei Fragebögen, die beide die Extraversion messen sollen. Demgegenüber betrachtet die divergente Validität die Zusammenhänge des Fragebogens mit konstruktfremden Fragebögen (Bühner, 2021). Angelehnt an das vorherige Beispiel wären dies die Zusammenhänge zwischen der Extraversion und den anderen Big Five Dimensionen Gewissenhaftigkeit, Offenheit für Erfahrung etc., welche geringer als bei der konvergenten Validität ausfallen sollten. Das nomologische Netzwerk sollte dabei durch theoretische Überlegungen zuvor definiert werden (Bühner, 2021). Die letzte Art der Validität ist die Kriteriumsvalidität, welche die Zusammenhänge des Konstruktes mit verschiedenen Kriterien (zum Beispiel Berufserfolg) betrachtet (Bühner, 2021). Dabei können die Kriterien vor dem Konstrukt (retrospektive Validität), gleichzeitig mit dem Konstrukt (konkurrente Validität) oder nach dem Konstrukt (prädiktive Validität) erhoben werden (Bühner, 2021). Darüber hinaus kann die inkrementelle Validität betrachtet werden, welche misst, inwiefern ein Fragebogen über die Messung eines anderen Konstruktes hinweg zur Verbesserung der Vorhersage eines Kriteriums beiträgt (Bühner, 2021).

Unter die Nebengütekriterien können Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit gefasst werden (Bühner, 2021). Unter der Normierung ist die Möglichkeit der Einordnung eines Testergebnisses zwischen beispielsweise „*unterdurchschnittlich“* und „*überdurchschnittlich“* zu verstehen (Bühner, 2021). Die Normierung sollte dabei nicht älter

als acht Jahre alt sein, ca. eine Normstichprobe von 300 Personen umfassen und für verschiedene Personengruppen vorliegen (zum Beispiel Führungskräfte und Auszubildende). Die Vergleichbarkeit besteht, wenn der Test verschiedene Parallelversionen besitzt, die Ähnliches messen (Bühner, 2021). Dadurch kann zum Beispiel eine Person einen Intelligenztest mehrfach bearbeiten (Version A und B) oder in einer Gruppentestung kann so weniger abgeschrieben werden. Ein Fragebogen ist als ökonomisch zu bewerten, wenn er eine geringe Bearbeitungszeit beansprucht, wenig Material benötigt, einfach handhabbar ist, eine Gruppentestung ermöglicht und schnell auswertbar ist (Bühner, 2021). Die Nützlichkeit eines Fragebogens ist dann gegeben, wenn das gemessene Konstrukt ein praktisches Bedürfnis abdeckt (Bühner, 2021). Dies kann zum Beispiel nicht der Fall sein, wenn es das gleiche Merkmal wie drei andere Fragebögen erfasst.

Berufserfolg in der Eignungsdiagnostik

In der Personalauswahl wird die passende Person für eine gewisse Tätigkeit anhand von eignungsdiagnostischen Verfahren gesucht (Blickle, 2014). Wie der Trimodale Ansatz von Schuler (2006) zeigt, werden diese Verfahren angewendet, um zukünftiges Verhalten zu prognostizieren. Aus diesem Grund wird nachfolgend aufgezeigt, welche Persönlichkeitsmerkmale mit welchem Kriterium zusammenhängen. Dabei werden die oben genannten Big Five (Costa & McCrea, 1992) und deren Aspekte nach DeYoung et al. (2007) im Vordergrund der Studien stehen. Diese Analysen werden oft als die Betrachtung der Kriteriumsvalidität betitelt. Die Kriteriumsvalidität beantwortet die Frage, inwiefern ein Verfahren eine prognostische Vorhersage in Bezug auf ein Kriterium treffen kann (Moosbrugger & Kelvava, 2020). Ein in der Forschung oft verwendetes Kriterium ist der Berufserfolg (z.B. Schmidt & Hunter, 1998). Berufserfolg bezeichnet die Effektivität bei der Bearbeitung einer bestimmten Tätigkeit im Beruf (Seibert & Kraimer, 2001), also die gemessene Arbeitsleistung. Leistung sollte vom extrinsischen Erfolg abgegrenzt werden (Judge et al., 1995), da dieser sich primär auf den materiellen Gewinn oder die Anzahl an Beförderungen bezieht. Die Begrifflichkeit Berufserfolg ist nicht einheitlich definiert. Daher wird dies im nächsten Abschnitt anhand eines Job Performance Modells für diese Studie festgelegt.

Job Performance Model

Viele Forschende betrachteten in ihren wissenschaftlichen Arbeiten die Zusammenhänge zwischen Berufserfolg und Persönlichkeit, beispielsweise Barrick und Mount (1991) mithilfe einer Metaanalyse. Tett und Burnett (2003) fanden die reine

Betrachtung der Zusammenhänge zwischen Persönlichkeitsmerkmalen und Berufserfolg zu einfach. Sie plädierten dafür, dass die gegebene Situation mit einbezogen werden müsse und entwickelten daraufhin das Job Performance Modell. Dieses Modell unterscheidet sich in zwei wesentlichen Punkten von der vorherigen Forschung. Es betrachtet die Situationen, welche als Moderator zwischen den Ausprägungen der Persönlichkeitsmerkmale und der Bewertung des Berufserfolgs fungieren. Darüber hinaus zeigt dieses Modell einen Mechanismus zwischen Persönlichkeitsmerkmalen und Berufserfolg auf, mithilfe dessen jedes Merkmal mit Berufserfolg in Verbindung gebracht und die Forschung weiterentwickelt werden kann. Dieses Modell wurde in dieser Arbeit als Grundlage verwendet, da die verschiedenen Arten der Darbietungsart auch einen Einfluss auf das Arbeitsverhalten haben könnten und somit eine spezifischere Betrachtung von Berufserfolg zu bevorzugen wäre.

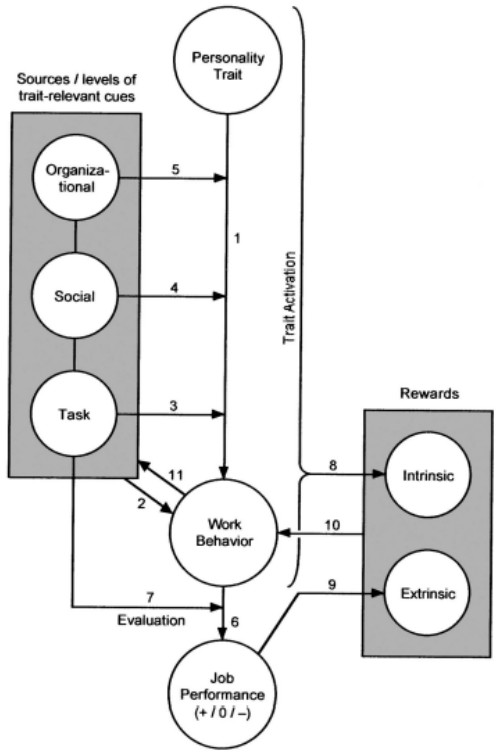
Abbildung 4 zeigt das Job Performance Model nach Tett und Burnett (2003). Die Kernaussage dieses Modells ist, dass Personen mit spezifischen Persönlichkeitseigenschaften während der Arbeit auf verschiedene situative Reize reagieren bzw. Persönlichkeitseigenschaften aktiviert werden und sich dies im Arbeitsverhalten zeigt. Dabei führt das gezeigte Arbeitsverhalten nicht unbedingt zu Berufserfolg. Der erste gekennzeichnete Pfad sagt aus, dass die Persönlichkeitsmerkmale einer Person zu einem gewissen Arbeitsverhalten führen. Diese Persönlichkeitsmerkmale können beispielsweise anhand eines Fragebogens erfasst werden und bilden zuerst einmal einen gewissen Wert in den Skalen ab. In dem zweiten Pfad wird der Haupteffekt einer Situation auf das Arbeitsverhalten dargestellt. Es wird davon ausgegangen, dass eine Situation immer die gleiche Eigenschaft aufweist, aber Personen unterschiedlich auf diese Situation reagieren. So reagieren Menschen mit einer hohen Ausprägung in der Dimension Offenheit für Erfahrung beispielsweise anders auf neue Arbeitsanforderungen als Personen mit einer niedrigeren Ausprägung in dieser Dimension, welche eher konservativ sowie traditionell orientierte agieren ggf. eher Routineaufgaben bevorzugen.

Die Pfade drei bis fünf zeigen drei verschiedene Eigenschaften der Situationen, welche den Pfad zwischen den Persönlichkeitsmerkmalen und dem Arbeitsverhalten moderieren. Alle drei Pfade können getrennt voneinander oder gemeinsam betrachtet werden. Eine Eigenschaft stellt die allgemeine Arbeitsaufgabe dar, welche beispielsweise Verantwortlichkeiten beinhaltet. Zusammengefasst kann als Arbeitsaufgabe alles gesehen werden, was in der Personalauswahl von einer Person erwartet wird, welche eine vakante Stelle besetzen soll. Der vierte Pfad betrachtet den Einfluss eines sozialen Faktors in der Arbeit mit anderen Menschen auf die Persönlichkeit und das Arbeitsverhalten. Diese

Interaktionen können mit Kolleg:innen, Vorgesetzten sowie Kund:innen stattfinden und beinhalten beispielsweise die Kommunikation sowie Teamfähigkeit. Deshalb könnte sich das Arbeitsverhalten in der Interaktion mit Kund:innen unterschiedlich zu der Interaktion mit Kolleg:innen gestalten. Als dritte Quelle wird die Organisation als Moderator zwischen Persönlichkeit und Arbeitsverhalten dargestellt. Unter diesen Punkt zählen beispielsweise die Strukturen, Kulturen oder das Klima der Organisation (Tett & Burnett, 2003). So kann auch ein angenehmes Arbeitsklima zu einer höheren Arbeitsmotivation und folglich einer besseren Arbeitsleistung führen (Day & Bedeian, 1991).

Abbildung 4

Modell des Mechanismus zwischen Persönlichkeitsmerkmalen und Berufserfolg nach Tett und Burnett (2003)



Der sechste Pfad zeigt den Zusammenhang zwischen dem Arbeitsverhalten und Berufserfolg auf, welcher von dem jeweiligen Kontext abhängt. So kann das Arbeitsverhalten in einer Situation geeignet und in einer anderen Situation ungeeignet sein, beispielsweise das Duzen von Kolleg:innen im Verhältnis zu Kund:innen. Der siebte Pfad kann als Evaluation betitelt werden und moderiert das Arbeitsverhalten und Berufserfolg. Die Bewertung des Berufserfolgs hängt von den Arbeitsanforderungen bzw. den Erwartungen ab, welche auf den drei Quellen (Aufgabe, Sozial und Organisation) basieren. Die nächsten drei Pfade (acht bis

zehn) thematisieren die Rolle der Motivation. Zum einen wird die intrinsische Motivation betrachtet. Darunter werden diese Motive verstanden, die in der Person entstehen und dadurch die Persönlichkeitsmerkmale aktivieren. Darüber hinaus besteht ein Zusammenhang zwischen Berufserfolg und der extrinsischen Motivation, indem andere Personen auf die Leistung reagieren und gesellschaftliche Normen bestehen. Extrinsische Motivation kann auch durch Lob, Akzeptanz oder Belohnung (finanziell oder Aufstiegschancen) gesteigert werden. Der zehnte Pfad sagt aus, dass Arbeitsverhalten, welches extrinsisch oder intrinsisch motiviert ist, wahrscheinlicher ausgeführt wird. Der Zusammenhang zwischen dem Arbeitsverhalten und den Quellen ist der letztbeschriebene Pfad. Dieser Pfad sagt, dass auch das Arbeitsverhalten einer Person die Aufgabe oder die Organisation beeinflussen kann. Dieser Einfluss kann sowohl negativ als auch positiv sein. So kann beispielsweise eine extrovertierte Person die Geselligkeit anderer Personen fördern, aber die Kreativität einer Person kann auch zum Beispiel durch veraltete Methoden eingeschränkt sein.

Dieses Modell verdeutlicht, dass es viele Einflussfaktoren gibt, welche neben der Persönlichkeitsausprägung den Berufserfolg beeinflussen. Für diese Arbeit ist die Aussage entscheidend, dass die Persönlichkeit nicht direkt zu Berufserfolg führt, sondern diese Beziehung mit dem Arbeitsverhalten interagiert, welches durch äußere Einflüsse gelenkt wird. Diese Aussage ist für diese Arbeit wichtig, da damit ggf. gezeigt werden kann, dass das Arbeitsverhalten in einem virtuellen Verfahren verschieden zu einem Präsenzverfahren sein kann und dennoch zu Berufserfolg führen könnte. Darüber hinaus wird in diesem Modell ersichtlich, dass nicht nur die Persönlichkeitsausprägungen für Berufserfolg entscheidend sind, sondern eher das Arbeitsverhalten in Fragebögen erfasst werden sollte, damit Berufserfolg besser vorhergesagt werden kann. Dies könnte der Fall sein, wenn die Persönlichkeitsfragebögen berufsbezogen formuliert sind und somit das Verhalten im Arbeitskontext bzw. das Arbeitsverhalten erfassen. Aus diesem Grund hat diese Arbeit zum Ziel, einen berufsbezogenen Fragebogen zu den DeYoung et al. (2007) Aspekten von Extraversion und Offenheit für Erfahrung zu entwickeln sowie zu validieren.

Berufserfolg und Persönlichkeit

Das zuvor erklärte Job Performance Model von Tett und Burnett (2003) zeigt, dass der Berufserfolg und die Persönlichkeitsmerkmale zusammenhängen. Viele Wissenschaftler:innen betrachteten vor allem den Zusammenhang zwischen den Big Five und Berufserfolg (z.B. Barrick & Mount, 1991; van Aarde et al., 2017). Dieser Abschnitt soll verdeutlichen, dass nicht nur Zusammenhänge zwischen Berufserfolg und der Persönlichkeit

bestehen, sondern auch für den Berufserfolg förderliche Ausprägungen der Dimensionen aufzeigen. So wird beispielsweise ersichtlich, dass eher eine hohe Ausprägung in der Extraversion mit Berufserfolg der Manager:innen zusammenhängt und dem gegenüber eine niedrige Ausprägung im Neurotizismus mit Berufserfolg korreliert (z.B. Barrick & Mount, 1991). Dies würde bedeuten, dass kontaktfreudige und wenig ängstliche Personen wahrscheinlicher einen hohen Wert in Berufserfolg aufweisen, als introvertierte und ängstliche Personen.

Eine viel zitierte Studie in diesem Gebiet ist die Metaanalyse von Barrick und Mount (1991), welche 117 Studien untersuchte. Sie teilten die Stichprobe in die fünf Berufsgruppen Fachkräfte, Manager:innen, Polizist:innen, Vertriebler:innen und angelernte Kräfte ein. Als Kriteriumsvariablen wurden drei Arten des Berufserfolgs (Leistungsniveau im Job, im Training und Personaldaten) gewählt. Die Ergebnisse zeigen, dass die Gewissenhaftigkeit mit allen drei Variablen des Berufserfolgs in allen Berufsgruppen zusammenhängt ($.06 < r < .14$). Dies bedeutet, dass Personen mit einer hohen Strukturiertheit und Sorgfalt in der Arbeitsweise, mit einer hohen Wahrscheinlichkeit einen höheren Berufserfolg besitzen. In der Dimension Extraversion zeigten sich auch positive Zusammenhänge, vor allem bei Manager:innen ($r = .18$) und Vertriebler:innen ($r = .15$). Bei den Manager:innen könnte dies an der benötigten Durchsetzungsfähigkeit liegen und Vertriebler:innen müssen zugewandt und stark im Kontaktverhalten sein, damit sie Kund:innen überzeugen können.

Barrick und Mount (1991) betrachteten den Zusammenhang zwischen der Persönlichkeit und dem Berufserfolg anhand von Studien im amerikanischen Raum. Diese Studie nahmen viele Forschende zum Anlass, die Ergebnisse für andere Teile der Welt zu replizieren. So führte beispielsweise Salgado (1997) eine Metaanalyse im europäischen Raum durch. Diese Analyse zeigte, dass die Ergebnisse in großen Teilen auf Europa replizierbar sind, da Salgado (1997) auch Zusammenhänge zwischen Gewissenhaftigkeit ($r = .25$) sowie Ausgeglichenheit ($r = .19$) und Berufserfolg fand. Darüber hinaus konnte aufgezeigt werden, dass Zusammenhänge in zwei Berufsgruppen (Manager:innen: $r = .20$; Polizisten sowie Polizistinnen: $r = .05$) zwischen Extraversion und Berufserfolg bestehen sowie Zusammenhänge zwischen Offenheit für Erfahrung ($r = .26$) sowie Verträglichkeit ($r = .30$) und Trainingsleistung sichtbar werden.

Nicht nur in Amerika oder Europa wurden Zusammenhänge erforscht und gefunden, sondern auch in Ostasien (Oh, 2009) und Südafrika (van Aarde et al., 2017) wurden Analysen auf die Studie von Barrick und Mount (1991) aufgebaut. Beide Studien erkannten ebenfalls die Bedeutung von Gewissenhaftigkeit (Oh, 2009: $r = .18$; van Aarde et al., 2017: $r = .22$)

und Ausgeglichenheit in Bezug auf Berufserfolg (van Aarde et al., 2017: $r = .11$). Jedoch fanden beide Studien auch einen größeren Zusammenhang zwischen der Extraversion und Berufserfolg (Oh, 2009: $r = .23$; van Aarde et al., 2017: $r = .15$). In Südafrika wurde ein negativer Zusammenhang zwischen der Extraversion und der akademischen Leistung gefunden, welcher mit der Ablenkbarkeit durch geselliges Verhalten begründet wurde ($r = -.38$). Darüber hinaus zeigte Oh (2009), dass Extraversion ($r = .23$) sogar einen stärkeren Zusammenhang mit Berufserfolg in Ostasien aufweist als Gewissenhaftigkeit ($r = .18$). Dies könnte durch die kulturbedingte zentrale Rolle der interpersonellen Beziehungen in der Arbeitswelt zustande kommen. Beide Studien lassen erkennen, dass die Persönlichkeit zur Vorhersage von Berufserfolg in Teilen valide ist, jedoch kleine kulturelle Unterschiede zwischen Personen unterschiedlicher Kulturkreise bestehen, welche berücksichtigt werden müssen.

Viele Studien kommen zu dem Ergebnis, dass Zusammenhänge zwischen den Big Five Dimensionen und Berufserfolg bestehen (z.B. Barrick & Mount, 1991). Nachfolgend stellt sich die Frage, ob dies auch bei der spezifischeren Erfassung der Persönlichkeit durch die Aspekte gegeben ist. Dies ist für die Arbeit entscheidend, da diese vor allem auf den Aspekten nach DeYoung et al. (2007) aufbaut und somit in diesem Abschnitt gezeigt werden soll, inwiefern die einzelnen Aspekte auch mit Berufserfolg korrelieren.

Dies ist Teil des Bandbreiten-Fidelitäts-Dilemma von Cronbach und Gleser (1957), welches besagt, dass bei jedem Verfahren abgewogen werden muss, inwiefern die schmalere Erfassung eines Kriteriums einen Mehrwert in Bezug auf die Validität liefert oder ob eine breite Erfassung ausreicht. So stellt sich hier die Frage, inwiefern die zehn Aspekte einen Mehrwert gegenüber der Erfassung der fünf Dimensionen liefern. Um dieses Dilemma genauer zu betrachten, werden zwei weitere Studien berichtet, die Zusammenhänge zwischen den genannten Aspekten und Berufserfolg untersuchten. Judge et al. (2013) verwendeten in ihrer Studie die Aspekte von DeYoung et al. (2007), konnten deren Struktur replizieren und fanden heraus, dass die Zusammenhänge zwischen den Aspekten und Berufserfolg generell signifikant größer sind als zu den fünf Dimensionen. Diese Studie konnte jedoch auch ähnliche Ergebnisse wie Barrick und Mount (1991) zeigen. So fanden Judge et al. (2013) Zusammenhänge zwischen den Aspekten der Gewissenhaftigkeit und Berufserfolg ($r = .26$). Als jedoch zusätzlich die Facettenebene betrachtet wurde, zeigte sich, dass die Aspekte gleich stark mit Berufserfolg korrelieren wie die Dimension ($r_{Ordung} = .24$; $r_{Fleiß} = .21$) (Judge et al., 2013). Dabei war auffällig, dass die Facette Positive Emotionen ($r = .20$) des Aspekts Enthusiasmus der Dimension Extraversion eine genauso große Korrelation zu Berufserfolg

wie die Dimension selbst aufweist (r = .20). Zusammengefasst wurde in dieser Studie deutlich, dass vor allem in den Dimensionen Verträglichkeit ($r_{Verträglichkeit}$ = .10; $r_{Höflichkeit}$ = .11) und Extraversion ($r_{Extraversion}$ = .12; $r_{Durchsetzungsfähigkeit}$ = .15) die Aspekte einen größeren Zusammenhang mit Berufserfolg (Aufgabenleistung) aufweisen als die Dimension. Lediglich im Aspekt Offenheit konnte dies nicht gefunden werden ($r_{Offenheit\ für\ Erfahrung}$ = .12; $r_{Intellekt}$ = .09). Judge et al. (2013) vermuten, dass dies jedoch an der sehr kulturlastigen und berufsunspezifischen Darstellung der Items liegen könnte, was durch einen berufsbezogen formulierten Fragebogen ggf. revidiert werden könnte. Eine Studie von Moldzio et al. (2021) konnte die Ergebnisse von Judge et al. (2013) stützen. Sie untersuchten die Zusammenhänge von Berufserfolg zu den Aspekten der Belastbarkeit (zum Beispiel technische Auszubildende: $r_{soziale\ Belastbarkeit}$ = .44) und Gewissenhaftigkeit (zum Beispiel Linienführungskräfte: $r_{Fleiß}$ = .27) anhand eines berufsbezogenen Fragebogens, welcher diese Aspekte erfasste. Sie fanden nicht nur Korrelationen zwischen den Aspekten und Berufserfolg, sondern konnten darüber hinaus zeigen, dass die Erhebung dieser Aspekte die inkrementelle Validität zur Vorhersage von Berufserfolg steigern kann (15% gegenüber dem NEO-FFI). Daraus lässt sich schließen, dass eine spezifischere Erfassung der Persönlichkeit durch die Aspekte in der Personalauswahl hilfreich sein könnte, um Berufserfolg detaillierter vorhersagen zu können.

Judge et al. (2013) und Moldzio et al. (2021) befürworten die spezifischere Erhebung der einzelnen Dimensionen durch die jeweiligen zwei Aspekte und konnten den Mehrwert für die Eignungsdiagnostik aufzeigen, beispielsweise für die Kriteriumsvalidität. Sie begründen dies durch die starken Zusammenhänge der Aspekte und Berufserfolg und durch die Erhöhung der inkrementellen Validität über die Big Five Dimensionen hinaus. Es gibt jedoch auch Forschende, welche die breitere Erfassung der Persönlichkeit empfehlen. Beispielsweise argumentieren Salgado et al. (2014), dass in ihrer Studie die engen Dimensionen sowohl breite als auch enge Leistungsmaße (Berufserfolg und kontraproduktive akademische Verhaltensweisen) signifikant vorhersagen, die Facetten jedoch nicht. Außerdem konnte gezeigt werden, dass die Facetten keine weitere Varianz über die Dimensionen hinaus aufklären und somit ein Einsatz der Facetten keinen Mehrwert in der Varianzaufklärung bringen würde. Einen weiteren Kritikpunkt der schmalen Erfassung lieferten Murphy und Dzieweczynski (2005) indem sie erklärten, dass die Korrelationen zwischen der Persönlichkeit und Berufserfolg in der Studie von Barrick und Mount (1991) nur gering bis moderat ausfallen und daher eine spezifischere Betrachtung keinen praktischen Mehrwert aufweist. Tett et al. (1991) befürworten demgegenüber die spezifischere Erfassung der Persönlichkeit, da sie anhand einer Metaanalyse zeigen konnten, dass die engen

Persönlichkeitsmerkmale Berufserfolg besser vorhersagen können. Sie schränkten diese Aussage jedoch dahingehend ein, dass sie diese bessere Vorhersagekraft nur dann sehen, wenn zuvor eine gründliche Anforderungsanalyse geleistet wurde und die Faktoren zudem den passenden Leistungskriterien zugeordnet werden. Rothstein und Jelley (2003) fassten in ihrer Übersichtsarbeit zusammen, dass in der Wahl der breiten versus schmalen Erfassung kein richtig oder falsch besteht, es jedoch wichtig ist, sich vor jedem neuen Auswahlverfahren die speziellen Umstände sowie Anforderungen anzuschauen und danach zu entscheiden.

Zusammengefasst zeigt die bisherige Forschung, dass Zusammenhänge zwischen der Persönlichkeit und Berufserfolg bestehen, jedoch die Stärke der Korrelationen von den einzelnen Dimensionen und den Berufsgruppen abhängt (z.B. Barrick & Mount, 1991). So ist der Zusammenhang mit der Dimension Extraversion bei Manager:innen und Vertriebler:innen höher als in anderen Berufsgruppen (z.B. Barrick & Mount, 1991). Es gibt jedoch auch Unterschiede in der Art der Leistungserhebung. Beispielsweise kann Offenheit für Erfahrung nicht generell Berufserfolg vorhersagen, jedoch Trainingserfolg. Die Studien von Judge et al. (2013) und Moldzio et al. (2021) zeigten, dass die differenziertere Betrachtung der Persönlichkeit anhand der zehn Aspekte fortgeführt und weiter erforscht werden sollte, um zu betrachten, inwieweit eine schmalere Erfassung der Persönlichkeit wünschenswert ist. Aufgrund des Alters der meisten Studien (beispielsweise Barrick & Mount, 1991) wurde die wachsende Bedeutung der Dimension Offenheit für Erfahrung noch nicht vollends berücksichtigt und sollte daher in der weiteren Forschung mehr im Fokus stehen. So gehen Schermuly et al. (2019) davon aus, dass durch Agilität und New Work die Mitarbeitende immer flexibler und lernbereiter sein müssen, da die Anforderungen der Arbeitswelt sich ständig ändern. Diese beiden Eigenschaften sind Teile der Dimension Offenheit für Erfahrung, welche daher zunehmend stärker mit Berufserfolg korrelieren könnte.

Weitere Berufskriterien und der Zusammenhang mit Persönlichkeitsfaktoren

Es ist nicht nur Berufserfolg als Berufskriterium interessant, sondern es gibt noch weitere Kriterien zur Erfolgsmessung. Ein Kriterium kann beispielsweise die Lebenslaufzufriedenheit sein, welche nach Judge et al. (1995) in objektive, extrinsische und subjektive sowie intrinsische Elemente unterteilt werden kann. Ziel dieses Abschnittes war es zu zeigen, inwiefern dieses Kriterium mit den Big Five Dimensionen korreliert und zu schauen, ob es vor allem für die Extraversion und die Offenheit für Erfahrung ein relevantes

Kriterium sein könnte, welches in dieser Arbeit berücksichtigt werden sollte. Als objektiv werden alle Elemente gewertet, anhand deren der Erfolg verglichen werden kann. Dies ist beispielsweise das Gehalt oder die Anzahl der Beförderungen. Die subjektive Bewertung beinhaltet demgegenüber die Zufriedenheit mit dem aktuellen Job sowie dem gesamten Karriereweg. Ng et al. (2005) fanden in ihrer Studie eine moderate Korrelation zwischen den objektiven und subjektiven Elementen, die sie mithilfe der Attributionstheorie begründeten (Johns, 1999). Diese Theorie sagt, dass Menschen bei Erfolg die Gründe bei sich selber sehen während sie bei Misserfolg externe Gründe dafür verantwortlich machen. Aus diesem Grund kann ein objektiver Erfolg, zum Beispiel eine Beförderung zu einem positiveren Selbstbild beitragen und die Zufriedenheit mit der eigenen Karriere steigern. Die Begrifflichkeit Karriere beinhaltet die Aufreihung aller geleisteten Jobs über die komplette Laufbahn hinweg (Ng et al., 2005). Sie wird im Nachgang als Lebenslauf benannt, da der Begriff Karriere schnell mit etwas Positiven assoziiert wird (Ng et al., 2005). Judge et al. (1995) definierten die Zufriedenheit im Job als einen emotionalen Zustand, der durch eigene positive Bewertungen in diesem Job entwickelt wird. Seibert und Kraimer (2001) definierten darauf aufbauend die Lebenslaufzufriedenheit als einen positiven Zustand, welcher aus der Bewertung der beruflichen Stationen entsteht. Diese Bewertungen wurden jedoch noch in Relation zu den eigenen Zielen und Erwartungen gesetzt. Grundlagen der Bewertungen waren hierbei zum Bespiel das Gehalt, Beförderungen und Entwicklungschancen. Obwohl die Laufbahn den aktuellen Job miteinschließt, wurden Hinweise darauf gefunden, dass die Lebenslaufzufriedenheit nicht mit denselben Variablen wie die Jobzufriedenheit zusammenhängt (Seibert & Kraimer, 2001). Es kann davon ausgegangen werden, dass die beiden Formen der Zufriedenheiten im Zusammenhang stehen, jedoch unterschiedliche Bewertungsschemata besitzen. Judge et al. (1995) begründeten diese unterschiedlichen Bewertungsschemata damit, dass die Orientierungen verschieden sein könnten. Beispielsweise ist die Lebenslaufzufriedenheit eher ergebnisorientiert und die Jobzufriedenheit wird eher prozessorientiert bewertet (Judge et al., 1995). Es werden jedoch einzelne Bestandteile der Lebenslaufzufriedenheit in der Jobzufriedenheit nicht mit einbezogen, beispielsweise vergangene Erfolge, welche für die aktuelle Tätigkeit keine Relevanz besitzen. Stattdessen werden bei der Jobzufriedenheit andere Aspekte (zum Beispiel Charakteristika der Tätigkeit) in die Bewertung mit einbezogen. Anhand dieser Abgrenzung sprachen Seibert und Kraimer (2001) von distinkten Konstrukten, welche auch getrennt voneinander betrachtet werden sollten.

Um eine Aussage über die Zusammenhänge des Kriteriums der Lebenslaufzufriedenheit und der Persönlichkeit treffen zu können, müssen auch in diesem Fall die Korrelationen betrachtet werden. Eine Reihe von Forschenden haben dies in Studien untersucht, welche den Fokus auf extrinsische und objektive Elemente legten. Einige Forschungsarbeiten fanden einen positiven Zusammenhang zwischen Gewissenhaftigkeit und Lebenslaufzufriedenheit, was dafürspricht, dass gewissenhaftere Personen mit einer höheren Wahrscheinlichkeit mit ihrem Lebenslauf zufrieden sind (Judge et al., 1999; Lounsbury et al., 2003: r = .11; Ng et al., 2005: r = .14). Auch Extraversion weist einen positiven Zusammenhang mit Laufbahnzufriedenheit auf, wohingegen Neurotizismus einen negativen Zusammenhang zeigt (Lounsbury et al., 2003: r = .37; Ng et al., 2005: r = -.36; Seibert & Kraimer, 2001: β = -.20). Dies zeigt, dass eher emotional instabile Personen mit einer höheren Wahrscheinlichkeit die Zufriedenheit negativer bewerten. Wenn jedoch der positiv gepoolte Begriff Belastbarkeit betrachtet wird, wie in der Studie Lounsbury et al. (2003), dann sind positive Zusammenhänge mit der Lebenslaufzufriedenheit erkennbar (Lounsbury et al., 2003; Ng et al., 2005). Für die anderen beiden Dimensionen Verträglichkeit und Offenheit für Erfahrung fanden die bisherigen Forschungsarbeiten keine einheitlichen Zusammenhänge. Ein positiver Zusammenhang zur Offenheit für Erfahrung wurde in den Arbeiten von Lounsbury et al. (2003) sowie Ng et al. (2005) sichtbar (Lounsbury et al., 2003: r = .15; Ng et al., 2005: r = .12). Seibert und Kraimer (2001) konnten in ihrem Artikel einen negativen Zusammenhang zwischen der Dimension Verträglichkeit und Lebenslaufzufriedenheit aufweisen (β = -.09), wohingegen Ng et al. (2005) mit einem schwachen positiven Zusammenhang genau das gegenteilige Ergebnis aufzeigten (r = .11). Die Forschungsarbeiten zeigen, dass die Lebenslaufzufriedenheit ein moderates Kriterium sein könnte, welches im Zusammenhang mit der Persönlichkeit bei der Personalauswahl oder Potenzialerkennung berücksichtigt werden könnte. Demgegenüber zeigte jedoch auch die Forschungsarbeit von Lounsbury et al. (2003), dass 85% der Varianz von dem Kriterium Lebenslaufzufriedenheit durch Neurotizismus, Optimismus und Arbeitswille aufgeklärt wurde. Da die untenstehende Studie 1 hauptsächlich auf den Dimensionen Extraversion und Offenheit für Erfahrung aufbaut und Ng et al. (2005) eine Moderation des Geschlechtes und der Zeit fand, wurde das Kriterium Lebenslaufzufriedenheit nicht weiter berücksichtigt.

Gerechtigkeitswahrnehmung in der beruflichen Eignungsdiagnostik

Die vorherigen Abschnitte machen deutlich, wie wichtig die richtige Auswahl der Eignungsdiagnostikinstrumente ist, um die Passung zwischen einer Person und einer

vakanten Stelle zu überprüfen und so die bestmöglich passenden Bewerbenden erkennen zu können. Jedoch ist nicht nur die Güte der Verfahren für eine erfolgreiche Personalauswahl wichtig, sondern auch die Gerechtigkeitswahrnehmung der Bewerbenden. Hausknecht et al. (2004) stellten fest, dass Unternehmen die Fairness des Auswahlprozesses fördern und die eignungsdiagnostischen Methoden eher berufsbezogen wählen sollten, damit die Bewerbenden eine geringere Chance haben, das Unternehmen aufgrund von beispielsweise Diskriminierung zu verklagen. Auch Kersting (2018) warnt, aufgrund der höheren Akzeptanz nur kurze unstrukturierte Vorstellungsgespräche anstatt standardisierte und detaillierte Eignungsdiagnostik durchzuführen. Darüber hinaus ist die Beachtung der Akzeptanz wichtig, damit die Bewerbenden nicht kurz vor Abschluss der Besetzungsphase die vakante Stelle ablehnen und so hohe Kosten entstehen (Moldzio, 2014). Aus diesem Grund hat dieser Abschnitt das Ziel, die Gerechtigkeitswahrnehmung zu definieren, Modelle sowie deren Bestandteile zu erklären und dadurch wichtige Aspekte für die Personalauswahl herauszufiltern. So stellt sich aktuell die Frage, ob virtuelle Verfahren genauso gerecht wahrgenommen werden, wie die Verfahren vor Ort (z.B. Basch & Melchers, 2020). Darüber hinaus hat sich auch gezeigt, dass die Dimensionen Extraversion und Offenheit für Erfahrung positiv mit der Gerechtigkeitswahrnehmung korrelieren (z.B. Brenner et al. 2016; Moldzio, 2014). Aus diesem Grund untersucht diese Arbeit nicht nur die Validität von virtuellen Simulationsaufgaben, sondern auch die Gerechtigkeitswahrnehmung der Bewerbenden.

Um jedoch tiefer auf die Forschung zur Gerechtigkeitswahrnehmung, auch Akzeptanz genannt, eingehen zu können, muss dieser Begriff zunächst definiert werden. Unter dem Begriff Gerechtigkeit „ist ein Idealzustand ausgeglichener Interessen ohne Benachteiligung von Einzelnen (Individuum) oder Gruppen“ zu verstehen (Schmitt, 2019). Auf die Eignungsdiagnostik projiziert bedeutet dies, dass alle Bewerbenden die gleichen Chancen besitzen sollten, eine vakante Stelle zu besetzen und somit keine bestimmte Gruppe von Bewerbenden bevorzugt werden würde. In einer Übersichtsarbeit erklärten Cropanzano et al. (2001), dass die wahrgenommene Gerechtigkeit aus drei Perspektiven betrachtet werden kann. Zum einen beschreibt die distributive Gerechtigkeit die Wahrnehmung in Bezug auf die Verteilung von Gegenständen materieller oder symbolischer Art. Dies könnten die Verteilungen von Gehalt oder Beförderungen sein, jedoch auch die gerechte Verteilung von Wertschätzung im Team. Die prozedurale Gerechtigkeit befasst sich mit der Wahrnehmung eines kompletten Prozesses. Das bedeutet, dass die prozedurale Gerechtigkeit die wahrgenommene Fairness in einem Auswahlprozess bewertet. Die interaktionale Gerechtigkeit befasst sich mit der wahrgenommenen Fairness in allen Interaktionen.

Beispielsweise kann die Interaktion zwischen zwei Kolleg:innen oder zwischen Vorgesetzten und Mitarbeitende stattfinden. Es stellt sich in Bezug auf die interaktionale Gerechtigkeit in der Personalauswahl die Frage, ob Bewerbende von Beobachtenden bevorzugt werden oder nicht.

Gerechtigkeitswahrnehmung von Auswahlprozessen

In der Personalauswahl ist die Gerechtigkeitswahrnehmung gegenüber einzelnen Auswahlverfahren nicht ausschließlich entscheidend. Die Gerechtigkeitswahrnehmung verändert sich über die verschiedenen Schritte eines Auswahlprozesses hinweg, was es auch zu beachten gilt (Konradt et al., 2020). Konradt et al. (2020) zeigten in ihrer Metaanalyse, in der nur Daten mit Messwiederholung verwendet wurden, dass bei einem dreistufigen Auswahlprozess die Gerechtigkeitswahrnehmung von dem ersten zum zweiten Schritt um 6% und dann vom zweiten zum dritten noch einmal um 1,1% abnahm. Die Abnahme der Gerechtigkeitswahrnehmung hängt auch von dem Ausgangswert zu Beginn des Prozesses ab. So wurde in dieser Metaanalyse ersichtlich, dass die Gerechtigkeitswahrnehmung mit einem höheren Ausgangswert zu Beginn stärker sinkt, als bei einem niedrigeren ($\beta = -.93$, $p <$ .001). Dies wurde damit begründet, dass sich die Bewerbenden eher auf die negativen bzw. nicht erfüllten Aspekte konzentrieren, als auf die positiven Dinge, die in dem Auswahlprozess erfüllt wurden. Eine weitere Erkenntnis war, dass die Gerechtigkeitswahrnehmung zwischen dem Pretest vor dem Verfahren und dem Posttest bei einer ungerechten Behandlung signifikant abnahm (d = -2,28, *CI* [-2,74 | -1,82]). Dies war bei einer gerechten Behandlung nicht der Fall (d = -,17, *CI* [-,43 | .09]). Eine wichtige Erkenntnis war zudem, dass sich neu erhaltende Informationen zwischen dem Posttest und der Postentscheidung noch positiv auf einen ungerecht wahrgenommenen Prozess auswirken können (d = .60, *CI* [.41 | .79]). Dies spricht dafür, dass die Bewerbenden ggf. eine Erklärung für das ungerechte Verfahren oder ein konstruktives Feedback erhalten haben. Eine weitere Erkenntnis war, dass die Gerechtigkeitswahrnehmung zwischen dem Posttest und der Postentscheidung stärker in einem kürzeren Zeitraum sank anstatt in einem längeren, was dafürspricht, dass die Bewerbenden in der längeren Zeit durch Kausalattributionen die Gerechtigkeitswahrnehmung anders bewerten, da sie ggf. das Verfahren besser verstehen (d_0 = -.53, *CI* [-.79 | -.26]; d_7 = -.11, *CI* [-.22 | -.01]; d_{30} = .17, *CI* [-.01| .35]). Damit die Bewerbenden nicht mit einer zu hohen Erwartung an das Verfahren starten und somit die Gerechtigkeitswahrnehmung vereinzelt stark abnimmt, empfehlen die Autor:innen, dass die Personaler:innen eines Unternehmens den Prozess vorab transparent und detailliert den

Bewerbenden erklären (Konradt et al., 2020). Des Weiteren sollten die Personaler:innen zeitnah ein konstruktives Feedback auch zur Einstellungsentscheidung geben, damit die Gerechtigkeitswahrnehmung nicht weiter sinkt, sondern eventuell sogar gesteigert werden kann.

Diese Metaanalyse zeigte, dass die Gerechtigkeitswahrnehmung abnimmt und diese Abnahme von verschiedenen Faktoren abhängt. Demgegenüber hat jedoch auch die direkte Gerechtigkeitswahrnehmung der Verfahren einen Einfluss auf die Abnahme. Aus diesem Grund ist es wichtig, auf einzelne Verfahren und deren Gerechtigkeitswahrnehmung zu schauen. Es sollten jedoch auch andere Aspekte wie beispielsweise die Persönlichkeit betrachtet werden, um dort eventuell verschiedene Einflüsse zu erkennen.

Gerechtigkeitswahrnehmung einzelner Auswahlverfahren

In den vorgenannten Abschnitten wurde gezeigt, dass starke Zusammenhänge zwischen den Persönlichkeitseigenschaften und Berufserfolg zu finden sind (z.B. Barrick & Mount, 1991). Daraus kann geschlossen werden, dass der Einsatz von Persönlichkeitsfragebögen empfehlenswert ist. Studien zur Anwendungshäufigkeit von eignungsdiagnostischen Verfahren belegen jedoch eine weitaus niedrigere Einsatzhäufigkeit von 20-40% im Vergleich zu weniger validen Verfahren (Armoneit, 2019). Es besteht somit in der Auswahl von eignungsdiagnostischen Verfahren eine Diskrepanz zwischen der Wissenschaft und der Praxis. Darüber hinaus zeigte sich, dass die Gerechtigkeitswahrnehmung über die Zeit hinweg während des Prozesses abnimmt, vor allem bei ungerecht wahrgenommenen Verfahren (Konradt et al., 2020). Aus diesem Grund betrachtet der nächste Abschnitt die Gerechtigkeitswahrnehmung einzelner Verfahren, um zu zeigen, welche eher als gerecht bzw. ungerecht wahrgenommen werden und welchen Einfluss die virtuelle Darbietungsart auf die Gerechtigkeitswahrnehmung besitzt.

Die Diskrepanz zwischen der Wissenschaft und der Praxis im Einsatz von Persönlichkeitsfragebögen versuchten König et al. (2010) aufzuklären. Sie baten Personalverantwortliche aus der deutschsprachigen Schweiz, in einer Umfrage die Einsatzhäufigkeiten von verschiedenen Verfahren in ihrer Personalauswahl anzugeben und das halbstrukturierte Interview, Leistungstests, Persönlichkeitsfragebögen, Assessment Center sowie die Graphologie mithilfe von sechs Variablen zu bewerten. Diese sechs Variablen leiteten die Autor:innen anhand von vorherigen Forschungsarbeiten ab und gingen davon aus, dass folgende Variablen die Einsatzhäufigkeit der Auswahlverfahren beeinflussen: Die Nutzung der Verfahren in der Praxis; die Wahrscheinlichkeit, dass der Einsatz der

Verfahren zu rechtlichen Problemen führt; die Bewerbendenreaktionen auf die Verfahren; die Möglichkeit der Selbstdarstellung der Organisation durch das Verfahren; die Abwägung der Vorhersagekraft sowie die mit dem Verfahren verbundenen Kosten. Die Auswertung ergab, dass die Vorhersagekraft bzw. die prädiktive Validität der Verfahren ein signifikanter Prädiktor für die Einsatzhäufigkeit eines Verfahrens ist, dieser jedoch nur eine untergeordnete Rolle in der tatsächlichen Einsatzhäufigkeit von Verfahren spielt (r = .26). Ein Beispiel dafür ist die Einsatzhäufigkeit der Fähigkeitstests von 18,6%, obwohl diese nach Schmidt und Hunter (1998) die höchste prädiktive Validität aufweisen. Die Zusammenhänge zwischen der Bewerbendenreaktionen (r = .37), den Kosten (r = -.24) sowie der Nutzung der Verfahren (r = .19) in der Praxis und der Einsatzhäufigkeit sind jedoch in Summe viel gravierender (König et al., 2010). Dies würde beispielsweise bedeuten, dass ein Verfahren, welches besser akzeptiert wird, aber weniger valide ist, in der Personalauswahl eher eingesetzt wird. Dies führt zu der Erkenntnis, dass die Bewerbendenreaktionen eine zentrale Rolle bei der Entscheidung für oder gegen bestimmte Auswahlverfahren spielt und somit nicht vernachlässigt werden sollte.

Der Begriff der Bewerbendenreaktionen ist eher im englischsprachigen Raum angesiedelt und wird im deutschsprachigen Raum beispielsweise auch als soziale Akzeptanz oder soziale Validität bezeichnet (Kersting, 1998). Soziale Akzeptanz beschreibt die Gerechtigkeitswahrnehmung der Bewerbenden während eines Auswahlverfahrens anhand von verschiedenen Aspekten und die Reaktion der Bewerbenden darauf. Kersting (1998) stellte fest, dass die soziale Akzeptanz auf verschiedene Arten mit dem Auswahlverfahren zusammenhängen kann. Die Wahl von nicht akzeptierten Verfahren kann Bewerbenden durch deren Ankündigung vor dem Verfahren abschrecken und verursachen, dass Bewerbenden die Bewerbung zurückziehen. Darüber hinaus lässt sich von der Wahl des Auswahlverfahrens auf die Organisation schließen. Dies könnte dazu führen, dass sie die Einstellung sowie Verhaltensweisen sowohl der Angestellten als auch der abgelehnten Bewerbenden beeinflusst. So können sich abgelehnte Bewerbenden beispielsweise in Gesprächen mit potenziellen Bewerbenden abwertend über die Organisation äußern und eine mögliche Bewerbung verhindern. 2018 verschärfte Kersting seine Aussagen zur sozialen Akzeptanz aufgrund des Fachkräftemangels noch einmal. Er war der Meinung, dass der Personalmangel dazu führt, dass die Personalauswahl so gestaltet wird, dass die Akzeptanz der Kandidat:innen an erster Stelle steht und so aussagekräftige Verfahren vernachlässigt werden. Jedoch zeigte Kersting (2018) in seiner Übersichtsarbeit auch, dass valide Verfahren von den Bewerbenden akzeptiert werden, wenn dennoch gewisse Regeln zur Akzeptanz (zum

Beispiel Umgang mit den Bewerbenden, konstruktives sowie zeitnahes Feedback) berücksichtigt werden. Somit zeigte er, dass sich die akzeptierten Verfahren und valide Verfahren nicht gegenseitig ausschließen und somit auch valide Verfahren eingesetzt werden können.

Da diese sichtbare Diskrepanz zwischen der Validität von eignungsdiagnostischen Verfahren und der sozialen Akzeptanz besteht, wurde dieses Thema weit erforscht und das Forschungsfeld erfuhr vor allem durch die Digitalisierung neuen Aufschwung (z.B. Basch et al., 2020; Blacksmith et al, 2016). Zahlreiche Forschungsarbeiten beschäftigten sich mit der Frage, ob online oder virtuell durchgeführte Verfahren genauso akzeptiert sind wie die herkömmlichen paper-pencil Verfahren oder Präsenzvarianten. 2016 betrachteten Blacksmith et al. die Frage, inwiefern technologievermittelte Interviews, beispielsweise Video- oder Telefoninterviews, zu den gleichen Bewertungen der Beobachtende wie ein Face-to-Face Interview führen. Dies erforschten sie anhand einer Meta-Analyse. Die Ergebnisse zeigten, dass die Reaktion der Bewerbenden auf die technologievermittelten Verfahren geringer war als im Face-to-Face Interview (d = -.36). Dabei wurden die Telefoninterviews (d = -.26) von den Bewerbenden besser bewertet, als die Videointerviews (d = -.36). Dies könnte daran liegen, dass auch die Bewerbenden den Eindruck hatten, nicht ihr gesamtes Potential zeigen zu können und die Situation als zu unpersönlich ansahen. Darüber hinaus fanden die Autor:innen eine geringere Bewertung der Interviewer:innen in dem virtuellen Verfahren im Vergleich zum Präsenzverfahren (d =-.41). Dieses Ergebnis wurde jedoch durch das Versuchssetting moderiert (Feldstudien: d = -.59; Laborstudien: d = -.22), da die Bewertungen in den realen Interviews geringer waren. Basch et al. (2020) replizierten die Ergebnisse anhand eines Experimentes und fanden auch eine schlechtere Bewertung der Interviewer:innen in dem virtuellen Verfahren im Vergleich zum Face-to-Face Verfahren ($F_{(2, 110)}$ = 4,60, p < .01). Ein weiteres Ergebnis war, dass die Darbietungsart negativ mit der Gerechtigkeitswahrnehmung der Bewerbenden zusammenhing (r = -.21). Dieses Ergebnis bedeutet, dass in dem Videointerview die Gerechtigkeitswahrnehmung mit einer höheren Wahrscheinlichkeit negativ bewertet wird im Vergleich zum Face-to-Face Interview. Darüber hinaus gab es bei der Videointerview Versuchsgruppe größere Bedenken zum Thema Privatsphäre. Die Ergebnisse zeigten, dass durch die unterschiedlichen Reaktionen der Bewerbenden auf die Darbietungsart sowie die Bewertung der Interviewer:innen ein Wechsel zwischen den Darbietungsarten innerhalb eines Auswahlprozesses vermieden werden sollte, damit die Auswahlentscheidung nicht hinsichtlich der Bewertung der Interviewer:innen verfälscht wird. In einer Übersichtsarbeit schlugen Basch und Melchers (2020) vor, dass in

weiterführenden Forschungsarbeiten die Validität von Face-to-Face Interviews und Videointerviews verglichen werden sollte, um zu überprüfen, ob beide Darbietungsarten die gleiche Aussagekraft in Bezug auf den Berufserfolg aufweisen. Unabhängig von dem Vergleich der Validität schlugen Kersting und Ziegler (2020) vor, dass sowohl die Interviewer:innen als auch die Bewerbende bei dem Einsatz neuerer Darstellungsarten zuvor besser geschult werden müssen, damit die Akzeptanz und die Durchführungsgenauigkeit verbessert werden kann. So sollte bei einem Videointerview den Bewerbenden zuvor die Möglichkeit gegeben werden, die Technik ausreichend auszuprobieren, um Sicherheit zu geben und Stress zu reduzieren.

Im Bereich der Intelligenztests und der Persönlichkeitsfragebögen sieht die Entwicklung ähnlich aus. Viele Forschungsarbeiten thematisieren die Reaktionen der Bewerbenden und ziehen einen Bezug zu Onlineverfahren (z.B. Petri et al., 2019). Wie Hossiep et al. (2015) zeigen konnten, werden zwar Persönlichkeitsfragebögen vermehrt eingesetzt, häufig jedoch nicht standardisierte und wissenschaftsfundierte Verfahren. So wurde der stark wissenschaftlich erforschte Fragebogen NEO-FFI nur in 5% der Fälle eingesetzt. Dies zeigt, dass es der Wissenschaft nicht gelingt, die Vorteile der wissenschaftlichen Fundierung der Praxis nahe zu bringen. Darüber hinaus werden Persönlichkeitsfragebögen oft aufgrund von Vorurteilen nicht eingesetzt (Beermann et al., 2013). In ihrer Forschungsarbeit präsentierten sie den Versuchspersonen einen Persönlichkeitsfragebogen sowie einen Intelligenztest. Im Anschluss sollten diese einen Akzeptanzfragebogen ausfüllen. Die Ergebnisse zeigten, dass der Persönlichkeitsfragebogen und der Intelligenztest in der Akzeptanz der Teilnehmenden auf einer sechsfach abgestuften Skala gut bewertet wurden (Persönlichkeitsfragebogen: $M = 4.48$ $SD = .83$; Intelligenzfragebogen: $M = 4.49$ $SD = 1.06$). Auffällig war, dass diese beiden Verfahren als sehr kontrollierbar eingestuft wurden ($M = 5.53$), da die Teilnehmenden zu jeder Zeit wussten, was von ihnen gefordert wird. Zum berufsbezogen formulierten Persönlichkeitsfragebogen gaben die Teilnehmenden an, dass sie sich nicht in ihrer Privatsphäre gestört fühlten (Wahrung der Privatsphäre: $M = 5.13$ $SD = .92$). Als Fazit aus der Erhebung folgt, dass die geringe Akzeptanz eher ein Vorurteil oder einen Mythos darstellt. Da vor allem der Berufsbezug positiv bewertet wurde, sollte darauf geachtet werden, dass ein berufsbezogener Persönlichkeitsfragebogen in der Praxis verwendet wird. Auch Petri et al. (2019) hinterfragten, ob die Gerechtigkeitswahrnehmung bei online oder paper-pencil durchgeführten Leistungstests vergleichbar ist. Die Ergebnisse ihrer Studie zeigten, dass die unbeaufsichtigten Onlinetestungen eine höhere Gerechtigkeitswahrnehmung

bei den Bewerbenden hervorruft ($M_{beaufsichtigt}$ = 4.57 & $M_{unbeaufsichtigt}$ = 4.79). Jedoch sollte auf eine möglichst lärmfreie Umgebung geachtet werden. Lawrence et al. (2009) fanden heraus, dass die Personen, die eine negative Störung, beispielsweise durch eine schlechte Internetverbindung oder Lärm während der Testung, hatten, den Test negativer wahrnahmen als Personen, welche die Testung in Ruhe und ungestört zuhause bearbeiteten (F = 5.134, p < .01). Daraus folgt, dass ein Hinweis zur ruhigen Raumatmosphäre und auf eine gesicherte Internetverbindung für die Wahrnehmung der Bewerbenden hilfreich wären.

Darüber hinaus wurde in diesen Studien deutlich, dass die Rahmenbedingungen einen wichtigen Faktor für die Gerechtigkeitswahrnehmung der Bewerbenden darstellen (Lawrence et al., 2009). Zu beachten ist, dass die Gerechtigkeitswahrnehmung lediglich in Interviews, Intelligenztests und Persönlichkeitsfragebögen untersucht wurde, während Simulationsaufgaben oder Assessment Center bisher unbeachtet blieben.

Modell der Bewerbendenreaktionen in Auswahlprozessen

Der vorherige Abschnitt hat gezeigt, dass unterschiedliche eignungsdiagnostische Verfahren verschieden gerecht wahrgenommen werden und die Gerechtigkeitswahrnehmung über die Zeit hinweg abnimmt (z.B. Konradt et al., 2020; Lawrence et al., 2009). Jedoch beantworteten die aufgeführten Studien nur bedingt die Hintergründe zu diesen variierenden Bewerbendenreaktionen. Aus diesem Grund wird in diesem Abschnitt auf die verschiedenen Determinanten der Gerechtigkeitswahrnehmung eingegangen, um zu verdeutlichen, welche Determinanten in der Praxis berücksichtigt werden müssen, da sie zu einem ungerecht wahrgenommenen Verfahren führen können und somit die Gerechtigkeitswahrnehmung über die Zeit hinweg mehr abnimmt als bei einem gerecht empfundenen Verfahren (Konradt et al., 2020).

In den frühen 1990er Jahren begannen Forscher:innen, wie beispielsweise Gilliland (1993) Determinanten der Bewerbendenreaktionen zu betrachten und Theorien zu entwickeln. Dabei wurde der Frage nachgegangen, welche Determinanten auf Seiten sowohl der Unternehmen als auch der Bewerbenden die Bewerbendenreaktionen beeinflussen könnten. Ryan und Ployhart (2000) unterstrichen die Wichtigkeit dieser Forschungsarbeiten, da sie davon ausgingen, dass die Wahrnehmung der Bewerbenden einen Einfluss auf die Einstellung zum Unternehmen und auf die Annahme einer potenziellen Stelle besitzt. Daher war es ihr Forschungsziel herauszufinden, wann und warum Bewerbenden Auswahlverfahren positiver oder negativer bewerten, damit dies bei der Konzipierung der Verfahren von Seiten der Unternehmen Berücksichtigung finden kann.

Die Theorie zur Bewerbendenreaktion in Auswahlverfahren von Gilliland (1993) war die erste und weitreichend zitierte Forschungsarbeit, welche von Ryan und Ployhart (2000) sowie Hausknecht et al. (2004) erweitert wurde. Gilliland (1993) betrachtete die Bewerbendenreaktion anhand einer theoretischen Herleitung von Forschungsarbeiten des Bereiches der organisationalen Gerechtigkeitswahrnehmung. Er unterteilte die Gerechtigkeitswahrnehmung von Auswahlverfahren in die zwei verschiedenen Bereiche, die prozedurale Gerechtigkeit (während des Auswahlprozesses) sowie die distributive Gerechtigkeit (während der Auswahlentscheidung). Gilliland (1993) entwickelte zehn Gerechtigkeitsregeln, anhand derer Bewerbende die Gerechtigkeit des Auswahlverfahrens bewerteten (s. Abbildung 5). Diese Regeln beinhalten drei verschiedene Themengebiete. Zum einen werden formelle Aspekte des Verfahrens bewertet, welche beispielsweise den Aufgabenbezug zu der Tätigkeit, jedoch auch die Möglichkeit einer zweiten Chance der Bewerbenden beinhalten. Der zweite Themenkomplex beschreibt die Transparenz des Verfahrens sowie die Art und Weise der Rückmeldungen an die Bewerbenden. Mittels der letzten drei Regeln bewerten die Bewerbenden die Gerechtigkeitswahrnehmung anhand der zwischenmenschlichen Behandlung. Diese besagt, dass beispielsweise eine wechselseitige Kommunikation bestehen sollte und die Beobachtende des Unternehmens sozial kompetent sein sollten.

Das Modell von Gilliland (1993) besagt, dass die Bewerbenden das Auswahlverfahren anhand der prozeduralen und distributiven Gerechtigkeit bewerten und so ein Urteil über den gesamten Auswahlprozess und das gesamte Auswahlergebnis gefällt wird. Die beiden Bewertungen führen wiederum zu drei verschiedenen Ergebnissen. Die ersten Ergebnisse zeigen sich schon während des Auswahlverfahrens anhand von Testmotivation und einer eventuellen rechtlichen Auseinandersetzung. Es führt jedoch auch in diesem Stadium schon dazu, dass die Bewerbenden das Stellenangebot reflektieren und sich unter Umständen gegen einen weiteren Bewerbungsprozess entscheiden. Die Ergebnisse nach dem Auswahlprozess beschreiben Aspekte nach der Annahme des Stellenangebotes. Durch den Auswahlprozess können Schlüsse auf das Unternehmensklima gezogen werden. Gerechtigkeitswahrnehmungen können auch die Arbeitsleistung oder -zufriedenheit beeinflussen. Als drittes Ergebnis nannte Gilliland (1993) die Selbstwahrnehmung, welche hauptsächlich von der Fairness des Auswahlergebnisses und nur bedingt von der Fairness des Prozesses beeinflusst wird. Die Selbstwahrnehmung beschreibt die Selbstwirksamkeit und das Selbstwertgefühl, jedoch auch zukünftige Absichten in Hinblick auf die Arbeitssuche.

Angelehnt an das Modell von Gilliland (1993) entwickelten Ryan und Ployhart (2000) als Teil einer Übersichtsarbeit ein Modell, welches die prozedurale und distributive Gerechtigkeitswahrnehmung beinhaltete (s. Abbildung 6). Ryan und Ployhart (2000) gingen davon aus, dass vier verschiedene Aspekte die Wahrnehmung der Bewerbenden beeinflussen. Sie benannten die Persönlichkeit der Bewerbenden, die Eigenschaften der Stelle (zum Beispiel Anforderungen), die Prozesseigenschaften und den Organisationskontext. Die Wahrnehmungen der Bewerbenden führten auch in diesem Modell zu Ergebnissen, welche jedoch im Vergleich zu Gilliland (1993) anders benannt wurden. Teile der Ergebnisse sind die Verfahrensleistung, Selbstwahrnehmung, Wahrnehmung zur Stelle/ Organisation sowie das Verhalten der Bewerbenden.

Abbildung 5

Die Regeln der Bewerbendenreaktionen nach Gilliland (1993), Abbildung und Übersetzung aus Böge (2016)

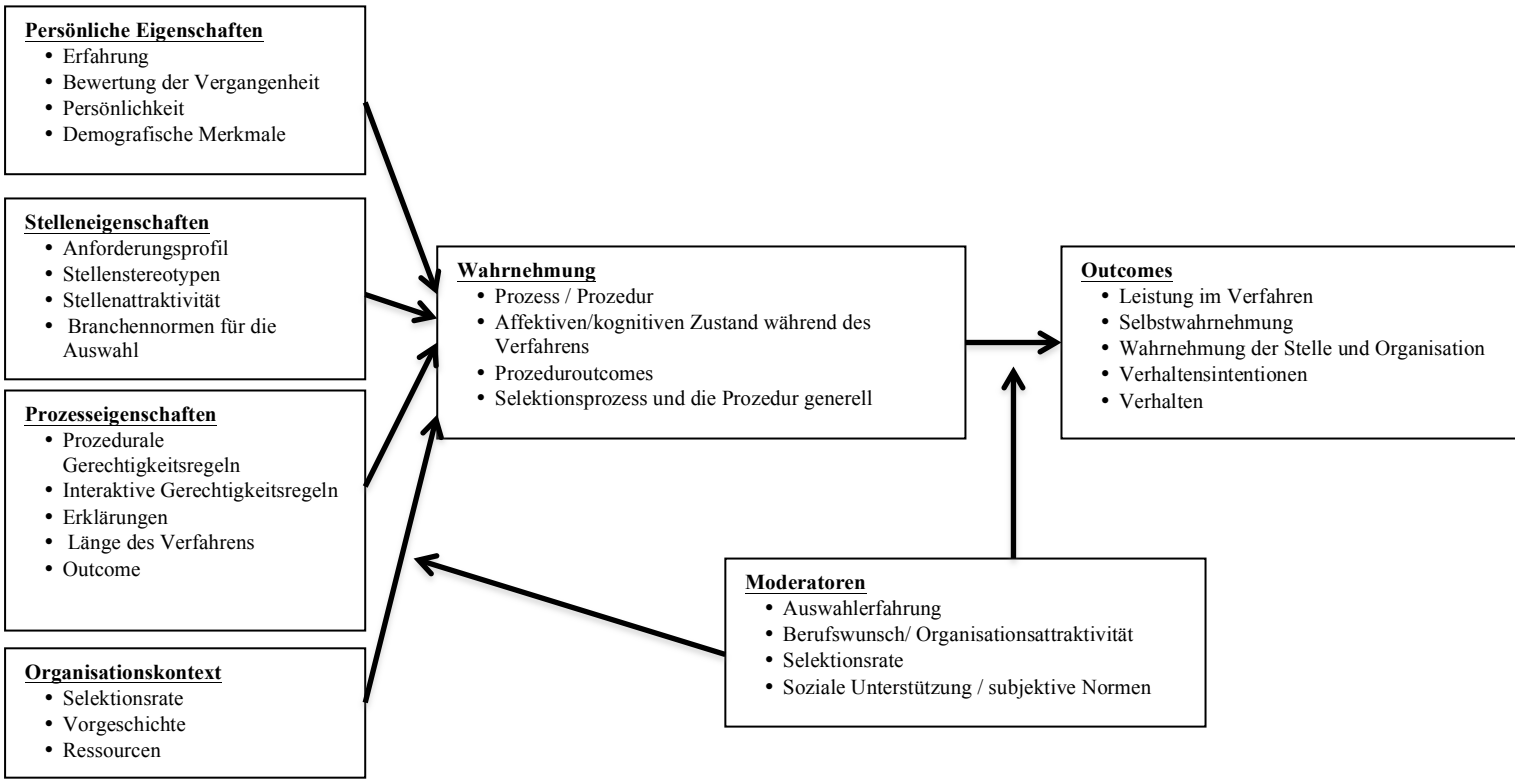
Formelle Aspekte des Verfahrens	
1. Regel:	Die erste Regel hebt den Anforderungsbezug des jeweiligen Test- bzw. Auswahlverfahrens zu der in Frage stehenden Tätigkeit hervor. Dabei wird dieser Anforderungsbezug von der wahrgenommenen Augenscheinvalidität beeinflusst.
2. Regel:	Test- bzw. Auswahlverfahren werden von Bewerber als gerecht erlebt, wenn sie sich einbringen können und das Gefühl haben, dass ihre Fähigkeiten, Fertigkeiten und Kenntnisse berücksichtigt werden.
3. Regel:	Die dritte Regel beschreibt, dass Test- bzw. Auswahlverfahren dann als gerecht erlebt werden, wenn Bewerber den Eindruck haben, dass das Verfahren die Möglichkeiten einer zweiten Chance und des Nachbesserns vermitteln.
4. Regel:	Als gerecht empfundene Auswahlverfahren zeichnen sich durch eine standardisierte Handhabung des Auswahlprozesses unabhängig von Zeit und Bewerber aus.
2. Erklärungen der Vorgehensweisen und des Ergebnisses	
5. Regel:	Gerechtigkeit in der Vorgehensweise wird erlebt, wenn Bewerbern zeitnah eine aussagekräftige Rückmeldung gegeben wird.
6. Regel:	Test- bzw. Auswahlverfahren werden von Bewerber als gerecht erlebt, wenn ihnen die Vorgehensweise transparent erläutert wird und ihnen Informationen über die Bewertungsregeln vorliegen.
7. Regel:	Die Beteiligten seitens des Unternehmens sollten aufrichtig und glaubwürdig wirken, um die wahrgenommene Gerechtigkeit durch die Bewerber zu steigern.
3. Art und Weise der zwischenmenschlichen Behandlung	
8. Regel:	Beteiligte am Auswahlverfahren sollten sozial kompetent wirken und den Bewerbern interessiert sowie freundlich begegnen.
9. Regel:	Bewerber wollen das Gefühl einer „wechselseitigen Kommunikation“ haben – daher sollte das Test- bzw. Auswahlverfahren Bewerbern die Gelegenheit bieten, Fragen stellen und eigene Ansichten einbringen zu können.
10. Regel:	Die zehnte Regel beschreibt, dass Fragen der Organisationsbeteiligten an die Bewerber angemessen und mit Bezug auf die in Frage stehende Tätigkeit angemessen sein sollen.

Anhand des Modells erarbeiteten Ryan und Ployhart (2000) Verhaltensempfehlungen für Unternehmen, um die positive Wahrnehmung der Bewerbenden des Auswahlverfahrens

zu fördern. Da die Einstellung der Bewerbenden zu einem Test deren Motivation beeinflussen kann, sollte vor dem Einsatz des Verfahrens geprüft werden, ob das Verfahren Bewerbenden demotiviert oder ihre Angst steigert. Bewerbenden eine geeignete Absage zu geben, ist zumeist schwierig, da sie nicht zu rational oder verzerrt sein sollte. Zu diesem Thema gibt es jedoch noch keine geeigneten Forschungen. Vermutlich aus diesem Grund geben die meisten Unternehmen eine Standardaussage (zum Beispiel: Es gab qualifiziertere Bewerbenden.) bei der Absage an. Als dritte Empfehlung nannten die Autor:innen die regelmäßige Evaluation der Wahrnehmung der Bewerbenden, So können schnell Kritikpunkte erkannt und verbessert werden. Darüber hinaus forderten die Autor:innen die Forschung auf, genauere Zusammenhänge zwischen der Wahrnehmung und dem Verhalten der Bewerbenden zu untersuchen, beispielsweise Gründe für die Annahme des Stellenangebotes, damit die Wirkgrößen differenziert erfasst und so konkretere Handlungsempfehlungen für Unternehmen hergeleitet werden können. Zuletzt mahnen die Autor:innen die Tatsache an, dass die Wahrnehmung nicht ausschließlich das Verfahren betrifft, sondern auch die Bewertung. So sollte bei der Auswahl des Verfahrens nicht nur die Gerechtigkeitswahrnehmung der Bewerbenden ein ausschlaggebender Faktor sein, sondern auch die Gütekriterien dürfen nicht vernachlässigt werden.

Abbildung 6

Modell der Bewerbendenreaktionen nach Ryan und Ployhart (2000) (eigene Darstellung).



Im Jahr 2004 untersuchten auch Hausknecht et al. die Bewerbendenreaktionen mithilfe einer Metaanalyse und entwickelten das Modell von Ryan und Ployhart (2000) weiter. Diese Metaanalyse beinhaltete neben den Forschungsarbeiten von Ryan und Ployhart (2000) sowie Gilliland (1993) 84 weitere Studien. Die Ergebnisse zeigten, dass das Alter und das Geschlecht nicht signifikant mit der Bewerbendenreaktionen korrelieren (r = -.02 bzw. r = .02). Die Betrachtung der Persönlichkeitsmerkmale verdeutlichten, dass Gewissenhaftigkeit (r = .08) und Neurotizismus (r = -.04) nicht signifikant mit der prozeduralen Gerechtigkeit sowie Gewissenhaftigkeit nicht signifikant mit Testmotivation zusammenhängen (r = .20). Ryan und Ployhart (2000) zeigten jedoch, dass die Offenheit für Erfahrung positiv mit den Bewerbendenreaktionen auf innovative Auswahlverfahren korrelieren könnte. Dies müsste jedoch weiter erforscht werden. Diese Forschungsarbeit konnte die Gerechtigkeitsregeln von Gilliland (1993) stützen, da sie Korrelationen zwischen den Verfahrensaspekten (zum Beispiel Bezug zur Tätigkeit) und der prozeduralen Gerechtigkeit oder Testmotivation finden konnte. Auch in Bezug auf die Ergebnisse fanden die Autor:innen zumeist moderate bis große Zusammenhänge mit der Wahrnehmung, jedoch waren die Zusammenhänge mit der Verfahrensleistung und der Selbstwirksamkeit geringer ausgeprägt. Die Autor:innen warnten davor, diese Forschungsgebiete nicht weiter zu betrachten, da die vorherigen Studien hauptsächlich nur Intentionen erfassten und so die Zusammenhänge verzerrt worden sein könnten. Hausknecht et al. (2004) setzten in ihrer Forschungsarbeit einen weiteren Schwerpunkt, welcher die Wahrnehmung der Bewerbenden in Hinblick auf die Auswahlinstrumente beinhaltete. Sie zeigten, dass Interviews (M = 3.84) und Arbeitsproben (M = 3.63) von den Bewerbenden positiver bewertet wurden als Test- und Fragebogenverfahren ($M_{\text{Testverfahren}}$= 3.14; $M_{\text{Persönlichkeitsfragebogen}}$= 2.88). Zusammengefasst konnte diese Forschungsarbeit viele Aspekte von Gilliland (1993) sowie Ryan und Ployhart (2000) bestätigen und die Wichtigkeit der Betrachtung von Bewerbendenreaktionen unterstreichen.

In einer neueren Metaanalyse betrachteten Anderson et al. (2010) weitere Aspekte der Bewerbendenreaktion in dem sie diese zwischen verschiedenen Auswahlmethoden und in unterschiedlichen Ländern verglichen. Dies war wichtig, damit die Unternehmen noch genauer wissen, welche Methoden von den Bewerbenden als gerecht empfunden werden und ob in internationalen Prozessen länderspezifische Eigenschaften berücksichtigt werden sollten. Die Ergebnisse zeigten, dass die Auswahlmethoden in drei Gruppen unterteilt werden konnten. Am meisten wurden die Arbeitsprobe (M = 5.38, CI [6.47 | 4.29]) und das Interview (M = 5.22, CI [5.43 | 5.00]) bevorzugt. Als positiv wurden die Lebensläufe (M = 4.97, CI

[4.97 | 4.97]), kognitiven Tests (*M* = 4.59, *CI* [5.30 | 3.89]), Referenzen (*M* = 4.36, *CI* [4.76 | 3.96]) und Persönlichkeitsfragebögen (*M* = 4.08, *CI* [4.61 | 3.54]) von den Bewerbenden bewertet. Am schlechtesten in der Bewerbendenreaktion wurden die Ehrlichkeitstests (*M* = 3.69, *CI* [4.41 | 2.96]), persönlichen Kontakte (*M* = 2.59, *CI* [3.07 | 2.10]) und die Graphologie (*M* = 2.33, *CI* [2.70 | 1.96]) von den Bewerbenden wahrgenommen. Es konnten nicht nur die Auswahlmethoden in drei Gruppen geclustert werden, sondern die Ergebnisse zeigten auch, dass die Unterschiede zwischen den Ländern in den Vertrauensintervallen sehr gering sind, da nur 2,5% der Ergebnisse im Mittelwert unter dem 95%-Vertrauensintervall lagen. Daraus kann geschlossen werden, dass die Ergebnisse der Studie von Anderson et al. (2010) in verschiedenen Ländern verallgemeinerbar sind und somit keine Länderunterschiede bei internationalen Verfahren berücksichtigt werden müssen. Weitere Erkenntnisse waren in der Studie von Anderson et al. (2010), dass Bewerbende Auswahlgespräche bzw. Interviews (*M* = 5.54, *CI* [6.40 | 4.67]) aussagekräftiger fanden, als viele andere Methoden. Dies widerspricht der Erkenntnis von Schmidt & Hunter (1998), welche die Validität von Interviews nicht hoch einschätzten. Dies stellt eine Diskrepanz zwischen der Forschung zur Validität und dem Empfinden der Bewerbenden dar.

Truxillo et al. (2009) betrachteten einen weiteren Aspekt der Gerechtigkeitswahrnehmung anhand einer Metaanalyse. Sie schauten, inwiefern gegebene Erklärungen mit der Gerechtigkeitswahrnehmung der Bewerbenden zusammenhängen. Die Ergebnisse verdeutlichten, dass Erklärungen nicht nur die Gerechtigkeitswahrnehmung der Bewerbenden erhöht (*M* = .12; *CI* [.06 |.18]), sondern auch die Attraktivität des Unternehmens stärkt (*M* = .06; *CI* [.01 |.1]). Dies würde für die Praxis bedeuten, dass man durch kleine und kostengünstige Erklärungen die Attraktivität des Unternehmens erhöhen kann (Truxillo et al., 2009). Des Weiteren bestehen auch Zusammenhänge zwischen den Erklärungen und der Testmotivation (*M* = .21; *CI* [.14 | .28]) sowie kognitiven Fähigkeiten (*M* = .09; *CI* [.03 | .14]). Truxillo et al. (2009) schlossen aus diesen Ergebnissen, dass Erklärungen die Motivation und die Testleistungen erhöhen können und dadurch die wahren Fähigkeiten besser eingeschätzt werden können. Dies könnte auch auf eine Erhöhung der Validität von kognitiven Fähigkeitstests hindeuten. Für die Replizierbarkeit dieser Ergebnisse war jedoch interessant, dass Truxillo et al. (2009) herausfanden, dass die Zusammenhänge in Feldstudien stärker waren als in Laborstudien und auch Studien mit Studierendenstichproben geringere Zusammenhänge fanden. Dies sollte in der Konzipierung von nachfolgenden Studien bedacht werden.

Zusammenhänge zwischen der Gerechtigkeitswahrnehmung und der Persönlichkeit

Gilliland (1993), Ryan und Ployhart (2000) sowie Hausknecht et al. (2004) berücksichtigten die Persönlichkeit als einen Aspekt, welcher mit der Gerechtigkeitswahrnehmung von Bewerbenden einhergeht. Nachfolgend betrachteten viele Studien die Zusammenhänge mit den einzelnen Big Five Dimensionen detaillierter. Dies ist entscheidend für die Unternehmen, da die verschiedenen Persönlichkeitseigenschaften unterschiedlich mit der Gerechtigkeitswahrnehmung zusammenhängen und darüber hinaus unterschiedliche Medien bevorzugt werden (z.B. Hertel et al., 2008; Moldzio, 2014). Diese Aspekte sollten für ein gerecht wahrgenommenes Verfahren von den Unternehmen berücksichtigt werden. Ryan und Ployhart (2000) mutmaßten, dass Offenheit für Erfahrung positiv mit der Einstellung zu innovativen Auswahlverfahren zusammenhängen könnte. Van Vianen et al. (2004) griffen dies auf und untersuchten die Entwicklung der Gerechtigkeitswahrnehmung von Bewerbenden. Dies fand in einem realen Auswahlverfahren statt, bei dem Bewerbende vor und während des Verfahrens, welches aus Test- und Fragebogenverfahren bestand, sowie nach dem Feedback die Gerechtigkeitswahrnehmung bewerten sollten. Die Ergebnisse zeigten, dass ein hoher Wert im Persönlichkeitsmerkmal Offenheit für Erfahrung ($d = .04$, $p < .05$), ein hoher Arbeitsbezug ($d = .04$, $p < .05$) sowie eine hohe wahrgenommene eigene Leistung ($d = .06$, $p < .001$) positiv mit der Gerechtigkeitswahrnehmung zusammenhängen. Dies bedeutet, dass Personen mit einer hohen Ausprägung im Merkmal Offenheit für Erfahrung mit einer höheren Wahrscheinlichkeit ein Verfahren als fair erachten als Personen mit niedrigeren Ausprägungen, was die Erkenntnisse von Ryan und Ployhart (2000) stärken würde. Dieser Zusammenhang wurde zu allen drei Testzeitpunkten sichtbar. Auch Brenner et al. betrachteten im Jahr 2016 das Merkmal Offenheit für Erfahrung, jedoch im Zusammenhang mit der Technologieakzeptanz und asynchronen Videointerviews. Die Ergebnisse deuteten darauf hin, dass Offenheit für Erfahrung ein Moderator zwischen der wahrgenommenen Nützlichkeit und der Einstellung zu asynchronen Interviews sein könnte ($\beta = .22$, $p < .05$). Demgegenüber fanden die Autor:innen auch einen positiven Zusammenhang zwischen Gewissenhaftigkeit und der Einstellung zum Verfahren ($\beta = .23$, $p < .05$), jedoch nicht mehr mit Einbezug der wahrgenommenen Nützlichkeit ($\beta = .13$, $p > .05$) und Benutzerfreundlichkeit ($\beta = .14$, $p > .05$). Dies würde bedeuten, dass Personen mit einer hohen Ausprägung im Merkmal Gewissenhaftigkeit asynchrone Interviews eher positiv bewerten würden.

Moldzio (2014) untersuchte die Zusammenhänge zwischen der Akzeptanz und der Persönlichkeit in realen Auswahlprozessen anhand von einer Stichprobe mit Bewerbenden in

der Haushaltgerätebranche, welche Persönlichkeitsfragebögen als Teil des Auswahlprozesses bearbeiteten. Die Ergebnisse zeigten, dass ein signifikant negativer Zusammenhang zwischen der Big Five Dimension Neurotizismus und Akzeptanz besteht (r = -.15, p <.05), was darauf schließen lässt, dass emotional stabilere Bewerbende mit einer höheren Wahrscheinlichkeit die Akzeptanz des Auswahlverfahrens positiver bewerten als emotional instabile Bewerbende. Darüber hinaus waren positive Signifikanzen zwischen Extraversion (r = .25, p <.01) sowie Offenheit für Erfahrung (r = .23, p <.01) und Akzeptanz sichtbar. Dies würde bedeuten, dass diese beiden Persönlichkeitsdimensionen nicht nur für die Kompetenz in der Digitalisierung immer wichtiger werden, sondern auch für die Akzeptanz von Auswahlverfahren eine bedeutsame Rolle spielen.

Konradt et al. (2016) hinterfragten in ihrer Studie neben dem Verlauf der Gerechtigkeitswahrnehmung in einem Auswahlverfahren auch die Rolle der Big Five Dimensionen. Sie stellten die Hypothesen auf, dass die Gerechtigkeitswahrnehmung über die Zeit hinweg abnimmt, dass ein hohes Anfangsniveau für eine geringere Abnahme der Gerechtigkeitswahrnehmung spricht sowie dass drei verschiedene Bewerbergruppen bestehen, welche unterschiedliche Anfangs- und Endniveaus in der Gerechtigkeitswahrnehmung aufweisen. Die Stichprobe wurde in einem Auswahlverfahren anhand von drei Messzeitpunkten (vor, während und nach dem Verfahren) betrachtet. Die Ergebnisse deuten darauf hin, dass die ersten beiden Hypothesen bestätigt werden können, da die Gerechtigkeitswahrnehmung kontinuierlich abgenommen hat (M = –.15, p < .05) und Personen mit einer hohen anfänglichen Gerechtigkeitswahrnehmung auch einen geringeren Abfall über die Zeit verzeichneten (M = –.22, p < .001). Bei der Betrachtung der Persönlichkeit wurde deutlich, dass eine niedrigere Ausprägung in den Merkmalen Extraversion (F = 3.23, p < .05) und Verträglichkeit (F = 3.26, p < .05) im Zusammenhang mit der Gerechtigkeitswahrnehmung steht und somit die Fairnesswahrnehmung eher geringer eingeschätzt wird.

Zuvor wurde ersichtlich, dass die Gerechtigkeitswahrnehmung mit einzelnen Persönlichkeitsmerkmalen zusammenhängt (z.B. Konradt et al., 2016; Moldzio, 2014). In Bezug auf die Digitalisierung könnte jedoch noch entscheidend sein, ob einzelne Darbietungsarten mit der Gerechtigkeitswahrnehmung in Bezug auf die Persönlichkeit zusammenhängen. Aus diesem Grund wird nachfolgend die Studie von Hertel et al. (2008) betrachtet, welche eine etwas andere Herangehensweise nutzte. Sie untersuchten, welches Medium Personen mit unterschiedlichen Persönlichkeitsmerkmalen eher präferieren und somit mehr akzeptieren würden. So stellten sie die Hypothese auf, dass extravertiertere

Personen eher reichhaltigere Medien bevorzugen, wohingegen Personen mit einer hohen Ausprägung im Merkmal Neurotizismus ein Medium mit einem niedrigeren Level in der Reichhaltigkeit bevorzugen (zum Beispiel Textantworten und kein Video). Die Stichprobe bestand aus $N = 228$ Personen, welche in zwei Gruppen unterteilt wurden. Die eine Gruppe kommunizierte mit einem reichhaltigen Medium, dem Face-to-Face Gespräch, und die zweite Gruppe anhand von E-Mails. Analog zu den Hypothesen konnte gezeigt werden, dass Personen mit einer hohen Ausprägung im Merkmal Extraversion ($\beta = .22$, $p < .001$) und einer niedrigen Ausprägung im Merkmal Neurotizismus ein reichhaltiges Medium bevorzugten ($\beta = -.13$, $p < .05$). Demgegenüber wählten eher introvertierte und ängstliche Personen die Kommunikation per Mails.

Als Fazit der Studien lässt sich festhalten, dass nicht nur die Eigenschaften des Verfahrens für die Bewerbendenreaktionen wichtig sind, sondern ebenfalls die Persönlichkeitsausprägungen der Bewerbenden. Dies lässt sich jedoch schwer berücksichtigen, da das Auswahlverfahren nicht für verschiedene Bewerbendengruppen aufgrund von deren Persönlichkeit variabel gehalten werden kann. Beim Einsatz von neuen und reichhaltigen Medien sollte darauf geachtet werden, dass introvertierte Personen und Personen mit einer hohen Ausprägung im Merkmal Neurotizismus auch ihre bestmögliche Leistung zeigen können.

Entwicklung und Validierung eines Fragebogens zu den Aspekten der Extraversion und Offenheit für Erfahrung

In den nachfolgenden Abschnitten werden die Fragestellung, Methoden sowie die Ergebnisse der ersten Studie dieser Arbeit dargestellt. Die Studie beinhaltet die Entwicklung sowie die Validierung eines berufsbezogen formulierten Fragebogens zu den Aspekten der Persönlichkeitsmerkmale Extraversion und Offenheit für Erfahrung.

Gegenstand der Fragestellung

In dem theoretischen Hintergrund wurde ersichtlich, dass Persönlichkeitsfragebögen ergänzend zu anderen Verfahren ein nützliches Instrument der Eignungsdiagnostik darstellen (Schmidt & Hunter, 1998). Dennoch werden diese aufgrund vieler Vorbehalte, beispielsweise der Gerechtigkeitswahrnehmung der Bewerbenden, nicht eingesetzt (Schuler et al., 2007). Ziel dieser ersten Studie war es, einen berufsbezogenen Fragebogen zur Extraversion und Offenheit für Erfahrung (AEOS; Wedemeyer et al., in Vorbereitung) zu entwickeln und zu überprüfen, ob dieser eine gute Validität aufweist, zudem gewisse weitere Gütekriterien erfüllt und somit in der Eignungsdiagnostik verwendet werden kann.

Kurzabriss der theoretischen und empirischen Grundlagen

Vorangegangene Studien konnten zeigen, dass die Persönlichkeitsmerkmale der Big Five Dimensionen (Neurotizismus, Extraversion, Gewissenhaftigkeit, Verträglichkeit und Offenheit für Erfahrung) nach Costa und McCrae (1992) mit Berufserfolg korrelieren und somit Kriteriumsvalidität aufweisen (z.B. Barrick & Mount, 1991; Salgado, 1997). Außerdem wird durch den Einsatz von Persönlichkeitsfragebögen über Intelligenztests hinaus die prädiktive Validität in eignungsdiagnostischen Verfahren um 18% erhöht, was auch für einen Einsatz dieser Fragebögen spricht (Schmidt & Hunter, 1998). Der Einsatz von psychometrisch fundierten Instrumenten in eignungsdiagnostischen Situationen erhöht somit nachweislich die Wahrscheinlichkeit, passende Kandidat:innen für eine bestimmte Stelle auszuwählen (Schmidt & Hunter, 1998). Demgegenüber scheint eine Diskrepanz zwischen den Erkenntnissen der Wissenschaft und der Einsatzhäufigkeit dieser Fragebögen in der Praxis zu bestehen (Kanning, 2022). Diese Diskrepanz zwischen der Güte der Verfahren und der Einsatzhäufigkeit könnte aufgrund von Vorbehalten der Unternehmen gegenüber der Gerechtigkeitswahrnehmung der Bewerbenden bestehen (Beermann et al., 2013), die allerdings inhaltlich nicht begründet erscheinen. Beermann et al. (2013) zeigten, dass Personen, die Persönlichkeitsfragebögen bearbeiteten, diese als gerecht wahrnahmen und die geringe Gerechtigkeitswahrnehmung eher auf Vorurteilen beruht. Auch neuere Forschungsarbeiten stellten fest, dass die Gerechtigkeitswahrnehmung der Bewerbenden bezüglich Persönlichkeitsfragebögen zwar im Vergleich zu anderen Verfahren nur im mittleren Bereich lag, jedoch als positiv von den Bewerbenden bewertet wurde (Anderson et al., 2010). Um diese Diskrepanz aufzuheben, empfehlen Hausknecht et al. (2004) sowie Beermann et al. (2013) berufsbezogene Verfahren einzusetzen, um die Gerechtigkeitswahrnehmung gegenüber Persönlichkeitsfragebögen zu erhöhen.

Neben der Diskrepanz zwischen der Einsatzhäufigkeit in der Praxis und der Forschung zur Validität von Persönlichkeitsfragebögen wurde in den letzten Jahren auch die Struktur des Big Five Modells von Costa und McCrae (1992) hinterfragt. DeYoung et al. (2007) kamen in ihrer Forschungsarbeit zu dem Ergebnis, dass zwischen den Dimensionen und ihren jeweils sechs Facetten eine weitere Hierarchieebene mit zwei Aspekten pro Dimension besteht. Judge et al. (2013) griffen diese Herangehensweise auf und untersuchten den Zusammenhang dieser Aspekte mit Berufserfolg. Die Ergebnisse zeigten, dass die Facetten sowie Aspekte mehr Varianz gegenüber Berufserfolg aufklärten als die Dimensionen. Somit kann daraus geschlossen werden, dass eine breitere Erfassung der Merkmale eine höhere Aussagekraft für Berufserfolg besitzt.

Nachdem ein psychologischer Persönlichkeitsfragebogen in der Diagnostik entwickelt wird, muss dieser anschließend anhand von Kriterien überprüft werden, um zu betrachten, ob der Fragebogen für einen Einsatz geeignet ist (Bühner, 2021). Um diese Überprüfung zu standardisieren und Richtwerte zu erschaffen haben Westhoff et al. (2010) die DIN 33430 entwickelt, mithilfe derer man einen diagnostischen Prozess gerecht und standardisiert entwickeln und durchführen kann. Diese Norm beinhaltet Richtwerte für diagnostische Verfahren zur Objektivität, Reliabilität, Validität, Eichung und Verfälschbarkeit. Dabei werden die Objektivität, Validität und Reliabilität als Hauptkriterien und die Eichung sowie Verfälschbarkeit als Nebenkriterien kategorisiert, welche an dieser Stelle nicht weiter thematisiert werden (Bühner, 2021). Die Objektivität umfasst die Frage, inwiefern die Ergebnisse eines Tests unabhängig von Testleitenden sind (Bühner, 2021). Dabei kann die Objektivität anhand des Prozesses in drei Arten unterteilt werden, in die Durchführungs-, Auswertungs- und Interpretationsobjektivität. Damit eine hohe Objektivität entstehen kann, ist eine möglichst standardisierte Durchführung (zum Beispiel online Fragebogen), Auswertung (zum Beispiel ein Algorithmus wertet die Fragebögen aus) und Interpretation (zum Beispiel standardisierte Profile) notwendig (Bühner, 2021). Das zweite Gütekriterium ist die Reliabilität, welche den Grad der Messgenauigkeit des Fragebogens untersucht (Bühner, 2021). Unter dem Grad der Messgenauigkeit ist die Frage zu verstehen, ob der Fragebogen auch das Merkmal misst, was er messen soll. Auch die Reliabilität kann auf drei verschiedenen Arten betrachtet werden. Zum einen gibt die Interne Konsistenz, auch Halbierungsreliabilität genannt, an, inwiefern zwei Hälften des Fragebogens miteinander korrelieren (Bühner, 2021). Die Retest-Reliabilität betrachtet die Frage, inwiefern ein Fragebogen zeitunabhängig ist und zu zwei verschiedenen Zeitpunkten miteinander korreliert (Bühner, 2021). Eine weitere Art der Reliabilität ist die Paralleltestreliabilität, welche die Korrelationen von zwei Fragebögen mit derselben Eigenschaft betrachtet (Bühner, 2021). Bühner (2021) definierte die Validität als das Ausmaß, in dem der Fragebogen das misst (zum Beispiel ein Konstrukt), was er messen soll. So wird bei einem Fragebogen zur Extraversion beispielsweise überprüft, ob er die Extraversion misst. Validität wird auf unterschiedliche Art und Weise betrachtet. Es wird unterschieden nach Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität. Inhaltsvalidität betrachtet, ob ein Item das zu messende Konstrukt erfasst (Bühner, 2021). Mit Konstruktvalidität wird die Frage beantwortet, inwiefern der Fragebogen ein Merkmal wirklich erfasst (Bühner, 2021). Die faktorielle Validität betrachtet dabei, ob die Faktorenstruktur des Fragebogens gegeben ist. Außerdem werden Zusammenhänge zu verschiedenen anderen Fragebögen auf zwei Arten

untersucht. Bei der konvergenten Validität wird betrachtet, ob der Fragebogen hoch mit einem Fragebogen korreliert, der ähnliche Konstrukte erfasst (konstruktverwandte Fragebögen). Darüber hinaus wird überprüft, ob der Fragebogen gering mit Fragebögen oder Skalen korreliert, die andere Konstrukte messen (konstruktfremde Fragebögen; Bühner, 2021). Diese Art der Validität wird divergente Validität genannt. Die divergente sowie konvergente Validität können als nomologisches Netzwerk zusammengefasst werden (Moosbrugger & Kelava, 2020). In der Eignungsdiagnostik bzw. Personalauswahl verfolgt die Kriteriumsvalidität das Ziel, zu überprüfen, ob ein Verfahren (hier Persönlichkeitsfragebögen) berufsbezogene Kriterien, beispielsweise Berufserfolg, vorhersagt (Bühner, 2021). Dies kann auch als prädiktive Validität benannt werden (s. Haupt- und Nebenkriterien eines Persönlichkeitsfragebogens).

Ableitung der Hypothesen

Studien konnten vermehrt zeigen, dass das Persönlichkeitsmerkmal Extraversion wichtige Zusammenhänge mit dem Kriterium Berufserfolg aufweist, vor allem bei Manager:innen sowie Vertriebler:innen (z.B. Barrick & Mount, 1991; Salgado, 1997). In dem Persönlichkeitsmerkmal Offenheit für Erfahrung wurde neben dem Zusammenhang mit Berufserfolg ersichtlich, dass dieses Persönlichkeitsmerkmal immer bedeutsamer wird, da Prozesse einer Beschleunigung von Veränderungen unterliegen und somit ein lebenslanges Lernen immer bedeutsamer sowie notwendig wird (Schermuly et al., 2019). Aus diesen Gründen wurde in dieser Studie die Entwicklung eines Fragebogens auf Grundlage dieser beiden Dimensionen fokussiert.

Neben der Bedeutsamkeit der Dimensionen wurde bei der Erfassung dieser ein neuartiger Weg verwendet. So wurde nicht nur aufbauend auf der Studie von Beermann et al. (2013) der Fragebogen berufsbezogen formuliert, da dies die Gerechtigkeitswahrnehmung der Bewerbenden steigern soll, sondern es wurde zudem eine neue Hierarchieebene auf Grundlage von DeYoung et al. (2007) berücksichtigt. Diese Grundlage besagt, dass beide Big Five Dimensionen jeweils zwei untergeordnete Aspekte besitzen. Extraversion beinhaltet die Aspekte Enthusiasmus und Durchsetzungsfähigkeit und Offenheit für Erfahrung die Aspekte Offenheit und Intellekt. Dieser Fragebogen erfasst somit erstmals die Big-Five Dimensionen berufsbezogen sowie die Einzelaspekte getrennt voneinander. Die Überprüfung der Gütekriterien des Fragebogens erfolgte auf Grundlage der klassischen Testtheorie sowie angelehnt an die Norm DIN 33430 von Westhoff et al. (2010). Demnach wurde auf Grundlage der faktoriellen Validität davon ausgegangen, dass der Fragebogen vier Skalen

(die eben genannten vier Aspekte) misst und die Modellpassung der Aspekte besser als die der Dimensionen ist. Darüber hinaus wurde die Reliabilität überprüft, um zu zeigen, dass der Fragebogen die Konstrukte zuverlässig misst (Gäde, Schermelleh-Engel & Werner, 2020). Um diese Aussage zu prüfen, wurde die Arbeit von George und Mallery (2002) als Grundlage zur Bewertung der Reliabilität verwendet, welche bei einem $\alpha \geq .80$ von einer guten Reliabilität ausgehen.

Bei der Betrachtung der Konstruktvalidität wurde zwischen zwei Arten unterschieden. Einerseits wurde die konvergente Validität anhand der konstruktverwandten Big Five Dimensionen Extraversion bzw. Offenheit für Erfahrung betrachtet (Bühner, 2021). Des Weiteren wurde die divergente Validität mit Hilfe der konstruktfremden Big Five Dimensionen geprüft. Die konstruktfremden Dimensionen wären in diesem Fall die vier anderen Big Five Dimensionen, auf welche die Aspekte nicht aufbauen. Bei den Aspekten Enthusiasmus und Durchsetzungsfähigkeit wären dies dann alle Big Five Dimensionen außer Extraversion und bei Offenheit und Intellekt alle Big Five Dimensionen außer Offenheit für Erfahrung.

Im Bezug zur divergenten Validität wurde darüber hinaus auch betrachtet, ob die zehn Skalen der Aspekte nach DeYoung et al. (2007) des zusammengefassten Fragebogens Business Big 5 (Integrität von ABGS, AVS, AEOS; Moldzio, Wedemeyer& Böge, in Vorbereitung) voneinander abgrenzbar sind oder sich ggf. Skalen ähneln und so der Einsatz beider Skalen redundant ist (Bühner, 2011). Aus diesen Anforderungen wurden Hypothesen abgeleitet, welche in Tabelle 1 abgebildet sind. Als Richtwert für die Güte der konvergenten und divergenten Validität wurden die Regeln nach Bühner (2011) gewählt, welche besagen, dass bei einer konvergenten Validität die Korrelation $r > |.5|$ und bei einer divergenten Validität $r < |.4|$ sein sollte.

Tabelle 1

Hypothesen zur Überprüfung der Gütekriterien sowie der Konstruktvalidität

Nummer	Annahme
H1	Die AEOS bildet die vier Aspekte nach DeYoung et al. (2007) ab. Die Passung des vier Faktoren Modells ist besser als die des zwei Faktoren Modells.
H2	Der Fragebogen AEOS weist eine hohe Reliabilität ($\alpha \geq .80$) nach George und Mallery (2002) auf.
H3	Der Fragebogen AEOS weist mittlere positive Korrelationen zu konvergenten Dimensionen ($r > |.5|$) bzw. Skalen anderer Fragebögen und Korrelationen $r < |.4|$ für divergente Dimensionen auf. **H3a:** Die Aspekte Enthusiasmus und Durchsetzungsfähigkeit korrelieren positiv im mittleren Bereich mit der Dimension Extraversion. **H3b:** Die Aspekte Offenheit und Intellekt korrelieren positiv im mittleren Bereich mit der Dimension Offenheit für Erfahrung. **H3c:** Die Aspekte Enthusiasmus und Durchsetzungsfähigkeit korrelieren $r < |.4|$ mit den Dimensionen Neurotizismus, Gewissenhaftigkeit, Offenheit für Erfahrung sowie Verträglichkeit. **H3d:** Die Aspekte Offenheit und Intellekt korrelieren $r < |.4|$ mit den Dimensionen Neurotizismus, Gewissenhaftigkeit, Extraversion sowie Verträglichkeit. **H3e:** Alle zehn Aspekte der Business Big 5 korrelieren $r < |.4|$ mit einander.

Viele Studien betrachteten zur Überprüfung der Kriteriumsvalidität den Zusammenhang zwischen den Big Five Dimensionen und Berufserfolg und fanden einen positiven Zusammenhang zwischen Berufserfolg und den beiden Dimensionen Extraversion sowie Offenheit für Erfahrung (z.B. Barrick & Mount, 1991; Salgado, 1997). Bei der Betrachtung der jeweiligen Aspekte der beiden Dimensionen in Bezug auf Berufserfolg wurde ersichtlich, dass signifikant positive Korrelationen zwischen Enthusiasmus, Durchsetzungsfähigkeit sowie Intellekt und Berufserfolg bestehen (Judge et al., 2013). Lediglich der Aspekt Offenheit weist keine signifikanten Korrelationen zu Berufserfolg auf. Dies könnte jedoch auf den fehlenden Berufsbezug und die kulturlastigen Formulierung der Items des NEO-FFI (Borkenau & Ostendorf, 2008) zurückzuführen sein (zum Beispiel „Ich probiere oft neue und fremde Speisen aus“). Da der vorliegende Fragebogen jedoch berufsbezogen formuliert wurde und andere Studien (zum Beispiel Barrick & Mount, 1991; Salgado, 1997) von einem Zusammenhang mit der Dimension Offenheit für Erfahrung ausgehen, wurde auch in dieser Studie ein positiver Zusammenhang zwischen dem Aspekt Offenheit und den Kriterien für Berufserfolg angenommen (s. Tabelle 2).

Tabelle 2

Hypothesen zur Kriteriumsvalidität

Nummer	Annahme
H4	Die Aspekte der Dimensionen Extraversion und Offenheit für Erfahrung weisen positive Korrelationen mit dem Kriterium Berufserfolg auf. **H4a:** Der berufsbezogene Aspekt Enthusiasmus weist eine positive Korrelation mit den Kriterien Berufserfolg, Ausbildungserfolg und schlussfolgerndes Denken auf. **H4b:** Der berufsbezogene Aspekt Durchsetzungsfähigkeit weist eine positive Korrelation mit den Kriterien Berufserfolg, Ausbildungserfolg und schlussfolgerndes Denken auf. **H4c:** Der berufsbezogene Aspekt Offenheit weist eine positive Korrelation mit den Kriterien Berufserfolg, Ausbildungserfolg und schlussfolgerndes Denken auf. **H4d:** Der berufsbezogene Aspekt Intellekt weist eine positive Korrelation mit den Kriterien Berufserfolg, Ausbildungserfolg und schlussfolgerndes Denken auf.

Ziel eines neuen Fragebogens ist es, einen Mehrwert gegenüber anderen Fragebögen zu schaffen. In diesem Fall würde der neu entwickelte Fragebogen einen Mehrwert zu den üblichen Fragebögen, welche die Big Five Dimensionen messen, bieten, wenn die prädiktive Validität erhöht wird, dadurch folglich mehr inkrementelle Validität entsteht und somit Varianz über die bisherigen Fragebögen hinweg aufgeklärt wird (Bühner, 2011). Judge et al. (2013) betrachteten sowohl die Zusammenhänge mit dem Kriterium Berufserfolg bezüglich der Dimensionen als auch der Aspekte sowie Facetten. Die Ergebnisse zeigten, dass die Aspekte einen signifikant höheren Zusammenhang zu Berufserfolg gegenüber den Dimensionen aufweisen. Die Studie von Moldzio et al. aus dem Jahr 2021 konnte die Ergebnisse von Judge et al. (2013) stützen. Moldzio et al. (2021) betrachtete jedoch lediglich die Dimensionen Gewissenhaftigkeit sowie Neurotizismus. Angelehnt an die Studien wurde in dieser Forschungsarbeit davon ausgegangen, dass der Fragebogen inkrementelle Validität aufweist und die prädiktive Validität in Bezug auf die berufsbezogenen Kriterien erhöht wird (s. Tabelle 3). Das schlussfolgernde Denken wurde auf Grundlage der Metaanalyse von Schmidt und Hunter (1998) gewählt. Sie fanden heraus, dass die kognitiven Fähigkeiten eine hohe Validität besaßen und untersuchten daher die inkrementelle Validität der anderen Verfahren über Fähigkeitstests hinweg. Da sie herausfanden, dass Persönlichkeitsfragebögen über die kognitiven Fähigkeiten hinweg inkrementelle Validität aufweisen, geht diese Arbeit auch davon aus.

Tabelle 3
Hypothesen zur inkrementellen Validität und dem Vergleich der prädiktiven Validität zwischen den Aspekten und der Dimension

Nummer	Annahme
H5	Die Aspekte weisen eine inkrementelle Validität über deren Dimensionen hinaus auf. **H5a:** Die vier Aspekte weisen inkrementelle Validität im Kriterium Berufserfolg über das schlussfolgernde Denken hinaus auf. **H5b:** Die vier Aspekte weisen inkrementelle Validität im Kriterium Berufserfolg über die Big Five Dimensionen hinaus auf.

Da Studien zeigten, dass beispielsweise Manager:innen eine höhere Durchsetzungsfähigkeit benötigen im Vergleich zu Auszubildenden (Barrick & Mount, 1991), wurden die Hypothesen 4a-d in den verschiedenen Stichproben geprüft. Ziel war es, zu untersuchen, ob in allen Stichproben alle Aspekte mit Berufserfolg korrelieren oder nicht. Da der Zusammenhang zwischen den berufsbezogenen Aspekten und Berufserfolg in den verschiedenen Gruppen noch nicht überprüft wurde, wurden die Richtungen der Hypothesen vorerst beibehalten.

Methoden

Stichprobe

Der neu entwickelte Fragebogen zur Erfassung von arbeitsbezogener Extraversion und Offenheit für Erfahrung wurde in vielen Unternehmen aus unterschiedlichen Branchen in realen Personalauswahlverfahren und Potenzialerkennungsverfahren eingesetzt, um die Gütekriterien in der Praxis zu prüfen. Dabei wurden die Versuchspersonen in Auszubildende, Expert:innen ohne Führungsverantwortung, Linienführungskräfte sowie hochrangige Führungskräfte unterteilt, damit eine detaillierte Betrachtung gewährleistet ist. Zuvor wurde eine Vorstudie an unterschiedlichen Schulen durchgeführt, um die einzelnen Items des Fragebogens auf Verständlichkeit zu prüfen. Darüber hinaus sollten im Anschluss Faktoranalysen berechnet werden, um nicht trennscharfe Items zu identifizieren und den Fragebogen zu kürzen. Eine Übersicht der Stichproben, welche Teil der Hauptstudie sind, ist in der Tabelle 4 dargestellt.

Tabelle 4
Darstellungen der verschiedenen Stichprobengruppen

Stichprobengruppe	Stichprobengröße	Alter *M* (*SD*)	Geschlechterverteilung m/w/d
Hochrangige Führungskräfte	90	44.22 (8.26)	87 / 13 / 0
Linienführungskräfte	475	33.67 (9.65)	71 / 29 / 0
Expert:innen	325	33.88 (8.56)	60 / 40 / 0
Kaufmännische Auszubildende	378	18.49 (2.59)	50 / 50 / 0
Technische Auszubildende	282	17.48 (2.69)	92 / 8 / 0

Anmerkung. Das Durchschnittsalter sowie die Standardabweichung sind in Jahren berechnet worden. Die Geschlechterverteilung wird in Prozent (%) angegeben.

Vorstudie. Für die Vorstudie wurde der Fragebogen von Schüler:innen, welche entweder die Oberstufe eines Gymnasiums oder die Abschlussklasse einer beruflichen Schule besuchten, sowie Studierenden bearbeitet. Diese Stichprobe, welche wenig bis keine Berufserfahrung aufwies, wurde gewählt, um das Verständnis berufsbezogen formulierter Items zu überprüfen. Es wurde davon ausgegangen, dass die Formulierungen des Fragebogens für alle Berufsgruppen verständlich sind, wenn dies in der berufsunerfahrenen Stichprobe der Fall ist. Die Stichprobe bestand aus insgesamt $N = 101$ Teilnehmenden, von denen sich $n = 62$ das weibliche, $n = 38$ das männliche und $n = 1$ das diverse Geschlecht zuschrieben. Da der Anteil der weiblichen Teilnehmerinnen bei 61% lag, war die Stichprobe nicht gleichverteilt. Die Proband:innen waren durchschnittlich $M = 20{,}02$ ($SD = 3{,}12$) Jahre alt. Von Berufserfahrung konnte gesprochen werden, wenn die Teilnehmer:innen mindestens ein mehrwöchiges Praktikum absolviert hatten. Eine fehlende Berufserfahrung war in dieser Studie jedoch kein Ausschlusskriterium, da in der Instruktion explizit enthalten war, das sich diese Proband:innen in Situationen in der Schule hineinversetzen sollen und zum Beispiel die Items auf Grundlage von Zusammenarbeiten mit Mitschüler:innen bewerten sollten. Lediglich $n = 4$ Teilnehmende gaben an, dass sie keine Berufserfahrung aufweisen. In Bezug auf einen vorhandenen Schulabschluss gaben $n = 11$ Proband:innen an, dass sie noch keinen Schulabschluss besitzen. Darüber hinaus verfügten $n = 3$ über einen Hauptschulabschluss, $n = 39$ über eine mittlere Reife, $n = 11$ über eine Fachhochschulreife und $n = 37$ über ein Abitur.

Auszubildende. Die Bewerbenden auf einen Ausbildungsplatz oder ein duales Studium nahmen im Rahmen einer Onlinetestung während der Vorauswahl, welche durch eine Unternehmensberatung unterstützt worden ist, an dieser Studie teil. Es wurden Auszubildende und duale Studierende für kaufmännische und technische Berufe sowie Techniker:innen im Kundendienst, welche vertriebliche Aufgaben in ihrer Ausbildung bearbeiten, gesucht. Die Unternehmen, welche Auszubildende und duale Studierende für Vakanzen suchten, sind dem Chemikalienhandel sowie der Elektrogeräte-, Versicherungs- und Dienstleistungsbranche zuzuordnen.

Kaufmännische Auszubildende und duale Studierende. Unter der Stichprobe der kaufmännischen Auszubildenden werden alle Bewerbenden gefasst, die im Auswahlprozess eine kaufmännische Testbatterie bearbeitet haben. Diese Testbatterie bestand aus einem Test zum logisch-schlussfolgernden Denken, Rechenaufgaben, einem Diktat bzw. einem Test zur deutschen Rechtschreibung sowie mehreren Persönlichkeitsfragebögen. Die Teilnehmenden bewarben sich sowohl auf klassische Ausbildungsberufe, wie zum Beispiel eine Ausbildung im Groß- und Außenhandel oder ein duales Studium in der Betriebswirtschaftslehre als auch auf neuartige Ausbildungsberufe wie beispielsweise „Kaufleute zur Marketingkommunikation“.

Wie in der Tabelle 4 ersichtlich wird, bestand die Gruppe der kaufmännischen Auszubildenden und dualen Studierenden aus insgesamt n = 378 Versuchspersonen. Die Hälfte der Bewerbenden schrieben sich das weibliche Geschlecht und die andere Hälfte das männliche Geschlecht zu. Das Durchschnittsalter lag bei M = 18.49 Jahren (SD = 2.59), was darauf zurückzuführen sein könnte, dass lediglich 4.27 % der Bewerbenden einen Realschul- bzw. Hauptschulabschluss besaßen, während die restlichen Bewerbenden eine Fachhochschulreife bzw. die allgemeine Hochschulreife erreichten.

Technische Auszubildende und duale Studierende. Die Stichprobe der technischen Auszubildenden und dualen Studierenden beinhaltete Bewerbenden, die im Auswahlprozess eine technische Testbatterie bearbeiteten. Diese Testbatterie bestand im Vergleich zu den kaufmännischen Test- und Fragebogenverfahren zusätzlich aus einem technikbasierten Leistungstest sowie zwei Aufgaben zum räumlichen Vorstellungsvermögen. Zu besetzende Ausbildungsplätze waren zum Beispiel Mechatroniker:innen sowie technische Produktdesigner:innen. Darüber hinaus wurden duale Studierende für die Berufe Elektrotechnik oder Maschinenbau gesucht. Auch die Stichprobe der vertrieblichen Berufe wie zum Beispiel Techniker:innen im Kundendienst wurden in diese Stichprobe

miteinbezogen. Sie bearbeiteten die gleiche Testbatterie, lediglich wurden Persönlichkeitsmerkmale unterschiedlich gewichtet. So wurde zum Beispiel bei den Techniker:innen im Kundendienst das Merkmal Extraversion stärker gewichtet.

Die Stichprobengröße der technischen Auszubildenden und dualen Studierenden lag mit $n = 282$ unter der Stichprobengröße der kaufmännischen Auszubildenden und dualen Studierenden. Auch die Geschlechterverteilung unterschied sich deutlich, da der Anteil der weiblichen Bewerbenden nur bei 8 % lag. Das Durchschnittsalter lag mit $M = 17.48$ ($SD = 2.69$) knapp ein Jahr unter dem der kaufmännischen Stichprobe. Dieser Unterschied ist dadurch zu erklären, dass in der technischen Stichprobe 13.96 % der Bewerbenden einen Haupt- bzw. Realschulabschluss besaßen. In der kaufmännischen Stichprobe waren dies lediglich 4.27 %.

Expert:innen ohne Führungsverantwortung. Teilnehmende wurden in der Studie als Expert:innen bezeichnet, wenn sie einen Berufsabschluss besaßen und als Arbeitnehmende tätig waren, jedoch keine Führungsverantwortung innehatten. Diese Gruppe beinhaltete sowohl Bewerbende auf eine Trainee-Position als auch erfahrene Fachkräfte. Die Datenerhebung wurde im Zusammenhang mit Auswahlverfahren und Potenzialerkennungsverfahren unterschiedlicher Unternehmen aus verschiedenen Branchen beispielsweise Elektrogeräte-, Versicherungs-, Pharma-, Elektrotechnik- sowie Dienstleistungsbranche durchgeführt. Die Datenerhebung fand sowohl in der Vorauswahl, in der die Bewerbenden nur Test- und Fragebogenverfahren bearbeitet haben, als auch in der Endauswahl, zum Beispiel in Assessment Centern statt. Die Stichprobe der Expert:innen umfasste $n = 325$ Personen. Die Personen nahmen in 56.92 % der Fälle an einem Assessment oder an einem Entwicklungsworkshop und nicht nur an der Vorauswahl teil. Die Geschlechterverteilung lag bei 60% männliche Teilnehmende und 40% weibliche Teilnehmende. Das Durchschnittsalter lag bei $M = 33.88$ ($SD = 8.56$) Jahre.

Führungskräfte. Auch die Daten von den Führungskräften basierten auf Personalauswahl- und Potenzialerkennungsverfahren, welche von einer Unternehmensberatung in der Durchführung unterstützt wurden. Die Stichprobe der Führungskräfte wurde in hochrangige sowie Linienführungskräfte unterteilt, da sich die persönlichen Anforderungen unterscheiden. Als Linienführungskräfte wurden diejenigen Bewerbenden definiert, welche die Arbeit von Auszubildenden oder Expert:innen leiteten. Dies können beispielsweise Ausbildungsleiter:innen oder Schichtleiter:innen sein. Zu dieser Gruppe zählten auch junge Arbeitnehmende, die eine erste Führungsposition besetzen sollten

oder darauf vorbereitet wurden. Als hochrangige Führungskräfte wurden die Teilnehmenden definiert, welche Linienführungskräfte leiteten. Zu dieser Gruppe gehörten beispielsweise Standortleiter:innen oder Abteilungsleiter:innen.

Linienführungskräfte. Die Stichprobe der Linienführungskäfte bestand aus $n = 475$ Teilnehmenden, welche in der Elektrogeräte-, Fördertechnik-, Versicherungs- sowie Elektrotechnikbranche arbeiteten. Die Daten wurden zu 46,95 % innerhalb eines Assessments oder Potenzialworkshops in Kombination mit anderen Übungen erhoben. Die Geschlechterverteilung lag bei 71% männlichen Teilnehmern und 29% weiblichen Teilnehmerinnen. Auch in dieser Stichprobe fühlte sich niemand dem diversen Geschlecht zugehörig. Das Durchschnittsalter lag mit $M = 33.67$ ($SD = 9.65$) im Bereich der Expert:innen.

Hochrangige Führungskräfte. Die letzte aufgeführte Stichprobe der hochrangigen Führungskräfte war mit $n = 90$ Teilnehmenden die kleinste Stichprobe. Das Durchschnittsalter lag mit $M = 44.22$ ($SD = 8.26$) höher als bei den Linienführungskräften, was aufgrund der größeren Verantwortung und somit der erforderlichen Erfahrung zu erklären ist. Diese Stichprobe bestand aus 87% männlichen Teilnehmenden. Dieser Anteil ist sehr hoch, scheint jedoch in der Wirtschaft realitätsgetreu zu sein, da die Datenerhebung nicht experimentell, sondern in realen Personalauswahl- sowie Potenzialerkennungsverfahren in vielen verschiedenen Branchen stattfand.

Operationalisierung der Variablen

Operationalisierung der Prädiktoren. In dieser Studie wurden zwei Prädiktoren erhoben. Einerseits die Persönlichkeit, welche als persönliche Voraussetzung der Teilnehmenden erachtet wurde, da die Big Five Dimensionen zeitstabile Persönlichkeitseigenschaften sind (Ostendorf & Angleitner, 2004). Darüber hinaus wurde auch die zeitstabile kognitive Fähigkeit des schlussfolgernden Denkens als Subdimension der Intelligenz betrachtet.

Persönlichkeit. Sowohl bei der Betrachtung der Konstruktvalidität als auch der Kriteriumsvalidität gilt die Persönlichkeit als Prädiktor. Diese Variable wurde durch die neu entwickelten arbeitsbezogenen Extraversions- und Offenheit für Erfahrungsskalen (Wedemeyer & Moldzio, in Vorbereitung) erhoben. Diese Skalen wurden auf Basis der Forschung von DeYoung et al. (2007) entwickelt, welche davon ausgingen, dass zwischen den Dimensionen und deren untergeordneten Facetten noch eine Hierarchieebene mit zwei Aspekten liegt.

Zu Beginn der Forschung bzw. der Entwicklung des Fragebogens wurden 16 Expert:innen aus der Wirtschaft sowie Wissenschaft, die in ihrem Arbeitsalltag Themen der Eignungsdiagnostik und vor allem Persönlichkeitsdiagnostik behandeln, in einem Schreiben gebeten, geeignete berufsbezogene Items zu den Aspekten zu formulieren. Dieses Schreiben bestand aus einer kurzen Herleitung sowie der Definitionen der vier Aspekte und Platz für die zu generierenden Items. Im Nachgang wurden diese Items überarbeitet, den einzelnen Aspekten zugeordnet und um Ideen der Autorin ergänzt.

Im nächsten Schritt wurde der Fragebogen, der zu diesem Zeitpunkt aus 119 Items bestand, in der Vorstudie auf Verständlichkeit geprüft. Die Bewerbenden bewerteten diesen Fragebogen auf einer fünffach abgestuften Likert-Skala von „*Starke Ablehnung*“ bis „*Starke Zustimmung*“. Anhand der $N = 101$ bearbeiteten Fragebögen wurde der Fragebogen nach den Berechnungen der explorativen Faktoranalyse (EFA) und der Betrachtung der Itemvarianz (Maß der Differenzierung der Versuchspersonen (Kelava & Moosbrugger, 2020) sowie der Faktorladungen der Items nach den folgenden Kriterien gekürzt:

- Ein Item wurde nur mit drei der fünf Ausprägungen bewertet.
- Faktorladungen lagen unterhalb von $\lambda = .5$ (Brown, 2015)

Diese Kriterien wurden gewählt, damit einerseits eine ausreichend hohe Itemvarianz gegeben ist, da die Items von den Teilnehmenden in mindestens vier Ausprägungen bewertet wurden. Als Kriterium der Faktorladung wurden $\lambda \geq .5$ gewählt, da Brown (2015) die Faktorladung ab .3 oder .4 als akzeptabel bewertet. Zudem sollte der Fragebogen ökonomisch sein. Daher wurde ein konservativerer Cut-Off-Wert von $\lambda \geq .5$ gewählt. Anhand dieser Kriterien wurde der Fragebogen auf 39 Items gekürzt. Er bestand abschließend aus neun Items für den Aspekt Enthusiasmus, acht für die Durchsetzungsfähigkeit, acht für die Offenheit und vierzehn für den Intellekt. Die Reliabilität des Fragebogens, welche im Anschluss in der Praxis weiter untersucht wurde, lag in der Vorstudie zwischen $\alpha = .69$ (Offenheit) und $\alpha = .89$ (Intellekt). Diese Reliabilitäten sind laut George und Mallery (2002), welche die Reliabilität von Persönlichkeitsfragebögen interpretierten, als kritisch bis gut zu bewerten. Die Objektivität des Fragebogens wurde von den Autor:innen als gegeben identifiziert, da der Fragebogen für alle Teilnehmenden mit einer identischen Instruktion standardisiert präsentiert wurde. Die Auswertungsobjektivität wurde als gegeben erachtet, da die Ausprägungen für alle Teilnehmenden mithilfe der Skalenmittewerte ausgewertet wurden.

Schlussfolgerndes Denken. Das schlussfolgernde Denken wurde unter Betrachtung der Kriteriumsvalidität erhoben, da untersucht werden sollte, inwiefern Berufserfolg von der

Persönlichkeit über das schlussfolgernde Denken hinaus vorhergesagt werden kann. Das schlussfolgernde Denken wurde anhand des Intelligenz-Struktur-Test Screening (Liepmann et al. 2012), kurz IST-Screening genannt, erhoben. Dieses Screening basiert auf dem Intelligenz-Struktur-Test 2000 R (Liepmann et al., 2007) und wurde aufgrund der Ökonomie in dieser Studie verwendet. Wegen der Kürze findet es auch in der Praxis gerne Anwendung. Das IST-Screening beinhaltet drei Aufgabengruppen mit je 20 Übungen; die Bearbeitungszeit beträgt ca. 26 Minuten. Das Testverfahren startet mit der Aufgabengruppe „Analogien“, in der das verbale schlussfolgernde Denken betrachtet wird. Nachfolgend beinhaltet der Test die Aufgabengruppen „Zahlenreihen“ sowie „Matrizen“, die das numerische und figurale Denken operationalisieren. Die Reliabilität des IST-Screenings wurde in den Studien zur Erstellung des Handbuches untersucht und lag nach der Interpretation von Bühner (2011) für Intelligenztests zwischen kritisch (α = .72) und gut (α = .90) (Liepmann et al., 2007). Dieses Testverfahren wurde in den Personalauswahl- sowie Potenzialerkennungsverfahren in der Vorauswahl oder zur Vorbereitung auf ein Assessment Center von den Bewerbenden online bearbeitet. Das IST-Screening wurde in allen Stichprobengruppen verwendet. Da alle Bewerbenden diesen Test auf der Onlineplattform des Hogrefe Verlages bearbeitet haben, ist die Bearbeitung sowie Auswertung standardisiert und identisch abgelaufen. Daher kann davon ausgegangen werden, dass die Objektivität gegeben war.

Operationalisierung der Kriterien

Operationalisierung der konvergenten Validität. Zur Betrachtung der konvergenten Validität wurden Fragebögen erhoben, die ähnliche Konstrukte wie Extraversion und Offenheit für Erfahrung sowie deren Aspekte Durchsetzungsfähigkeit, Enthusiasmus, Intellekt und Offenheit betrachten (s. Tabelle 5). Daher wird im Nachfolgenden ein Fragebogen zu den Big Five erklärt.

Im folgenden Abschnitt wird das NEO-Fünf-Faktoren Inventar (Borkenau & Ostendorf; 2008), kurz NEO-FFI genannt, betrachtet, welches die Big Five Persönlichkeitsmerkmale misst. Dieser Fragebogen ist eine Kurzform des weit verbreiteten NEO-Persönlichkeitsinventars von Costa und McCrae (1992). Das NEO-FFI ist ein faktoranalytisch konstruiertes Selbstbeschreibungsinventar mit 60 Items, welche mit je zwölf Items die Ausprägungen auf den Dimensionen Neurotizismus, Gewissenhaftigkeit, Verträglichkeit, Extraversion und Offenheit für Erfahrung messen. In dieser Erhebung sollten mittels NEO-FFI die Persönlichkeitsmerkmale der Teilnehmenden erfasst werden. Der Fragebogen ist so konzipiert worden, dass die 60 Items von den Teilnehmenden anhand einer

fünffach abgestuften Likert-Skala, welche von „*starke Ablehnung“* bis „*starke Zustimmung“* ging, bewertet wurden. Die Bearbeitungszeit des NEO-FFI beträgt inklusive Instruktion etwa zehn Minuten und ist somit sehr ökonomisch. Dieser Fragebogen erfährt einen breiten Einsatzbereich von der klinischen Psychologie über die Studienberatung bis hin zur Arbeits- und Organisationspsychologie. Wie die Studien im theoretischen Teil zeigen (s. Berufserfolg und Persönlichkeit & Zusammenhänge zwischen der Gerechtigkeitswahrnehmung und der Persönlichkeit), kann dieser Fragebogen trotz des nicht vorhandenen Berufsbezugs als sehr gut erforscht und für die Eignungsdiagnostik geeignet betrachtet werden.

Tabelle 5
Zuordnung der Fragebogenskalen zur konvergenten und divergenten Validität

Fragebogen	Konvergente Validität	Divergente Validität
NEO-FFI	Extraversion	Gewissenhaftigkeit
	Offenheit für Erfahrung	Neurotizismus
		Verträglichkeit
ABGS		Fleiß
		Ordnung
		Soziale Belastbarkeit
		Dauerbelastbarkeit
AVS		Einfühlungsvermögen
		Bescheidenheit

Da in diesem Abschnitt die Operationalisierung der konvergenten Validität beschrieben wird und somit die der inhaltlich ähnlichen Konstrukte, wird in diesem Teil lediglich auf die Skalen Offenheit für Erfahrung und Extraversion des NEO-FFI eingegangen. Diese beiden Dimensionen sind inhaltlich konvergent zu dem Fragebogen AEOS, da dieser die berufsbezogenen Aspekte der beiden Dimensionen misst. Die zwölf Items zur Extraversion des NEO-FFI, vier davon wurden negativ formuliert, messen die Geselligkeit und Aktivität einer Person. Ein Beispiel-Item lautet: „Ich habe gerne viele Leute um mich herum.“. Die Offenheit für Erfahrung wird auch anhand von zwölf Items erfasst, von denen sieben Items negativ formuliert wurden. Die Dimension erfragt das Interesse an Kunst sowie Kultur und die Bereitschaft, neue Dinge auszuprobieren. Dies wird beispielsweise durch das Item „Ich probiere oft neue und fremde Speisen aus.“ erhoben.

Die Objektivität war bei diesem Fragebogen gegeben, da er standardisiert online durchgeführt und die Auswertung für alle Teilnehmenden anhand der Skalenmittelwerte und einer Normtabelle betrachtet wurde. Die interne Konsistenz der Extraversionsskala lag bei α =.81 und die der Offenheit für Erfahrungsskala bei α =.75. Diese Werte sind als akzeptabel bis gut zu bewerten (George & Mallery, 2002). Um die Retest-Reliabilität für das

Fragebogenhandbuch zu untersuchen, wurde der Fragebogen nach zwei Jahren erneut von *N* = 146 Proband:innen ausgefüllt (Borkenau & Ostendorf; 2008). Die Retest-Reliabilität liegt bei r_{tt} = .81 in der Extraversion und bei r_{tt} = .76 für die Offenheit. Auch diese Werte liegen im akzeptablen bis guten Bereich (George & Mallery, 2002).

Operationalisierung der divergenten Validität. Zur Überprüfung der divergenten Validität wurden Fragebögen betrachtet, von denen Skalen zur konvergenten sowie einige zur divergenten Validität zugeordnet wurden (s. Tabelle 5). Da die allgemeine Konstruktion dieser Fragebögen im vorherigen Abschnitt ausführlich erklärt wurde, werden zur Beschreibung der Skalen zur divergenten Validität hier nur die einzelnen Skalen erklärt. Darüber hinaus werden aber auch andere Fragebögen beschrieben. Zu dem Abschnitt der divergenten Validität gehört auch das Intelligenz-Struktur-Test Screening (Liepmann et al.,2012). Es dient der Überprüfung, ob die Aspekte der Offenheit für Erfahrung mit dem schlussfolgernden Denken korrelieren. Da dieser Test im Abschnitt „Schlussfolgerndes Denken„ ausführlich erklärt wurde, wird in diesem Abschnitt nicht weiter darauf eingegangen.

Das NEO-Fünf-Faktoren Inventar (NEO-FFI) von Borkenau und Ostendorf (2008) wird auch in diesem Abschnitt betrachtet, da sich die drei Dimensionen Neurotizismus, Gewissenhaftigkeit und Verträglichkeit inhaltlich von den Dimensionen Extraversion und Offenheit für Erfahrung unterscheiden. Da der NEO-FFI im letzten Abschnitt ausführlich erklärt worden ist, werden hier nur die restlichen drei Dimensionen und deren Gütekriterien beschrieben.

Die Dimension Neurotizismus erfasst anhand von zwölf Items und ihren zugehörigen sechs Facetten den Grad der Ängstlichkeit, Nervosität sowie Unsicherheit der Teilnehmenden (Ostendorf & Angleitner, 2004). Ein Beispiel-Item lautet: „Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.“. Vier Items dieser Dimension wurden negativ formuliert. Die interne Konsistenz liegt für diese Dimension mit α = .87 im guten Bereich (George & Mallery, 2002). Auch die Retest-Stabilität nach zwei Jahren liegt im hohen Bereich bei r_{tt} = .80 (Borkenau & Ostendorf, 2008).

Die Dimension Verträglichkeit erfasst die Frage, inwiefern eine Person zwischenmenschlich agiert (Ostendorf & Angleitner, 2004). Eine hohe Ausprägung in der Dimension Verträglichkeit beschreibt eine Person, die zu Kooperation, zwischenmenschlichen Vertrauen und Harmonieorientierung neigen könnte (Ostendorf & Angleitner, 2004). Die Harmonieorientierung zeigt zum Beispiel das Item: „Ich würde lieber

mit anderen zusammenarbeiten, als mit ihnen zu wetteifern.“. Diese Dimension wird anhand von zwölf Items, von denen acht Aussagen negativ formuliert wurden, erfasst. Die Verträglichkeit ist die Persönlichkeitsdimension, welche die geringste interne Konsistenz und Retest-Reliabilität aufweist (Borkenau & Ostendorf, 2008). Beide Werte werden als kritisch (r_{tt} = .65) bis akzeptabel (α = .72) bewertet (George & Mallery, 2002).

Die Dimension Gewissenhaftigkeit hinterfragt, inwiefern die Teilnehmenden ein ordentliches, zuverlässiges und systematisches Verhalten besitzen (Ostendorf & Angleitner, 2004). Darüber hinaus werden auch die Pünktlichkeit und der Ehrgeiz betrachtet. Ein Item lautet: „Ich kann mir meine Zeit recht gut einteilen, so dass ich meine Angelegenheiten rechtzeitig beende.“. In dieser Dimension wurden vier Items negativ invertiert. Gewissenhaftigkeit weist die zweithöchste Reliabilität auf. Die interne Konsistenz liegt im guten Bereich bei α = .84 und die Retest-Reliabilität bei r_{tt} = .81(Borkenau & Ostendorf, 2008).

Moldzio et al. entwickelten 2019 einen berufsbezogenen Fragebogen zur Erfassung von Belastbarkeit und Gewissenhaftigkeit. Sie lehnten sich in der Entwicklung an die Forschung von DeYoung et al. (2007) an und konzipierten jeweils zwei Aspekte zu den beiden Big Five Dimensionen Neurotizismus und Gewissenhaftigkeit im beruflichen Kontext. Ziel dieser Fragebogenentwicklung war, durch den Berufsbezug die Akzeptanz der Bewerbenden zu steigern und durch die Aspekte eine differenziertere Prognose in der Eignungsdiagnostik abgeben zu können. Da durch die Konzipierung des Fragebogens ein großer Bezug zu den beiden Big Five Dimensionen besteht, wird dieser Fragebogen auch im Zusammenhang mit der divergenten Validität betrachtet. Ein weiterer Grund ist, dass die Arbeitsbezogenen Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS), die unten näher beschriebenen Arbeitsbezogenen Verträglichkeitsskalen (AVS; Moldzio, Böge & Wedemeyer, in Vorbereitung) sowie der neu entwickelte Fragebogen zu den Arbeitsbezogenen Extraversions- und Offenheit für Erfahrung Skalen (AEOS) nach Veröffentlichung dieser Arbeit als zusammengefasster Fragebogen „*Business Big 5*“ erscheinen sollen (Moldzio, Wedemeyer & Böge, in Vorbereitung). Daher ist es wichtig, zu überprüfen, ob die einzelnen Aspekte miteinander korrelieren.

Der Fragebogen ABGS wurde ökonomisch aufgebaut und misst mit nur insgesamt 35 Items, von denen elf Items rekodiert wurden, die zwei Aspekte Soziale Belastbarkeit und Dauerbelastbarkeit der Dimension Neurotizismus sowie die beiden Gewissenhaftigkeitsaspekte Fleiß und Ordnung. Die Proband:innen bewerteten diese Aussagen anhand einer fünffach abgestuften Likert-Skala, welche von „*starke Ablehnung“*

bis „*starke Zustimmung“* ging. Die Bearbeitung dieses Fragebogens dauerte zwischen fünf und zehn Minuten. Die Reliabilität liegt zwischen $\alpha = .79$ (Dauerbelastbarkeit) im akzeptablen und $\alpha = .84$ (Fleiß) im akzeptablen bis guten Bereich (George & Mallery, 2002). Die Retest-Reliabilität nach maximal einem Jahr liegt bei Arbeitnehmenden zwischen kritisch ($r_{tt} = .64$, $r_{tt} = .67$ und $r_{tt} = .69$) und akzeptabel ($r_{tt} = .76$). Bei den Auszubildenden, die den Fragebogen nach 2,5 Jahren noch einmal ausgefüllt haben, liegt die Retest-Reliabilität deutlich geringer zwischen $r_{tt} = .34$ und $r_{tt} = .71$($M_{rtt} = .50$) (Moldzio et al., 2019).

Personen, die eine hohe Ausprägung in der Sozialen Belastbarkeit angeben, beschreiben sich im Arbeitskontext als leistungsstark und unaufgeregt in sozial anspruchsvollen Situationen, wie zum Beispiel bei einer Präsentation (Moldzio et al., 2019). Dieser Aspekt wurde anhand von fünf Items, von denen drei negativ formuliert waren, gemessen. Ein Item lautet zum Bespiel wie folgt: „Bei Vorträgen zeige ich keine äußeren Anzeichen von Unsicherheit.“. Die Dauerbelastbarkeit beinhaltet den Umgang mit allen beruflichen Belastungen, die über einen längeren Zeitraum andauern. Sie wurde in dem Fragebogen mit fünf Items, von denen vier rekodiert wurden, erfasst. Ein Beispielitem ist „Ich kann auch nach einem längeren, intensiven Arbeitseinsatz weitere Aufgaben übernehmen.“.

Fleiß und Ordnung sind die berufsbezogenen Aspekte der Gewissenhaftigkeit. Arbeitnehmende schreiben sich eine hohe Ausprägung in dem Fleißaspekt zu, wenn sie ehrgeizig und zielstrebig arbeiten (Moldzio et al., 2019). Der Aspekt wurde anhand von 15 positiv gepolten Items erfasst. Ein Beispielitem für diesen Aspekt ist „Mein Ziel ist es, möglichst viel bei der Arbeit zu leisten.“. Der Ordnungsaspekt beinhaltet die sorgfältige und strukturierte Arbeitsweise. Zehn Items, von denen vier rekodiert wurden, erfassen diesen Aspekt. Ein Beispielitem lautet wie folgt: „Es ist mir wichtig, einen strukturierten Zeitplan während der Arbeit zu haben.“.

Die Arbeitsbezogenen Verträglichkeitsskalen (Moldzio, Böge & Wedemeyer, in Vorbereitung) wurden angelehnt an die obengenannten ABGS für die Big Five Dimension Verträglichkeit entwickelt. Auch in diesem Fragebogen wurden mit dem Einfühlungsvermögen sowie der Bescheidenheit zwei Aspekte dieser Dimension berufsbezogen erfasst. Dabei betrachtet das Einfühlungsvermögen die Fähigkeit, Kolleg:innen mit Sympathie und Empathie zu begegnen. Dies wird beispielsweise in dem ersten Item dieses Aspekts verdeutlicht: „Ich frage bei anderen aus meinem beruflichen Umfeld nach, wenn ich den Eindruck habe, dass sie etwas belastet.“ Die Bescheidenheit erfasst, inwiefern man die Wünsche der anderen berücksichtigt und eine gemeinsame Lösung

in den Vordergrund stellt. Ein Item lautet zum Beispiel: „Wenn jemand aus meinem beruflichen Umfeld dominant auftritt, lasse ich ihm den Vortritt.“. Auch dieser Fragebogen wurde einerseits für die Betrachtung der divergenten Validität genutzt, da die Verträglichkeit Teil der Big Five Dimensionen ist und die Abgrenzung zur Extraversion durch die zwei Aspekte noch genauer betrachtet werden kann. Andererseits soll auch dieser Fragebogen Teil der „*Business Big 5*“ werden.

Die AVS besteht aus 19 Items, welche auf einer fünffach abgestuften Likert-Skala zwischen „*starke Ablehnung*“ und „*starke Zustimmung*“ bewertet wurden. Die Skala für das Einfühlungsvermögen beinhaltet neun Items, von denen zwei negativ formuliert wurden. Die Skala Bescheidenheit besteht aus zehn Items, von denen vier rekodiert wurden. Die Reliabilität wurde zuvor nicht veröffentlicht. Daher wurde dieser Fragebogen anhand der oben angegebenen Stichprobe in dieser Studie überprüft. Die Faktorladungen waren in beiden Aspekten sehr gering ausgeprägt, sodass der Fragebogen anhand von Brown (2015) eingekürzt wurde. Die Einkürzung erfolgte durch Löschung aller Items $\lambda < .3$. Somit bestand der Fragebogen anschließend aus $n = 6$ Items für den Aspekt Einfühlungsvermögen und $n = 4$ für den Aspekt Bescheidenheit. Die Betrachtung der Reliabilität zeigte, dass die interne Konsistenz mit $\alpha_{\text{Einfühlungsvermögen}} = .60$ und $\alpha_{\text{Bescheidenheit}} = .65$ nach George und Mallery (2002) als kritisch zu bewerten ist. Auch Bühner (2011) setzte die Faustregel auf $\alpha > .7$. Somit wurde dieser Fragebogen nur bei der Betrachtung der Konstruktvalidität (nomologisches Netzwerk) mit einbezogen, da die Messgenauigkeit dieses Fragebogens nicht gegeben war.

Operationalisierung der Kriteriumsvalidität. In dieser Studie wurde die Kriteriumsvalidität anhand des Kriteriums Berufserfolg betrachtet, da dieses in der Eignungsdiagnostik essentiell ist (Barrick & Mount, 1991). Nachfolgend kann die Operationalisierung der Kriteriumsvalidität in die Potenzialaussage, die Eingruppierung der Vorauswahlergebnisse und die aktuelle Durchschnittsnote in der Schule unterteilt werden. Die erste Variable wurde in den Stichprobengruppen der Expert:innen, Linienführungskräften sowie hochrangigen Führungskräften erhoben, die zweite und dritte Variable in der Gruppe der Auszubildenden.

Die Potenzialaussage als Bestandteil eines Assessment Centers wurde am Ende eines Verfahrens durch die Beobachtenden getätigt und soll den Auftraggeber:innen einen Hinweis darauf geben, ob die Bewerbenden Potenzial für die vakante Stelle besitzen. Auch bei Potenzialerkennungsverfahren wurde diese Aussage getätigt, jedoch in Hinblick auf eine potentielle Beförderung oder den nächsten beruflichen Entwicklungsschritt. Da diese

Potenzialaussage Teil von Assessment Centern oder Potenzialerkennungsverfahren war, wurde diese Variable in allen Stichproben exklusiv der der Auszubildenden erhoben. Sie wird in dieser Studie zur Betrachtung der Kriteriumsvalidität genutzt. Es sollte untersucht werden, inwiefern Zusammenhänge zwischen den Persönlichkeitsmerkmalen und der Potenzialaussage bestehen. Diese Aussage wurde anhand einer vierfach abgestuften Skala getätigt (s. Tabelle 6). Wurde eine Eins oder eine Zwei vergeben, bedeutete dies, dass die Bewerbenden Potenzial für die Position besitzen und eine sofortige Empfehlung erhalten. Die beiden Ausprägungen wurden wie folgt definiert: „*Hat Potenzial für die anvisierte Positionsebene. Es sind keine begleitenden Fördermaßnahmen notwendig.*“ bzw. „*Hat Potenzial für die anvisierte Positionsebene. Es sind begleitend noch einzelne Fördermaßnahmen notwendig.*“. Wenn eine Drei vergeben wurde, wurde den Bewerbenden aktuell nicht zugetraut, diese Position zu besetzen. Jedoch ist eine Veränderung der Potenzialaussage nach persönlicher wie fachlicher Weiterentwicklung möglich. Dies kann zum Beispiel der Fall sein, wenn Potenzialträger:innen eines Unternehmens eine erste Führungsposition besetzen sollen. Diese Beförderung wird erst einmal zurückgestellt, die Bewerbenden werden gefördert und durchlaufen ein paar Jahre später wieder ein Verfahren zur Potenzialerkennung oder Personalauswahl. Die dritte Ausprägung wurde wie folgt formuliert: „*Weist zum momentanen Zeitpunkt noch nicht das erforderliche Potenzial für die anvisierte Positionsebene auf. Es sind mehrere Fördermaßnahmen erforderlich. Eine erneute Überprüfung der Potenzialeinschätzung sollte frühestens in einem Jahr erfolgen.*“. Die letzte und vierte Ausprägung besagt, dass die Bewerbenden aktuell und zu einem späteren Zeitpunkt kein Potenzial für die vakante Stelle aufweisen. Diese Aussage lautete: „*Weist zum momentanen Zeitpunkt nicht das erforderliche Potenzial für die anvisierte Positionsebene auf. Auch die Prognose für die Zukunft ist ungünstig.*“.

Diese Potenzialaussage wurde nicht empirisch entwickelt und erforscht, fand jedoch seit 2010 bei $N = 1.703$ Auswahl- und Potenzialerkennungsverfahren einer Unternehmensberatung Anwendung. Da diese Aussage sowohl für das Handbuch der ABGS (Moldzio et al. 2019) wie auch für einen weiteren Artikel zur Überprüfung der Kriteriumsvalidität der ABGS (Moldzio et al. 2021) verwendet wurde, wird in dieser Studie davon ausgegangen, dass dies eine geeignete Variable zur Erhebung der Kriteriumsvalidität ist.

Eine weitere Variable, die für die Betrachtung der Kriteriumsvalidität des Fragebogens genutzt wurde, stellt die Eingruppierung der Vorauswahlergebnisse dar. Auch sie soll den Zusammenhang zwischen den Persönlichkeitsmerkmalen und dem Berufserfolg

aufzeigen. Die Eingruppierung wurde ebenfalls nicht empirisch entwickelt. Sie bildet jedoch seit Jahren einen großen Anteil der Rückmeldungen zu Test- und Fragebogenverfahren in der Vorauswahl, welche eine Unternehmensberatung den Unternehmen als Feedback zu Leistungen und der Persönlichkeit von Kandidat:innen erteilt. In allen Vorauswahlverfahren für Auszubildende und duale Studierende wurde diese Eingruppierung der Vorauswahl getätigt, damit die Kund:innen eine übersichtliche Aufstellung erhalten, welche Teilnehmenden für den nächsten Bewerbungsschritt eingeladen werden sollten. Darüber hinaus wurde diese Aussage auch in einzelnen Personalauswahlverfahren bei den anderen Stichproben (zum Beispiel Expert:innen) erhoben, wenn diese lediglich an einer Vorauswahl mit Test- und Fragebogenverfahren teilgenommen haben.

Unter der Eingruppierung versteht man eine sechsfach abgestufte Variable, die bei einigen Verfahren noch einmal in eine Leistungseinschätzung und eine Persönlichkeitseinschätzung unterteilt worden ist. Da diese Unterteilung lediglich bei den Auszubildenden und dualen Studierenden vorgenommen wurde, wird im Anschluss allein die Gesamteingruppierung betrachtet, damit die Vergleichbarkeit der Stichproben gewahrt werden kann. Die Eingruppierung wurde anhand der Test- und Fragebogenausprägungen erstellt. Die Eins bzw. ein „*C*“ wurde vergeben, wenn die Bewerbenden ausschließlich unterdurchschnittliche oder weit unterdurchschnittliche Ergebnisse erzielt hatten (s. Tabelle 6). Dies bedeutet, dass die Bewerbenden ungeeignet für die anvisierte Position sind. Den Bewerbenden wurde eine Zwei bzw. ein „*B*-“ zugewiesen, wenn sie überwiegend mittlere Ergebnisse erreicht hatten, jedoch auch einige unter- bzw. weit unterdurchschnittliche Ergebnisse vorzufinden waren. Auch in diesem Fall wurde Kund:innen abgeraten, diese Bewerbenden zu weiteren Schritten einzuladen. Die Drei oder das „*B*“ wurde bei Bewerbenden vergeben, die in Summe durchschnittliche Ergebnisse oder mittlere Persönlichkeitsausprägungen erzielt hatten. Ab dieser Eingruppierung wurde die Einladung zu einem weiteren Bewerbungsschritt empfohlen. Die oben beschriebene Systematik wurde auch bei den Bewertungen Vier bis Sechs bzw. „*B*+“ bis „*A*“ verfolgt. So wurde eine Vier bei überwiegend mittleren und einigen weit bzw. überdurchschnittlich hohen Ergebnissen in den Testverfahren bzw. Ausprägungen in den Persönlichkeitsfragebögen, eine Fünf bei überwiegend weit bzw. überdurchschnittlich hohen und wenigen mittleren Resultaten und eine Sechs bei ausschließlich überdurchschnittlich bis weit überdurchschnittlich hohen Leistungen und Persönlichkeitsausprägungen vergeben.

Um eine Aussage über den Zusammenhang zwischen dem Ausbildungserfolg und den Persönlichkeitsmerkmalen tätigen zu können, wurde die aktuelle Durchschnittsnote in der

Schule von allen Bewerbenden für eine Ausbildung oder ein Duales Studium erhoben. Die Betrachtung der Schulnoten ist ein guter Prädiktor für den Ausbildungserfolg (r = .41), wie empirisch nachgewiesen werden konnte (Schuler & Marcus, 2006).

Tabelle 6
Darstellung und Beschreibung der Abstufungen der Kriteriumsvariablen

Variablen	Abstufung	Beschreibung
Potenzialaussage	1	Hat Potenzial für die anvisierte Positionsebene. Es sind keine begleitenden Fördermaßnahmen notwendig.
	2	Hat Potenzial für die anvisierte Positionsebene. Es sind begleitend noch einzelne Fördermaßnahmen notwendig.
	3	Weist zum momentanen Zeitpunkt noch nicht das erforderliche Potenzial für die anvisierte Positionsebene auf. Es sind mehrere Fördermaßnahmen erforderlich. Eine erneute Überprüfung der Potenzialeinschätzung sollte frühestens in einem Jahr erfolgen.
	4	Weist zum momentanen Zeitpunkt nicht das erforderliche Potenzial für die anvisierte Positionsebene auf. Auch die Prognose für die Zukunft ist ungünstig
Eingruppierung	A	Ausschließlich überdurchschnittliche oder weit überdurchschnittliche Ergebnisse
	A-	Überwiegend weit bzw. überdurchschnittlich hohe und wenig mittlere Resultate
	B+	Überwiegend mittlere Ergebnisse, vereinzelt überdurchschnittliche Resultate
	B	Mittlere Ergebnisse
	B-	Überwiegend mittlere Ergebnisse, vereinzelt unterdurchschnittliche Resultate
	C	Ausschließlich unterdurchschnittliche oder weit unterdurchschnittliche Ergebnisse
Durchschnittsnote	1 - 6	Die Durchschnittsnote wurde aus Zeugnissen entnommen und eine 1 steht für eine sehr gute und eine 6 für eine ungenügende Leistung
Schlussfolgerndes Denken	1 - 60	Gesamtwertes des IST-Screenings (Summe aller Rohwerte)

Die Schulnote wurde anhand einer sechsfach abgestuften Skala erhoben. Dabei stellt die Eins, wie im deutschen Schulsystem üblich, die beste und die Sechs die schlechteste Note dar. Da in der Oberstufe ein Punktesystem von eins bis fünfzehn existiert, bei dem die Eins aber die schlechteste Note ist, wurden diese Noten in eine Eins bis Sechs Systematik umkodiert. Aufgrund der Erhebung einer Durchschnittsnote war eine Nachkommerstelle möglich. Da diese Angaben den beigefügten Zeugnissen aus den Bewerbungsunterlagen

entnommen wurden, ist davon auszugehen, dass diese Angaben korrekt waren. Die Richtigkeit wurde nicht überprüft, es mussten auch keine Daten ausgeschlossen werden.

Untersuchungsdurchführung

Da die Idee der AEOS angelehnt an die ABGS (Moldzio et al. 2019) entstanden ist, wurde sich sehr stark an diesem Vorgehen orientiert. Ziel bei der Entwicklung der AEOS war es, einen berufsbezogenen Fragebogen mit den vier Aspekten der Dimensionen Extraversion und Offenheit für Erfahrung von DeYoung et al. (2007) zu entwickeln. Im ersten Schritt wurden zunächst die vier Aspekte Enthusiasmus, Durchsetzungsfähigkeit, Offenheit und Intellekt genauer definiert. Dieser Schritt war wichtig, da Expert:innen der Eignungsdiagnostik um Mithilfe bei der Itemgenerierung gebeten wurden. So konnte eine Vielzahl von Items generiert und somit ein möglichst breites Spektrum abgedeckt werden. Das Anschreiben an die Expert:innen beinhaltete das Ziel und den Nutzen dieser Studie sowie eine Herleitung und Definition dieser vier Aspekte. Ergänzend dazu wurde den Expert:innen anhand eines Beispiels erklärt, wie Personen mit einer hohen oder niedrigen Ausprägung in den einzelnen Aspekten handeln könnten. Personen mit einer niedrigen Ausprägung in der Durchsetzungsfähigkeit wurden beispielsweise als Mitarbeitende beschrieben, die sich schnell von der Meinung anderer umstimmen lassen. Es wurden 16 Expert:innen angeschrieben, von denen 15 geantwortet haben. Als Expert:innen wurden einerseits wissenschaftliche Mitarbeitende sowie Professor:innen aus der Arbeits- und Organisationspsychologie verschiedener Universitäten bezeichnet. Diese waren langjährige Kooperationspartner:innen einer Unternehmensberatung, mit deren Zusammenarbeit viele Forschungsarbeiten vorangetrieben wurden. Hinzu kamen Expert:innen, die mit einem psychologischen Background in der Personalabteilung eines Unternehmens arbeiteten und in Praxisprojekten Partner:innen einer Unternehmensberatung waren.

Neben der Itemgenerierung der Expert:innen formulierte auch die Autorin Items. Insgesamt wurden $N = 439$ Items formuliert, von denen $n = 108$ Enthusiasmus, $n = 126$ Durchsetzungsfähigkeit, $n = 104$ Intellekt sowie $n = 101$ dem Aspekt Offenheit zugeordnet wurden. Diese Items wurden im Anschluss gesichtet und aufgrund der Praktikabilität und der Verständlichkeit nach den folgenden Kriterien sondiert (Schuler et al., 2007):

- Items passten nicht zu den Definitionen der Aspekte.
- Items waren uneindeutig formuliert und hätten zu mehreren Aspekten gepasst.

- Items waren schwer verständlich (zum Beispiel Auf mich passt die Beschreibung: Wenn ich durch die Tür hinausfliege, komme ich durch das Fenster wieder herein.).
- Items waren doppelt negativ formuliert.

Einige Items wurden nicht berufsbezogen formuliert. Bei diesen wurde betrachtet, ob durch eine Umformulierung das Item hätte geeignet sein können. Es wurden final 119 Items als geeignet bewertet und für die Vorstudie in einem Fragebogen zusammengefasst. Dieser Fragebogen bestand aus den ausgewählten Items sowie Fragen zu Geschlecht, Alter, Muttersprache, Schulabschluss sowie zur Berufserfahrung. Ziel dieser Vorstudie war es, den Fragebogen auf Verständlichkeit zu prüfen und zu erkennen, ob die Itemschwierigkeit gegeben ist, d.h. dass keine Boden- oder Deckeneffekte auftreten, sondern die gesamte Breite der Likert-Skala verwendet wurde. Außerdem wurde geprüft, ob die Items das Konstrukt erfassen. Daher wurden Schüler:innen aus der Oberstufe eines Gymnasiums sowie zweier beruflicher Schulen ausgewählt, da diese Schüler:innen eher Berufserfahrung aufweisen und davon ausgegangen wurde, dass sie ein guter Maßstab für die Testung der Verständlichkeit sind. Final haben $N = 101$ Teilnehmende den Fragebogen bearbeitet. Die genaue Stichprobenzusammensetzung ist in dem Abschnitt Vorstudie zu finden. Der Fragebogen wurde zunächst in einem paper-pencil Format erhoben. Mit Beginn der Covid-19 Pandemie wurde er auch online erfasst.

Anschließend wurde der Fragebogen nach den Berechnungen der explorativen Faktorenanalyse und der Betrachtung der Itemvarianz sowie Faktorladung, welche im nächsten Abschnitt genauer erklärt werden, nach den folgenden Kriterien gekürzt:

- Bei diesem Item nutzten die Teilnehmenden nicht die gesamte Skala.
- Faktorladungen lagen unter $\lambda = .5$ (Brown, 2015)
- Die Reliabilität könnte stark verbessert werden, wenn dieses Item gelöscht wird.

Falls die Teilnehmenden nicht die gesamte Skala in gewissen Items nutzen, würde dies für eine geringe Itemvarianz sprechen, welche als Maß der Differenzierungsfähigkeit eines Items verstanden wird (Kelava & Moosbrugger, 2020). Eine geringe Itemvarianz würde dazu führen, dass Personen mit verschiedenen Persönlichkeiten dennoch eine ähnliche Ausprägung des Items ankreuzen und somit die Persönlichkeit nur schwer differenziert werden kann. Anhand dieser Kriterien wurde der Fragebogen auf 39 Items, resultierend aus neun Items für den Aspekt Enthusiasmus, acht für die Durchsetzungsfähigkeit, acht für die Offenheit sowie vierzehn für den Intellekt, gekürzt.

Abschließend wurden die Daten ab Sommer 2020 in allen Personalauswahl- und Potenzialerkennungsverfahren einer Unternehmensberatung erhoben. Die AEOS wurde den Bewerbenden in die AVS integriert (Moldzio, Böge & Wedemeyer, in Vorbereitung) präsentiert. Außerdem wurden die anderen Variablen zur Erfassung der Konstrukt- und Kriteriumsvalidität im gleichen Kontext erhoben. Neben der Datenerhebung in Personalauswahl- und Potenzialerkennungsverfahren bildeten die AEOS, der NEO-FFI sowie die ABGS/ AVS einen Bestandteil der zweiten Studie dieser Promotion (s. Die Validierung von virtuellen Dialogübungen im eignungsdiagnostischen Kontext). Sie wurden in die Berechnung der Gütekriterien in diese Studie mit einbezogen.

Datenanalyse

Vorstudie. Die Daten der Vorstudie wurden zuerst auf fehlende Werte gesichtet. Dies war nur bei den deskriptiven Daten und nicht bei der AEOS der Fall. Vereinzelt haben Personen ein Item doppelt bewertet oder ein Kreuz zwischen zwei Ausprägungen (zum Beispiel zwischen „*starke Ablehnung*“ und „*Ablehnung*“) gesetzt. In diesen Fällen wurde ein Mittelwert für die Ausprägung des Items gebildet. Dies war nur in den paper-pencil Varianten möglich. Im nächsten Schritt wurden alle Items betrachtet und untersucht, ob ein Item eventuell nur anhand von drei anstatt von fünf Ausprägungen der Likert-Skala bewertet wurde. Dies würde für eine geringe Itemvarianz sowie Boden- und Deckeneffekte sprechen, welche auf eine geringe Differenzierungsfähigkeit eines Items hindeuten würden (Kelava & Moosbrugger, 2020). Diese Itemvarianz wurde in der Analyse der nächsten Schritte berücksichtigt. Anschließend wurde in SPSS (IBM Corp, 2016) eine explorative Faktorenanalyse (EFA) berechnet und die Faktorladungen betrachtet. Anhand dieser Berechnung sollte überprüft werden, auf wie vielen Faktoren bzw. latenten Variablen die einzelnen Items laden (Brandt, 2020). Obwohl dieser Fragebogen auf einem theoretischen Hintergrund beruhte und somit der Aufbau durch ein strukturprüfendes Verfahren hätte analysiert werden müssen (hier: konfirmatorische Faktorenanalyse), entschied sich die Autorin in der Vorstudie dafür, den neuartigen theoretischen Ansatz (Aspekte nach DeYoung, 2007 et al. und die berufsbezogene Formulierung) erst einmal hypothesensuchend zu begutachten. Dies wurde durchgeführt, um sicher zu stellen, dass die Items, beispielsweise durch den Berufsbezug, nicht auf mehr als vier Faktoren laden. Die EFA wurde zu Beginn ohne eine festgesetzte Faktoranzahl gerechnet und die Faktorenanzahl anhand eines Screeplots betrachtet. Die Elbow-Regel besagt, dass die Anzahl der Faktoren durch den Screeplot des Eigenwertes bestimmt wird und somit die Faktoren links des Knicks der Kurve

gezählt werden (Brandt, 2020). Im Anschluss wurde eine weitere Faktorenanalyse mit der festgesetzten Faktorenanzahl gerechnet und alle Items mit einer Faktorladung unter $\lambda = .5$ aus dem Fragebogen entfernt (Brown, 2015). Um nicht unendlich viele Faktorlösungen zu erhalten und eine Einfachstruktur schaffen zu können, wurde die Analyse mit einer Rotation gewählt (Brandt, 2020). Einfachstruktur bedeutet in diesem Fall, dass jedes Item primär auf einem Faktor lädt und ggf. noch einen Sekundärfaktor besitzt, bei dem die Ladung jedoch geringer ist. Als Rotationsverfahren wurde die orthogonale Rotation des Varimax-Verfahrens gewählt, da so die individuelle und unabhängige Betrachtung der Items bestehen bleibt (Brandt, 2020). Im Nachgang wurden in weiteren Schritten die nicht ausgeschöpften Items mit geringer Faktorladung ausgeschlossen, bevor die Faktorenanalyse neu berechnet wurde. Im Anschluss fand die Beurteilung der Reliabilitäten statt. Es wurde die Reliabilität in den einzelnen Skalen betrachtet und dabei untersucht, ob die Löschung eines Items diese verbesserte und dies zu einer höheren Messgenauigkeit führte. Die Items in dem reduzierten Fragebogen wurden final auf die Verständlichkeit und ggf. doppelte Verneinung kontrolliert, was zum Ausschluss weiterer Items auf dem Fragebogen führte.

Hauptstudie. Nach der Erhebung der Daten der Hauptstudie wurde erneut die faktorielle Validität berechnet und somit Hypothese 1 getestet. Zur Überprüfung der Hypothese 1 wurde in dem Datenanalyseprogramm R Studio (RStudio Team, 2020) eine konfirmatorische Faktorenanalyse (CFA) mit Hilfe des Zusatzprogramms „*lavaan*“ (Rosseel, 2012) berechnet. Hieraus folgte die Beurteilung, ob der Fragebogen die vorgegebene Faktorenstruktur besitzt (Gäde, Schermelleh-Engel & Brandt, 2020). Dies wurde anstelle einer explorativen Faktorenanalyse durchgeführt, da dieser Fragebogen auf einem theoretischen Hintergrund basierte und die Anzahl der Faktoren analog zu den vier Aspekten auf vier gesetzt werden konnte (Gäde, Schermelleh-Engel & Brandt, 2020). Die CFA überprüfte die Zusammenhänge einer beobachtbaren Variablen (hier beispielsweise ein Item) mit einer latenten Variablen (zum Beispiel der Aspekt Offenheit, welcher nicht beobachtbar ist). Die latenten Variablen können auch als Faktoren benannt werden, auf die verschiedene beobachtbare Variablen (Items) laden. Um die Faktorstruktur erfassen zu können, fand ein Vergleich verschiedener Modelle miteinander statt. Zu Beginn wurde die Faktorenstruktur auf Eindimensionalität überprüft, was bedeuten würde, dass alle beobachtbaren Variablen auf einem latenten Faktor laden (Gäde, Schermelleh-Engel & Brandt, 2020). In den nächsten Modellen erfolgte die Prüfung auf Mehrdimensionalität. Sie besagt, dass einzelne beobachtbare Variablen verschiedenen Faktoren zuzuordnen sind, jedoch nur auf einem

Faktor laden. Da aufgrund des theoretischen Hintergrunds zwei verschiedene Faktorstrukturen möglich waren (zwei Dimensionen nach Costa und McCrae, 1992 oder vier Aspekte nach DeYoung et al., 2007), wurden diese verglichen. Zu diesem Zweck erfolgte in dieser CFA die Berechnung der folgenden drei Modelle:

- Modell 1: Alle Items laden auf einem Faktor
- Modell 2: Die Items laden auf zwei Faktoren (Extraversion und Offenheit für Erfahrung)
- Modell 3: Die Items laden auf vier Faktoren: Extraversion wird in Enthusiasmus und Durchsetzungsfähigkeit und Offenheit für Erfahrung wird in Offenheit und Intellekt gesplittet

Die Berechnung der CFA sollte zeigen, dass die Modellpassung bei der 4-Faktorenlösung am besten ist und der Fragebogen ein mehrdimensionales anstatt eines eindimensionalen Konstrukts misst (Gäde, Schermelleh-Engel & Brandt, 2020). Dies wurde an den vier verschiedenen Fitwerten sowie mit einem χ^2-Test betrachtet. Die beiden Kennwerte Comparative Fit Index (*CFI*) und Tucker-Lewis Index (*TLI*) prüften, wie gut ein Wert passt (West et al., 2012). Die beiden anderen Werte Root Mean Square Error of Approximation (*RMSEA*) und Standardized Root Mean Square Residual (*SRMR*) gaben die „*Schlechte*" der Passung an (West et al., 2012). Angelehnt an diese Forschung wurde von einem guten Modellfit gesprochen, wenn die *CFI* und der *TLI* Werte $\geq .95$, der *RMSEA* $\leq .06$ und der *SRMR* $\leq .08$ waren. Der χ^2-Test überprüfte dabei die Nullhypothese und schaute, ob sich die Kovarianzmatrix der Indikatorvariablen nicht von der modellimplizierten Matrix unterscheidet. Dies wird erkennbar, wenn der χ^2-Wert klein ist und der Test nicht signifikant wird. In dieser Studie wurde daher davon ausgegangen, dass der χ^2-Wert bei der 4-Faktorenlösung am geringsten ist (Gäde, Schermelleh-Engel & Brandt, 2020). Im ersten Schritt wurde erneut die Faktorladung aller Items betrachtet, die Items mit einer Ladung $\lambda < .4$ aus dem Modell entfernt und im Fragebogen gestrichen. Dadurch konnte der Fragebogen hinsichtlich der Gütekriterien verbessert und durch die Kürzung ökonomischer werden. Nach der unten beschriebenen Itemanalyse wurde die CFA nochmals mit dem Ziel berechnet, zu untersuchen, ob die vier Faktorenstruktur einen besseren Modellfit als die zweifaktorielle Lösung besitzt. Dies wurde neben den oben beschriebenen Fitwerten auch mithilfe des Akaike Information Criterion (*AIC*) betrachtet. Die Modelle wurden als nicht geschachtelt identifiziert, da nicht nur Parameter im neuen Modell hinzugefügt wurden, sondern die Struktur ganzheitlich geändert wurde (Werner, 2012). Durch diese nicht vorhandene

Schachtelung sind die Freiheitsgrade der Modelle verschieden. Dies berücksichtigt der *AIC* Wert. Es wird von einer guten Modellpassung gesprochen, wenn der *AIC* Wert möglichst klein ist. Somit wurden in dieser Studie die *AIC* Werte der Modelle verglichen und das Modell mit dem kleinsten Wert als das bestpassendste Modell identifiziert.

Vor der Berechnung der konfirmatorischen Faktorenanalyse wurden zunächst die Daten auf multivariate Normalverteilung getestet. Dies gilt als Voraussetzung der CFA (Bühner, 2021). Die Überprüfung der Normalverteilung erfolgte mithilfe des Mardia Tests im Zusatzpaket „*psych*“ (Revelle, 2022). Dieser Test betrachtete die Schiefe der Daten und deren Kurtosis. Die Faustregel nach West et al. (1995) zur Schiefe der Daten besagt, dass von keiner Normalverteilung ausgegangen werden kann, wenn die Werte im Betrag kleiner als zwei sind. In diesem Datensatz lagen alle Werte unter dem Grenzwert und somit konnte für die Schiefe von keiner Normalverteilung ausgegangen werden. In dem Bereich der Kurtosis sahen die Daten ähnlich aus, da alle Items unter dem Grenzwert < 7 (West et al., 1995) lagen. Insgesamt kann somit von keiner Normalverteilung ausgegangen werden. Daher fand der Maximum-Likelihood Robust (*MLR*) Schätzer in den nachfolgenden Analysen Anwendung (Bühner, 2021). Dieser Schätzer ist robust gegenüber einer fehlenden Normalverteilung und führt somit zu keiner Verzerrung in der CFA.

Um die Modellpassung verbessern zu können, wurden auch Modification Indices (*MI*) betrachtet, welche Hinweise darauf liefern, ob Items untereinander zusammenhängen bzw. korrelieren (Weiber & Mühlhaus, 2014). Dabei schätzen die *MI*, inwiefern der χ^2-Wert sinkt, falls ein Parameter festgesetzt wird. Diese Korrelationen können durch verschiedene Methodeneffekte entstehen, beispielsweise durch ähnlich formulierte Items (Gäde, Schermelleh-Engel & Brandt, 2020). Es wurden nicht alle *MI*s ins Modell aufgenommen, sondern gemäß der Empfehlung von Weiber und Mühlhaus (2014) bei jedem *MI* die theoretische und sachlogische Grundlage betrachtet. Aus diesem Grund wurden die Items, welche kovariierten, auf gleiche Satzbausteine hin untersucht und danach entschieden, ob das Modell modifiziert wird oder nicht.

Im nächsten Schritt folgte die Itemanalyse, welche kontrolliert, ob die generierten Items gut konstruiert sind und verschiedene Testpersonen voneinander unterscheiden können (Kelava & Moosbrugger, 2020). Dazu wurde zu Beginn die Itemschwierigkeit berechnet, welche betrachtet, inwiefern eine Bejahung der Items auch zu einer hohen Merkmalsausprägung führt (Kelava & Moosbrugger, 2020). Eine kleine Itemschwierigkeit würde somit bedeuten, dass den Personen eine Bejahung eher schwerer fallen würde. Die

Itemschwierigkeit wird berechnet, indem eine Differenz aus den Personen, welche dem Item zustimmen, und der Gesamtstichprobe gebildet wird (Beauducel & Leue, 2014):

$$P_i = \frac{n_r}{n}$$

Die Itemschwierigkeit P_i liegt zwischen Null und Eins. Ein Wert nahe Null beschreibt eine hohe Schwierigkeit des Items. Nach Bühner (2011) wird von einer hohen Schwierigkeit gesprochen, wenn $P_i < .2$ beträgt. Eine mittlere Schwierigkeit ist gegeben, wenn die Schwierigkeit im Bereich $.2 < P_i < .8$ liegt. Ab $P_i > .8$. wird von einer niedrigen Itemschwierigkeit gesprochen. Die Trennschärfe als weiterer Kennwert in der Itemanalyse zeigt auf, inwiefern ein Item mit der Merkmalsausprägung übereinstimmt (Kelava & Moosbrugger, 2020). Mittels dieser wird der Grad bestimmt, in dem eine hohe Ausprägung eines Items auch zu einer hohen Ausprägung im Skalenmittelwert führt. Die Trennschärfe r_{it} kann zwischen Minus Eins und Eins liegen. Ein Wert nahe Null deutet darauf hin, dass das Item von dem Merkmalsmittelwert stark abweicht und kein Zusammenhang festzustellen wäre. Weist ein Item eine negative Trennschärfe auf, dann korreliert ein niedriger Itemwert mit einer hohen Merkmalsausprägung und deutet auf einen Mangel hin (Kelava & Moosbrugger, 2020). Angelehnt an Bühner (2011) wurde die Trennschärfe in die drei folgenden Gruppen einsortiert:

- $r_{it} \leq .30$: niedrig bzw. nicht trennscharf
- $.3 < r_{it} \leq .50$: mittel bzw. akzeptable Trennschärfe
- $r_{it} \geq .51$: hoch bzw. trennscharf

Eine zu hohe Trennschärfe sagt aus, dass viele Items sehr identisch sein können. Daher wurde eine Trennschärfe von ca. $r_{it} = .5$ angestrebt (Kelava & Moosbrugger, 2020). Darüber hinaus kann auch eine hohe Reliabilität ein Hinweis darauf sein, dass die Items sich zu wenig unterscheiden (Gäde, Schermelleh-Engel & Brandt, 2020). In diesem Fall wurde die Itemkorrelation einer Skala betrachtet. Bei einer zu hohen Itemkorrelation kann davon ausgegangen werden, dass sich die Items kaum inhaltlich unterscheiden, folglich der Fragebogen weiter gekürzt werden kann und somit noch ökonomischer wird. Diese Kürzung wurde anhand der inhaltlichen Betrachtung und der Faktorladungen vorgenommen. Bei der Kürzung der Items wurde darauf geachtet, dass pro Skala mehr als vier Items enthalten waren, um die Inhaltsvalidität zu garantieren (Bühner, 2021).

Zur Überprüfung der Hypothese zwei wurde die Reliabilität bestimmt. Als Teil der Reliabilität wurde die interne Konsistenz mit dem Koeffizienten Cronbachs Alpha (α)

berechnet, da die Retest-Reliabilität, d.h. die Messgenauigkeit, nach einer gewissen Zeit Teil der weiterführenden Forschung sein wird. In der vorliegenden Studie, basierend auf Praxisdaten, war es nicht möglich die Retest-Reliabilität zu erheben, da zumeist die Teilnehmende frühestens nach zwei Jahren ein weiteres Verfahren durchlaufen und erst dann der Fragebogen ein zweites Mal ausgefüllt werden kann. Die Bewertung der Reliabilität wird anhand der Kriterien von George und Mallery (2002) für Persönlichkeitsfragebögen wie folgt bewertet:

- $\alpha \leq .59$: unzureichend
- $.60 < \alpha \leq .69$: kritisch
- $.70 < \alpha \leq .79$: akzeptabel
- $.80 < \alpha \leq .89$: gut
- $\alpha \geq .90$: sehr gut

Neben der faktoriellen sind auch die konvergente und divergente Validität Teil der Konstruktvalidität, welche anhand der Hypothesen 3a-e überprüft wurden. Es wurde bei der Betrachtung der konvergenten Validität davon ausgegangen, dass die Skalen des Fragebogens moderat mit Skalen korrelieren, welche auf ähnlichen Konstrukten aufbauen. Als moderat definiert Bühner (2011) eine Korrelation von $r > |.5|$. Demgegenüber betrachtet die divergente Validität Zusammenhänge zwischen der AEOS und Fragebögen mit unterschiedlichen Konstrukten. Es wird von einer guten Gültigkeit der AEOS ausgegangen, wenn nur geringe bis keine Korrelationen bestehen. Von einer guten Passung der divergenten Validität wurde bei einer Korrelation von $r < |.4|$ ausgegangen (Bühner, 2011). Zur Untersuchung der Hypothesen 3a-e und somit die Berechnung der divergenten und konvergenten Validität erfolgte anhand einer CFA mittels des Datenprogramms R Studio (RStudio Team, 2020). Die CFA wurde anstatt der klassischen Korrelationsanalyse gewählt, da diese die Trait-, Methoden- und Messfehleranteile getrennt voneinander betrachtet (Schermelleh-Engel et al., 2020). Darüber hinaus kann die Modellgüte überprüft werden, wohingegen die Korrelationsanalyse nur jeden einzelnen Pfad separat analysieren kann. Teil dieser Analyse waren die Items der AEOS, des NEO-FFI, ABGS und AVS Items. Es wurden die folgenden Modelle berechnet:

- Modell 1: Korrelationen zwischen den vier Aspekten und den fünf Dimensionen des NEO-FFI
- Modell 2: Korrelationen zwischen den zehn Aspekten (ABGS, AVS und AEOS)

Dieses Verfahren wurde nicht nur gewählt, um die Korrelationen zwischen den verschiedenen Skalen zu betrachten, sondern auch um die AEOS von häufig eingesetzten Fragebögen abzugrenzen und einen Mehrwert durch den Einsatz des Fragebogens aufzuzeigen.

Zur Untersuchung der Hypothese 4a-d sowie der Kriteriumsvalidität wurde eine Korrelationsanalyse durch ein Strukturgleichungsmodell in R Studio (RStudio Team, 2020) mithilfe des Zusatzpakets „*lavann*“ (Rosseel, 2012) berechnet. Aus den oben genannten Gründen wurde sich in dieser Analyse ebenfalls für ein Strukturgleichungsmodell und nicht für eine Korrelationsanalyse entschieden. Als praxisrelevante Kriterien für Berufserfolg fanden in dieser Studie Potenzialaussagen, Eingruppierungen, die Werte des IST-Screenings als Operationalisierung des schlussfolgernden Denkens und die Durchschnittsnoten der Auszubildenden aus der Schule Verwendung. Diese vier Kriterien können in zwei Perspektiven geteilt werden. Zum einen spricht man bei der Durchschnittsnote und den Werten des IST-Screenings von retrospektiven Kriterien. Dies besagt, dass die beiden Kriterien vor der Bearbeitung des Fragebogens gegeben waren und somit die Vorhersagekraft der Persönlichkeit durch diese beiden Kriterien betrachtet wird (Beauducel & Leue, 2014). Die anderen beiden Kriterien der Potenzialaussage und Eingruppierung wurden prognostisch, das heißt nach der Persönlichkeit, erhoben (Beauducel & Leue, 2014). Somit wurde die Vorhersagekraft der Persönlichkeit auf diese Kriterien berechnet. Die Größe der Korrelationen wurde nach Cohen (1988) bewertet, welcher die Korrelation wie folgt klassifizierte:

- $.1 < r < .3$: klein bis moderate Korrelation
- $.3 < r < .5$: moderate bis große Korrelation
- $r \geq .5$: große Korrelation

Im letzten Schritt dieser Studie sollte die inkrementelle Validität untersucht und gezeigt werden, dass der Einsatz des Fragebogens einen Mehrwert für die Eignungsdiagnostik bietet und die üblichen Test- und Fragebogenverfahren um diesen Fragebogen ergänzt werden sollten. Zur Überprüfung der Hypothese 5a-b wurde eine hierarchischen Regressionsanalyse in einem Strukturgleichungsmodell berechnet und betrachtet, ob der Anteil der aufgeklärten Varianz (R^2) an der Kriteriumsvariable durch Hinzunahme der AEOS erhöht werden kann. Wenn dies der Fall ist, wird von einer Verbesserung der Vorhersagekraft gesprochen. Nach Chin (1998) kann auch das Bestimmtheitsmaß bewertet werden, um zu erkennen, wie hoch die Erklärungskraft ist. Dabei

wird von einer substanziellen Erklärungskraft gesprochen, wenn $R^2 \geq .67$ ist, einer mittelguten bei einem Wert zwischen $.66 \leq R^2 > .33$ und von einem schwachen Bestimmtheitsmaß bei einem Betrag von $R^2 \leq .33$ (Chin, 1998). Es wurde sich für ein Strukturgleichungsmodell entschieden, da das Strukturgleichungsmodell neben den manifesten Variablen auch die latenten Variablen berücksichtigt und darüber hinaus die komplette Modellgüte betrachtet (Backhaus et al., 2015).

Nachfolgend werden die vier Schritte der hierarchischen Regressionsanalyse erklärt (s. Tabelle 7). Im ersten Schritt wurde eine Regressionsanalyse mit dem Wert des IST-Screenings für das schlussfolgernde Denken und einem Kriterium (Eingruppierung und Potenzial getrennt) berechnet. Anschließend erfolgte die Aufnahme der vier Aspekte in die Analyse. In einer zweiten Analyse wurden die zwei Schritte analog mit dem NEO-FFI als Prädiktor zur Vorhersage der Kriterien durchgeführt. Die Betrachtung der inkrementellen Validität über das schlussfolgende Denken und die Big Five Dimensionen hinaus wurde separat berechnet, da es der Autorin wichtig war, den Mehrwert des neuen Fragebogens bzw. der Erfassung der Aspekte gegenüber den Big Five Dimensionen zu verdeutlichen.

Tabelle 7
Erklärung des Vorgehens zur Betrachtung der inkrementellen Validität

Schritt	Hierarchische Regression
Erste Analyse erster Schritt	Ist-Screening sagt das Kriterium vorher
Erste Analyse zweiter Schritt	IST-Screening und AEOS sagen das Kriterium vorher
Zweite Analyse erster Schritt	NEO-FFI sagt das Kriterium vorher
Zweite Analyse zweiter Schritt	NEO-FFI und AEOS sagen das Kriterium vorher

Weitere Analysen. Um zu überprüfen, ob die Kriteriumsvalidität in allen Stichproben für alle Aspekte besteht, wurden die Hypothesen 4a-d in jeder Stichprobe noch einmal separat berechnet. Dabei wurde das gleiche Vorgehen verwendet, was zuvor beschrieben wurde. Da die Stichprobe der hochrangigen Führungskräfte mit $n = 90$ relativ klein ausfiel, wurde diese mit der Stichprobe der Linienführungskräfte zusammengelegt, sodass die weiteren Analysen mit vier verschiedenen Stichproben berechnet wurden (kaufmännische Auszubildende, technische Auszubildende, Expert:innen und Führungskräfte).

Insgesamt wurden in dieser Studie viele Hypothesen mit demselben Datensatz berechnet. Diese Mehrfachtestung kann zur Kumulation des 5% Alpha-Niveaus führen und somit den Fehler erster Art erhöhen (Bender & Lange, 2001). Dies würde bedeuten, dass die Ergebnisse ggf. signifikant werden, obwohl diese in einer Einzeltestung nicht signifikant

geworden wären. Eine Art dagegen zu wirken ist die Adjustierung, die das Alpha durch die Anzahl an Testungen dividiert. Diese Adjustierung hat jedoch zur Folge, dass die Hypothesen sehr konservativ getestet werden und somit eventuell Hypothesen nicht signifikant werden, obwohl sie eigentlich in der Einzeltestung signifikant sind (Bender & Lange, 2001). Dies wird als Fehler zweiter Art bezeichnet. Aus diesem Grund sind sich Forschende nicht einig, ob eine Alpha-Adjustierung notwendig ist oder nicht. Bender und Lange (2001) empfehlen daher, bei jeder Fragestellung abzuwägen, welcher Fehler kontrolliert werden soll und sich vor diesem Hintergrund für oder gegen eine Adjustierung zu entscheiden. In dieser Studie entschied sich die Autorin gegen eine Adjustierung, da die Ergebnisse erste Ansätze zur Validität und Reliabilität liefern sollten und somit die Kontrolle des Fehlers zweiter Art im Vordergrund stand. Darüber hinaus war die Analyse in den einzelnen Stichproben nur ein Zusatz, um ggf. weitere Forschungsthemen zu erkennen. Somit konnte auch hier die Kontrollierbarkeit des Fehlers zweiter Art gerechtfertigt werden. Bei einer umfänglichen Validität besteht jedoch das Problem, dass die Hypothesenbildung sehr komplex ist und somit zu einer hohen Multiplizität führt. So wurden alleine in Hauptstudie vier Kriterien mit vier Aspekten überprüft und dies würde zu einem Alpha-Niveau von 0,3% führen (5%-Niveau durch 16 Testungen dividieren), was vermuten lässt, dass kaum Ergebnisse signifikant werden können.

Ergebnisse

Der Ergebnisteil besteht aus zwei Teilen, da der Fragebogen AEOS (Wedemeyer et al., in Vorbereitung), welcher in dieser Studie validiert wurde, in einer Vorstudie gekürzt wurde. Zu dieser Vorstudie folgt im nächsten Abschnitt lediglich eine kurze Zusammenfassung, da das Hauptaugenmerk auf der Hauptstudie lag.

Vorstudie

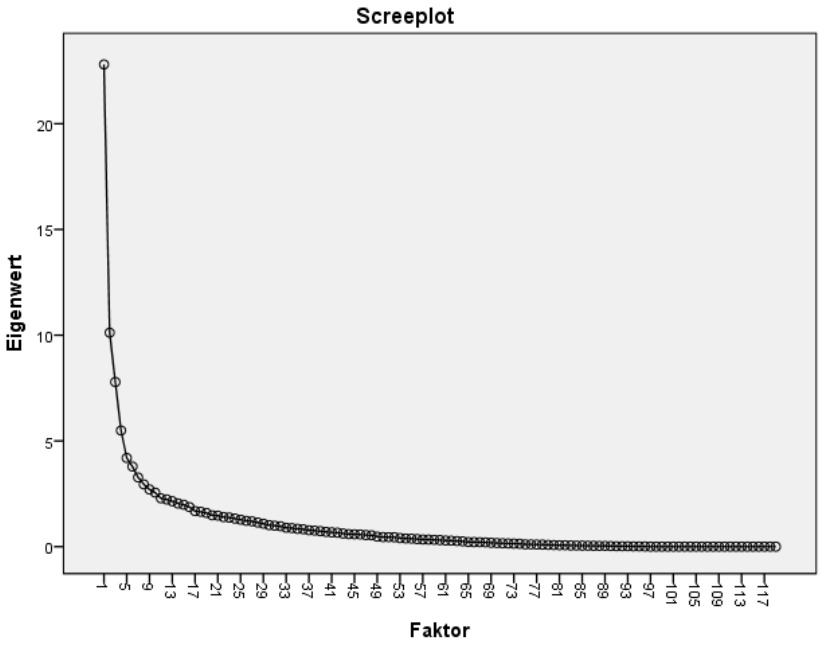
Die Vorstudie wurde an beruflichen Schulen durchgeführt und verfolgte den Zweck, den Fragebogen einzukürzen, damit dieser in Auswahl- und Potenzialerkennungsverfahren neben herkömmlichen Fragebögen eingesetzt werden kann. Der ursprüngliche Fragebogen umfasste 119 Items, welche wie folgt zu den vier Aspekten zuordbar waren:

- 30 Enthusiasmus
- 30 Durchsetzungsfähigkeit
- 30 Offenheit
- 29 Intellekt

Die erste berechnete explorative Faktorenanalyse zeigte im Screeplot einen Knick bei sechs Faktoren (s. Abbildung 7). Daher wurde die nächste Faktorenanalyse festgesetzt auf sechs Faktoren berechnet.

Abbildung 7

Screeplot der ersten explorativen Faktorenanalyse mit 119 Items

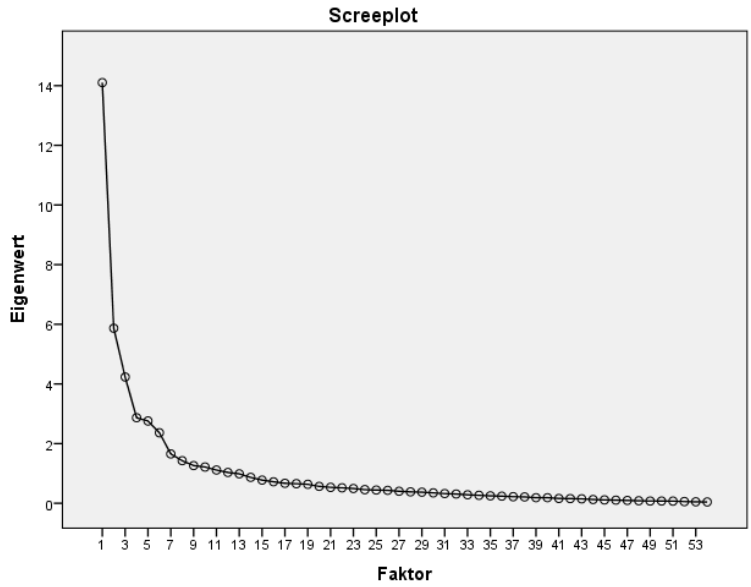


In der nächsten Faktorenanalyse mit den sechs festgelegten Faktoren wurden die Faktorladungen betrachtet und untersucht, welche Items eine Ladung unter $\lambda = .5$ besaßen. Dies war der Fall bei 66 Items und so wurde die Faktorenanalyse mit 53 Items weiterberechnet. Der Screeplot ließ eine 4-Faktoren-Struktur erkennen (s. Abbildung 8).

Abschließend wurden noch diejenigen Items entfernt, welche von den Teilnehmenden mit nur drei Ausprägungen bewertet wurden oder auf mehreren Faktoren hoch luden. Alle Items wurden ein weiteres Mal auf Verständlichkeit und die inhaltliche Passung zu den Aspekten kontrolliert. Daraus resultierte ein aus 39 Items bestehender abschließender Fragebogen, die mit 9 Items auf dem Aspekt Enthusiasmus, 8 Items auf den Aspekten Durchsetzungsfähigkeit und Offenheit sowie 14 Items auf dem Aspekt Intellekt luden. Die berechneten Reliabilitäten lagen zwischen $\alpha = .69$ und $\alpha = .89$ und wurden nach George und Mallery (2002) als kritisch bis gut bewertet. Obwohl die Reliabilität in dem Offenheitsaspekt geringer ausfiel, wurden keine weiteren Items vor der Hauptstudie ausgeschlossen, um die Inhaltsvalidität der Skala nicht einzuschränken.

Abbildung 8

Screeplot der explorativen Faktorenanalyse mit 53 Items



Hauptstudie

Nach Erhebung der Daten wurden zu Beginn der Auswertung noch einmal die Kennwerte der Items berechnet und kontrolliert, ob Items aus den Skalen entfernt werden müssen. Zunächst wurde die konfirmatorische Faktorenanalyse berechnet und die Faktorladungen in dem 4-Faktoren-Modell betrachtet. Als Cut-Off-Wert für eine zu niedrige Ladung wurde nach Brown (2015) $\lambda = .3$ gewählt. Die Skala Enthusiasmus bestand aus 9 Items. Das Item „ Die Zusammenarbeit mit anderen in meinem Arbeitsumfeld macht mir Spaß.“ besaß eine Faktorladung von $\lambda = .22$ und wurde daher entfernt. Die Itemanalyse dieser Skala wird in Tabelle 8 dargestellt. Nach Bühner (2011) war die Trennschärfe als akzeptabel bis hoch zu bewerten und die Itemschwierigkeit lag im mittleren Bereich. Daher wurden keine weiteren Items aus der Skala entfernt. Die Reliabilität lag für diese Skala bei $\alpha = .82$. Somit ist für diese Skala die zweite Hypothese bestätigt.

Tabelle 8
Skalenkennwerte für die Skala Enthusiasmus

Item	Trennschärfe	Itemschwierigkeit
Es fällt mir leicht, im Arbeitsleben auf Andere zuzugehen.	.36	.39
Ich freue mich darauf, neue Kollegen kennenzulernen.	.67	.41
Ich tausche mich bei der Arbeit gern mit Kollegen aus.	.39	.40
Ich bin neuen Kollegen gegenüber eher verschlossen.	.68	.37
Gegenüber neuen Kollegen öffne ich mich nur langsam.	.58	.34
In Pausenzeiten ziehe ich mich meistens aus dem Team zurück.	.52	.36
Ich bin ein fröhlicher Mitarbeiter.	.44	.41
Im Arbeitsalltag habe ich eine positive Ausstrahlung.	.64	.38

Anmerkung. Die Trennschärfe (r_{it}) und die Itemschwierigkeiten (P_i) sind für jedes Item dargestellt.

Die zweite Skala der Dimension Extraversion war die des Aspekts Durchsetzungsfähigkeit, welche mit 8 Items erfasst wurde. Alle Items besaßen Faktorladungen über $\lambda = .48$ und lagen daher über dem Cut-Off-Wert. Bei der Betrachtung der Itemanalyse wurde ersichtlich, dass auch die Trennschärfe im akzeptablem bis hohen Bereich und die Itemschwierigkeit im mittleren Bereich lag. In dieser Skala musste kein weiteres Item entfernt wurde (s. Tabelle 9). Die Reliabilität lag bei $\alpha = .81$. Dies ist nach George und Mallery (2002) als hoch zu bewerten. Daher wurde auch für diese Skala die zweite Hypothese bestätigt.

Tabelle 9
Skalenkennwerte für die Skala Durchsetzungsfähigkeit

Item	Trennschärfe	Itemschwierigkeit
In Arbeitsgruppen bestimme ich gern die Richtung.	.52	.30
Innerhalb meines Teams kann ich mich nur schwer durchsetzen.	.52	.39
Ich kann meine Meinung bei der Arbeit leicht durchsetzen.	.61	.31
Bei der Arbeit gelingt es mir, eigene Interessen durchzusetzen.	.44	.34
Bei der Arbeit kann ich meine eigenen Ziele auch gegen Widerstand durchsetzen.	.49	.31
Bei der Arbeit besitze ich ein hohes Durchsetzungsvermögen.	.56	.33
Bei der Arbeit kann ich mit Argumenten meinen eigenen Standpunkt gut vertreten.	.54	.39
In beruflichen Situationen gebe ich ungern den Ton an.	.57	.32

Anmerkung. Die Trennschärfe (r_{it}) und die Itemschwierigkeiten (P_i) sind für jedes Item dargestellt.

Die dritte Skala ist die Offenheitsskala, welche der Dimension Offenheit für Erfahrung zuzuordnen ist. Diese Skala besaß zu Beginn der Hauptanalyse 8 Items, von denen jedoch drei eine Faktorladung zwischen $\lambda = .15$ und $\lambda = .25$ besaßen. Diese wurden aus dem Fragebogen ausgeschlossen. Die Itemanalyse zeigte auch für diese Skala mittlere bis hohe Trennschärfen und eine mittlere Itemschwierigkeit (s. Tabelle 10). Die zweite Hypothese konnte für die Offenheitsskala bestätigt werden, da das Cronbachs Alpha bei $\alpha = .81$ lag.

Tabelle 10
Skalenkennwerte für die Skala Offenheit

Item	Trennschärfe	Itemschwierigkeit
Neue Aufgaben bei der Arbeit finde ich sehr reizvoll.	.30	.64
Im Arbeitsleben bin ich nicht kreativ.	.61	.58
Bei der Arbeit probiere ich gern kreative Vorgehensweisen aus.	.72	.54
Ich arbeite gern in einem Unternehmen, das meine Kreativität zulässt und fördert.	.69	.61
In meinem Arbeitsumfeld arbeite ich gern mit kreativen Methoden.	.72	.56

Anmerkung. Die Trennschärfe (r_{it}) und die Itemschwierigkeiten (P_i) sind für jedes Item dargestellt.

Als letzte Skala wurde die Intellektskala mit 14 Items auf die beschriebenen Kriterien überprüft. Sie war der Dimension Offenheit für Erfahrung zuzuordnen. Die Betrachtung der Faktorladungen zeigte, dass zwei Items unter dem Cut-Off-Wert nach Brown (2015) von λ = .3 lagen und somit auszuschließen waren. Die Betrachtungen der Trennschärfe und der Itemschwierigkeit zeigten, dass ein Item eine negative Trennschärfe aufwies und 11 Items in einem hohen Bereich lagen (.54 < r_{it} <.66). Da eine negative Trennschärfe für einen Mangel des Items spricht, wurde dieses Item entfernt (Kelava & Moosbrugger, 2020). Eine hohe Trennschärfe spricht dafür, dass sich die Items sehr ähnlich sind und daher wurden in dieser Skala Interkorrelationen der Items gerechnet. Die Berechnung zeigte auf, dass drei Items mit einem *r* zwischen .68 < *r* < .69 stark unter einander korrelierten. Daraufhin wurden die drei Items inhaltlich überprüft. Es stellte sich heraus, dass diese auch inhaltlich sehr ähnlich formuliert waren, wie folgendes Beispiel zeigt:

- In meinem Job suche ich gern nach neuen intellektuellen Herausforderungen.
- In meinem Arbeitsumfeld interessiere ich mich für intellektuelle Herausforderungen.

Daher wurde das verständlichste Item ausgewählt und die anderen beiden Items gelöscht. Somit bestand der Fragebogen nun aus 9 Items. Eine erneute Betrachtung der Trennschärfe und Itemschwierigkeit zeigte, dass die Itemschwierigkeit von einem hohen (.25 < P_i < .28) in den mittleren Bereich gesunken war (.33 < P_i < .37), die Trennschärfe aber immer noch im mittleren Bereich lag (s. Tabelle 11). Die Reliabilität lag nach Ausschluss des Items mit der negativen Trennschärfe bei α = .89. Nach dem Ausschluss der beiden inhaltlich ähnlichen und hoch korrelierenden Items sank die Reliabilität auf α = .86. Die zweite Hypothese konnte bestätigt werden.

Tabelle 11
Skalenkennwerte für die Skala Intellekt

Item	Trennschärfe	Itemschwierigkeit
Meine Neugierde hilft mir bei der Arbeit.	.60	.36
In meinem Beruf bin ich wissbegierig.	.63	.35
Um weiter zu lernen, probiere ich neue Arbeitsweisen aus.	.46	.33
Auch über mein Fachgebiet hinaus möchte ich gern beruflich hinzulernen.	.58	.37
Ich setze mich gern mit neuen, beruflichen Themen auseinander.	.66	.36
Ich widme mich gern komplizierten Arbeitsaufgaben.	.60	.34
In meinem Job suche ich gern nach neuen intellektuellen Herausforderungen.	.58	.33
Intellektuell anspruchsvolle Arbeitsaufgaben reizen mich.	.62	.33
Komplexe Aufgaben versuche ich bei meinem Job zu vermeiden.	.61	.35

Anmerkung. Die Trennschärfe (r_{it}) und die Itemschwierigkeiten (P_i) sind für jedes Item dargestellt.

Konstruktvalidität. Im Anschluss an die Itemanalyse wurde die Konstruktvalidität betrachtet. Dies wurde auf zwei Weisen durchgeführt. Zum einen wurde die faktorielle Validität anhand einer konfirmatorischen Faktorenanalyse berechnet und zum anderen wurde das nomologische Netzwerk anhand der konvergenten und divergenten Validität erfasst. Im nächsten Berechnungsschritt wurde die Hypothese 1 überprüft und betrachtet, ob der Fragebogen die vier Aspekte nach DeYoung et al. (2007) erfasst oder die zwei Faktorenlösung mit dem Big Five Ansatz eine bessere Passung aufweist, wodurch die faktorielle Validität gegeben wäre. Für diese Berechnung wurden drei Modelle anhand einer konfirmatorischen Faktorenanalyse berechnet und diese anhand der Fit-Werte und den *AIC*-Wert mit einander verglichen. Im Modell 1 luden alle Items auf einem Faktor, Modell 2 beinhaltete zwei Faktoren (Extraversion und Offenheit für Erfahrung). Das dritte Modell erfasste die vier Aspekte (Enthusiasmus, Durchsetzungsfähigkeit, Offenheit und Intellekt) nach DeYoung et al. (2007). Erkennbar war, dass sich die Fit-Werte verbesserten und zu den Richtwerten nach West et al. (2012) tendierten (s. Tabelle 12).

Tabelle 12

Ergebnisse konfirmatorische Faktorenanalyse

Modell	χ^2 (df)	*CFI*	*TLI*	*RMSEA*	*SRMR*	*AIC*
Modell 1	5137.88 (405)***	.69	.67	.10	.08	96105.08
Modell 2	4523.04 (404)***	.73	.71	.09	.07	95263.35
Modell 3	3285.38 (400)***	.81	.80	.08	.07	93688.68
Modell 4	3915.40 (398)***	.77	.75	.09	.07	94497.50
Modell 5	3508.57 (397)***	.80	.78	.08	.06	93948.80
Modell 6	2855.62 (393)***	.84	.83	.07	.06	93109.15

Anmerkung. Modell 1 beschreibt die einfaktorielle Lösung. Modell 2 erfasst die 2-Faktoren-Lösung. Modell 3 betrachtet die 4-Faktoren-Lösung. Modell 4 bis 6 sind identisch aufgebaut, es wurden jedoch sieben Modification Indices zugelassen. $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Die Fit-Werte *CFI* und *TLI* sollen nach West et al. (2012) ≥ .95 sein. Dies ist in keinem Modell der Fall, die Werte sind im dritten Modell jedoch am höchsten. Des Weiteren besteht ein guter Modellfit, wenn die *RMSEA* ≤ .06 und der *SRMR* ≤ .08 sind. Auch dies ist in diesen Modellen nicht gegeben, jedoch ist die richtige Tendenz erkennbar. Es wird von einer besseren Modellpassung ausgegangen, wenn der χ^2-Wert klein ist (Gäde, Schermelleh-Engel & Brandt, 2020). Da der χ^2-Wert im dritten Modell kleiner ist als im zweiten Modell, wurde die Hypothese 1 laut der Fit-Werte bestätigt. Zur finalen Bestätigung der Hypothese 1 wurde der *AIC*-Wert betrachtet, welcher die unterschiedlichen Anzahlen der Freiheitsgrade in nicht geschachtelten Modellen mitberücksichtigt (Werner, 2012). Ein Modell ist umso besser, wenn der *AIC*-Wert klein ist. Der Wert war bei Modell 3 am geringsten und somit konnte die Hypothese 1 bestätigt werden (s. Tabelle 12).

Im nächsten Schritt wurden die Modification Indices betrachtet, um zu schauen, ob zwischen den Items Zusammenhänge aufgrund von Methodeneffekten bestanden. Alle Modification Indices wurden inhaltlich verglichen und nach gleichen Wörtern überprüft. Es wurden sieben Zusammenhänge für die weiteren Berechnungen zugelassen. Die sprachlichen Ähnlichkeiten waren beispielsweise wie folgt:

- Ich bin **neuen Kollegen gegenüber** eher verschlossen.
- **Gegenüber neuen Kollegen** öffne ich mich nur langsam.
- Ich kann meine Meinung **bei der Arbeit** leicht **durchsetzen**.
- **Bei der Arbeit** kann ich meine eigenen Ziele auch gegen Widerstand **durchsetzen**.
- **Bei der Arbeit** besitze ich ein hohes **Durchsetz**ungsvermögen.

Nachdem die oben beschriebenen Modification Indices zugelassen wurden, erfolgte wiederholt eine Berechnung einer CFA für die drei Modelle. Die Fit-Werte aus Modell 3

waren am besten (s. Tabelle 12). Durch die Modellanpassung lagen die *CFI*- und *TLI*-Werte noch näher an dem Richtwert ≤ .95 (West, et al., 2012). Auch der χ^2 -Wert sank weiter, was für eine bessere Modellpassung spricht (Gäde, Schermelleh-Engel & Brandt, 2020). Obwohl die Fit-Werte für die Bestätigung der Hypothese 1 sprachen, wurde auch in diesen drei Modellen der *AIC* betrachtet. Die Berechnungen zeigten, dass der *AIC* im Modell 6 geringer war, als in den Modellen 4 und 5. Dadurch konnte die Hypothese 1 bestätigt werden. Das Modell 6 mit den Modification Indices wies einen besseren *AIC*-Wert auf als das Modell 3 ohne Modification Indices. Aus diesem Grund wurden in allen weiteren Berechnungen die Modification Indices zugelassen.

Nachdem die Hypothese 1 und somit die faktorielle Validität bestätigt war, wurde anhand einer CFA das nomologische Netzwerk betrachtet. Dazu wurde zu Beginn die konvergente und divergente Validität im Zusammenhang mit den Big Five Dimensionen des NEO-FFI berechnet. Hypothese 3a-b gingen davon aus, dass der Fragebogen eine mittlere positive Korrelation zu ähnlichen Konstrukten aufweist ($r > |.5|$; Bühner, 2011). Daraus folgte, dass Intellekt und Offenheit im mittleren Bereich positiv mit der Offenheit für Erfahrung korrelieren müssten (Hypothese 3b) und Enthusiasmus und Durchsetzungsfähigkeit mit Extraversion (Hypothese 3a). Bei der Berechnung der CFA wurde eine Fehlermeldung ausgewiesen (Varianz-Kovarianzmatrix der geschätzten Parameter scheint nicht positiv definiert zu sein), was dafürsprach, dass das Programm das vordefinierte Modell nicht replizieren konnte ($\chi^2(3872) = 23667.00$; $p < .001$; *CFI* = .61; *TLI* = .59; *RMSEA* = .06; *SRMR* = .07). Da dies für eine zu große Komplexität des Modells sprach, wurde entschieden, das Modell dahingehend zu vereinfachen und im Folgenden mit Skalenmittelwerten als manifeste Variablen zu rechnen. So konnten zwar die verschiedenen Werte jedes Items (zum Beispiel Fehlervarianz) nicht mehr erfasst werden, jedoch wurde die Struktur des Modells im gesamten betrachtet. In dieser Studie korrelierten Enthusiasmus und Extraversion mit $r = .52$ ($p < .001$) signifikant mit einander und Durchsetzungsfähigkeit und Extraversion korrelierten signifikant mit $r = .39$ ($p < .001$), womit sie unter dem Richtwert nach Bühner (2011) lagen. Dies bedeutete, dass Hypothese 3a für den Aspekt Enthusiasmus, jedoch nicht für die Durchsetzungsfähigkeit bestätigt wurde. In der Dimension Offenheit für Erfahrung sah dies anders aus. Zwar korrelierten die Aspekte Offenheit und Intellekt auf dem 5%-Niveau signifikant mit der Dimension Offenheit für Erfahrung, allerdings waren die Korrelation sehr gering ($r_{Offenheit} = .27$; $r_{Intellekt} = .31$). In diesem Fall konnte die Hypothese 3b nicht bestätigt werden.

Im nächsten Schritt wurde die divergente Validität gegenüber den NEO-FFI Dimensionen betrachtet. Die Hypothese 3c besagte, dass die Aspekte Durchsetzungsfähigkeit und Enthusiasmus $r < |.4|$ mit allen Big Five Dimensionen außer Extraversion korrelieren (Bühner, 2011). Die Korrelation zwischen dem Aspekt Enthusiasmus und der Dimensionen Offenheit für Erfahrung betrug $r = .12$. Die anderen Dimensionen lagen über dem Richtwert von $r < |.4|$ (Verträglichkeit $r = .42$; Gewissenhaftigkeit $r = .45$; Ausgeglichenheit $r = -.47$). Somit konnte die Hypothese 3c für den Aspekt Enthusiasmus lediglich für die Zusammenhänge mit der Dimension Offenheit für Erfahrung bestätigt werden und für die anderen drei Dimensionen wurde die Hypothese 3c verworfen. Die Ergebnisse zum Aspekt Durchsetzungsfähigkeit waren anders. Dieser Aspekt korrelierte unterhalb des Richtwertes mit den Dimensionen Offenheit für Erfahrung ($r = .18$), Verträglichkeit ($r = .12$) und Gewissenhaftigkeit ($r = .34$). Somit konnte für diese drei Dimensionen die Hypothese 3c bestätigt werden. Die Dimension Ausgeglichenheit korrelierte jedoch hoch negativ ($r = -.48$) mit der Durchsetzungsfähigkeit und bestätigte die Hypothese 3c nicht.

Die Hypothese 3d betrachtete analog zur Hypothese 3c die divergente Validität der Big Five Dimensionen gegenüber der Aspekte Offenheit und Intellekt der Dimension Offenheit für Erfahrung. Diese Hypothese besagte, dass die zwei Aspekte der Offenheit für Erfahrung mit den anderen vier Dimensionen $r < |.4|$ korrelieren. Der Aspekt Offenheit korrelierte unterhalb des Richtwertes mit den Dimensionen Ausgeglichenheit ($r = -.32$), Verträglichkeit ($r = .28$), Gewissenhaftigkeit ($r = .36$) sowie Extraversion ($r = .33$) und bestätigte somit die Hypothese 3d in allen Korrelationen des Aspekts Offenheit. Der Aspekt Intellekt korrelierte lediglich mit der Dimension Verträglichkeit ($r = .33$) unterhalb des Richtwertes, womit die Hypothese 3d für diesen Aspekt bestätigt worden ist. Die anderen Dimensionen Ausgeglichenheit ($r = -.44$), Gewissenhaftigkeit ($r = .45$) und Extraversion ($r = .43$) korrelierten höher mit dem Aspekt Intellekt und somit konnte die Hypothese 3d nicht bestätigt werden.

Tabelle 13

Korrelationstabelle der Business Big 5

	1	2	3	4	5	6	7	8	9	10
1 SB	1									
2 Db	.40***	1								
3 En	.36***	.52***	1							
4 D	.56 ***	.48***	.65***	1						
5 Of	.34***	.44***	.81***	.66***	1					
6 I	.37***	.51***	.78***	.67***	.89***	1				
7 Ei	.18***	.15***	-.05	.26***	-.02	.03	1			
8 B	-.15***	-.12***	-.20***	-.32***	-.20***	-.31***	.24***	1		
9 F	.37***	.50***	.50***	.47***	.57***	.68***	.22***	-.16***	1	
10 Or	.12***	.23***	.30***	.13***	.20***	.20***	.01	.06*	.38***	1

Anmerkung. Folgende Abkürzungen wurden verwendet: 1 SB (Soziale Belastbarkeit); 2 Db (Dauerbelastbarkeit); 3 En (Enthusiasmus); 4 D (Durchsetzungsfähigkeit); 5 Of (Offenheit); 6 I (Intellekt); 7 Ei (Einfühlungsvermögen); 8 B (Bescheidenheit); 9F (Fleiß) und 10 Or (Ordnung). $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Neben den Zusammenhängen zu den Big Five sollten auch die Zusammenhänge zwischen den zehn Aspekten der Business Big 5 (ABGS, AVS und AEOS) überprüft werden, um zu prüfen, ob die Skalen endogen und somit abgrenzbar voneinander sind. Die Hypothese 3e besagte, dass alle Aspekte $r < |.4|$ miteinander korrelieren. Für diese Analyse wurde eine CFA berechnet, bei der die gewünschten Fit-Werte (*CFI* und *TLI* ≥ .95) nicht erreicht werden konnten ($\chi^2(2092) = 9679.89$; $p < .001$; *CFI* = .80; *TLI* = .78; *RMSEA* = .05; *SRMR* = .08), jedoch die Werte vom *RMSEA* bei < .06 und *SRMR* bei <.08 lagen.

Die Betrachtung der Korrelationen zeigte für den Aspekt Soziale Belastbarkeit, dass der Aspekt lediglich zu dem Aspekt Durchsetzungsfähigkeit eine Korrelation $r > |.4|$ aufweist (s. Tabelle 13). Somit konnte Hypothese 3e für den Aspekt Soziale Belastbarkeit in Bezug auf alle Aspekte außer Durchsetzungsfähigkeit bestätigt werden. Demgegenüber war der Aspekt Dauerbelastbarkeit weniger von den anderen Aspekten abgrenzbar und wies nur mit den Aspekten Ordnung (r = .23), Bescheidenheit (r = -.12) und Einfühlungsvermögen (r = .15) geringe Zusammenhänge unter dem Cut-Off-Wert auf. Somit konnte nur für die Zusammenhänge zwischen dem Aspekt Dauerbelastbarkeit und den oben aufgeführten Aspekten die Hypothese 3e bestätigt werden und für die anderen Aspekte musste sie verworfen werden. Bei dem Aspekt Enthusiasmus waren vor allem die sehr hohen Korrelationen zu den Aspekten Offenheit (r = .81) und Intellekt (r = .78) auffällig, wodurch die divergente Validität nicht gegeben war und die Hypothese 3e nicht bestätigt werden konnte. Die Aspekte Einfühlungsvermögen, Bescheidenheit und Ordnung besaßen zu keinem

anderen Aspekt eine Korrelation $r > |.4|$, womit für diese Skalen die Hypothese 3e bestätigt wurde. Diese Ergebnisse zeigten, dass die Hypothese für einzelne Aspekte bestätigt wurde, sich andere Aspekte sehr ähnelten, hoch korrelierten und die Hypothese in Bezug auf diese nicht bestätigt wurde (s. Tabelle 14).

Tabelle 14
Übersicht der Hypothesenprüfung 3e der Business Big 5

	1	2	3	4	5	6	7	8	9	10
1 Soziale Belastbarkeit										
2 Dauerbelastbarkeit	+									
3 Enthusiasmus	+	-								
4 Durchsetzungsfähigkeit	-	-	-							
5 Offenheit	+	-	-	-						
6 Intellekt	+	-	-	-	-					
7 Einfühlungsvermögen	+	+	+	+	+	+				
8 Bescheidenheit	+	+	+	+	+	+	+			
9 Fleiß	+	-	-	-	-	-	+	+		
10 Ordnung	+	+	+	+	+	+	+	+	+	

Anmerkung. Das „+“ bedeutet, dass die Hypothese 3e bestätigt wurde und das „-“, dass die Hypothese 3e nicht bestätigt wurde.

Kriteriumsvalidität. Die Kriteriumsvalidität wurde aus zwei Blickwinkeln mit unterschiedlichen Variablen betrachtet. Auf der einen Seite wurde sie retrospektiv erfasst und somit überprüft, ob die zuvor erhobene Durchschnittsnote bzw. das schlussfolgernde Denken mit den Aspekten korrelieren. Auf der anderen Seite wurde die prognostische Validität berechnet, um zu betrachten, inwiefern die verschiedenen Aspekte mit dem Kriterium Berufserfolg korrelieren. Die Hypothese 4 wurden in vier Unterhypothesen für jeden Aspekt separat geprüft. Dabei bezog sich die Hypothese 4a auf den Aspekt Enthusiasmus, die Hypothese 4b auf die Durchsetzungsfähigkeit, Hypothese 4c auf Offenheit und die Hypothese 4d auf den Aspekt Intellekt.

Zunächst wurde in der Stichprobe der Auszubildenden die Korrelation der Aspekte und der Durchschnittsnote mithilfe eines Strukturgleichungsmodells berechnet. Die Hypothese 4a-d besagte, dass eine geringe Note mit einer hohen Ausprägung der Aspekte korreliert. Für die Berechnungen wurden alle fehlenden Werte in der Note aus dem Datensatz entfernt, sodass 626 Fälle übrig blieben. Bei der Betrachtung der Modellpassung wurde ersichtlich, dass die *RMSEA* = .05 und *SRMR* = .05 Werte im Normbereich nach West et al. (2012) lagen. Demgegenüber passten die Fit-Werte *CFI* = .90 und *TLI* = .89 nicht ganz. Sie lagen unter dem Cut-Off-Wert von ≥ .95. Der χ^2-Test wurde signifikant, was wiederum auch für keine optimale Passung der Daten zu dem Modell sprach ($\chi^2(418) = 992.37$; $p < .001$).

Die Betrachtung der Korrelationen in dem Strukturgleichungsmodell zeigt, dass die Note mit allen Aspekten außer Enthusiasmus signifikant korreliert (s. Tabelle 15).

Tabelle 15
Kennwerte der Korrelationen zwischen der Durchschnittsnote und den Aspekten

Aspekte	*r*	*95% CI*
Enthusiasmus	-.09	[- .08 | .08]
Durchsetzungsfähigkeit	-.26***	[.19 | .33]
Offenheit	-.13*	[-.21 | -.05]
Intellekt	-.21***	[-.28 | -.13]

Anmerkung. $N = 626$ Auszubildende; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Die negativen Werte im *r* bedeuten, dass eine geringe Note mit einer hohen Ausprägung in den Aspekten korreliert. Da in Deutschland eine geringe Note besser als eine hohe Note ist, wurde in diesem Fall die Hypothese 4b-d bestätigt, lediglich für den Aspekt Enthusiasmus konnte die Hypothese 4a nicht bestätigt werden. Die Größe der Korrelationen lag nach Cohen (1988) im kleinen bis moderaten Bereich.

Das zweite retrospektive Kriterium war das schlussfolgernde Denken. Es wurde in den Hypothesen 4a-d davon ausgegangen, dass ein hoher Wert im schlussfolgernden Denken mit hohen Ausprägungen in den einzelnen Aspekten korreliert. Diese Analyse wurde anders als bei der Note mit allen Stichproben berechnet, da alle Teilnehmenden einen Test zum schlussfolgernden Denken bearbeiteten. Auch in dieser Berechnung wurde vor der Erstellung des Strukturgleichungsmodells der Datensatz auf alle Personen, die keine fehlenden Daten in der Kriteriumsvariable besaßen, gekürzt. Der Datensatz beinhaltete somit 1218 Fälle. Die Betrachtung des Modells ergab, dass die Fit-Werte *RMSEA* und *SRMR* nach West et al. (2012) für eine gute Modellpassung sprachen, jedoch die *CLI* und *TLI* von den Richtwerten abwichen und der χ^2-Test signifikant wurde ($\chi^2(418) = 1596.11$; $p < .001$; $CFI = .90$; $TLI = .89$; $RMSEA = .05$; $SRMR = .05$). Die Korrelation wurde für alle Aspekte auf dem 5% Niveau signifikant und es wurde deutlich, dass ein hoher Wert im schlussfolgernden Denken mit hohen Ausprägungen in den Aspekten korreliert (s. Tabelle 16). Wie bei dem Kriterium der Note sind die Korrelationen klein ausgeprägt. Aus den Analysen konnte geschlossen werden, dass die Hypothesen 4a bis 4d für dieses Kriterium bestätigt wurden.

Tabelle 16

Kennwerte der Korrelationen zwischen dem schlussfolgernden Denken und den Aspekten

Aspekte	*r*	*95% CI*
Enthusiasmus	.08*	[.02 | .14]
Durchsetzungsfähigkeit	.17***	[.11 | .22]
Offenheit	.12**	[.06 | .18]
Intellekt	.22***	[.17 | .27]

Anmerkung. $N = 1218$ Bewerbende; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Neben den retrospektiven Kriterien wurden für die prognostische Validität die Eingruppierung und das Potenzial für das Kriterium Berufserfolg erhoben. Die beiden Variablen unterschieden sich dahingehend, dass die Eingruppierung in allen Stichprobengruppen in der Vorauswahl erhoben wurde und das Potenzial nicht in der Stichprobe der Auszubildenden erfasst wurde, da diese kein Assessment Center durchliefen. Somit umfasste der Datensatz 979 gültige Fälle, welche die Variable der Eingruppierung besaßen. Hypothese 4a-d besagten, dass die Ausprägungen in den Aspekten positiv mit der Eingruppierung korrelieren. Zur Erfassung der Kriteriumsvalidität wurde auch in diesem Fall ein Strukturgleichungsmodell berechnet, welches ähnliche Kennwerte zu den vorherigen Modellen aufwies und keine optimale Passung zu den Daten besaß ($\chi^2(418) = 1389.103$; $p < .001$; *CFI* = .90; *TLI* = .89; *RMSEA* = .05; *SRMR* = .05). Die Betrachtung der Korrelationen zur Überprüfung der Hypothese zeigte, dass die Hypothese für alle Aspekte auf einem 5%-Niveau bestätigt werden konnte und die Korrelationen im moderaten Bereich lagen (Cohen, 1998) (s. Tabelle 17).

Tabelle 17

Kennwerte der Korrelationen zwischen der Eingruppierung und den Aspekten

Aspekte	*r*	*95% CI*
Enthusiasmus	.27***	[.21 | .33]
Durchsetzungsfähigkeit	.30***	[.24 | .36]
Offenheit	.36***	[.30 | .41]
Intellekt	.38***	[.33 | .43]

Anmerkung. $N = 979$ Bewerbende; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Das zweite Kriterium von Berufserfolg das Potenzial wurde bei 459 Personen angegeben, da diese ein Assessment durchliefen. Es wurde in Hypothese 4a-d davon ausgegangen, dass eine hohe Ausprägung in den Aspekten mit einem niedrigen Wert im Potenzial korreliert, da ein niedriger Wert im Potenzial für eine gute Leistung sprach. Die *RMSEA*- und *SRMR*-Werte lagen wieder in dem angegebenen Bereich von West et al. (2012), die Modellpassung war jedoch bezogen auf die *CFI*- und *TLI*-Werte schlechter als in den

vorherigen Modellen ($\chi^2(418) = 887.35$; $p < .001$; *CFI* = .89; *TLI* = .87; *RMSEA* = .05; *SRMR* = .05). Der χ^2-Wert hingegen war niedriger, was auf eine bessere Modellpassung hindeutete. Wie in Tabelle 18 erkennbar ist, wurde die Korrelation nur für den Aspekt Offenheit signifikant und somit konnte nur die Hypothese 4c bestätigt werden. Die Analyse zeigte, dass die Fehlervarianz des Kriteriums bei $s^2 = 1.05$ lag und somit ein großer Teil der Varianz nicht auf die latente Variable zurückzuführen ist (Bühner, 2021).

Tabelle 18

Kennwerte der Korrelationen zwischen dem Potenzial und den Aspekten

Aspekte	*r*	*95% CI*
Enthusiasmus	-.03	[-.12 | .06]
Durchsetzungsfähigkeit	.07	[-.02 | .16]
Offenheit	.12*	[.03 | .21]
Intellekt	.08	[-01 | .17]

Anmerkung. $N = 459$ Bewerbende; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Ein weiterer Bestandteil der Kriteriumsvalidität ist die inkrementelle Validität. Die Hypothese 5a besagte, dass die vier Aspekte inkrementelle Validität im Kriterium Berufserfolg über das schlussfolgernde Denken hinaus aufweisen. Zur Bestimmung der inkrementellen Validität wurde ein Strukturgleichungsmodell mit dem IST-Screening berechnet, welches die Eingruppierung vorhersagt. Im Anschluss wurden die Aspekte im zweiten Modell hinzugenommen. Analog wurde dies auch mit dem NEO-FFI durchgeführt, um die Hypothese 5b zu betrachten, welche besagt, dass die vier Aspekte inkrementelle Validität im Kriterium Berufserfolg über die Big Five Dimensionen hinaus aufweisen. Alle Modellpassungen sind in Tabelle 19 zusammengefasst. Die Regression zeigte, dass das IST-Screening die Eingruppierung signifikant vorhersagte ($\beta = .69$; $p < .001$) und das Bestimmtheitsmaß $R^2 = .47$ nach Chin (1998) als mittelgut zu bewerten war.

Tabelle 19

Ergebnisse Strukturgleichungsmodelle zur inkrementellen Validität mit dem Kriterium Eingruppierung

Modell	χ^2 (df)	*CFI*	*TLI*	*RMSEA*	*SRMR*
Modell 1	2164.43 (455)***	.85	.84	.06	.07
Modell 2	1614.62 (444)***	.90	.89	.05	.05

Anmerkung. Modell 1 beschreibt die Vorhersage der Eingruppierung durch das IST-Screening und Modell 2 durch das IST-Screening und die Aspekte der AEOS. $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Die Betrachtung der Modellpassung für das Modell 2 unter Hinzunahme der Aspekte zeigte bessere Fit-Werte. Zudem lagen der *RMSEA*- und *SRMR*-Wert im Normbereich nach West et al. (2012). Die Eingruppierung wurde neben dem IST-Screening lediglich von dem Aspekt Enthusiasmus signifikant vorhergesagt ($\beta = .10$; $p < .05$), das Bestimmtheitsmaß wuchs jedoch auf $R^2 = .53$. Dies bedeutet, dass der Aspekt die Varianzaufklärung um 5,9% gegenüber dem IST-Screening erhöhte und somit konnte die Hypothese 5a nur teilweise bestätigt werden, da die anderen Aspekte die Varianzaufklärung nicht erhöhten.

Die gleichen Berechnungen wurden mit dem NEO-FFI durchgeführt. Bei der Berechnung der Strukturgleichungsmodelle erfolgte eine Fehlermeldung, dass die Varianz bzw. Kovarianz nahe Null ist. Daher entschied sich die Autorin dazu, multiple Regressionen zu berechnen, welche für die Vorhersage der Eingruppierung durch den NEO-FFI signifikant wurden ($F(5,804) = 31.53$; $p < .001$). Das Bestimmtheitsmaß mit dem $R^2 = .16$ war als sehr schwach zu bewerten (Chin, 1998). Diese Berechnung wurde unter Hinzunahme der Skalenmittelwerte der AEOS erneut berechnet und wiederum signifikant ($F(9,800) = 19.05$; $p < .001$). Das Bestimmtheitsmaß wuchs auf $R^2 = .18$ an und somit konnte der Fragebogen AEOS die Varianzaufklärung gegenüber dem NEO-FFI um 1,2% erhöhen. Die Hypothese 5b wurde auch in diesem Fall bestätigt.

Die inkrementelle Validität wurde nicht nur in dem Kriterium Eingruppierung betrachtet, sondern auch in dem Kriterium Potenzial. Es wurde die gleiche Vorgehensweise wie bei der Eingruppierung gewählt. Die Hypothesen 5a-b besagten, dass die vier Aspekte inkrementelle Validität im Kriterium Berufserfolg über das schlussfolgernde Denken (Hypothese 4a) bzw. über die Big Five Dimensionen (Hypothese 4b) hinaus aufweisen.

Die Fit-Werte des Strukturgleichungsmodells mit dem IST-Screening zeigten, dass die Modellpassung bei den *RMSEA*- und *SRMR*-Werten als gut zu bewerten war, die *CFI*- und *TLI*-Werte jedoch zu niedrig waren (s. Tabelle 20). Das IST-Screening sagte das Potenzial signifikant vorher ($\beta = -.15$; $p < .01$) und das Bestimmtheitsmaß betrug $R^2 = .02$. Die Hinzunahme der Aspekte in dem nächsten Modell verbesserte die Fit-Werte, die *CFI*- und *TLI*-Werte lagen jedoch unter dem Cut-Off-Wert von West et al. (2012). Die Regression zeigte, dass sich das Bestimmtheitsmaß auf $R^2 = .07$ erhöhte, aber kein Aspekt das Kriterium Potenzial signifikant auf dem 5% Niveau vorhersagte. Somit konnte die Hypothese 5a nicht bestätigt werden. Die Berechnungen der Strukturgleichungsmodelle für den NEO-FFI erreichten in keinem der Fit-Werte den Cut-Off-Wert. Daher wurde sich gegen eine Interpretation des Modells und für eine Regressionsanalyse entschieden (s. Tabelle 20). Die erste Berechnung, in welcher der NEO-FFI das Potenzial vorhersagte, wurde nicht signifikant

($F(5,450) = 1.84$; p = n.s.) und das Bestimmtheitsmaß betrug $R^2 = .02$. Die Hinzunahme der AEOS Aspekte führte ebenfalls nicht zu einer Signifikanz der Regression und somit konnte die Hypothese 5b nicht bestätigt werden ($F(9,446) = 1.22$; p = n.s.).

Tabelle 20
Ergebnisse Strukturgleichungsmodelle zur inkrementellen Validität mit dem Kriterium Potenzial

Modell	χ^2 (df)	*CFI*	*TLI*	*RMSEA*	*SRMR*
Modell 1	1157.02 (455)***	.84	.83	.06	.06
Modell 2	977.96 (444)***	.88	.87	.05	.05
Modell 3	18737.28 (3994)***	.32	.30	.09	.09
Modell 4	18571.81 (3983)***	.32	.31	.09	.09

Anmerkung. Modell 1 beschreibt die Vorhersage des Potenzials durch das IST-Screening und Modell 2 durch das IST-Screening und die Aspekte der AEOS. Modell 3 und 4 wurden analog aufgebaut. Das IST-Screening wurde durch den NEO-FFI ersetzt. $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Weitere Analysen

Teil dieser weiteren Analyse war die Betrachtung der Kriteriumsvalidität in den einzelnen Stichproben, um zu prüfen, ob die Aspekte mit den Kriterien in allen Gruppen ähnlich korrelieren oder in einzelnen Stichproben gewisse Aspekte besonders hoch bzw. gar nicht korrelieren. Auch in diesem Abschnitt wurden die Hypothesen 4a-d betrachtet, welche besagen, dass die Aspekte der Dimensionen Extraversion und Offenheit für Erfahrung positive Korrelationen mit dem Kriterium Berufserfolg aufweisen. Zu Beginn wurde die Korrelation der Durchschnittsnote mit den Aspekten in den beiden Stichproben der Auszubildenden betrachtet (s. Tabelle 21).

Tabelle 21
Kennwerte der Korrelationen zwischen der Durchschnittsnote und den Aspekten

Aspekte	Kfm. Auszubildende		Techn. Auszubildende	
	r	*95% CI*	*r*	*95% CI*
Enthusiasmus	.02	[-.09 | .12]	-.10	[-.21 | .02]
Durchsetzungsfähigkeit	-.20*	[-.30 | -.10]	-.22**	[-.33 | -.11]
Offenheit	.02	[-.09 | .12]	-.23*	[-.34 | -.12]
Intellekt	-.13	[-.23 | -.03]	-.22***	[-.33 | -.11]

Anmerkung. N = 348 kfm. Auszubildende und N = 278 techn. Auszubildende wurden in die Analyse mit einbezogen; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Das Strukturgleichungsmodell für die Korrelationsanalyse zwischen der Durchschnittsnote und den Aspekten der kaufmännischen Auszubildenden wurde nur in dem

Aspekt Durchsetzungsfähigkeit signifikant und zeigte einen kleinen negativen Zusammenhang. Somit wurde nur Hypothese 4b bestätigt. Die Modellgüte war in den Kennwerten *RMSEA* und *SRMR* nach West et al. (2012) als gut zu bewerten, die anderen Kennwerte lagen unter den Cut-Off-Werten ($\chi^2(418) = 711.39$; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .05; *SRMR* = .06). In der Stichprobe der technischen Auszubildenden sahen die Korrelationen anders aus und es wurden alle Aspekte bis auf Enthusiasmus auf dem 5%-Niveau signifikant. Die Korrelationen konnten nach Cohen (1998) als klein bis moderat bewertet werden. Die Modellpassung war jedoch im Vergleich zu den kaufmännischen Auszubildenden schlechter ($\chi^2(418) = 772.72$; $p < .001$; *CFI* = .86; *TLI* = .85; *RMSEA* = .06; *SRMR* = .06). Somit wurden die Hypothesen 4b-d in Bezug auf die Durchschnittsnote bei den kaufmännischen Auszubildenden bestätigt.

Als nächstes Kriterium wurde das schlussfolgernde Denken im Zusammenhang mit den Aspekten betrachtet (s. Tabelle 22). Es wurde in den Hypothesen 4a-d davon ausgegangen, dass ein hoher Wert im schlussfolgernden Denken mit hohen Ausprägungen in den einzelnen Aspekten korreliert. Dieser Zusammenhang wurde für alle Stichproben separat untersucht. Die zuvor berechnete Analyse mit allen Stichproben zusammen ergab eine positive Korrelation zwischen allen Aspekten und dem schlussfolgernden Denken auf dem 5-% Niveau.

Tabelle 22

Kennwerte der Korrelationen zwischen dem schlussfolgernden Denken und den Aspekten

Aspekte	Kfm. Auszubildende		Techn. Auszubildende		Expert:innen		Führungskräfte	
	r	*95% CI*	*r*	*95% CI*	*r*	*95% CI*	*r*	*95% CI*
Enthusiasmus	-.13*	[.03 | .23]	.09	[-.03 | .21]	.01	[-.12 | .15]	.16*	[-06 | .26]
Durchsetzungs-fähigkeit	.04	[-.06 | .14]	.21***	[.10 | .32]	.09	[-.05 | .22]	.16*	[.06 | .26]
Offenheit	.03	[-.07 | .13]	.29**	[.18 | .39]	-.10	[-.04 | .23]	.03	[-.07 | .13]
Intellekt	.08	[-.02 | .18]	.32***	[.21 | .42]	.02	[-.12 | .16]	.19**	[.09 | .29]

Anmerkung. $N = 377$ kfm. Auszubildende, $N = 280$ techn. Auszubildende, $N = 210$ Expert:innen und $N = 351$ Führungskräfte wurden in die Analyse mit einbezogen; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

In der Betrachtung der getrennten Stichproben wurden die verschiedenen Anforderungen der Stichproben sichtbar und so unterschieden sich die Korrelationen zwischen den Subgruppen. Es zeigte sich für die Gruppe der kaufmännischen Auszubildenden eine bessere Modellgüte als für die der technischen Auszubildenden (Kfm: $\chi^2(418) = 744.01$; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .05; *SRMR* = .06; Techn: $\chi^2(418)$

= 775.25; $p < .001$; *CFI* = .86; *TLI* = .85; *RMSEA* = .06; *SRMR* = .06). Demgegenüber zeigte sich nur eine negative signifikante Korrelation zwischen dem Aspekt Enthusiasmus und dem schlussfolgernden Denken, wohingegen in der technischen Subgruppe alle Aspekte außer Enthusiasmus positiv mit dem schlussfolgernden Denken korrelierten. Dies bedeutet, dass für die kaufmännische Stichprobe Hypothese 4a-d nicht bestätigt werden konnten und bei den technischen Auszubildenden die Hypothesen 4b-d bestätigt wurden. Die Modellpassung in der Expert:innen Stichprobe war nach West et al. (2012) in den Fit-Werten *RMSEA* und *SRMR* als gut zu bewerten, es scheint jedoch in dieser Stichprobe keinen Zusammenhang zwischen dem schlussfolgernden Denken und den Aspekten zu geben ($\chi^2(418)$ = 608.25; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .05; *SRMR* = .06). Dies bedeutet, dass in der Stichprobe der Expert:innen die Hypothesen 4a-d nicht bestätigt werden konnten. In der Gruppe der Führungskräfte wurden die Korrelationen zwischen den Aspekten Enthusiasmus, Durchsetzungsfähigkeit sowie Intellekt positiv mit dem schlussfolgernden Denken signifikant ($\chi^2(418)$ = 687.12; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .04; *SRMR* = .05). Daraus kann geschlossen werden, dass die Hypothesen 4a, 4b und 4d bestätigt sind.

Das nächste zu betrachtende Kriterium war die Eingruppierung, welche in der Gesamtstichprobe mit allen Aspekten positiv korrelierte. Hypothese 4a-d besagten, dass die Ausprägungen in den Aspekten positiv mit der Eingruppierung korrelieren. In der Stichprobe der technischen Auszubildenden waren die Ergebnisse vergleichbar mit der Gesamtstichprobe und die Korrelation zwischen allen Aspekten und der Eingruppierung lagen im moderaten bis guten Bereich (Cohen, 1988) (s. Tabelle 23). Damit konnten die Hypothesen 4a-d bestätigt werden. Dabei lag die Modellgüte in den Fit-Werten *CFI* und *TLI* unter der Stichprobe der kaufmännischen Auszubildenden, welche nur schwach in den Aspekten Durchsetzungsfähigkeit, Offenheit und Intellekt mit der Eingruppierung korrelierten (Kfm: $\chi^2(418)$ = 740.64; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .05; *SRMR* = .05; Techn: $\chi^2(418)$ = 769.16; $p < .001$; *CFI* = .87; *TLI* = .85; *RMSEA* = .06; *SRMR* = .06). Für die Stichprobe der kaufmännischen Auszubildenden wurden somit die Hypothesen 4b-d bestätigt. Auch in den Stichproben der Expert:innen und Führungskräfte korrelierten alle Aspekte signifikant positiv mit der Eingruppierung auf einem moderaten bis guten Niveau (Cohen, 1988). Die Modellpassung der Expert:innen war jedoch deutlich besser als die der Führungskräfte (Expert:innen: $\chi^2(418)$ = 568.37; $p < .001$; *CFI* = .91; *TLI* = .90; *RMSEA* = .05; *SRMR* = .07; FK: $\chi^2(418)$ = 644.03; $p < .001$; *CFI* = .88; *TLI* = .87; *RMSEA* = .06; *SRMR* = .06). Daraus kann abgeleitet werden, dass für die beiden Stichproben die Hypothesen 4a-d bestätigt wurden.

Tabelle 23

Kennwerte der Korrelationen zwischen der Eingruppierung und den Aspekten

Aspekte	Kfm. Auszubildende		Techn. Auszubildende		Expert:innen		Führungskräfte	
	r	*95% CI*	*r*	*95% CI*	*r*	*95% CI*	*r*	*95% CI*
Enthusiasmus	.11	[.00 | .21]	.34***	[.23 | .44]	.31*	[.15 | .45]	.41**	[.28 | .53]
Durchsetzungs-fähigkeit	.17**	[.07 | .27]	.45***	[.35 | .54]	.43***	[.29 | .56]	.50***	[.38 | .60]
Offenheit	.19**	[.09 | .29]	.44***	[.34 | .53]	.45**	[.31 | .57]	.54***	[.43 | .64]
Intellekt	.28***	[.18 | .37]	.49***	[.40 | .57]	.44***	[.30 | .56]	.46***	[.34 | .57]

Anmerkung. $N = 378$ kfm. Auszubildende, $N = 280$ techn. Auszubildende, $N = 143$ Expert:innen und $N = 178$ Führungskräfte wurden in die Analyse mit einbezogen; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Zuletzt wurde das Kriterium Potenzial im Zusammenhang mit den Aspekten überprüft, welches lediglich in den beiden Gruppen der Expert:innen und Führungskräfte während eines Assessments erhoben wurde. Es wurde in Hypothese 4a-d davon ausgegangen, dass eine hohe Ausprägung in den Aspekten mit einem niedrigen Wert im Potenzial korreliert, da ein niedriger Wert im Potenzial für eine gute Leistung sprach. Die vorangegangene Untersuchung in der Gesamtstichprobe zeigte einen signifikanten Zusammenhang zwischen dem Aspekt Offenheit und dem Potenzial. Die separaten Analysen zeigten, dass diese Signifikanz ausschließlich auf der Stichprobe der Führungskräfte beruhte, da in der Gruppe der Expert:innen keine signifikanten Korrelationen zwischen den Aspekten und dem Potenzial zu erkennen waren ($\chi^2(418) = 660.81$; $p < .001$; *CFI* = .88; *TLI* = .86; *RMSEA* = .06; *SRMR* = .07) (s. Tabelle 24). Daraus lässt sich ableiten, dass in dieser Stichprobe die Hypothesen 4a-d nicht bestätigt wurden. In der Stichprobe der Führungskräfte war hingegen eine schwach positive Korrelation zwischen Offenheit und Intellekt gegenüber dem Potenzial zu erkennen. Diese Korrelation widerspricht der Hypothese, da das Potenzial eine invers kodierte Variable war und somit eine positive Korrelation dafürsprach, dass eine hohe Merkmalsausprägung mit einem geringen Potential zusammenhängt. Die Modellpassung war nach West et al. (2012) in den *CFI*- und *TLI*-Werten nicht gegeben ($\chi^2(418) = 711.19$; $p < .001$; *CFI* = .87; *TLI* = .86; *RMSEA* = .05; *SRMR* = .06). Somit konnte auch für die Stichprobe keine der Hypothesen 4a-d bestätigt werden.

Tabelle 24

Kennwerte der Korrelationen zwischen dem Potenzial und den Aspekten

Aspekte	Expert:innen		Führungskräfte	
	r	*95% CI*	*r*	*95% CI*
Enthusiasmus	-.04	[-.18 | .10]	-.03	[-.15 | .09]
Durchsetzungs-fähigkeit	.02	[-.12 | .16]	.07	[-.05 | .19]
Offenheit	.03	[-.11 | .17]	.16*	[.04 | .27]
Intellekt	-.03	[-.17 | .11]	.14*	[.02 | .25]

Anmerkung. $N = 185$ Expert:innen und $N = 274$ Führungskräfte wurden in die Analyse mit einbezogen; $p < .05$ (*), $p < .01$ (**), $p < .001$ (***)

Zusammengefasst führten die Untersuchungen zu den Erkenntnissen, dass sich der Fragebogen als reliabel und faktoriell valide erweist. Demgegenüber zeigte sich, dass der Fragebogen nur bedingt divergent und konvergent valide ist und vor allem die Aspekte der Business Big 5 teilweise nicht trennbar sind sowie hoch miteinander korrelieren. Der Fragebogen scheint auch nicht für alle Kriterien valide zu sein und die Stichproben unterscheiden sich in den Signifikanzen. Insgesamt besitzt der Fragebogen jedoch inkrementelle Validität über das schlussfolgernde Denken und die Big Five Dimensionen hinweg in Bezug auf das Kriterium der Eingruppierung.

Diskussion

Ziel der Studie war es, den neu entwickelten sowie berufsbezogen formulierten Fragebogen zu den vier Aspekten der Extraversion und Offenheit für Erfahrung nach DeYoung et al. (2007) auf seine Gütekriterien zu untersuchen. Dabei lag der Fokus auf der Konstruktvalidität (faktorielle Validität und nomologisches Netzwerk), der Reliabilität sowie der Kriteriumsvalidität – mit der inkrementellen Validität inbegriffen.

Konstruktvalidität

Die erste Hypothese besagte, dass die vier Aspekte-Struktur nach DeYoung et al. (2007) eine bessere Modellpassung aufweist als die zweifaktorielle Lösung nach den Big Five Dimensionen Extraversion und Offenheit für Erfahrung (Costa & McCrae, 1992). Diese Untersuchung wurde mittels einer konfirmatorischen Faktorenanalyse durchgeführt, anhand derer die Hypothese bestätigt werden konnte. Aufgrund dessen kann davon ausgegangen werden, dass der Fragebogen die vier Aspekte Enthusiasmus, Durchsetzungsfähigkeit sowie Offenheit und Intellekt besser abbildet als die beiden Big Five Dimensionen und somit die Faktorenstruktur nach DeYoung (2007) repliziert werden konnte. Diese Ergebnisse deuten darauf hin, dass in der Praxis die spezifischere Erfassung der Persönlichkeit in beispielsweise Auswahlverfahren eine bessere Konstruktvalidität ausweist. Somit können die Konstrukte

genauer erfasst werden. Für die Theorie bedeutet die Bestätigung der ersten Hypothese, dass neben Judge et al. (2013) eine weitere Studie die differenzierte Big Five Struktur nach DeYoung et al. (2007) replizieren konnten und somit in weiteren Studien diese Struktur nicht vernachlässigt werden sollte. So könnten weitere Forschungsarbeiten noch vertiefender die Zusammenhänge zwischen den Aspekten und weiteren Kriterien betrachten und somit zur Kriteriumsvalidität der berufsbezogenen Aspekte einen Beitrag leisten. Dies könnte zum Beispiel der Zusammenhang zwischen den Aspekten und den objektiven Kriterien Gehalt und Anzahl der Beförderungen sein.

Die zweite Hypothese betrachtete die Reliabilität und besagte, dass der Fragebogen reliabel ist. Die Reliabilität lag zwischen $\alpha = .81$ (Durchsetzungsfähigkeit und Offenheit) und $\alpha = .86$ (Intellekt) in einem guten Bereich und somit konnte bestätigt werden, dass der Fragebogen reliabel ist. Dies bedeutet, dass das Konstrukt der vier Aspekte nach DeYoung et al (2007) durch diesen Fragebogen messgenau erfasst wird und somit die Messgültigkeit des Fragebogens gegeben ist (Gäde, Schermelleh-Engel & Werner, 2020). In dieser Studie wurde als Reliabilität ausschließlich die interne Konsistenz betrachtet. Ergänzend zur internen Konsistenz sollte nachfolgend noch die Retest-Reliabilität untersucht werden, um zu prüfen, ob der Fragebogen auch über die Zeit hinweg messgenau ist. Diese Ergebnisse deuten darauf hin, dass die AEOS reliabel die vier Aspekte der Extraversion und Offenheit für Erfahrung nach DeYoung et al. (2007) berufsbezogen erfassen kann und somit ein Einsatz des Fragebogens in der Theorie und Praxis bei Fragestellungen empfehlenswert ist. Dies könnte zum Beispiel in der Personalpraxis der Vertriebsauswahl oder in der Potenzialerkennung von Nachwuchsführungskräften zu differenzierteren Vorhersagen und passgenauen Trainingsmaßnahmen führen. In der Theorie könnte der Fragebogen in Fragestellungen, welche die spezifischere Betrachtung der Big Five betrachten, hilfreich sein.

Ein weiterer Teil der Konstruktvalidität sind neben der faktoriellen Validität die konvergente sowie divergente Validität, welche das nomologische Netzwerk abbilden. Zum einen wurden die Korrelationen der Fragebogenaspekte mit den Big Five Dimensionen anhand des NEO-FFI betrachtet. Zum anderen fand die Betrachtung mit den anderen berufsbezogen formulierten acht Aspekten nach DeYoung et al. (2007) durch die Business Big 5 statt. Die Hypothese 3a nimmt an, dass die Aspekte Enthusiasmus und Durchsetzungsfähigkeit positiv im mittleren Bereich mit der Dimension Extraversion korrelieren. In der Berechnung der CFA mit den Items des NEO-FFI und der AEOS wies das Datenanalyseprogramm Fehlermeldungen aus. Das vordefinierte Modell konnte nicht identifiziert werden. Daher wurde das Modell vereinfacht und nur Skalenmittelwerte als

manifeste Variablen wurden verwendet. Die Hypothese 3a konnte nur für den Aspekt Enthusiasmus bestätigt werden. Dies bedeutet, dass eine hohe Ausprägung im Aspekt Enthusiasmus mit einer hohen Ausprägung in der Dimension Extraversion korreliert. Der Aspekt Durchsetzungsfähigkeit korreliert zwar signifikant mit der Dimension Extraversion, unterschritt jedoch die Faustregel nach Bühner (2011) und somit konnte die Hypothese nicht bestätigt werden. Dies könnte zwei Ursachen haben. Zum einen könnte der Aspekt in den sehr heterogenen Stichproben unterschiedlich beantwortet worden sein, was zu einer Verzerrung der Korrelation führen kann (Brandt & Moosbrugger, 2020). Zum anderen könnte auch die berufsbezogene Formulierung zu einer niedrigeren Korrelation geführt haben. Dieser Faktor wurde vor allem bei der Hypothese 3b ersichtlich, die annahm, dass die Aspekte Offenheit und Intellekt positiv im mittleren Bereich mit der Dimension Offenheit für Erfahrung korrelieren. Dies konnte in beiden Aspekten nicht bestätigt werden. Da die Items im NEO-FFI sehr kulturell (zum Beispiel Theater, Lieblingsessen und Kunst) formuliert waren, wichen die Items zu den Aspekten im beruflichen Kontext sehr stark ab. Die Ergebnisse zeigen, dass das Konstrukt der berufsbezogenen Aspekte doch weiter als gedacht von den NEO-FFI Items abzuweichen scheint und die konvergente Validität nur in dem Aspekt Enthusiasmus bestätigt werden konnte. Daraus lässt sich schließen, dass in der Praxis der Einsatz empfehlenswert sein könnte, um einen differenzierteren Blick auf die Persönlichkeit zu erlangen. Darüber hinaus sollte in der Wissenschaft jedoch erforscht werden, ob der Unterschied in weiteren Stichproben auch besteht und wodurch diese Unterschiede entstehen.

Die Hypothesen 3c und d zur Betrachtung der divergenten Validität besagten, dass die Aspekte der Dimensionen Extraversion (Hypothese 3c) und Offenheit für Erfahrung (Hypothese 3d) niedrig nach der Faustregel von Bühner (2011) mit den anderen vier Dimensionen korrelieren. Bei dem Aspekt der Extraversion wurde ersichtlich, dass die Hypothese zum Teil bestätigt werden konnte, da der Zusammenhang zwischen dem Aspekt und der Dimension Offenheit für Erfahrung gering war, jedoch die anderen Dimensionen höher als die Faustregel mit Enthusiasmus korrelierten. Diese Ergebnisse bedeuten, dass das Konstrukt Enthusiasmus nur von der Offenheit für Erfahrung abgrenzbar ist und mit den Dimensionen Neurotizismus, Gewissenhaftigkeit und Verträglichkeit hohe Zusammenhänge besitzt. Dies würde den Nutzen des Einsatzes des Fragebogens herabsetzen, da die Zusammenhänge darauf hindeuten, dass die Erfassung des Aspektes inhaltlich keinen Mehrwert zur Aufklärung der Persönlichkeitsmerkmale in der Praxis liefert. Die Betrachtung der divergenten Validität in dem Aspekt Durchsetzungsfähigkeit zeigt hingegen eine klarere

Tendenz in Richtung der Bestätigung der Hypothese 3c. Lediglich im Zusammenhang zwischen diesem Aspekt und der Ausgeglichenheit konnte die Hypothese nicht bestätigt werden. Diese Analyse zeigte, dass der Aspekt Durchsetzungsfähigkeit von den Dimensionen Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit abgrenzbar ist und somit ein Einsatz des Fragebogens über den NEO-FFI hinweg sinnvoll zu sein scheint, da der Fragebogen Konstrukte über die Dimensionen hinweg erfasst. Die divergente Validität wurde auch anhand der Aspekte der Offenheit für Erfahrung analog zur Hypothese 3c betrachtet. Der Aspekt Offenheit korrelierte mit den Dimensionen Ausgeglichenheit, Extraversion, Verträglichkeit und Gewissenhaftigkeit mit $r < |.4|$ und somit konnte die Hypothese 3d nach der Faustregel von Bühner (2011) bestätigt werden. In dem Aspekt Intellekt lag die Korrelation nur mit der Dimension Verträglichkeit unter dem Wert $r < |.4|$ und somit konnte die Hypothese 3d nur für diese Dimension bestätigt werden. Die Ergebnisse der Hypothesen 3c und 3d deuten darauf hin, dass weiter untersucht werden sollte, ob durch den Berufsbezug die Dimensionszuordnung der Aspekte noch besteht, oder ob dadurch die Dimensionen eher überlappend erfasst werden. Darüber hinaus sollten auch die Zusammenhänge der Aspekte in weiteren Stichproben untersucht werden. So könnten beispielsweise die divergente und konvergente Validität in einzelnen Berufsgruppen separat betrachtet werden, um zu sehen, ob die Validitäten in allen Stichproben nicht vollständig bestehen oder ob der Fragebogen nur für einzelne Stichproben valide ist. Dies könnte auch einen Mehrwert für die Praxis liefern, welche dann mehr Informationen besitzen, für welche Stichproben der Fragebogen empfehlenswert ist und für welche der Einsatz keinen Mehrwert liefert.

Im nächsten Schritt wurde die divergente Validität in dem Fragebogen Business Big 5 mit den zehn Aspekten nach DeYoung et al. (2007) betrachtet. Die Hypothese besagte, dass alle zehn Aspekte $r < |.4|$ miteinander korrelieren. Dies wurde wieder anhand einer konfirmatorischen Faktorenanalyse betrachtet. Die Analyse zeigte, dass die Hypothese 3e nicht vollständig bestätigt werden konnte. Dies war zwar für die Korrelationen der Aspekte Einfühlungsvermögen, Bescheidenheit und Ordnung mit allen anderen Aspekten der Fall, demgegenüber gab es aber auch starke Korrelationen zwischen den Aspekten der Dimensionen Extraversion und Offenheit für Erfahrung. Auch in diesen Fällen können Methodeneffekte, vor allem die ähnlichen Formulierungen, Auswirkungen auf das nomologische Netzwerk haben (Gäde, Schermelleh-Engel & Werner, 2020). Darüber hinaus können auch die heterogenen Gruppen zu Korrelationsverzerrungen führen (Brandt & Moosbrugger, 2020).

Zum nomologischen Netzwerk kann festgehalten werden, dass die Hypothesen zur konvergenten und divergenten Validität zum Teil bestätigt werden konnten und somit die Messgültigkeit bzw. die Abgrenzung zu inhaltlich fernen Konstrukten nur zum Teil besteht. Aufgrund der Methodeneffekte sollte jedoch im Nachgang das nomologische Netzwerk für jede Stichprobengruppe separat betrachtet werden, damit die Korrelationen durch die starke Stichprobenheterogenität nicht beeinflusst werden. In der weiteren Forschung könnte sich erneut den Items gewidmet werden, um die inhaltlichen Formulierungen dahingehend zu betrachten, ob ggf. Items umformuliert oder einzelne Items aus dem Fragebogen entfernt werden sollten, wenn sich die Items zwischen den Aspekten zu ähnlich sind. Eine weitere Annahme zu den abweichenden Korrelationen könnte die berufsbezogene Formulierung der Items sein, die zu anderen Korrelationen in Bezug auf die Aspekte bzw. Dimensionen führt. So kann es sein, dass die Gewissenhaftigkeit und Offenheit für Erfahrung nicht miteinander korrelieren, aber die Aspekte Fleiß und Intellekt zusammenhängen, weil ggf. die Zielstrebigkeit und Offenheit für Weiterbildungen in Zusammenhang gebracht werden können. Dies sollte nachfolgend weiter untersucht werden, um Zusammenhänge der berufsbezogenen Aspekte noch genauer erklären zu können.

Insgesamt lässt sich über die Konstruktvalidität festhalten, dass der Fragebogen faktoriell valide und reliabel ist. Dies bedeutet, dass die vier Aspekte durch den Fragebogen messgenau erfasst werden und somit ein Einsatz in der Praxis empfehlenswert ist. Demgegenüber scheint der Berufsbezug jedoch für eine schlechtere Abgrenzbarkeit der Konstrukte zu führen und es sollte in weiteren Untersuchungen überprüft werden, ob dies nur in der Stichprobe besteht oder das Problem generell auffindbar ist. Darüber hinaus könnte auch die separate Betrachtung der Stichproben einen Aufschluss darüber liefern, inwiefern der Fragebogen für alle Stichproben einen Mehrwert im Einsatz liefert. Generell kann jedoch gesagt werden, dass die Ergebnisse zur faktoriellen Validität die Annahmen von DeYoung et al. (2007) replizieren und der Fragebogen die vier Aspekte besser erfasst als die zwei Dimensionen Extraversion und Offenheit für Erfahrung.

Kriteriumsvalidität

Neben der Konstruktvalidität war auch die Überprüfung der Kriteriumsvalidität Teil dieser Studie. Dies wurde auf zweierlei Arten betrachtet, zum einen retrospektiv mit der Durchschnittsnote und dem schlussfolgernden Denken und zum anderen prognostisch mit den Berufserfolg Kriterien Eingruppierung und Potenzial. Retrospektiv bedeutete hier, dass die Kriterien vor dem Fragebogen erhoben wurden und prognostische Variablen im Nachgang

erfasst wurden (Beauducel & Leue, 2014). Die Hypothesen 4a-d besagten, dass die Aspekte einen positiven Zusammenhang mit den Kriterien aufweisen. Diese Hypothesenprüfung fand wieder anhand eines Strukturgleichungsmodells mit Korrelationsanalyse statt.

Hypothese 4a besagte, dass der berufsbezogen formulierte Aspekt Enthusiasmus eine positive Korrelation mit den Kriterien Berufserfolg, schlussfolgerndes Denken und Ausbildungserfolg aufweist. In der Gesamtstichprobe wurden die Zusammenhänge zwischen dem Aspekt Enthusiasmus und der Eingruppierung sowie dem schlussfolgernden Denken signifikant. Damit korrelierte eine hohe Merkmalsausprägung mit einem hohen Wert in der Eingruppierung und im schlussfolgernden Denken und die Hypothese wurde für die beiden Kriterien bestätigt. Die Korrelationen waren dagegen nach Cohen (1988) als schwach zu bewerten. Die Aufschlüsselung der einzelnen Stichprobengruppen zeigten eine signifikante Korrelation zwischen Enthusiasmus und dem Kriterium schlussfolgerndes Denken in den Stichproben der kaufmännischen Auszubildenden und Führungskräfte. Eine schwach negative Korrelation der Auszubildenden widerlegt die Hypothese 4a für diese Stichprobe. Somit scheint eine niedrige Merkmalsausprägung im Aspekt Enthusiasmus mit einem hohen Wert im schlussfolgernden Denken zusammenzuhängen. In der Betrachtung der Eingruppierung wurde ersichtlich, dass der Aspekt in allen Stichprobengruppen bis auf die kaufmännischen Auszubildenden moderat bis gut mit der Eingruppierung korreliert (Cohen, 1988). In keiner Stichprobe wurden Zusammenhänge zwischen dem Potenzial sowie der Durchschnittsnote und dem Aspekt Enthusiasmus gefunden. Somit konnte Hypothese 4a nur vereinzelt für verschiedene Stichproben und unterschiedliche Kriterien bestätigt werden und die Erkenntnis von Judge et al. (2013), dass Enthusiasmus mit Berufserfolg korreliert, nur teilweise repliziert werden. Diese Ergebnisse zeigen, dass die Stichproben unterschiedliche Kriteriumsvaliditäten in Bezug auf den Aspekt Enthusiasmus aufweisen. Vor allem in der Stichprobe der Führungskräfte scheint es entscheidend für den Berufserfolg zu sein, ob eine Person gesellig ist oder nicht. Dies würde die Ergebnisse von Barrick und Mount (1991) replizieren, welche vor allem bei Manager:innen eine hohe Korrelation zwischen der Extraversion und dem Berufserfolg fanden. Für die Forschung kann daraus geschlossen werden, dass verschiedene Kriterien betrachtet werden sollten, da nicht alle Kriterien mit den Persönlichkeitsmerkmalen zusammenhängen.

Im nächsten Schritt wurde der Zusammenhang zwischen dem Aspekt Durchsetzungsfähigkeit und den Kriterien betrachtet. Hypothese 4b besagte, dass der berufsbezogene Aspekt Durchsetzungsfähigkeit eine positive Korrelation mit den Kriterien Berufserfolg, Ausbildungserfolg und schlussfolgerndes Denken aufweist. Dies bedeutet, dass

eine hohe Ausprägung in diesem Aspekt mit einem hohen Wert in den Kriterien korreliert. Dies konnte in der Gesamtstichprobe für die Kriterien Ausbildungserfolg, schlussfolgerndes Denken und die Eingruppierung des Kriteriums Berufserfolg bestätigt werden. Daraus folgt, dass eine hohe Ausprägung des Aspekts mit den Kriterien zusammenhängt, somit im Arbeitsleben wünschenswert sein könnte und der Aspekt valide ist. In der separaten Betrachtung der Stichproben wurde ersichtlich, dass die Durchsetzungsfähigkeit nicht mit dem schlussfolgernden Denken in den Stichproben kaufmännische Auszubildende und Expert:innen korreliert. In diesen Stichproben scheint das Kriterium in Zusammenhang mit dem Durchsetzungsfähigkeit nicht von entscheidender Bedeutung bzw. unabhängig voneinander zu sein. Somit scheint nicht jede Person mit einem hohen Wert im schlussfolgernden Denken automatisch auch eine hohe Durchsetzungsfähigkeit zu besitzen. Auffällig war darüber hinaus, dass die Eingruppierung mit der Durchsetzungsfähigkeit eine hohe Korrelation von $r = .5$ in der Führungskräftegruppe aufwies. Demgegenüber lag die Korrelation bei den kaufmännischen Auszubildenden nur bei $r = .167$. Dies würde die Studie von Barrick und Mount (1991) stützen, die vor allem bei den Manager:innen einen Zusammenhang zwischen Berufserfolg und der Dimension Extraversion fanden. In dem Kriterium Potenzial entstanden keine signifikanten Ergebnisse. Dadurch konnte auch die Hypothese 4b nur bedingt bestätigt werden. Diese Ergebnisse zeigen, dass es zukünftig sinnvoll sein dürfte, die Kriterien in den einzelnen Stichproben separat zu betrachten, um aussagekräftige Erkenntnisse zu erlangen, da die Stichproben sehr heterogen zu sein scheinen und es somit zu Korrelationsverzerrungen kommen kann. Darüber hinaus zeigt auch dieser Aspekt, dass es in der Forschung interessant sein könnte, in Persönlichkeitsstudien den Aspekt vor allem nicht bei Führungskräften außer Acht zu lassen.

Die Hypothese 4c betrachtete den Zusammenhang zwischen dem Aspekt Offenheit und den Kriterien. Dieser Aspekt war der einzige, bei dem Judge et al. (2013) keinen Zusammenhang mit Berufserfolg fanden. Trotzdem wurde die Hypothese gewählt, da Salgado (1997) einen Zusammenhang zwischen Berufserfolg und der übergeordneten Dimension Offenheit für Erfahrung erkannt hat. Ausschließlich dieser Aspekt korrelierte in dieser Studie bei der Betrachtung der Gesamtstichprobe mit allen Kriterien signifikant und somit konnte die Hypothese 4c bestätigt werden. Dies könnte auch die Annahme von Schermuly et al. (2019) stützen, dass die Dimension Offenheit für Erfahrung durch die Digitalisierung und Agilität immer bedeutsamer in der Arbeitswelt wird. Die separate Betrachtung in den einzelnen Gruppen ließ erkennen, dass auch in dem Aspekt Offenheit Gruppenunterschiede bestehen. So scheint nur in der Stichprobe der technischen und nicht in

der Gruppe der kaufmännischen Auszubildenden eine kleine bis moderate Korrelation mit dem Aspekt Offenheit zu bestehen (Cohen, 1988). Die Signifikanz der Korrelation zwischen dem Aspekt und dem schlussfolgernden Denken in der Gesamtstichprobe scheint nur auf die technischen Auszubildenden zurückzuführen zu sein. Dies würde bedeuten, dass vor allem in technischen Berufen ein hoher Wert im schlussfolgernden Denken mit einer hohen Offenheit zusammenhängt und dadurch auch in der Personalauswahl berücksichtigt werden sollte. In allen anderen Stichproben waren keine Signifikanzen erkennbar. Demgegenüber zeigten die Ergebnisse einen kleinen (kaufmännische Auszubildende) bis großen (Führungskräfte) positiven Zusammenhang zwischen dem Aspekt Offenheit und der Eingruppierung in allen Stichproben. Diese Unterschiede lassen den Schluss zu, dass für den Erfolg von Führungskräften der Aspekt einen größeren Bestandteil für sich in Anspruch zu nehmen scheint als bei den Auszubildenden. Dies könnte auch durch inhaltliche Aspekte der Positionen begründet sein. Die Führungskraft muss zum Beispiel zuerst offen gegenüber Neuerungen sein, damit sie diese auch an ihre Mitarbeitenden weitergibt. Des Weiteren muss eine Führungskraft eine große Themenrepräsentanz besitzen, damit sie die vielen verschiedenen Themen der Mitarbeitenden bedienen kann. Dies würde auch für eine erforderliche hohe Ausprägung in der Offenheit für Erfahrung sprechen. Auch der Zusammenhang zwischen dem Potenzial und dem Aspekt wurde lediglich in der Gruppe der Führungskräfte signifikant, was diese These stützen würde. Insgesamt bestätigen die positiven Zusammenhänge des berufsbezogenen Aspektes die Aspektebetrachtung und die bisher schwachen Ergebnisse der Offenheit für Erfahrung, da der berufsbezogene Aspekt Offenheit valide ist und somit einen entscheidenden Beitrag in der Personalauswahl liefern kann. Da die Offenheit auch durch die Agilität immer mehr in den Vordergrund zu rücken scheint, sollten auch nachfolgend Zusammenhänge betrachtet werden, um der Praxis mehr Wissen und Handlungsempfehlungen in Bezug auf die Offenheit für Erfahrung an die Hand geben zu können.

Der letzte zu betrachtende Aspekt in Verbindung mit den Kriterien ist der Intellekt, welcher die Offenheit gegenüber intellektueller Weiterbildung thematisierte. Da beispielsweise Judge et al. (2013) einen positiven Zusammenhang herausfanden, stützte sich die Hypothese 4d auch auf diese Vermutung. In der Gesamtstichprobe wurden kleine bis große signifikante Zusammenhänge zwischen dem Aspekt und den Kriterien Durchschnittsnote, schlussfolgerndes Denken sowie der Eingruppierung gefunden. Lediglich für das Potenzial konnte die Hypothese nicht bestätigt werden. Dies sprach für eine teilweise Replizierung der Ergebnisse von Judge et al. (2013) sowie Barrick und Mount (1991) und die

teilweise Bestätigung der Hypothesen. Die Betrachtung des Zusammenhangs zwischen der Durchschnittsnote und dem Aspekt in den separaten Gruppen zeigte nur ein signifikantes Ergebnis in der Gruppe der technischen Auszubildenden. Dies spricht dafür, dass vor allem in dieser Stichprobe ein Zusammenhang zwischen einer guten Schulnote und einer hohen Merkmalsausprägung besteht. Dieser Zusammenhang zu dem schlussfolgernden Denken ist lediglich in den Stichproben der technischen Auszubildenden sowie der Führungskräfte ersichtlich. Die Ergebnisse deuten darauf hin, dass es vor allem bei den technischen Auszubildenden und Führungskräften entscheidend für den Berufserfolg ist, dass sie offen für intellektuelle Weiterbildungen sind, um sich immer weiter zu entwickeln. In den beiden Kriterien der Eingruppierung und des Potenzials waren die gleichen Tendenzen wie in dem Aspekt Offenheit erkennbar, was die bereits oben genannte Annahme von Schermuly et al. (2019), dass die Aspekte der Dimension Offenheit für Erfahrung im Wandel der Arbeitswelt wichtiger werden, weiter stützt. Vor allem im Bereich des lebenslangen Lernens könnte der Aspekt Intellekt eine wichtige Bedeutung einnehmen und sollte daher weiterführend in der Wissenschaft berücksichtigt werden. Die bestehenden Zusammenhänge der Kriterien und dem Aspekt Intellekt sprechen dafür, dass dieser bisher eher vernachlässigte Aspekt in der Personalauswahl auch mehr an Bedeutung gewinnen und einen Teil der Eignungsdiagnostik darstellen sollte.

Die Ergebnisse zeigen, dass vor allem in Bezug auf die Kriteriumsvalidität große Unterschiede zwischen den Aspekten und den einzelnen Stichproben bestehen. Dies spricht dafür, dass es für die Wissenschaft entscheidend sein könnte, die Zusammenhänge zwischen den Stichproben und den Aspekten in Bezug auf Kriterien zu untersuchen, damit die Praxis noch genauer weiß, in welcher Stichprobe der Einsatz der AEOS entscheidend sein kann und welche Aspekte in welchen Stichproben eher in den Hintergrund treten. Eine Erstellung von Mustern oder Profilen könnte dabei hilfreich sein. Die oben aufgeführten Ergebnisse sprechen jedoch durch die unterschiedlichen Zusammenhänge der Aspekte mit den Kriterien für eine differenzierte Betrachtung der Persönlichkeit durch die Aspekte.

Die Ergebnisse zeigen deutlich, dass nicht jede Hypothese für jedes Kriterium bestätigt werden konnte und somit nicht jeder Aspekt gleich wichtig für die Kriterien ist und die Kriterien unterschiedliche Qualität in Bezug auf Berufserfolg besitzen. Genauso scheinen einzelne Merkmalsausprägungen in unterschiedlichen Stichprobengruppen entscheidender im Zusammenhang mit den Kriterien zu sein als andere. Daher sollten nachfolgende Analysen in den einzelnen Stichproben separat betrachtet werden, um Verzerrungen vorzubeugen.

Die Hypothese 5a besagte, dass die vier Aspekte inkrementelle Validität über das schlussfolgernde Denken hinaus aufweisen. Darüber hinaus wurde auch die Hypothese 5b untersucht, welche besagt, dass die vier Aspekte inkrementelle Validität über die Big Five Dimensionen hinaus aufweisen. Alle Analysen sollten mittels eines Strukturgleichungsmodells mit den Kriterien Eingruppierung und Potenzial berechnet werden. Es entstanden in den Modellberechnungen der Hypothese 5b Fehlermeldungen oder die Modellpassung war nicht gut. Daher wurde auf die Berechnung multipler Regressionen zurückgegriffen. Die Betrachtung der Hypothese 5a zeigte, dass im Kriterium der Eingruppierung die Varianzaufklärung durch die Hinzunahme der Aspekte um 5,9% gegenüber dem schlussfolgernden Denken erhöht werden konnte. Dies beruhte jedoch nur auf dem Aspekt Enthusiasmus. Bei den anderen Aspekten wurde die Regression nicht signifikant. Dies könnte vor allem an den heterogenen Stichproben gelegen haben, da die zuvor beschriebenen Korrelationen starke Abweichungen zwischen den Gruppen aufzeigten. Deshalb sollte nachfolgend die inkrementelle Validität separat für jede Stichprobe betrachtet werden. Dies spricht dafür, dass Enthusiasmus eine Art Major Aspekt ist, der Relevanz für alle Berufsgruppen von Auszubildenden bis zu hochrangigen Führungskräften besitzt, während die anderen Aspekte für die jeweilige Berufsgruppen teilweise bedeutsam sind, teilweise aber auch nicht. Weitere Forschungsarbeiten könnten untersuchen, ob Enthusiasmus vielleicht auch vielmehr ein allgemeiner Antriebsaspekt anstatt ein wirklicher Persönlichkeitsaspekt und somit eher ein motivationaler Repräsentant ist. Dennoch sprechen die Ergebnisse dafür, dass der Einsatz einen Mehrwert zusätzlich zu dem IST-Screening schafft und in der Personalauswahl berücksichtigt werden sollte. Eine weitere Erforschung der herangetragenen Kritikpunkte sind jedoch zuvor notwendig, um sicher zu stellen, dass nicht wirklich nur ein Aspekt den Mehrwert liefert und ggf. in verschiedenen Stichprobengruppen der Mehrwert auch gering ausfällt. In dem Kriterium Potenzial wurden keine Regressionen signifikant, was auch auf die Fehlervarianz zurückzuführen sein könnte, welche in den Limitationen genauer erklärt wird. Die Varianzaufklärung wurde gegenüber den Dimensionen des NEO-FFI in dem Kriterium Eingruppierung um 1,2% erhöht. Dies würde für den Einsatz des Fragebogens über den NEO-FFI hinaus sprechen. In der Vorhersage des zweiten Kriteriums Potenzialaussage wurden die Regressionen nicht signifikant.

Limitationen

Die Modellpassung der CFA zur Überprüfung der Hypothese 1 war in den *CFI*- und *TLI*-Werten nach West et al. (2012) nicht ideal und beim χ^2-Test wurde die Nullhypothese

abgelehnt. Das bedeutet, dass die Daten nicht ideal zu dem Modell passen (Gäde, Schermelleh-Engel & Brandt, 2020). Zur Verbesserung der Modellpassung wurden die Modification Indices betrachtet und sieben Korrelationen zwischen den Items zugelassen, weil sie inhaltlich nah beieinander lagen oder mehrere gleiche Wörter besaßen. Dabei wurde konservativ vorgegangen, da die Modifikationen durch Stichprobeneffekte entstanden sein könnten und diese eventuell in weiteren Untersuchungen mit anderen Stichproben nicht zu finden wären (Gäde, Schermelleh-Engel & Brandt, 2020). Darüber hinaus könnten auch Methodeneffekte zu Fehlerkovarianzen geführt haben, welche nachfolgend mit einer schlechten Modellpassung einhergehen könnten (Gäde, Schermelleh-Engel & Brandt, 2020). Ein Methodeneffekt könnte die Verwendung von invertierten bzw. rekodierten Items sein (Gäde, Schermelleh-Engel & Werner, 2020), welche nachfolgend durch die Verwendung eines weiteren Faktors „Invers“ untersucht werden könnte. Die Invertierung von einzelnen Items wurde explizit in diesem Fragebogen vorgenommen, um Antworttendenzen beispielsweise die Zustimmungstendenz abzuschwächen (Kelava & Moosbrugger, 2020). Darüber hinaus können auch ähnlich formulierte Items zu Fehlerkovarianzen führen (Gäde, Schermelleh-Engel & Werner, 2020). Davon wurde in diesem Fragebogen ausgegangen, da die Betrachtung der Modification Indices zeigte, dass durch den Berufsbezug des Öfteren ähnliche Textbausteine verwendet wurden (zum Beispiel Arbeit, Kollegen und Kolleginnen sowie beruflich). Dies sollte in nachfolgenden Untersuchungen überprüft und ggf. Items umformuliert werden. Die Umformulierung würde zu einer neuen Datenerhebung und Analyse führen. Da der Fragebogen in drei der vier Skalen aus acht bis neun Items besteht, könnten in den Skalen auch die Itemkorrelationen betrachtet werden und ggf. weitere Items aus dem Fragebogen ausgeschlossen werden. Dies könnte ggf. weitere Fehlerkovarianzen auflösen, die Modellpassung verbessern und den Fragebogen noch ökonomischer gestalten. Im Aspekt Offenheit war die Trennschärfe von zwei Items hoch, was auf hohe Interkorrelationen hindeuten kann. Es wurde jedoch für die Inhaltsvalidität keine Löschung der Items durchgeführt, da die Skala mit nur fünf Items gerade eben dem Anspruch von mindestens vier Items genügte (Bühner, 2021). Aus diesem Grund sollte in nachfolgender Forschung die Offenheitsskala erneut untersucht und ggf. die Items noch einmal reduziert oder angepasst werden. Die weitere Kürzung des Fragebogens würde den Einsatz des Fragebogens durch seine Ökonomie vereinfachen, da dieser somit noch besser in Kombination mit anderen Fragebögen eingesetzt werden könnte.

Des Weiteren waren die Stichprobengruppen nicht nur im Alter und der Geschlechterverteilung sehr heterogen, sondern wahrscheinlich auch in den Ausprägungen

der Merkmale. Dies könnte zu Verzerrungen geführt haben, da sich beispielsweise hohe und niedrige Ausprägungen mitteln würden. Dies hätte zur Folge, dass Zusammenhänge und Ergebnisse verzerrt werden und die Aussagekraft abschwächen könnten. Daher ist es wichtig, in nachfolgenden Analysen die Stichproben ggf. getrennt voneinander zu betrachten, um detaillierter für jede Stichprobe den Mehrwert der AEOS aufzuzeigen. Darüber hinaus zeigen die unterschiedlichen Ergebnisse in den Stichproben, dass eine getrennte Normierung des Fragebogens entscheidend sein wird.

Diese Studie erhob Daten aus der Praxis, was vor allem in der Analyse zum Kriterium „*Potenzial*“ Schwierigkeiten bereitete und eine dritte Limitation darstellt. In den Analysen wurde ersichtlich, dass das Kriterium „*Potenzial*“ eine hohe Fehlervarianz aufweist, was darauf hindeutet, dass die Varianz der Variable nicht nur auf die latente Variable zurückzuführen ist, sondern auch von anderen latenten Variablen abhängen kann (Bühner, 2021). Dies kann zum Beispiel bedeuten, dass die Varianz in der Potenzialaussage auch durch das Geschlecht oder Alter der Versuchspersonen beeinflusst wird. Dies führt dazu, dass die Ergebnisse des Potenzials schwer zu interpretieren sind. Daher könnte es hilfreich sein, in nachfolgenden Untersuchungen andere Außenkriterien wie beispielsweise das Gehalt oder Vorgesetztenbeurteilungen zu wählen und dort die Zusammenhänge mit den Aspekten zu betrachten. Um keine Fehlinterpretationen vorzunehmen, wird dieses Kriterium nachfolgend in den Implikationen nicht weiter erwähnt. Dennoch könnte diese Limitation dazu führen, dass der Einsatz der Variable generell auch in der Praxis hinterfragt wird, da ggf. die Variable die Personalentscheidung auch falsch beeinflusst und somit ungeeignet ist. Da die Variable „Potenzial“ von Beobachtenden abgegeben wird, könnte diese auch auf Beobachtungsfehlern (zum Beispiel HALO-Effekt; ein Merkmal überstrahlt die anderen in der Wahrnehmung) entstehen und somit die Fehlervarianz erhöhen.

Eine weitere Limitation stellt die Alphafehler-Kumulation dar. Es wurden viele Modelle und Hypothesen ohne Alpha-Adjustierung auf dem 5% Niveau gerechnet. Dadurch kann das 5%-Niveau des Alphafehlers aufgebläht bzw. kumuliert werden und somit können zufällige Signifikanzen entstanden sein (Bender & Lange, 2001). Dies würde bedeuten, dass die Ergebnisse einen Grad an Interpretierbarkeit verlieren und diese Berechnungen noch einmal mit einer Alphaadjustierung durchgeführt werden sollten. Da diese Studie jedoch sehr komplex in der Vielfalt der Hypothesen war, wurde sich gegen die Adjustierung entschieden, um für die nachfolgenden Forschungsarbeiten Ansatzpunkte zu erhalten. Aus diesem Grund sollte jedoch vor allem in weiteren, weniger komplexen Studien die Adjustierung verwendet

werden, um zu schauen, welche Ergebnisse falsch signifikant geworden sind und somit auch den Fehler erster Art zu kontrollieren.

Implikation und weitere Forschung

Diese Studie konnte aufzeigen, dass der Fragebogen reliabel ist, die vier Aspekte nach DeYoung et al. (2007) misst und somit auch die Validität gegeben ist. Darüber hinaus konnte festgestellt werden, dass der Fragebogen in verschiedenen Stichproben mit unterschiedlichen Kriterien zusammenhängt und somit davon ausgegangen werden kann, dass die Kriteriumsvalidität gegeben ist. Darüber hinaus konnten die Erkenntnisse von Barrick und Mount (1991) sowie Judge et al. (2013) in vielen Teilen repliziert werden. Die Zusammenhänge zwischen den Aspekten Offenheit und Intellekt mit den Kriterien sprechen für die Vermutung von Schermuly et al. (2019), dass vor allem die Dimension Offenheit für Erfahrung während der Digitalisierung und in agilen Teams immer entscheidender wird. Diese Erkenntnisse sprechen für den Einsatz der AEOS in der Praxis, da dadurch eine differenziertere Persönlichkeitserfassung möglich ist. Daher sollte der Fokus in anschließenden Forschungsarbeiten stärker auf diese Dimension und ihre Aspekte gelenkt werden. Dabei könnte dieser Fragebogen mit dem Berufsbezug hilfreich sein, da die Dimension Offenheit für Erfahrung des NEO-FFI sehr kulturell gemessen wird. Die inkrementelle Validität zeigte für das Kriterium Berufserfolg (durch die Eingruppierung gemessen), dass der Fragebogen über das schlussfolgernde Denken 5,9% und über die Big Five Dimensionen 1,2% hinaus Varianz aufklären kann. Somit könnte der Fragebogen einen entscheidenden Beitrag in der Personalauswahl und Potenzialerkennung leisten. Er könnte durch seine Ökonomie zusätzlich in der Vorauswahl verwendet werden, um die Passung der Personen zu der vakanten Stelle zu überprüfen und somit weniger nicht geeignete Personen zum nächsten Schritt zuzulassen. Damit der vermehrte Einsatz des Fragebogens stattfindet, müssten jedoch nachfolgend mehr Analysen in den separaten Stichproben durchgeführt werden, um zu betrachten, wie die inkrementelle Validität in diesen Stichproben ist. Darüber hinaus scheinen die Itemformulierungen des Fragebogens zu ähnlich zu sein, sodass sich über eine weitere inhaltliche Itemkürzung Gedanken gemacht werden sollte. Das nomologische Netzwerk hat gezeigt, dass wahrscheinlich durch den Berufsbezug die Zusammenhänge unter den Aspekten bzw. zu anderen Dimensionen anders als erwartet auftreten und die Konstrukte schwieriger abgrenzbar sind. Dies sollte nachfolgend anhand einer Betrachtung, ob dies auf einen Stichprobeneffekt oder auf den Berufsbezug zurückzuführen ist, überprüft werden.

Für die Theorie kann aus diesen Ergebnissen abgeleitet werden, dass nachfolgend die Struktur der Aspekte nach DeYoung et al. (2007) in weiteren Forschungsarbeiten der Arbeits- und Organisationspsychologie zu empfehlen ist. Darüber hinaus zeigte die Studie, dass auch Offenheit für Erfahrung mit Berufserfolg korreliert und somit vor allem die Aspekte Offenheit und Intellekt angelehnt an die Forschungsarbeit von Schermuly et al. (2019) immer entscheidender zu sein scheinen und somit im Arbeitskontext mehr untersucht werden sollten. Da der Fragebogen reliabel und valide ist, könnte dieser Fragebogen auch in nachfolgenden Forschungsarbeiten erhoben werden, um die Konstrukte der vier Aspekte zu erfassen. Die inkrementelle Validität zeigte darüber hinaus, dass die Erhöhung der Varianzaufklärung hauptsächlich auf Grundlage des Aspekts Enthusiasmus zustande gekommen ist, was für Enthusiasmus als eine Art „*Major-Aspekt*“ spricht. Dies würde dafürsprechen, dass dieser Aspekt vor allem in der Personalauswahl zu beachten ist, jedoch auch weiter erforscht werden sollte, ob er nicht eher ein Antriebsaspekt ist und motivationale Komponenten besitzt.

Die Validierung von virtuellen Dialogübungen im eignungsdiagnostischen Kontext

Nachdem in dem vorangegangenen Teil der Arbeit (erste Studie) der Persönlichkeitsfragebogen zur Erfassung der arbeitsbezogenen Extraversions- und Offenheitsskalen (AEOS; Wedemeyer et al., in Vorbereitung) validiert wurde, wird in den nachfolgenden Abschnitten anhand einer experimentellen Studie die Validität von virtuell durchgeführten Dialogübungen sowie deren Gerechtigkeitswahrnehmung durch Teilnehmende mit der Dialogübung in Präsenz verglichen.

Gegenstand der Fragestellung

Die zweite Studie befasst sich mit der Forschungsfrage, ob virtuelle Dialogübungen die gleiche prädiktive Validität aufweisen wie Dialogübungen in Präsenz. Darüber hinaus wird der Frage nachgegangen, ob die Gerechtigkeitswahrnehmung in beiden Darbietungsarten gleich bewertet wird und die Aspekte der beiden Persönlichkeitsmerkmale Extraversion und Offenheit für Erfahrung nach DeYoung et al. (2007) mit der Gerechtigkeitswahrnehmung zusammenhängen. Die Forschungsfragen werden in Form eines experimentellen Designs mit Arbeitnehmenden untersucht.

Kurzabriss der theoretischen und empirischen Grundlagen

Vorangegangene Studien zeigten, dass die Digitalisierung der Personalauswahl immer mehr in den Vordergrund der Forschung rückt (z.B. Basch & Melchers, 2020; Blacksmith et al., 2016; Meade et al., 2007). Die Digitalisierung der Verfahren liefert Unternehmen viele Chancen, zum Beispiel durch die Verkürzung der Verfahren, die Minimierung der Kosten

und ggf. den Aufbau eines moderneren sowie besseren Images (Fellner, 2019; Konradt & Sarges, 2003). Dennoch sind viele Verfahren der Personalauswahl beispielsweise Postkorbübungen oder virtuelle Simulationsaufgaben noch nicht ausreichend erforscht. Somit kann noch nicht darauf geschlossen werden, dass die Validität in allen Verfahren unabhängig von der Darbietungsart gegeben ist (Hertel et al., 2003). Bühner (2021) definiert die Inhaltsvalidität als das Ausmaß, in dem ein Verfahren das misst (z.B. ein Konstrukt), was es messen soll. Viele Autor:innen sind sich einig, dass beispielsweise die Persönlichkeitsfragebögen in eine Onlineversion gebracht werden können, die Validität bestehen bleibt und die Objektivität durch die standardisierte Auswertung sogar erhöht wird (z.B. Chuah et al, 2006; Hertel et al., 2003). Die Objektivität umfasst dabei die Frage, inwiefern die Ergebnisse eines Tests unabhängig von Testleitenden sind (Bühner, 2021). Demgegenüber scheint die „einfache Übersetzung“ der paper-pencil Testung eines Leistungstests in eine virtuelle Version nicht empfehlenswert zu sein, da die Bearbeitung virtuell zum Teil weniger zeitintensiv ist und daher Normierungen (Bezugsystem zur Einordung des Ergebnisses; Bühner, 2021) angepasst werden müssten (Hertel et al., 2003). Auch bei einem Vergleich von Interviews in Präsenz mit virtuell durchgeführten Interviews war erkennbar, dass Bewerbende in einem virtuellen Interview schlechter bewertet wurden als in einem Interview in Präsenz (Basch et al., 2020; Blacksmith et al, 2016). Diese Ergebnisse zeigen, dass nicht alle Verfahren unabhängig von der Darbietungsart valide sind und somit die weiteren Verfahren, beispielsweise Simulationsübungen, welche durch eine starke Interaktion zwischen mindestens zwei Personen geprägt sind, dahingehend weiter erforscht werden sollten.

Ein weiterer Faktor, welchen die Unternehmen bei der Auswahl eines geeigneten Verfahrens in der Personalauswahl beachten müssen, ist die Gerechtigkeitswahrnehmung, auch Akzeptanz genannt, da eine geringe Gerechtigkeitswahrnehmung zur Ablehnung eines Stellenangebotes führen kann (Gilliland, 1993; Moldzio, 2014). Beauducel und Leue (2014) definierten die Akzeptanz als ein weiteres Nebenkriterium zur Testgüte, welche das spontane Werteurteil (positiv versus neutral versus negativ) der Probanden zu dem Verfahren beinhaltet. Unter dem Begriff Gerechtigkeit „ist ein Idealzustand ausgeglichener Interessen ohne Benachteiligung von Einzelnen (Individuum) oder Gruppen“ zu verstehen (Schmitt, 2019). Die Gerechtigkeitswahrnehmung zielt in dieser Arbeit auf die Wahrnehmung der Bewerbenden ab. Eine geringe Gerechtigkeitswahrnehmung durch Bewerbende kann beispielsweise zu einer Ablehnung der vakanten Stelle oder zu einem schlechten Image des Unternehmens führen, was bei dem heutigen Fachkräftemangel fatal wäre (z.B. Moldzio,

2014). Die Metaanalyse von Konradt et al. (2020) zeigte, dass bei einem ungerechten gegenüber einem gerechten Verfahren die Gerechtigkeitswahrnehmung zwischen dem Pre- und Posttest signifikant abnahm. Das bedeutet, dass nicht nur die Gerechtigkeit einer Auswahlmethode von den Bewerbenden bewertet wird, sondern eine schlechte Bewertung einen negativen Einfluss auf den gesamten Prozess besitzt. Es wurden zum Beispiel virtuelle Interviews schlechter hinsichtlich der Gerechtigkeit von den Bewerbenden bewertet als ein Interview in Präsenz (Basch et al., 2020; Blacksmith et al, 2016). Forschungsarbeiten zeigten, dass die Persönlichkeit mit der Gerechtigkeitswahrnehmung korreliert (Moldzio, 2014; Van Vianen et al., 2004). Beispielsweise korrelieren die Big Five Dimensionen Offenheit für Erfahrung sowie Extraversion positiv mit der Gerechtigkeitswahrnehmung (z.B. Konradt et al., 2016; Moldzio, 2014; Van Vianen et al., 2004).

Diese Erkenntnisse deuten darauf hin, dass die Validität verschiedener virtueller bzw. onlinebasierter eignungsdiagnostischer Verfahren erforscht werden muss, um für die Praxis notwendige Implikationen abzuleiten. Dabei sollte die Gerechtigkeitswahrnehmung der Bewerbenden berücksichtigt werden.

Neuere Forschungsarbeiten fanden heraus bzw. replizierten, dass die Persönlichkeit nicht nur anhand der Big Five Dimensionen (Neurotizismus, Verträglichkeit, Extraversion, Gewissenhaftigkeit und Offenheit für Erfahrung; Costa & McCrae, 1992) erfasst werden kann, sondern eine Hierarchieebene unterhalb mit jeweils zwei Aspekten pro Dimension eine spezifischere Betrachtung ermöglicht (DeYoung et al., 2007; Judge et al., 2013). DeYoung et al. (2007) benannten die Aspekte der Extraversion Enthusiasmus und Durchsetzungsfähigkeit. Enthusiasmus beinhaltet die Geselligkeit und Herzlichkeit einer Person. Die Aspekte der Offenheit für Erfahrung sind Offenheit und Intellekt. Dabei beinhaltet die Offenheit eher die Verwendung von neuen Methoden und Arbeitsweisen und Intellekt die Offenheit gegenüber intellektuellen Weiterbildungen.

Ableitung der Hypothesen

Durch die Digitalisierung sind neue Möglichkeiten in der Eignungsdiagnostik entstanden (zum Beispiel die orts- und zeitunabhängige Testung) (Kanning, 2022). Aufgrund der weltweiten Covid-19- Pandemie mussten viele Auswahlverfahren in den virtuellen Raum verlegt werden. Diese sind aber aktuell noch nicht gut erforscht (Kanning, 2022). Verschiedene Forschungsarbeiten setzten sich zuvor mit dem Vergleich der Darbietungsarten (in Präsenz versus virtuell) auseinander, jedoch mit anderen Übungen der Personalauswahl (z.B. Interview oder Test- und Fragebogenverfahren), auf welche die Hypothesen dieser

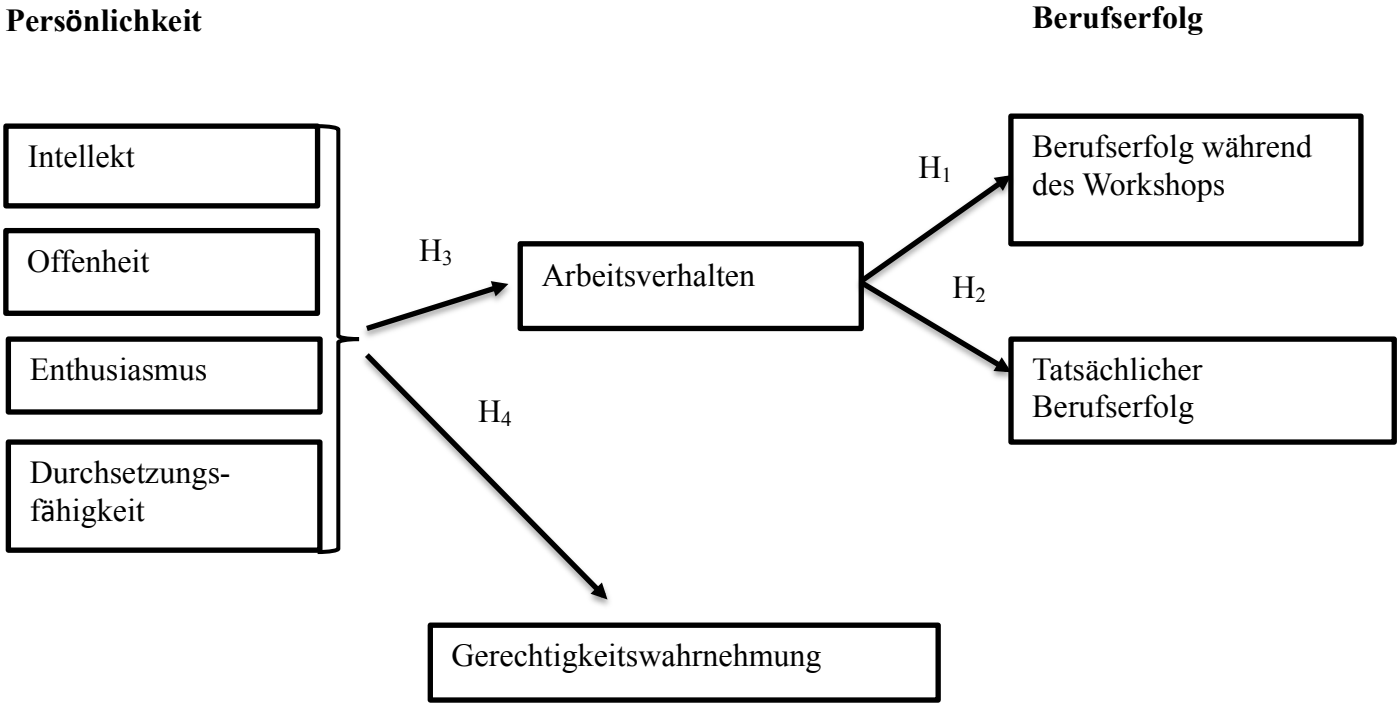
Studie gestützt wurden (z.B. Basch & Melchers, 2020; Blacksmith et al., 2016; Meade et al., 2007). Beispielsweise untersuchten Basch und Melchers (2020) die virtuellen Interviews und fanden heraus, dass die Bewerbenden im virtuellen Interview schlechter bewertet wurden als die in Präsenz.

Um die Hypothesen zu der prädiktiven Validität gegenüber dem Berufserfolg bilden zu können, wurde das Job Performance Modell von Tett und Burnett (2003) als Grundlage gewählt. Die prädiktive Validität ist Teil der Kriteriumsvalidität und betrachtet die Zusammenhänge mit zeitlich später erhobenen Kriterien (Bühner, 2021). Das Job Performance Model (Tett & Burnet, 2003) besagt, dass eine Persönlichkeitsausprägung nicht direkt zu Berufserfolg führt. Verschiedene Einflussfaktoren in der Aufgabe, der Organisation oder der sozialen Interaktion aktivieren bestimmte Persönlichkeitsausprägungen und führen zu einem bestimmten Arbeitsverhalten. Je nach Art der Aktivierung des Arbeitsverhalten einer Person kann es zu Berufserfolg führen oder nicht. So muss zum Beispiel eine sehr gewissenhafte Person abwägen, ob in der gegebenen Situation eine 100%ige Lösung zu Erfolg führt oder eine 80%ige Lösung ausreichend ist. Da diese Studie die prädiktive Validität in den beiden Darbietungsarten (in Präsenz versus virtuell) überprüfte, wurden als Einflussfaktoren die beiden verschiedenen Darbietungsarten unter der Annahme gewählt, dass die Darbietungsart das Arbeitsverhalten aktiviert und dadurch die Vorhersage des Arbeitsverhaltens zu Berufserfolg beeinflusst (s. Abbildung 9). Da Basch und Melchers (2020) von einer schlechteren Bewertung der Bewerbenden durch die Beobachtenden im virtuellen Interview berichteten, wurde in der Hypothese 1 dieser Studie davon ausgegangen, dass die Vorhersage von Berufserfolg während des Workshops durch das Arbeitsverhalten in der virtuellen Darbietungsart schlechter eingeschätzt/ bewertet wird als in Präsenz. Der Berufserfolg während des Workshops ist in dieser Studie als eine unabhängige Einschätzung während des Workshops anhand der Dialogübung zu verstehen.

Um nicht nur die Validität der Dialogübung während des Workshops zu erheben, sondern auch prädiktive Validität in einem übungsunabhängigen Kriterium zu erfassen, wurde die prädiktive Validität der Dialogübung auch für den tatsächlichen Berufserfolg betrachtet. Dies hatte zum Ziel, die Aussagekraft für die Praxis zu erhöhen. Da zu der Richtung dieser Hypothese noch keine vorangegangenen Forschungsarbeiten bestehen, wurde in Hypothese 2 auch keine bestimmte Richtung überprüft. Es wird davon ausgegangen, dass sich die Vorhersage von Berufserfolg durch das Arbeitsverhalten in den Darbietungsarten unterscheiden.

Abbildung 9

Darstellung der Hypothesen nach Tett und Burnett (2003) sowie Gilliland (1993) differenziert nach der virtuellen Darbietungsart und der Darbietungsart in Präsenz



Die Hypothese 3 betrachtete mit der Vorhersage des Arbeitsverhaltens durch die Persönlichkeit in den beiden Darbietungsarten einen weiteren Bestandteil des Job Performance Modells von Tett und Burnett (2003). Schermuly et al. (2019) behaupten, dass das Persönlichkeitsmerkmal Offenheit für Erfahrung immer wichtiger in der aktuellen Arbeitswelt bzw. in Bezug auf die Digitalisierung und die daraus entstehenden wechselnden Arbeitsanforderungen wird. Da die virtuelle Darbietungsart auch Teil der Digitalisierung ist, wurde dieses Persönlichkeitsmerkmal angelehnt an die vorherige Studie erfasst, um einen differenzierteren Blick auf die Big Five Dimensionen mit den berufsbezogenen Aspekten Offenheit und Intellekt nach DeYoung et al. (2007) zu erhalten. Da das Persönlichkeitsmerkmal Extraversion nachfolgend auch eine wichtige Rolle in der Gerechtigkeitswahrnehmung spielt (Gilliland, 1993; Moldzio, 2014), wurde in der Studie das Persönlichkeitsmerkmal Extraversion anhand der beiden Aspekte Enthusiasmus und Durchsetzungsfähigkeit betrachtet. Darüber hinaus wurde davon ausgegangen, dass in einer Dialogübung diese beiden Aspekte beobachtbar sind. Da es noch keine Forschungsarbeit zu den Unterschieden der Vorhersage des Arbeitsverhaltens durch die vier Aspekte in den Darbietungsarbeiten gab, wurde die Hypothese 3 in beide Richtungen überprüft. Es wurde dennoch davon ausgegangen, dass in der Vorhersage des Arbeitsverhaltens durch die vier

Aspekte der Offenheit für Erfahrung und der Extraversion Unterschiede zwischen den Darbietungsarten existieren.

Nachdem die Güte bzw. Validität der virtuellen Dialogübung betrachtet wurde, setzte diese Forschungsarbeit einen weiteren Schwerpunkt auf die Gerechtigkeitswahrnehmung, da bei einer geringen Gerechtigkeitswahrnehmung Bewerbende eine Stelle ablehnen oder ihre Bewerbung zurückziehen könnten, was in Zeiten des Fachkräftemangels fatal wäre (Kersting, 2018; Moldzio, 2014). Verschiedenste Arbeiten untersuchten die Gerechtigkeitswahrnehmung in unterschiedlichen Onlineverfahren (z.B. Basch & Melchers, 2020; Beermann et al., 2013; Meade et al., 2007), spezialisierten sich jedoch nicht auf Dialogübungen. Es wird davon ausgegangen, dass virtuelle Persönlichkeitsfragebögen und Intelligenztest von den Bewerbenden besser bewertet werden als die Verfahren in Präsenz (Beermann et al., 2013). Die Bewerbendenreaktionen sind jedoch in technikbasierten Interviews verhaltener und so werden die ursprünglichen Interviews in Präsenz gerechter wahrgenommen (Basch & Melchers, 2020; Blacksmith, 2016). Da die interpersonelle Interaktion in einer Dialogübung eher mit dem Interview anstatt mit einem Persönlichkeitsfragebogen vergleichbar ist, wurde auch in dieser Studie von einer geringeren Gerechtigkeitswahrnehmung der Bewerbenden in dem virtuellen Raum ausgegangen.

Der letzte Forschungsaspekt dieser Arbeit war der Zusammenhang zwischen der Persönlichkeit und der Gerechtigkeitswahrnehmung. Viele Forschungsarbeiten, darunter vor allem Gilliland (1993), untersuchten diese Frage und fanden heraus, dass Personen mit einer hohen Ausprägung in dem Persönlichkeitsmerkmal Offenheit für Erfahrung wahrscheinlicher ein Verfahren als gerecht wahrnehmen im Vergleich zu Personen mit einer niedrigen Ausprägung (Ryan & Ployheart, 2000; Van Vianen et al., 2004). Moldzio fand im Jahr 2014 die gleichen Zusammenhänge, jedoch mit der Dimension Extraversion. Diese Ergebnisse sollen in dieser Studie mithilfe der Aspekte nach DeYoung et al (2007) untersucht werden, um einen spezifischeren Blick auf die zuvor erkannten Zusammenhänge zu erhalten (s. Abbildung 9). Hypothese 4 besagte, dass sich die Vorhersage der Gerechtigkeitswahrnehmung durch die vier Aspekte in den Darbietungsarten unterscheidet. Auch hier wurde keine bestimmte Richtung der Hypothesentestung gewählt, da aus vorherigen Arbeiten nur ersichtlich ist, dass Verfahren in Präsenz besser bewertet werden (Basch & Melchers, 2020) und die Persönlichkeitsmerkmale Offenheit für Erfahrung sowie Extraversion mit der Gerechtigkeitswahrnehmung zusammenhängen (Gilliland, 1993; Moldzio, 2014). Die Kombination bzw. die Wechselwirkung zwischen der Darbietungsart

und der Vorhersage der Gerechtigkeitswahrnehmung durch die Aspekte wurde noch in keiner Forschungsarbeit berücksichtigt.

Die Hypothesen zu dem oben erklärten Modell bzw. zu den zwei Forschungsfragen der Studie können wie folgt zusammengefasst werden:

- H_1: Das Arbeitsverhalten sagt Berufserfolg während des Workshops in der virtuellen Darbietungsart geringer vorher als in Präsenz.
- H_2: Das Arbeitsverhalten sagt durch die Darbietungsart den tatsächlichen Berufserfolg unterschiedlich vorher.
- H_3: Die vier Aspekte der Offenheit für Erfahrung und Extraversion sagen durch die Darbietungsart das Arbeitsverhalten unterschiedlich vorher.
- H_4: Die vier Aspekte der Offenheit für Erfahrung und Extraversion sagen durch die Darbietungsart die Gerechtigkeitswahrnehmung unterschiedlich vorher.

Weitere Analysen

Neben den Fragen, ob die Persönlichkeit oder die Darbietungsart mit der Gerechtigkeitswahrnehmung zusammenhängen, wurde zusätzlich die Gerechtigkeitswahrnehmung angelehnt an die Forschung von Basch & Melchers (2020) und Blacksmith et al. (2016) betrachtet und untersucht, was den Teilnehmenden am wichtigsten in Bezug auf die Gerechtigkeitswahrnehmung ist. In den beiden Forschungsarbeiten wurde ersichtlich, dass virtuelle Interviews als weniger gerecht von den Bewerbenden wahrgenommen werden. Diese befürchten, dass sie nicht ihr gesamtes Potenzial zeigen können und die Privatsphäre ggf. nicht gegeben ist. Darüber hinaus war ersichtlich, dass die Bewerbenden in der virtuellen Darbietungsart von den Beobachtenden schlechter bewertet wurden. Die Betrachtung der untenstehenden Hypothesen hatte zum Ziel, die Ergebnisse auf die Dialogübung zu übertragen und zu prüfen, ob durch die Pandemie und die daraus resultierende Vielzahl von Videokonferenzen eine andere Tendenz der Gerechtigkeitswahrnehmung darstellt (Basch & Melchers, 2020; Erhebung vor der Pandemie). Ergänzend zu der Studie von Basch und Melchers (2020) wurde in dieser Arbeit nicht nur die Bewertung vom Berufserfolg, sondern auch die des Arbeitsverhaltens betrachtet. Die Hypothesen wurden angelehnt an die Forschung von Basch & Melchers (2020) und Blacksmith et al. (2016) wie folgt formuliert:

- H_5: Teilnehmende in der virtuellen Darbietungsart werden schlechter von den Beobachtenden bewertet als Teilnehmende der Präsenzstichprobe.

- H_6: Die Gerechtigkeitswahrnehmung ist in der virtuellen Stichprobe oder Gruppe schlechter als in der Stichprobe in Präsenz.

Forschungsarbeiten verwendeten zumeist die Regeln nach Gilliland (1993) zur Gerechtigkeitswahrnehmung, um zu betrachten, ob ein Verfahren von den Bewerbenden als gerecht wahrgenommen wird (z.B. Böge, 2016). Dabei hat noch keine Arbeit hinterfragt, ob in der Wichtigkeit dieser elf Regeln ein Unterschied besteht. In dieser Studie wurden die Wichtigkeit der Gilliland-Regeln zur Gerechtigkeit (1993) betrachtet, um aktuelle Handlungsempfehlungen für Unternehmen ableiten zu können. Da dies nur als Handlungsempfehlung bewertet wurde, wurde von Hypothesenbildungen abgesehen.

Methoden

Stichprobe

Die Stichprobe setzte sich aus unterschiedlichen Personengruppen zusammen. Die Rekrutierung wurde durch Kontakte von Kund:innen einer Unternehmensberatung, sowie in Zusammenarbeit mit einer Volkshochschule durchgeführt. Bei den Kund:innen wurde ein kostenloser Workshop zur Stärken- und Schwächen-Analyse für alle Arbeitnehmenden (zum Beispiel für Auszubildende sowie duale Studierende im letzten Lehrjahr als auch langjährige Mitarbeitende) angeboten. Die Volkshochschule bot diesen Workshop für diejenigen Interessenten an, welche Berufserfahrung aufwiesen und eine Einschätzung der eigenen Kompetenzen durch Dritte (zum Beispiel Kolleg:innen oder Führungskräfte) erbringen konnten.

Die Erhebung der Daten startete im April 2021 und es wurde eine Gesamtstichprobe von $N = 100$ Teilnehmenden erfasst. Die Hälfte der Teilnehmenden absolvierte den Workshop in Präsenz in den Büroräumen der Unternehmensberatung oder in den Räumlichkeiten der Kund:innen, die andere Hälfte nahm per Videokonferenz über ein handelsübliches System teil. Da die Erhebung im April 2021 startete und bei den Kund:innen durch Corona noch Zugangsbeschränkungen bestanden bzw. im Herbst 2021 die Homeoffice Pflicht eingeführt wurde (Bundesregierung, 2022), konnten zu Beginn nur die virtuellen Workshops durchgeführt werden. Im Frühjahr 2022 starteten die Workshops in Präsenz. Dadurch konnten die Personen nicht zufällig den Gruppen zugeordnet werden, sodass es Unterschiede in der Demografie der Teilnehmenden gab. Diese Unterschiede wurden in der Datenanalyse berücksichtigt.

In der Geschlechterverteilung unterschieden sich die zwei Stichproben kaum. So nahmen in der Gruppe in Präsenz 27 Frauen und in der virtuellen Darbietungsart 24 Frauen

teil. Dies bedeutete einen Anteil von 54 % bzw. 48% der Teilnehmenden in der jeweiligen Stichprobengruppe. Das Alter der Teilnehmenden sowie die daraus resultierende Berufserfahrung in Jahren unterschied sich dem gegenüber in den beiden Stichproben deutlich. So lag der Altersmittelwert in der Gruppe in Präsenz bei $M = 35{,}52$ ($SD = 11{,}00$) und die Berufserfahrung bei $M = 14{,}52$ ($SD = 10{,}74$). Demgegenüber lag der Altersmittelwert der Stichprobe im virtuellen Raum mit $M = 27{,}32$ ($SD = 7{,}79$) unter dem der Gruppe in Präsenz. Auch die Berufserfahrung war in der virtuellen Stichprobe geringer $M = 5{,}58$ ($SD = 6{,}48$). Im höchsten Bildungsgrad existierten leichte Unterschiede. So besaßen in der virtuellen Stichprobe mehr Personen das Abitur als höchsten Bildungsgrad und in der Gruppe in Präsenz mehr einen Bachelor bzw. Master (s. Tabelle 25). Dies könnte mit dem Alter zusammenhängen.

Tabelle 25
Übersicht zur Verteilung der höchsten Bildungsgrade in den beiden Darbietungsarten (in Präsenz vs. virtuell)

Höchster Bildungsgrad	Anzahl	
	In Präsenz	Virtuell
Hauptschulabschluss	0	0
Realschulabschluss / mittlere Reife	5	5
Fachhochschulreife/ Abitur	17	29
Bachelor	11	9
Meister (Handwerk)	1	1
Master bzw. Magister	6	4
Diplom	8	2
Promotion	2	0

Operationalisierung der Variablen

Operationalisierung der unabhängigen Variablen. In dieser Studie wurden zwei unabhängige Variablen betrachtet, von denen angenommen wurde, dass sie die abhängigen Variablen beeinflussen. Zum einen wurde der Workshop auf zwei verschiedene Arten dargeboten, welche verglichen wurden, und zum anderen wurde die Persönlichkeit, welche durch vier verschiedene Fragebögen erhoben wurde, auch als unabhängige Variable kategorisiert.

virtuell versus nicht virtuell. Um zu betrachten, ob die virtuelle Dialogübung das gleiche misst wie die Dialogübung in Präsenz, wenn alle beteiligten Personen an einem Ort sind, wurde der Workshop bei der Hälfte der Teilnehmenden virtuell per Videokonferenz über ein handelsübliches System und bei der anderen Hälfte im persönlichen Kontakt in Präsenz durchgeführt. In dieser Arbeit war die Versuchspersonengruppe in Präsenz die Kontrollgruppe. Jeweils nahmen zwei Beobachtende und ein:e Proband:in an dem Workshop

teil. Die Bestandteile des Workshops waren unabhängig von der Darbietungsform gleich und werden nachfolgend beschrieben.

Persönlichkeit. Um Persönlichkeitsunterschiede in Zusammenhang mit der Gerechtigkeitswahrnehmung sowie der prädiktiven Validität zu messen, wurde der Persönlichkeitsfragebogen AEOS (Wedemeyer et al., in Vorbereitung) eingesetzt (s. Abschnitt Persönlichkeit), welcher die vier Aspekte der Big Five Dimensionen Extraversion und Offenheit für Erfahrung nach DeYoung et al (2007) berufsbezogen erfasst. Darüber hinaus wurden drei weitere Fragebögen erhoben, welche die Big Five (NEO-FFI; Borkenau & Ostendorf, 2008), deren berufsbezogene Aspekte (AGBS; Moldzio et al. 2019; AVS; Moldzio, Böge & Wedemeyer, in Vorbereitung) und die Selbstwirksamkeit (Schyns & Collani, 2002) maßen. Diese drei Fragebögen wurden jedoch nur erhoben, damit die Teilnehmenden ihre Persönlichkeitsausprägungen kennenlernen und Handlungsempfehlungen für spätere Berufssituationen erhalten konnten. Aus diesem Grund wird im Nachgang nicht näher auf die drei Fragebögen eingegangen. Im Fokus der Forschungsarbeit stand der neu entwickelte Fragebogen AEOS der vorherigen Studie. Da diese Studie auf die vorherige aufbaut, wird im Folgenden dieser Fragebogen nur kurz erklärt.

Die arbeitsbezogenen Extraversions- und Offenheitsskalen AEOS (Wedemeyer et al., in Vorbereitung) wurden angelehnt an die Forschung zu den Aspekten der Big Five Dimensionen von DeYoung et al. (2007) entwickelt und stellen eine Erweiterung der oben genannten ABGS (Moldzio et al., 2019) und AVS (Moldzio, Böge & Wedemeyer, in Vorbereitung) dar. Die AEOS betrachtete anhand von 39 Items (in der Studie 1 nachfolgend gekürzt auf 30 Items) die berufsbezogenen Aspekte der Dimensionen Extraversion und Offenheit für Erfahrung. Extraversion wurde durch die beiden Aspekte Enthusiasmus und Durchsetzungsfähigkeit repräsentiert. Enthusiasmus beschrieb, inwiefern die Teilnehmenden gerne mit anderen zusammen oder lieber alleine arbeiten. Die Fähigkeit, andere durch gute Argumente zu überzeugen, wurde durch den Aspekt Durchsetzungsfähigkeit erfasst. Offenheit für Erfahrung wurde anhand der Aspekte Offenheit und Intellekt differenzierter aufgezeigt. So betrachtete die Offenheit eher das Ausprobieren von neuen Methoden oder Arbeitsweisen sowie die Zusammenarbeit mit Kolleg:innen aus anderen Kulturen, wohingegen der Intellekt die Offenheit gegenüber Weiterbildungen und neuen Themenfeldern beschrieb. Angelehnt an die zuvor entwickelten Fragebögen (ABGS, AVS)

wurde dieser auch mit derselben fünffach abgestuften Likert-Skala erhoben („*starke Ablehnung“* bis „*starke Zustimmung“*).

Die vorherige Validierungsstudie zeigte, dass der Fragebogen die vier Aspekte nach DeYoung et al. (2007) abbildet und die einzelnen Aspekte unterschiedlich mit Kriterien zum Berufserfolg zusammenhängen. Darüber hinaus lag die Reliabilität (interne Konsistenz) zwischen $\alpha = .81$ (Durchsetzungsfähigkeit und Offenheit) und $\alpha = .86$ (Intellekt), was nach George und Mallery (2002) als gut zu bewerten war.

Operationalisierung der abhängigen Variablen. Die abhängigen Variablen sind diejenigen Variablen, von denen angenommen wurde, dass sie von den unabhängigen Variablen beeinflusst werden. In dieser Studie wurde davon ausgegangen, dass Einschätzungen zum Berufserfolg und Gerechtigkeitswahrnehmungen abhängig von der Darbietungsart und der Persönlichkeit sind.

Berufserfolg. Ein Teil dieser Arbeit besteht aus der Erfassung der prädiktiven Validität. In diesem Fall wurde angelehnt an das Job Performance Modell nach Tett und Burnett (2003) der Berufserfolg mithilfe des Arbeitsverhaltens sowie der Berufserfolg während des Workshops und im Beruf erhoben. In den folgenden Abschnitten werden zunächst das Anforderungsprofil und im Anschluss die darauf aufbauenden Einschätzungen beschrieben, da das Anforderungsprofil die Grundlage der Erfassung des Arbeitsverhaltens ist (Schuler, 2000).

Vor Beginn der Durchführung der Workshops wurde ein Anforderungsprofil erstellt, damit die Workshops nah an den in der Praxis durchgeführten Auswahlverfahren konzipiert sind, da das Anforderungsprofil so die gewünschten Kompetenzen für den späteren beruflichen Erfolg definiert (Schuler, 2000). Das Anforderungsprofil diente als Orientierung für die einzelnen Einschätzungen der Kompetenzen. Es bestand in dieser Studie aus drei Kompetenzen, die jeweils mit vier Verhaltensankern bzw. Beschreibungen erfasst wurden.

Um anhand dieser Studie einen Mehrwert für die eignungsdiagnostische Praxis zu schaffen, wurde versucht, ein Anforderungsprofil mit den meist verwendeten Kompetenzen der Praxis zu erstellen. Dies war wichtig, damit diese Studie in der Praxis repliziert werden kann. Auf die Frage, welche Kompetenzen bei Assessment Centern im deutschsprachigen Raum (Deutschland, Österreich und die Schweiz) am häufigsten abgefragt wurden, fanden Höft und Obermann im Jahr 2010 eine Antwort. In ihrer Studie wurden 233 Organisationen online zu ihrer Assessment Center Praxis befragt, von denen 171 Organisationen angaben, diese regelmäßig durchzuführen, was einen Anteil von 73% darstellte. Die Analyse zeigte,

dass Assessment Center bei internen Verfahren eher zur Potenzialerkennung und bei externen Verfahren zur Personalauswahl verwendet wurden. Darüber hinaus fanden diese Verfahren hauptsächlich bei Führungskräften statt, aber auch Trainees, Auszubildende und Fachkräfte durchliefen die Assessment Center. Die Analyse der Verwendung erfasster Kompetenzen zeigte, dass in 90,5% der Assessment Center Kommunikationsfähigkeit getestet wurde. Sie war somit die meist erhobene Kompetenz. Die am zweit häufigsten gemessene Kompetenz war die Durchsetzungsfähigkeit. Sie wurde in 87% der Fälle erfasst. Die Analysefähigkeit wurde in 80,5% der Verfahren betrachtet und belegte somit den dritten Platz. Die vierthäufigste Kompetenz war die Konfliktfähigkeit, welche in 77,5% der Verfahren bewertet wurde.

Basis dieser Studie sind drei Kompetenzen, die in einer Dialogübung in der Regel beobachtbar sind und eingeschätzt werden können. Angelehnt an die Studie von Höft und Obermann (2010) wurden Kommunikationsfähigkeit, Durchsetzungsfähigkeit sowie Konfliktfähigkeit in dieser Studie erfasst. Die Analysefähigkeit wurde in dieser Studie nicht erhoben, da sie in einer Dialogübung schwer zu erfassen ist.

Kommunikationsfähigkeit setzt sich aus zwei Begriffsbestandteilen zusammen. Kommunikation wird laut Bierhoff (2021) als ein Prozess verstanden, in dem ein Individuum bzw. eine Gruppe von Individuen Informationen über Ideen, Gefühle und Absichten einer anderen Person bzw. einer Gruppe von Personen übermittelt. Dabei sind motivationale, emotionale sowie soziale Aspekte bedeutsam. Es geht über die reine Übermittlung der Botschaft hinaus. So können diese Informationen verbal (mündliche oder schriftliche Kommunikation) sowie nonverbal (zum Beispiel Mimik, Gestik, Stimme, persönliche Erscheinung) übermittelt werden. Kommunikation dient dazu, Informationen zu vermitteln, Entscheidungen vorzubereiten, Motivation zu erzeugen oder ein gewünschtes Image durch Eindrucksmanagement herzustellen (Selbstdarstellung, Bierhoff, 2021). Darüber hinaus wurde in dieser Studie der Begriff der Fähigkeit nach Häcker (2016) als die Gesamtheit der zur Ausführung einer bestimmten Leistung erforderlichen Bedingungen beschrieben. Anhand dieser Definitionen wurden vier Verhaltensanker für das Merkmal Kommunikationsfähigkeit definiert:

- Die Kandidatin bzw. der Kandidat ist in der Lage, ihre/ seine Ideen, ihre/ seine Meinung und Entscheidungen gezielt an sein Gegenüber zu vermitteln.
- Die Kandidatin bzw. der Kandidat setzt Mimik, Gestik und die Stimme überzeugend ein.

- Die Kandidatin bzw. der Kandidat drückt sich klar und verständlich aus.
- Die Kandidatin bzw. der Kandidat hört aktiv zu und lässt den Gesprächspartner ausreden.

Als Grundlage für die Erstellung der Verhaltensanker der Durchsetzungsfähigkeit wurde die Definition von Isenschmid aus dem Jahr 2013 verwendet. Diese beschreibt Durchsetzungsfähigkeit als Befähigung, andere von der eigenen Meinung zu überzeugen. Personen mit einer hohen Ausprägung können ihre Meinungen, Ideen, Vorstellungen oder Ziele gegenüber anderen (zum Beispiel den Vorgesetzten, dem Kollegium oder anderen Mitarbeitenden) durchsetzen. Dies schaffen sie, indem sie gegen Widerstand fair argumentieren, den anderen dabei jedoch nicht überreden, sondern überzeugen. Das Abwägen von Argumenten, Nachgeben und Einsehen, dass andere „Recht" haben, spricht ebenso für eine hohe Durchsetzungsfähigkeit. Mithilfe dieser Definition wurden für Durchsetzungsfähigkeit vier Verhaltensanker abgeleitet:

- Die Kandidatin bzw. der Kandidat ist in der Lage, ihr/ sein Gegenüber mit Argumenten von ihrer /seiner Meinung zu überzeugen.
- Die Kandidatin bzw. der Kandidat ist auch in der Lage andere Meinungen anzunehmen und den gegenüber nicht zu überreden.
- Die Kandidatin bzw. der Kandidat teilt ihre/ seine Meinung auch trotz Widerstände mit.
- Die Kandidatin bzw. der Kandidat gewinnt andere Personen für ihre/ seine Ideen.

Auch Konfliktfähigkeit unterliegt keiner einheitlichen Definition (Häcker, 2016). In dieser Studie wurde sich für eine Begriffszusammensetzung aus „Interpersonaler Konflikt" und „Fähigkeit" entschieden. Letzteres wurde zuvor zur Kommunikationsfähigkeit definiert (Häcker, 2016). Interpersonale Konflikte definierte Lewin (1935; zitiert nach Häcker & Stapf, 1998) als einen Interessenskonflikt oder eine Folge von diskrepanten Handlungsabsichten zwischen Personen oder gesellschaftlichen Gruppen. Diese Konflikte oder Diskrepanzen können auf drei verschiedene Weisen entstehen. Zum einen durch Ziele, die sich gegenseitig ausschließen. Zum Beispiel möchte ein Unternehmen die eigenen Produkte teuer verkaufen, aber die potentiellen Käufer:innen möchten ein günstiges Produkt erwerben. Des Weiteren können Personen oder Gruppen das gleiche Ziel verfolgen, jedoch gibt es nicht genügend Ressourcen. Dies passiert, wenn zum Beispiel zwei Mitarbeitende das Ziel haben, die gleiche Führungsposition besetzen zu wollen. Der letzte Punkt, der zu Konflikten führen kann, sind Verhandlungen, die einen Kompromiss erschweren können. Dies wäre zum Beispiel der Fall,

wenn man sich bei Preisverhandlungen immer mehr annähert, jedoch eine Verhandlungspartnerin oder ein Verhandlungspartner wieder einen Rückschritt macht. Aus diesen Definitionen sind vier Verhaltensanker entstanden, die wie folgt lauten:

- Die Kandidatin bzw. der Kandidat ist in der Lage ihre/ seine eigenen Ziele für einen gemeinsamen Kompromiss hintenanzustellen.
- Die Kandidatin bzw. der Kandidat zeigt sich unvoreingenommen gegenüber anderen Meinungen.
- Die Kandidatin bzw. der Kandidat erkennt Konfliktpunkte und versucht diese zu lösen.
- Die Kandidatin bzw. der Kandidat scheut sich nicht Konflikte einzugehen.

Die Teilnehmenden sollten im Workshop nach der Dialogübung die eigenen Fähigkeiten einschätzen. Dies war für die Teilnehmenden wichtig, um die zuvor abgeprüften Kompetenzen vor dem Feedback zu verstehen. Alle Teilnehmenden nahmen an einer fünf-minütigen Umfrage teil, welche die drei Kompetenzen aus dem Anforderungsprofil und deren vier Verhaltensanker beinhaltete. Diese Verhaltensanker waren in der Ich-Form formuliert und mussten von den Teilnehmenden auf einer sechsfach abgestuften Likert-Skala bewertet werden. Diese Skala ging von „*Trifft überhaupt nicht zu*“ bis „*Trifft vollständig zu*“. Die Teilnehmenden sollten diese Skala als normalverteilt verstehen, sodass ein breiter mittlerer Bereich existiert. Falls die Teilnehmenden diesen Verhaltensanker als Stärke sahen, kreuzten sie weiter rechts an („*Trifft vollständig zu*“; bei einer vermeintlichen Schwäche weiter links, „*Trifft überhaupt nicht zu*“). Neben den jeweils vier Verhaltensankern sollten die Teilnehmenden eine Gesamteinschätzung der Kompetenz anhand der gleichen Skala vornehmen. Die Verhaltensanker dienten dazu, dass sich die Teilnehmenden intensiv mit den Kompetenzen auseinandersetzen mussten, sie wurde jedoch nicht in die Auswertung mit einbezogen. Die Skala, angelehnt an die Normalverteilung, wurde gewählt, da die Art der Skala in einer Unternehmensberatung Jahre lang erprobt und als gut befunden wurde. Durch den gleichen Aufbau der Erfassung der Kompetenzeinschätzung konnte so eine bessere Vergleichbarkeit zur Praxis hergestellt werden.

An jedem Workshop nahmen zwei Beobachtende teil, um den Berufserfolg der Teilnehmenden zu bewerten und am Ende des Workshops ein Feedback abzugeben. Es wurde direkt im Anschluss der Dialogübung das Arbeitsverhalten in dem Gespräch anhand der drei Anforderungskriterien Kommunikationsfähigkeit, Durchsetzungsfähigkeit und Konfliktfähigkeit bewertet. Die Beobachtenden schätzten nicht die Verhaltensanker ein,

sondern gaben nur eine Gesamtbewertung pro Kriterium mithilfe der sechsfach abgestuften Likert-Skala ab, welche ebenfalls bei der Selbsteinschätzung und der nachfolgend erörterten Fremdeinschätzung verwendet wurde („*Trifft überhaupt nicht zu*“ bis „*Trifft vollständig zu*“).

Zusätzlich zu dieser Einschätzung des Arbeitsverhaltens wurde auch das Potenzial als Variable des Berufserfolgs eingeschätzt. Dies wurde im Feedback nicht besprochen, sondern diente der Prüfung der Hypothesen. Bei der Potenzialaussage nutzten die Beobachtenden vier Kategorien. Eine Eins wurde vergeben, wenn die Teilnehmenden die Dialogübung gut durchgeführt hatten und es kaum Anregungen zur Verbesserung gab. Das Item wurde wie folgt formuliert: „*Die Teilnehmerin/ der Teilnehmer hat das Potenzial für weitere berufliche Entwicklungsschritte und benötigt keine begleitenden Fördermaßnahmen*.“. Die Aussage „*Die Teilnehmerin/ der Teilnehmer hat das Potenzial für weitere berufliche Entwicklungsschritte, benötigt jedoch noch Fördermaßnahmen.*“ wurde ausgewählt, wenn die Teilnehmenden die Situation gut durchgeführt hatten, aber aufgrund von beispielsweise fehlender Erfahrung oder Qualifizierung Entwicklungsbedarf gesehen wurde. Diese Einschätzung wurde mit einer Zwei kodiert. Eine Drei wurde vergeben, wenn Potenzial noch nicht erkennbar war, jedoch gute Ansätze, die weiter ausgebaut werden müssen, sichtbar waren. Bei diesen Personen wurde davon ausgegangen, dass diese nach einer persönlichen Weiterentwicklung das Potenzial später aufweisen würden. Die Einschätzung lautete: „*Die Teilnehmerin/ der Teilnehmer weist zum aktuellen Zeitpunkt noch nicht das Potenzial für weitere berufliche Entwicklungsschritte auf, das Potenzial könnte sich durch Fördermaßnahmen entwickeln.*“. Die letzte Aussage war wie folgt formuliert und wurde mit einer Vier kodiert: „*Die Teilnehmerin/ der Teilnehmer weist nicht das Potenzial für weitere berufliche Entwicklungsschritte auf.*“. Dies war zum Beispiel der Fall, wenn Teilnehmende ein konfliktvermeidendes Verhalten gezeigt sowie die Konfliktpunkte in der Dialogübung nicht angesprochen hatten.

Um die Variable des tatsächlichen Berufserfolgs zu erfassen, wurde eine Fremdeinschätzung vor Beginn des Workshops eingeholt. Sie bestand aus einer fünfminütigen Umfrage, in der das Arbeitsverhalten bewertet sowie eine Potenzialaussage für den Berufserfolg abgegeben werden sollte. Die Bewertung des Arbeitsverhaltens anhand der drei Kompetenzen diente nur dazu, dass die fremd einschätzenden Personen die Kompetenzen verstanden und so den Berufserfolg bewerten konnten. Die Teilnehmenden erhielten die Umfrage per Link und sollten diesen an Kolleg:innen oder Führungskräfte weiter leiten. Die Teilnehmenden sollten Vertrauen zu dieser ausgewählten Person aufweisen und diese Person sollte die Teilnehmenden gut im Arbeitsverhalten kennen. Das Einholen dieser Einschätzung

war wichtig, damit eine weitere objektive Variable für die Validierung der Dialogübung herangezogen werden konnte.

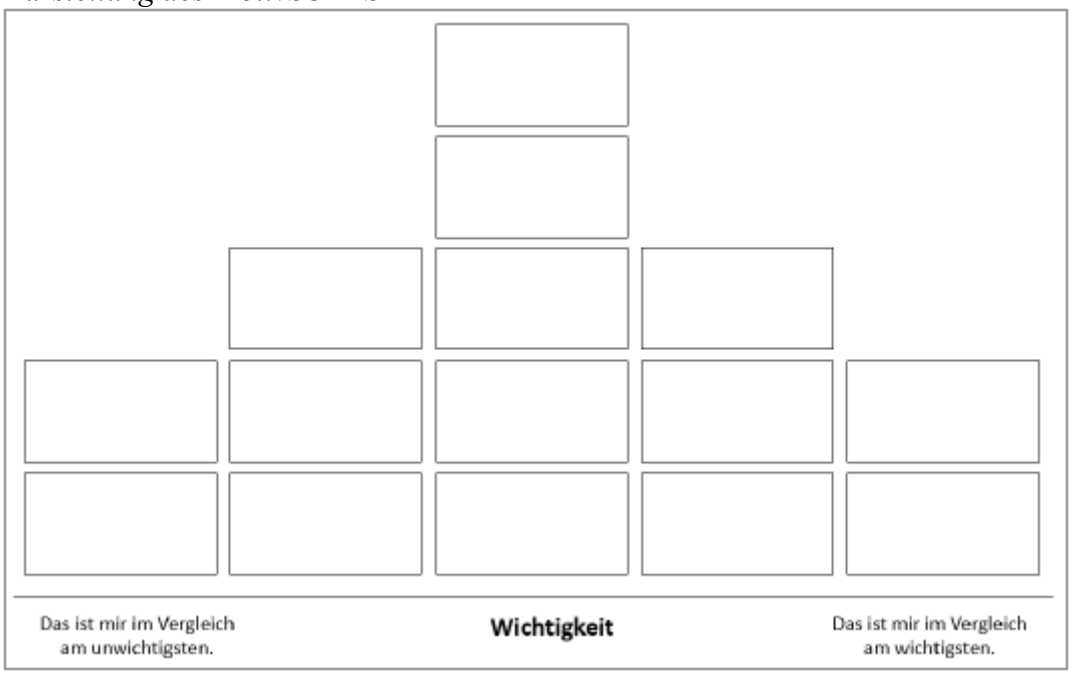
Gerechtigkeitswahrnehmung. Die zweite abhängige Variable ist die Gerechtigkeitswahrnehmung. Um die Gerechtigkeitswahrnehmung zu erheben, wurden die Items eines Fragebogens mit der Methodik eines anderen kombiniert, welche nachfolgend erklärt werden.

Der erste Fragebogen ist die Selection Procedure Justice Scale (Bauer et al, 2001), kurz SPJS genannt. Dieser Fragebogen wurde basierend auf den Gerechtigkeitsregeln von Gilliland (Gilliland, 1993) in der englischen Sprache entwickelt. Er besteht aus elf Subskalen, welche durch 39 Items repräsentiert werden. Nachfolgend übersetzte Böge (2016) diese Skala angelehnt an die Methodik von Brislin (1980) sowie Van de Vijver und Hambleton (1996) in die deutsche Sprache. Bei dieser Methodik wurden die Skalen und Items von Expert:innen ausgehend von der Originalsprache (Englisch) ins Deutsche übersetzt. Danach wurde die Übersetzung von englischen Muttersprachler:innen in die englische Sprache zurückübersetzt, um zu prüfen, ob der Inhalt gleichgeblieben war. Die Items mussten auf einer fünffach abgestuften Likert-Skala von „*starke Ablehnung*“ bis „*starke Zustimmung*“ beantwortet werden.

Die Methodik des Fragebogens stammt aus dem MotivSORT (Ellwart et al., 2018). Dieses Tool wurde entwickelt, um wichtige Arbeitsmotive abzufragen. Das MotivSORT besteht aus zwei Schritten. Zuerst müssen Teilnehmende die 15 Motive in Form von Legekarten nach ihrer Wichtigkeit sortieren. Dabei besteht die Sortierung aus fünf Spalten von „*Das ist mir im Vergleich am unwichtigsten*“ bis „*Das ist mir im Vergleich am wichtigsten*“. Diese Sortierung sieht jedoch vor, dass in der linken und rechten Spalte nur zwei Motive stehen dürfen, in den Spalten daneben nur drei und in der Mitte fünf (s. Abbildung 10).

Abbildung 10

Darstellung des MotivSORTS



Anmerkung. Die Darstellung stammt aus dem Artikel Ellwart et al. (2018)

Die Sortierung der Motive basiert auf der Einschätzung der Wichtigkeit für den einzelnen Mitarbeitenden und nicht auf der aktuellen Arbeitssituation. In einem zweiten Schritt müssen Teilnehmende diese Motive nach der aktuellen Gegebenheit in ihrem individuellen Beschäftigungskontext anhand einer vierstufigen Skala bewerten. Wenn das Motiv in der aktuellen Arbeitsstelle „*nicht gegeben*“ ist, dann wird das Motiv rot, bei „*eher nicht erfüllt*“ gelb, „*eher erfüllt*“ dunkelgrün und „*erfüllt“* hellgrün eingefärbt. Die Betrachtung der Validität zeigte, dass ein positiver Zusammenhang zwischen MotivSORT und Arbeitszufriedenheit (r = .67) sowie der affektiven Bindung an den Betrieb (r = .56) besteht (Ellwart et al., 2018). Darüber hinaus konnten Zusammenhänge zur Tatkraft, Kündigungsabsicht und emotionaler Erschöpfung gefunden werden (Ellwart et al., 2018). Darüber hinaus sagten 98% der Befragten, dass das Tool praktikabel, nützlich sowie verständlich sei (Ellwart et al., 2018). Dies zeigt, dass diese Art des Tools Zusammenhänge zu Motiven aufzeigen kann sowie verständlich und valide ist. Somit ist dieses Tool nicht nur in der Lage die Motive zu bewerten, sondern auch auf einer anderen Ebene nach Wichtigkeit zu bewerten. In dieser Forschungsarbeit wurde die Struktur des MotivSORTs leicht abgeändert verwendet, um die Gerechtigkeitswahrnehmung in eine Rangfolge zu bringen und somit einen ersten Ansatzpunkt zu erhalten, welche der elf Gilliland-Regeln (1993) am wichtigsten sind.

In dieser Forschung wurde der eigens entwickelte GerechtigkeitsSORT verwendet, der die Skala der SPJS (Bauer et al., 2001; Böge, 2016) mit der Methode des MotivSORTS (Ellwart et al., 2018) verknüpfte. Dazu wurden im ersten Schritt die Legekarten und die Struktur definiert. Angelehnt an die Gerechtigkeitsregeln von Gilliland (1993) und die SPJS sind elf Legekarten entstanden. Sie beinhalteten jeweils ein Beispielitem, um den Teilnehmenden die Regeln verständlich zu machen (s. Abbildung 11). Es wurde sich für nur jeweils ein Beispielitem entschieden, da sonst nicht genügend Legekarten auf dem Bildschirm sichtbar gewesen wären, was die Sortierung in der Onlineumfrage schwierig gemacht hätte.

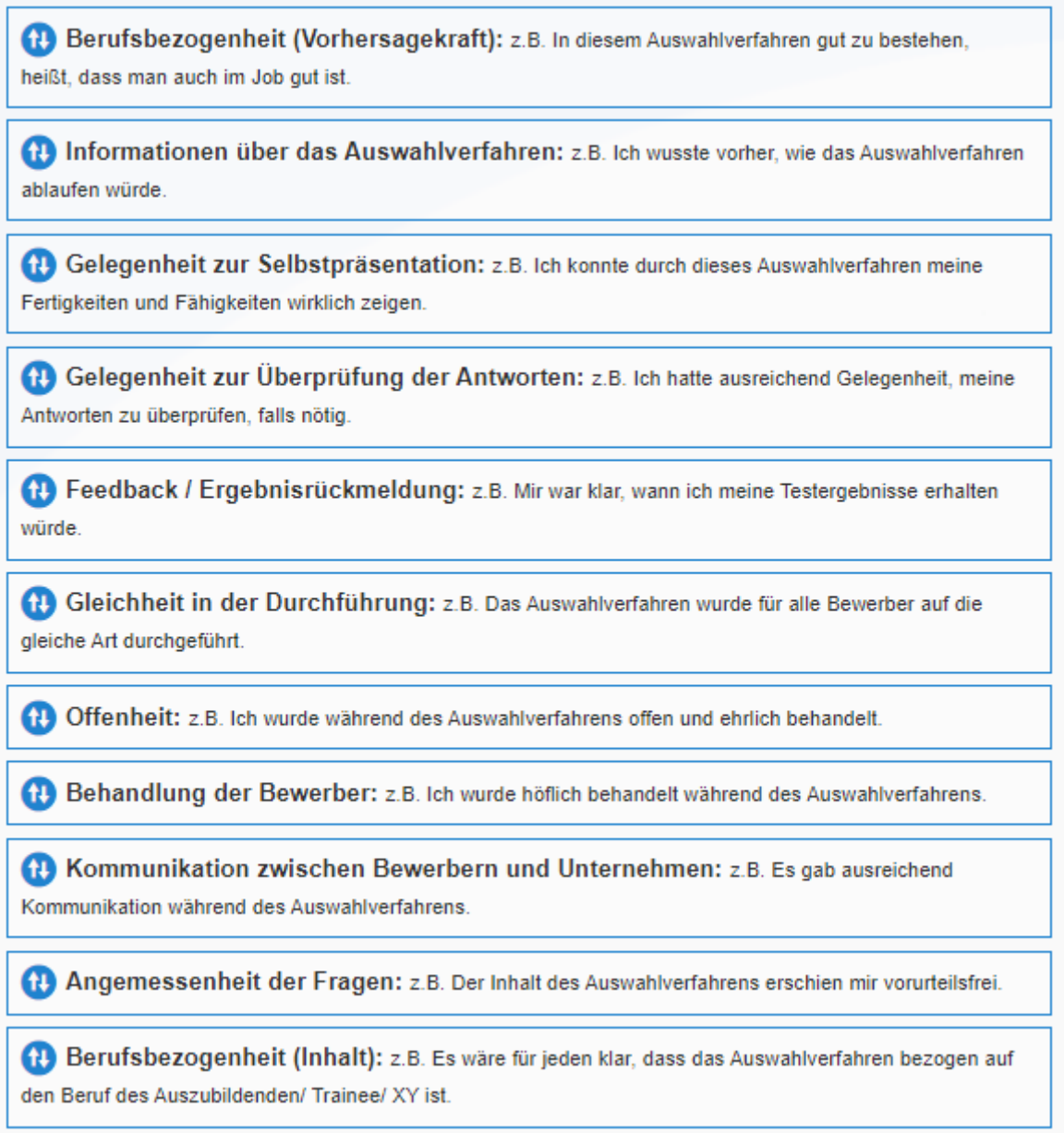
Die Teilnehmenden sortierten die Legekarten in der Reihenfolge von Eins bis Elf, in der virtuellen Darbietungsart anhand einer Onlineumfrage und in Präsenz im paper-pencil Format. Mit der Eins wurden Karten bewertet, die den Teilnehmenden für ein gerechtes Verfahren am wichtigsten waren, analog mit Elf die unwichtigsten.

Im nächsten Schritt mussten die Teilnehmenden bewerten, ob diese elf Gerechtigkeitsregeln in diesem Workshop gegeben waren. Es wurden alle elf Gerechtigkeitsregeln mit allen Items als Erklärung präsentiert, damit die Anlehnung an die SPJS (Bauer et al., 2001; Böge, 2016) gegeben war. Die Teilnehmenden mussten alle Gerechtigkeitsregeln auf einer vierfach abgestuften Skala bewerten. Diese Skala beinhaltete „*nicht gegeben*“, „*eher nicht gegeben*“, „*eher gegeben*“ und „*gegeben*“.

Durch diese Methode sind die im zweiten Schritt erhaltenen Informationen sehr ähnlich zu der Bearbeitung des Fragebogens SPJS (Bauer et al. 2001). Es wurde lediglich auf den Skalenmittelwert verzichtet, da nur die Gesamtskala bewertet wurde. Darüber hinaus wurde die Likert-Skala von einer fünfstufigen auf eine vierstufige Skala analog zum MotivSORT geändert. So konnte die Vergleichbarkeit beibehalten werden. Die Aussagekraft der Wichtigkeit der Gerechtigkeitsregeln konnte durch den ersten Teil (Sortierung) gesteigert werden. Dadurch wurde erfasst, was Bewerbenden in Auswahlverfahren im Allgemeinen wichtig ist. Diese Erkenntnis gibt die Möglichkeit durch Handlungsanweisungen in Auswahlverfahren gegensteuern zu können.

Abbildung 11

Aufbau der Legekarten des GerechtigkeitSORT in der Onlineumfrage



Durchführung

Die Versuchsdurchführung kann in vor, während und nach dem Workshop unterteilt werden. Vor dem Workshop wurde der Kontakt zu den Teilnehmenden von der Testleiterin aufgenommen, die Testdiagnostik wurde an die Teilnehmenden verschickt und die Fremdeinschätzung musste von Kolleg:innen ausgefüllt werden. Während des Workshops fand die Dialogübung, die Selbsteinschätzung sowie die Einschätzung der Beobachtenden statt und es wurde ein Feedback gegeben. Nach dem Workshop wurde die Gerechtigkeitsumfrage von den Teilnehmenden bearbeitet und die Zusammenfassung anhand eines Profilbogens von der Testleiterin an die Teilnehmenden versendet.

Vor dem Workshop. Die Teilnehmenden wurden bei der Volkshochschule und in Kund:innenprojekten einer Unternehmensberatung akquiriert. Die jeweiligen Ansprechpartner:innen schickten gesammelt die Mailadressen der Teilnehmenden an die Testleiterin oder die Teilnehmenden meldeten sich per Mail bei der Testleiterin. Die Teilnehmenden wurden per Mail von der Testleiterin nach einem Termin für ein ca. zehn minütiges Telefonat gefragt. In diesem Telefonat stellte sich die Versuchsleiterin kurz vor und erklärte den kompletten Ablauf. Darüber hinaus wurde ein Termin für einen individuellen Workshop vereinbart. Dieses Telefonat wurde geführt, damit mehr Verbindlichkeit entstand und die Teilnehmenden zu einer höheren Wahrscheinlichkeit alle Teile der Studie bearbeiten würden. Nach diesem Gespräch wurde eine Mail von der Testleiterin an die Teilnehmenden versendet, in der der Ablauf nochmals erklärt wurde. Diese E-Mail beinhaltete zusätzlich zwei Links zu dem Persönlichkeitsfragebogen sowie zur Fremdeinschätzung, welche online bearbeitet wurden. Der Bearbeitungszeitraum umfasste meist eine Woche. In dieser Umfrage generierten die Teilnehmenden einen Versuchspersonencode, um die Anonymität zu bewahren und dennoch die verschiedenen Umfragen zuordnen zu können. Da alle Umfragen mit diesem Code generiert wurden, musste zu der Person, welche die Fremdeinschätzung abgab, kein Kontakt aufgenommen werden. Darüber hinaus wurde auch eine Microsoft-Teams Einladung von der Testleiterin an die Teilnehmenden versendet, falls der Workshop virtuell stattfand.

In dieser Studie wurden die Fragebögen vor Beginn des Workshops von allen Teilnehmenden bearbeitet. Diese vier Fragebögen wurden den Teilnehmenden mit einer Datenschutzerklärung präsentiert. Vor Beginn jedes einzelnen Workshops wurden die Persönlichkeitsfragebögen ausgewertet und mit einer internen Norm der Unternehmensberatung für Expert:innen ohne Führungsverantwortung verglichen. Diese Ausprägungen wurden mit den Teilnehmenden in einem Feedbackgespräch besprochen. Die Norm ist für Bewerbende, die eine Berufsausbildung oder ein Studium absolviert haben, jedoch keine Führungsverantwortungen übernehmen, berechnet worden. Darüber hinaus haben die Teilnehmenden ihre Fragebogenergebnisse anhand eines Profilbogens am Ende der Studie ausgehändigt bekommen, welcher die Ergebnisse bzw. das Persönlichkeitsprofil detailliert darstellte (s. Anhang Profilbogen).

Während des Workshops. An dem Workshop nahmen zwei Beobachterinnen und die Teilnehmenden teil. Zu Beginn stellte sich die neben der Studienleiterin noch unbekannte zweite Beobachterin vor. Anschließend wurde das Ziel sowie der Mehrwert dieser Studie

erneut erklärt. Darüber hinaus wurde der Ablauf noch einmal erläutert und nach dem Versuchspersonencode gefragt. Dieser Code musste von der Testleiterin erfragt werden, damit Sie die Auswertung der Persönlichkeitsfragebögen zuordnen konnte. Danach wurde den Teilnehmenden die Instruktion zu der Dialogübung vorgelegt oder im virtuellen Workshop per Mail versendet. Diese Instruktion beinhaltete das Szenario, dass die teilnehmende Person sich vorstellen soll, vor drei Monaten die Leitung eines fünfköpfigen Projektteams übernommen zu haben. In diesem Team ist Frau Müller, welche früher leistungsstark auftrat. In diesem Team kam sie oftmals zu spät zu Besprechungen und die Kooperation mit anderen Teammitgliedern lief nicht optimal. Nun soll ein klärendes Gespräch zwischen der Leitung des Projektteams und Frau Müller stattfinden. Diese Instruktion sollten die Teilnehmenden aufmerksam lesen. Zudem wurde Zeit für Rückfragen gegeben. Nachdem dies geschehen war, ging die Versuchsperson in einen anderen Raum und bereitete sich 15 Minuten auf das Gespräch vor. Im virtuellen Kontext verließ die Person die Konferenz, um sich nach 15 Minuten erneut einzuloggen. Parallel dazu reflektierten die Beobachterinnen die Inhalte der Testdiagnostik.

Im nächsten Schritt wurde die Dialogübung ca. zehn Minuten lang durchgeführt. Als Gesprächspartnerin (Frau Müller) diente eine Beobachterin, die mithilfe einer Handlungsanweisung die Gespräche so standardisiert wie möglich gestaltete. Die zweite Beobachterin beobachtete das Verhalten der Versuchsperson und dokumentierte dies. Nach dieser Übung verließen die Teilnehmenden erneut den Raum, um den Selbsteinschätzungsbogen zu bearbeiten, dessen Aufbau die Testleiterin zuvor noch erklärte. Parallel besprachen die Beobachterinnen den Verlauf des Gesprächs und bewerteten das Arbeitsverhalten und den Berufserfolg anhand der Kompetenzen sowie der Potenzialaussage. Darüber hinaus besprachen sie Rückmeldungen und Hinweise zum Verhalten, welche im Feedbackgespräch angemerkt werden sollten.

Nachdem alle Bewertungen vorgenommen worden waren, ließen die Beobachterinnen die Versuchsperson wieder in den Raum. Danach fand das Feedback statt, welches in zwei Teile geteilt wurde. Als erstes wurde die Dialogübung besprochen. Zunächst wurden die Teilnehmenden gebeten, die eigene Leistung selber zu reflektieren. Darauf wurde von den Beobachterinnen Bezug genommen und anschließend das Gespräch mithilfe der Verhaltensanker betrachtet. Im Anschluss gab die Testleiterin Handlungsempfehlungen für weitere Konfliktgespräche. Im zweiten Teil des Feedbacks wurden die Persönlichkeitsfragebögen reflektiert. So wurden zum Beispiel bei einer gering ausgeprägten Belastbarkeit Stressbewältigungsstrategien diskutiert oder bei einer gering ausgeprägten

Gewissenhaftigkeit Hinweise gegeben, wie der Arbeitsalltag besser strukturiert werden kann. Wenn daraufhin keine Fragen bestanden, war der Workshop beendet.

Nach dem Workshop. Im Anschluss an den Workshop wurde den Kandidat:innen die Gerechtigkeitsumfrage per Mail geschickt. Sie wurden gebeten, diese schnellstmöglich zu bearbeiten, solange die Erinnerungen an den Workshop noch präsent waren. In dem Präsenzverfahren wurde den Teilnehmenden die Gerechtigkeitsumfrage in einem paper-pencil Format präsentiert, um die Quote der nicht bearbeiteten Umfragen zu minimieren. Nachdem die Teilnehmenden die Umfrage bearbeitet hatten, wurde ihnen noch eine letzte Mail zugesendet. Diese Mail beinhaltete den im Workshop besprochenen Profilbogen. Nachfolgend hatten die Teilnehmenden die Möglichkeit, Kontakt aufzunehmen, falls im Nachgang noch Fragen entstanden sind.

Datenanalyse

Bevor die Daten erhoben werden konnten und nachfolgend die Datenanalyse stattfand, wurden sich zu Beginn Gedanken über einen geeigneten Stichprobenumfang gemacht, um eine Aussage über die Power der Berechnungen treffen zu können (Janczyk & Pfister, 2020). Unter der Power ist die Wahrscheinlichkeit, dass ein Effekt tatsächlich besteht, zu verstehen (Janczyk & Pfister, 2020). Dies bedeutet, dass bei einer hohen Power davon ausgegangen werden kann, dass der Effekt wirklich besteht und somit die Ergebnisse interpretiert werden können. Es wirken drei Einflussfaktoren auf die Power (Janczyk & Pfister, 2020). Der erste Einflussfaktor ist die Wahl des Signifikanzniveaus. Wird das Signifikanzniveau groß gewählt, dann ist die Power auch groß. Ein weiterer Einflussfaktor ist die Größe des Effekts, welche bei einer Steigerung auch zu einer größeren Power führt (Janczyk & Pfister, 2020). Als letztes hat die Verringerung des Standardfehlers einen positiven Einfluss auf die Power. Den Standardfehler kann durch eine große Stichprobe reduziert werden, daher gilt, je größer der Stichprobenumfang, desto größer die Power (Janczyk & Pfister, 2020). Würde jedoch die Stichprobe unendlich groß sein, dann wird fast jeder kleinste Effekt signifikant, was für die Aussagekraft von Ergebnissen nicht förderlich ist (Janczyk & Pfister, 2020). Daher ist es entscheidend, den Stichprobenumfang nicht zu groß und nicht zu klein zu wählen, da eine kleine Stichprobe zu einer kleinen Power führt. Im Bereich der Strukturgleichungsmodelle existieren viele unterschiedliche Regelungen zu dem Stichprobenumfang. Die erste Regel besagt, dass der Stichprobenumfang fünfmal größer sein sollte als die Anzahl zu schätzender Parameter (z.B. Bagozzi & Yi, 1988). In dieser Arbeit werden acht Parameter geschätzt, von daher sollte das $N > 40$ pro Darbietungsart betragen.

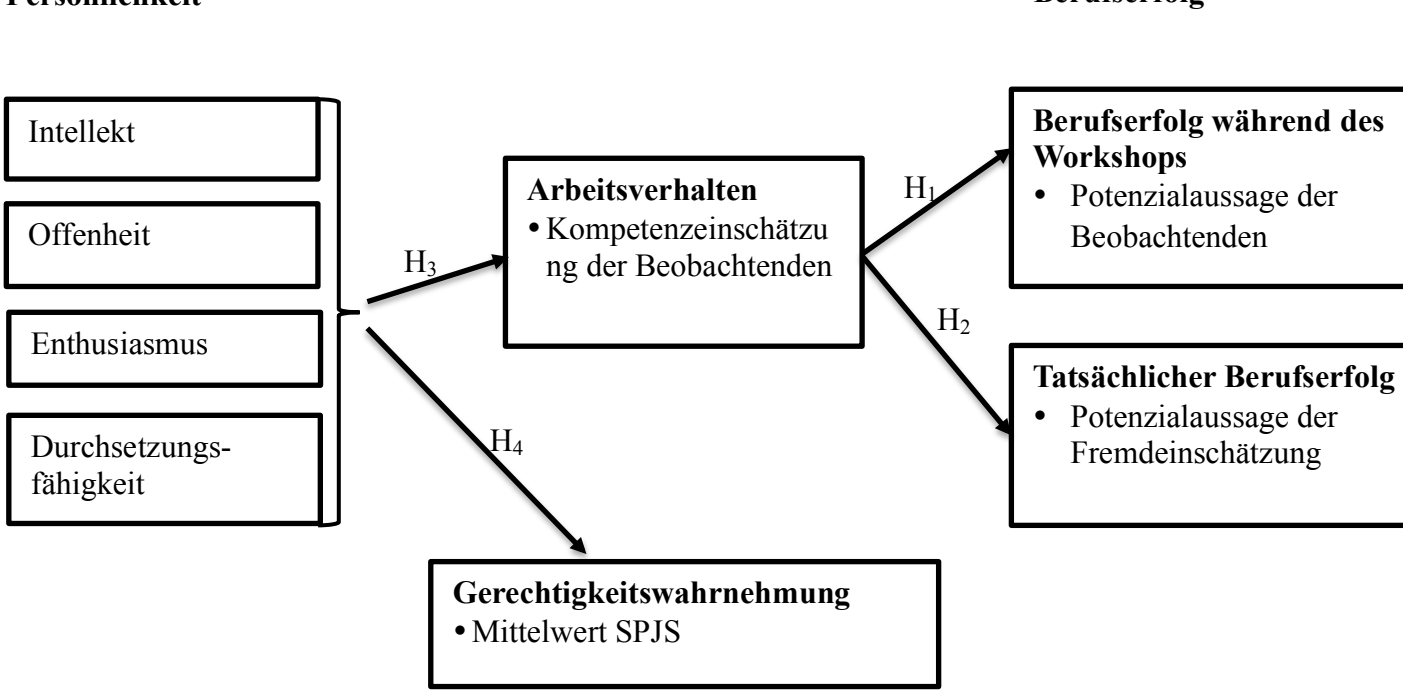
Eine andere Regel besagt, dass der Stichprobenumfang so gewählt werden sollte, dass die Division der Anzahl der zu schätzenden Parameter von der Stichprobengröße > 50 ist (Bagozzi, 1981). Dies wäre in dieser Arbeit eine Stichprobengröße von $N > 58$ pro Stichprobengröße. Die Arbeit besaß ein aufwändiges Setting und hatte zum Ziel, mithilfe eines experimentellen Aufbaus erste Erkenntnisse für weitere Forschungsarbeiten zu gewinnen. Aus diesem Grund wurde sich für eine Kombination aus beiden Regeln entschieden und ein $N = 50$ pro Darbietungsart (virtuell & Präsenz) erhoben.

Zu Beginn der Datenanalyse wurde ein Datensatz in SPSS (IBM Corp., 2016) erstellt, welcher in den nächsten Berechnungsschritten in R (RStudio Team, 2020) weiterverarbeitet wurde. Die Daten wurden in mehr als einem Jahr in verschiedenen Unternehmen von 100 Versuchspersonen nicht randomisiert erhoben, da die anfängliche Homeoffice Pflicht (Bundesregierung, 2022) zu Beginn der Pandemie die Durchführung in Präsenz verhinderte und somit eine zufällige Zuordnung der Teilnehmenden nicht ermöglicht werden konnte. Dies könnte zu einer Verzerrung der Aussagekraft von Darbietungsarten führen (Stone & Tang, 2013). Um dieser Verzerrung entgegenzuwirken, wurde das Propensity Score Matching angewendet, welches die Personen der beiden Darbietungsarten anhand der demografischen Daten (Alter, Geschlecht, Berufserfahrung, höchster Bildungsgrad) matcht. Durch dieses Matching der beiden Versuchsgruppen kann die Stichprobenzusammensetzung an die Randomisierung heranreichen (Stone & Tang, 2013). Als Grundlage des Verfahrens wurde die Vorgehensweise von Randolph et al. (2014) gewählt, welcher das Matching mit Hilfe des Zusatzpaketes „*matchit*“ (Ho et al., 2011) von R (RStudio Team, 2020) durchführte. Die Matchingpaare wurden anhand der demografischen Daten gematcht, welche aus Alter, Geschlecht, Berufserfahrung sowie dem höchsten Bildungsgrad bestanden. Es existieren sechs verschiedene Matching Methoden, die gemäß dem Vorschlag von Randolph et al. (2014) auf die beste Passung ausprobiert wurden. Die erste Methode ist das „*Exact Matching*“, bei dem jede Versuchsperson mit einer Person aus der Kontrollgruppe, welche den gleichen Wert in den Kovarianzen besitzt, gematcht wird. „*Subclassification*“ wird die Methode genannt, bei der die Daten in Subgruppen unterteilt werden, die ähnliche Kovariate in den demografischen Daten besitzen. Als Matching Methode „*Nearest Neighbour*“ wird das Vorgehen beschrieben, bei dem jede Versuchsperson mit ihrem nächsten Nachbarn (Eigenschaften am nächsten) gematcht wird. Das „*Optimal Matching*“ stellt die Matchingpaare so zusammen, dass der Mittelwertabstand am geringsten ist. Die „*Genetic Matching*“ Methode wird als rechenintensiv bezeichnet, da in dieser Methode ein Algorithmus zum Abgleich der beiden Gruppen gesucht wird. Als letzte Methode wird die

„*Coarsened Exact Matching*“ Methode beschrieben, die analog zum „*Exact Matching*“ vorgeht, jedoch das Gleichgewicht der anderen Kovariate ebenfalls berücksichtigt. Um sich für eine Matching-Methode entscheiden zu können, wurden die Kennwerte der Stichproben vor und nach dem Matching betrachtet. Die Entscheidung fiel auf die Methode mit den geringsten Mittelwertunterschieden zwischen den Gruppen, da somit die geringsten Stichprobeneffekte zu erwarten waren. Zusätzlich wurden der Jitter Plot und die Histogramme betrachtet, welche die oben aufgeführten Kennwerte grafisch darstellen, um die Stichprobenverteilung zu betrachten. Auch hier wurde sich für die Methode entschieden, bei der die beiden Gruppen der virtuellen Darbietungsart und der Darbietungsart in Präsenz grafisch am vergleichbarsten waren. Nach den Analysen wurde die Anwendung der „*Subclassification*“ Methode als sinnvoll erachtet, da die Passung der beiden Gruppen in diesem Matching am besten passte. Die genauen Kennwerte werden im Ergebnisteil erläutert. Alle weiteren Berechnungen wurden anhand dieses neu zusammengestellten Datensatzes berechnet.

Die Studie sollte überprüfen, ob die Zusammenhänge des Job Performance Models von Tett und Burnett (2003) in den beiden Darbietungsarten (virtuell und in Präsenz) verschieden sind. Darüber hinaus wurde der Zusammenhang zwischen der Persönlichkeit und der Gerechtigkeitswahrnehmung betrachtet. Zur Überprüfung aller Hypothesen wurde ein Strukturgleichungsmodell mit Regressionspfaden gerechnet, welches Zusammenhänge zwischen manifesten und latenten Variablen analysiert (Wentura & Pospeschill, 2015). Die latenten Variablen, welche nicht beobachtbar waren, wurden auf vier verschiedene Arten durch manifeste Variablen erfasst. Die Persönlichkeit wurde anhand der Items der AEOS (Wedemeyer et al., in Vorbereitung), das Arbeitsverhalten anhand des Gesamtwerts der Kompetenzen von den Beobachtenden, Berufserfolg durch die Potenzialaussage von den Beobachtenden sowie der Fremdeinschätzung und die Gerechtigkeitswahrnehmung durch den Mittelwert der SPJS (Bauer et al, 2001) betrachtet (s. Abbildung 12). Das Strukturgleichungsmodell wurde gegenüber den Regressionsanalysen bevorzugt, da Wechselwirkungen zwischen den Variablen betrachtet werden können (Weiber & Mühlhaus, 2014).

Abbildung 12
Vereinfachte Darstellung des Strukturgleichungsmodells zur Hypothesentestung



Anmerkung. Das Modell wurde in beiden Darbietungsarten gerechnet und verglichen.

Da die Vorhersagekraft der Persönlichkeit auf das Arbeitsverhalten und den Berufserfolg sowie die Zusammenhänge der Gerechtigkeitswahrnehmung in den beiden Gruppen virtuell und in Präsenz verglichen wurde, wurde eine Mehrgruppen-Kausalanalyse berechnet. Die Mehrgruppenkausalanalyse, auch MGKA genannt, betrachtet die Zusammenhänge eines Strukturgleichungsmodells, jedoch über mehrere Gruppen hinweg, und vergleicht diese (Weiber & Mühlhaus, 2014). Die Mehrgruppen-Kausalanalyse wurde mithilfe des Datenanalyse Programms R (RStudio Team, 2020) und dessen Zusatzprogramm „*lavaan*“ (Rosseel, 2012) berechnet. Bevor die Analyse durchgeführt wurde, wurde die Messvarianz anhand des Vorgehens nach Werner (2012) kontrolliert. Es ist wichtig, dass die Merkmale bzw. die Strukturmodelle in beiden Gruppen vergleichbar sind, damit die kausalen Unterschiede nur auf die Darbietungsart zurückzuführen sind (Weiber & Mühlhaus, 2014). Diese Prüfung wurde über „*semTools*“ (Jorgensen, 2022) berechnet, welches vier verschiedene Arten der Invarianz überprüft. Zum einem prüft es, ob eine konfigurale Varianz besteht, was bedeuten würde, dass beide Gruppen in den Merkmalen gleich sind (Weiber & Mühlhaus, 2014). Die schwache Invarianz war gegeben, wenn die Faktorladungen (loadings) in beiden Gruppen vergleichbar waren (Werner, 2012). Die nächste Form der Invarianz war die starke Invarianz, welche von der Vergleichbarkeit der Faktorladungen sowie Konstanten

(intercepts) zwischen den Gruppen ausging. Eine weitere Steigerung der Invarianz war die strikte Invarianz, welche die Vergleichbarkeit der Faktorladungen, Konstanten, Messfehlervarianz und Mittelwerte der latenten Variablen (means) voraussetzte. Nachdem die Messinvarianz überprüft wurde, wurde das Strukturgleichungsmodell, wie in Abbildung 12 beschrieben, für beide Gruppen getrennt mit dem Zusatzbefehl group = „*Variable*“ berechnet. Die Modellgüte wurde in der Mehrgruppen-Kausalanalyse genauso wie in dem ursprünglichen Strukturgleichungsmodell nach West et al. (2012) bewertet (*TLI* und *CFI* ≥.95, *RMSEA* ≤.06 und *SRMR* ≤.08). Die Kennwerte des Strukturgleichungsmodells wurden in beiden Gruppen getrennt voneinander dargeboten und mussten verglichen werden. Entscheidend für die Regressionsanalysen dieses Modells war das Bestimmtheitsmaß R^2, welches die Varianzaufklärung der abhängigen Variable angab. Das Bestimmtheitsmaß wurde nach Chin (1998) interpretiert, welcher von einer substanziellen Erklärungskraft spricht, wenn $R^2 \geq .67$ ist und von mittelgut bei einem Wert $.66 \leq R^2 > .33$. Bei einem Betrag von $R^2 \leq .33$ definierte Chin (1998) das Bestimmtheitsmaß als schwach. Neben dem Bestimmtheitsmaß konnten die unstandardisierten Regressionskoeffizienten verglichen werden, da standardisierte Koeffizienten nicht gruppenspezifisch kalibriert werden (Urban & Meyerl, 2013).

Weitere Analysen. Zur Überprüfung der Hypothesen 5 und 6 wurden Mittelwertvergleiche anhand von *t*-Tests für unabhängige Stichproben berechnet, da diese beiden Analysen das Strukturgleichungsmodell zu komplex gestaltet hätten. Darüber hinaus sollte die Überprüfung dieser Hypothesen nur einen ersten Anhaltspunkt für weitere Forschungsarbeiten liefern und somit wurde sich für einfache Mittelwertvergleiche entschieden. Da die Versuchspersonen nur einer Darbietungsart zugeordnet wurden und den Workshop virtuell oder in Präsenz durchliefen, sind die beiden Darbietungsarten unabhängig in der Stichprobe (Wentura & Pospeschill, 2015). Für die Mittelwertvergleiche wurde zu Hypothese 5 ein Mittelwert aus den drei Kompetenzeinschätzungen (Kommunikationsfähigkeit, Konfliktfähigkeit und Durchsetzungsfähigkeit) gebildet sowie separat die Potenzialaussage betrachtet. Der *t*-Test zur Überprüfung der Hypothese 6 wurde anhand der Mittelwerte aus den Bewertungen der elf Gerechtigkeitsregeln von Gilliland (1993) berechnet. Die Powerberechnung bei G*Power (Faul et al., 2009) für einen einseitigen *t*-Test zeigte, dass bei einer mittleren Effektstärke von $d = .3$, einem Signifikanzniveau $\alpha = .05$ und einer Power von .9 ein Stichprobenumfang von $N = 97$ gewählt werden sollte. Damit ist die Aussagekraft für diese Arbeit mit 100 Versuchspersonen gegeben. Es wurde auf eine

Power von .95 verzichtet, da ein Stichprobenumfang von $N = 122$ nötig gewesen wäre, die *t*-Tests jedoch nicht der Hauptanalyse dieser Studie dienten.

Zur Erstellung einer Handlungsanweisung für die Praxis, welche Hinweise auf die Bedeutung der einzelnen Gerechtigkeitsregeln nach Gilliland (1993) geben sollten, wurde eine grafische Analyse berechnet. Da die Teilnehmenden Rangfolgen erstellt hatten, wurden keine Mittelwertanalysen berechnet. Diese hätten keine Aussagekraft gehabt, da sich so die Rangfolgen gemittelt und somit verzerrt hätten (Stange, 2013). Stattdessen wurde pro Gerechtigkeitsregel ein Histogramm erstellt, um zu analysieren, welche Gerechtigkeitsregel eher niedrige (1 = am wichtigsten) und welche eher hohe Bewertungen besaß. Daraus wurden die drei wichtigsten und die drei unwichtigsten Gerechtigkeitsregeln identifiziert und anschließend als Handlungsempfehlung formuliert.

Ergebnisse

Da die Stichproben in der Demografie verschieden waren und die Teilnehmenden nicht randomisiert in die zwei Versuchsgruppen zugeordnet wurden, wurde zu Beginn ein Propensity Score Matching mit der Methode „*Subclassifikation*“ durchgeführt. Die Ergebnisse des Matchings zeigten, dass sich die beiden Stichproben in den demografischen Daten angenähert haben (s. Tabelle 26). Da die Geschlechterverteilung vor dem Matching fast identisch war, hat sich bei dem Matching diesbezüglich nichts verändert. Die größten Unterschiede existierten zwischen dem Durchschnittsalter beider Gruppen. Die Mittelwertdifferenz in den Subgruppen unterschied sich nun nur noch um $M_{Differenz} = -.07$, was bedeutet, dass das Alter in der virtuellen Gruppe unter dem der Präsenzgruppe lag. Bei den demografischen Daten Bildungsgrad und Berufserfahrung lag der Mittelwertunterschied bei $M_{Bildungsgrad\ Diff} = -.02$ und $M_{Berufserfahrung\ Diff} = -.14$.

Im nächsten Schritt wurde die grafische Vergleichbarkeit der beiden Matchinggruppen anhand der Histogramme und dem Jitter Plot betrachtet (s. Abbildung 13). Der Jitter Plot zeigte durch die fehlenden Kreise in der obersten und untersten Reihe, dass es keine Personen gab, für die kein Matching möglich war. Die beiden mittleren Zeilen zeigten die Verteilung der Matchings in den beiden Gruppen. In beiden Gruppen lagen viele Personen im Bereich zwischen .04 und .8. Weiter links unterschieden sich die Grafiken der Kontrollgruppe und der virtuellen Gruppe. In der Kontrollgruppe waren viele kleine Punkte zwischen .0 und .4 ersichtlich, wohingegen in der virtuellen Gruppe sechs größere Punkte zwischen ca. .1 und .3 lagen. Die großen Punkte bedeuteten, dass vielen Personen der gleiche Wert zugeschrieben wurde.

Tabelle 26

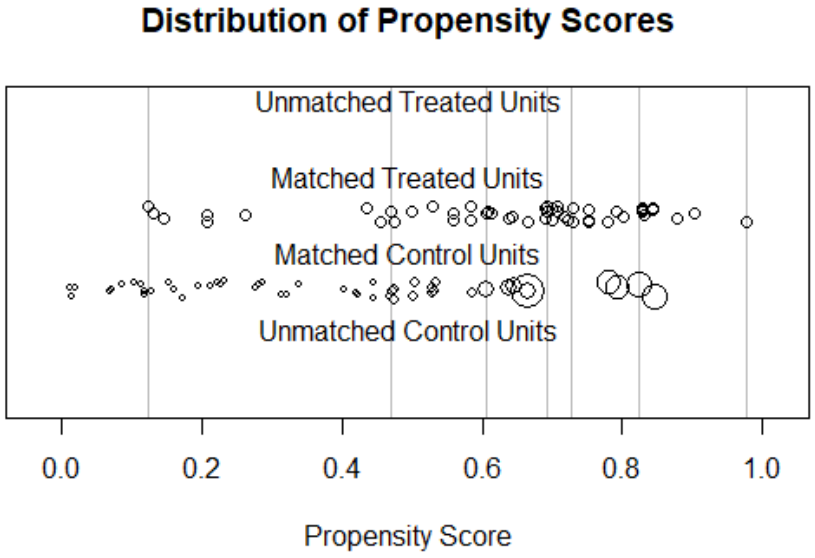
Mittelwertvergleich der demografischen Variablen vor und nach dem Matching

Demografie	Vor dem Matching		Nach dem Matching		
	$M_{(in\ Präsenz)}$	$M_{(virtuell)}$	$M_{(in\ Präsenz)}$	$M_{(virtuell)}$	$M_{(Differenz)}$
Geschlecht	1.54	1.48	1.51	1.48	-.07
Alter	35.52	27.32	27.87	27.32	-.07
Bildungsgrad	4.36	3.52	3.54	3.52	-.02
Berufserfahrung	14.52	5.58	6.51	5.58	-.14

Anmerkung. Die Mittelwertdifferenz in der letzten Spalte ist negativ, da der Mittelwert der Kontrollgruppe (in Präsenz) von dem der virtuellen Gruppe differenziert wurde.

Abbildung 13

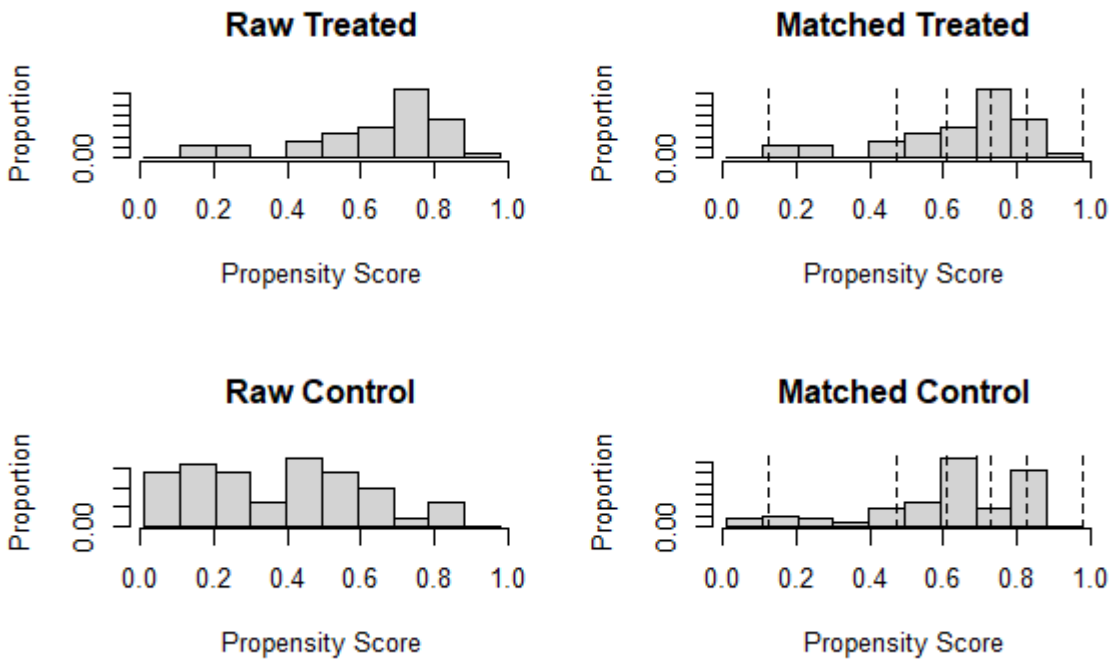
Jitter Plot des Propensity Score Matchings



Anmerkung. Jeder Kreis in der Abbildung stellt ein Matching dar. Je größer die Kreise sind, desto mehr Matchings liegen in dem Bereich.

Die Betrachtung der Histogramme zeigt, dass sich die Histogramme zu den beiden Gruppen vor dem Matching sehr unterschieden (s. Abbildung 14). Nach dem Matching waren die Histogramme vergleichbar, jedoch hatte die virtuelle Gruppe einen hohen Wert bei .7 wohingegen die in Präsenz Gruppe jeweils einen hohen Balken bei .6 und .8 besaß. Anhand der Mittelwerte, des Jitter Plots und der Histogramme wurde sich für die Verwendung des gematchten Datensatzes entschieden, welcher in allen weiteren Analysen verwendet wurde.

Abbildung 14
Histogramme des Propensity Score Matchings



Anmerkung. Die beiden Histogramme auf der linken Seite sind vor dem Matching entstanden, die rechts nach dem Matching. Oben ist die virtuelle Gruppe und unten die in Präsenz Gruppe abgebildet.

Nach Durchführung des Propensity Score Matchings wurde das Strukturgleichungsmodell berechnet. Bei der Berechnung des Strukturgleichungsmodells wurde eine Fehlermeldung ausgegeben, welche zur Schussfolgerung führt, dass das Modell für die Anzahl der Versuchspersonen zu komplex war. Daher wurde nachfolgend das Modell nur mit Mittelwerten als manifeste Variablen berechnet.

Vor der Berechnung des neuen Strukturgleichungsmodells wurde die Messinvarianz kontrolliert und betrachtet, ob die Variablen des Modells in beiden Gruppen (virtuell versus in Präsenz) vergleichbar sind. Die Berechnungen verdeutlichten, dass signifikante Unterschiede zwischen den Gruppen in den Konstanten (intercepts) existierten, was auf eine schwache Invarianz hindeutet. Da nur eine schwache Invarianz bestand, wurde das Strukturgleichungsmodell berechnet und interpretiert (s. Tabelle 27).

Tabelle 27
Ergebnisse der Überprüfung der Messinvarianz

	Df	*ACI*	*BCI*	χ^2	χ^2_{Diff}	Df_{Diff}	*p*
Configural	34	1370.80	1563.60	32.43			
Loadings	36	1369.30	1556.90	34.94	2.51	2	n.s.
Intercepts	38	1373.40	1555.80	43.02	8.07	2	< .05
Means	45	1370.10	1534.20	53.65	10.64	7	n.s

Anmerkung. n.s. bedeutet nicht signifikant

Nach der Überprüfung der Messinvarianz der Daten konnten die Hypothesen anhand der Mehrgruppenanalyse des Strukturgleichungsmodells untersucht werden. Die akzeptablen Fit-Werte eines Strukturgleichungsmodells nach West et al. (2012) sind gegeben, wenn *TLI* und *CFI* ≥.95, *RMSEA* ≤.06 und *SRMR* ≤.08 sind. In dem vereinfachten Modell war die akzeptable Modellgüte gegeben und der χ^2-Test war nicht signifikant. Daher schienen die Daten zum Modell zu passen ($\chi^2_{(18)}$ = 11.340; p >.5; *CFI* = 1.00; *TLI* = 1.11; *RMSEA* = .00; *SRMR* = .04). Die Teststatistiken unterschieden sich in den einzelnen Gruppen und der χ^2-Wert der Kontrollgruppe (in Präsenz) lag unter dem der virtuellen Subgruppe ($\chi^2_{\text{Präsenz}}$ = 2.17; χ^2_{virtuell} = 9.17).

Die Hypothese 1 besagte, dass das Arbeitsverhalten Berufserfolg während des Workshops in der virtuellen Darbietungsart geringer vorhersagt als in Präsenz. In der virtuellen Versuchsgruppe wurde diese Regressionsanalyse des Strukturgleichungsmodells signifikant. Daraus folgte, dass das Arbeitsverhalten den Berufserfolg während des Workshops vorhersagte (β = -.87; p <.001). Der unstandardisierte Regressionskoeffizient besagte, dass eine niedrige Potenzialaussage der Beobachtenden durch ein hohes Arbeitsverhalten vorhergesagt wird. Das Bestimmtheitsmaß für die Variable Berufserfolg während des Workshops (Potenzialaussage Beobachtende) betrug R^2 = .67, was nach Chin (1998) für einer substanzielle Erklärungskraft spricht. Da die Potenzialaussage negativ kodiert war, sagte ein hohes Arbeitsverhalten einen hohen Berufserfolg während des Workshops vorher. In der Kontrollgruppe wurde diese Regressionsanalyse des Strukturgleichungsmodells ebenfalls signifikant (β = -.80; p <.001). Im Vergleich zur virtuellen Stichprobe war der Regressionskoeffizient der Kontrollgruppe größer. Daraus folgte, dass bei einem gleichen Wert im Arbeitsverhalten in der Kontrollgruppe ein niedriger Berufserfolg während des Workshops vorhergesagt wurde, da die Potenzialaussage negativ kodiert war. Das Bestimmtheitsmaß für Berufserfolg lag mit R^2 = .84 höher als in der virtuellen Stichprobe und sorgte für eine substanzielle Erklärungskraft. Da der

unstandardisierte Regressionskoeffizient in der Kontrollgruppe größer als in der virtuellen Gruppe war, konnte die Hypothese 1 bestätigt werden.

Die Hypothese 2 nahm an, dass das Arbeitsverhalten durch die Darbietungsart den tatsächlichen Berufserfolg unterschiedlich vorhersagt. Die Analyse der Hypothese 2 wurde in der virtuellen Stichprobe nicht signifikant und somit schien kein Zusammenhang zwischen dem Arbeitsverhalten und dem Berufserfolg während des Workshops zu bestehen (β = -.12; p = n.s.). Das Bestimmtheitsmaß betrug R^2 = .03, was nach Chin (1998) als schwache Erklärungskraft bewertet wird. Auch die Analyse in der Kontrollgruppe wurde nicht signifikant und besaß ein schwaches Bestimmtheitsmaß (β = -.09; p = n.s.; R^2 = .04). Beide Analysen fanden keine Vorhersagekraft des Arbeitsverhaltens zum tatsächlichen Berufserfolg. Hypothese 2 konnte nicht bestätigt werden, da diese von einem Unterschied zwischen den Gruppen ausging.

Die Hypothese 3 überprüfte, ob die vier Aspekte der Offenheit für Erfahrung und Extraversion durch die Darbietungsart das Arbeitsverhalten unterschiedlich vorhersagen. Dieser Zusammenhang wurde in der virtuellen Stichprobe in keinem der vier Aspekte signifikant ($\beta_{Intellekt}$ = .25; p = n.s.; $\beta_{Offenheit}$ = -.14; p = n.s.; $\beta_{Enthusiasmus}$ = -.51; p = n.s.; $\beta_{Durchsetzungsfähigkeit}$ = -.26; p = n.s.). Das Bestimmtheitsmaß konnte als schwach bewertet werden (R^2 = .05). Diese Ergebnisse bedeuteten, dass das Arbeitsverhalten von keinem Aspekt vorhergesagt wurde. In der Kontrollgruppe wurde demgegenüber die Regression des Strukturgleichungsmodells vom Intellekt und Arbeitsverhalten signifikant (β = 1.42; p <.001). Der Regressionskoeffizient besagte, dass ein hoher Wert im Aspekt Intellekt auch eine hohe Bewertung des Arbeitsverhaltens vorhersagt. Darüber hinaus wurden neben dem Regressionspfad von Intellekt und Arbeitsverhalten auch in der Kontrollgruppe keine weiteren Pfade der drei anderen Aspekte signifikant ($\beta_{Offenheit}$ = .06; p = n.s.; $\beta_{Enthusiasmus}$ = -.47; p = n.s.; $\beta_{Durchsetzungsfähigkeit}$ = -.59; p = n.s.). Das Bestimmtheitsmaß lag mit R^2 = .32 deutlich über dem der virtuellen Stichprobe und zudem am Grenzwert zur mittleren Erklärungskraft (ab .33 nach Chin (1998)). Die Ergebnisse lassen auf die Bestätigung von Hypothese 3 schließen.

Die vierte und letzte Hypothese besagte, dass die vier Aspekte der Offenheit für Erfahrung und Extraversion durch die Darbietungsart die Gerechtigkeitswahrnehmung unterschiedlich vorhersagen. In der virtuellen Stichprobe sagten die Aspekte der Extraversion und Offenheit für Erfahrung das Arbeitsverhalten nicht signifikant vorher ($\beta_{Intellekt}$ = .08; p = n.s.; $\beta_{Offenheit}$ = -.03; p = n.s.; $\beta_{Enthusiasmus}$ = .03; p = n.s.; $\beta_{Durchsetzungsfähigkeit}$ = .23; p = n.s.). Das Bestimmtheitsmaß war als schwach zu bewerten R^2 = .05. Diese Ergebnisse bedeuteten, dass

keine Zusammenhänge zwischen den Aspekten und dem Arbeitsverhalten bestanden. In der Kontrollgruppe waren die Ergebnisse vergleichbar und es wurde auch hier keine Regression signifikant ($\beta_{Intellekt}$ = -.05; p = n.s.; $\beta_{Offenheit}$ = .06; p = n.s.; $\beta_{Enthusiasmus}$ = -.04; p = n.s.; $\beta_{Durchsetzungsfähigkeit}$ = .03; p = n.s.). Obwohl das Bestimmtheitsmaß der virtuellen Gruppe etwas größer war als das der Kontrollgruppe ($R^2_{virtuell}$ = .05, $R^2_{Präsenz}$ = .03), wurde sich aufgrund der nicht signifikanten Ergebnisse gegen eine Bestätigung der Hypothese 4 entschieden.

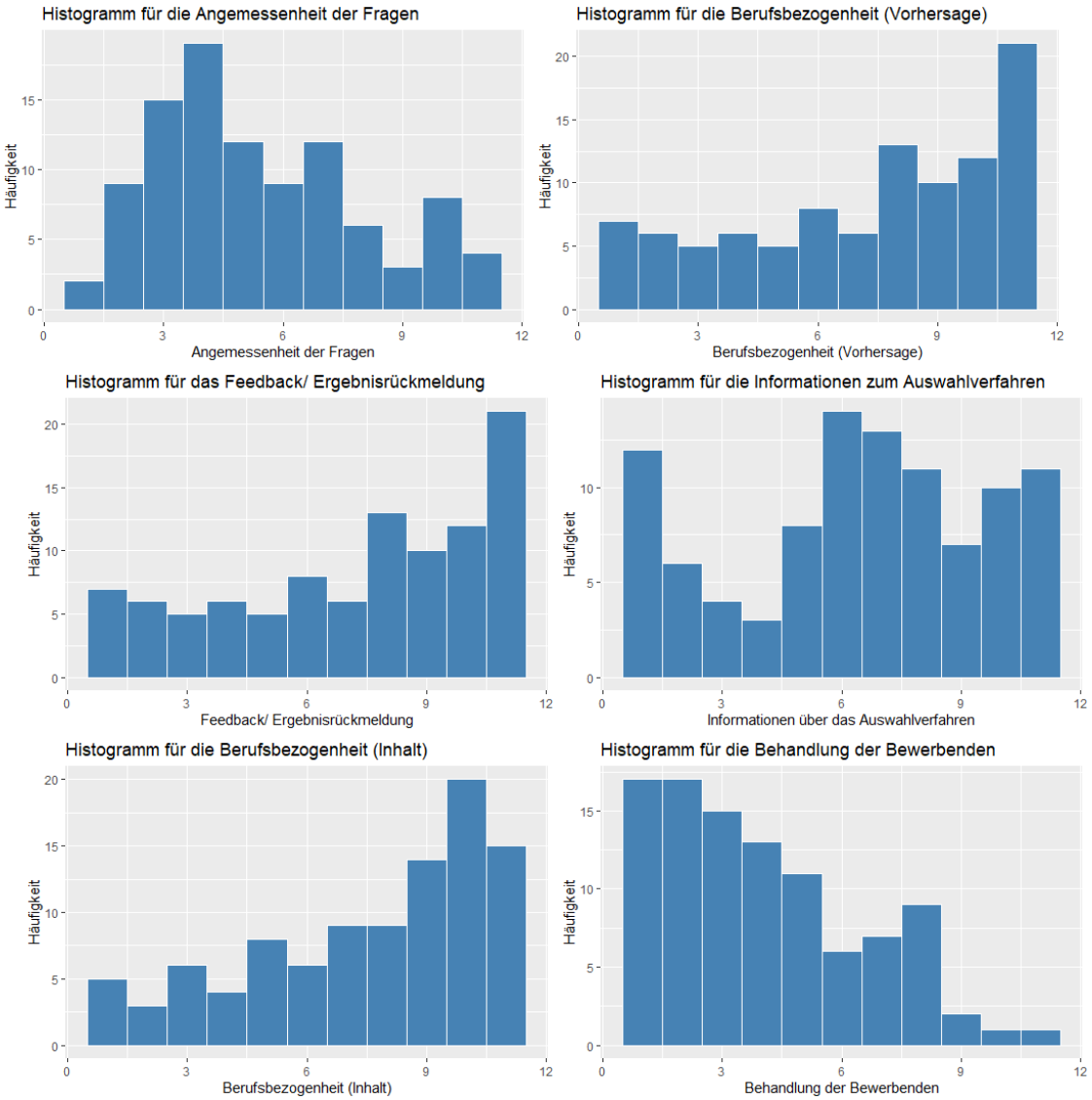
Im Anschluss wurden die Hypothesen 5 und 6 angelehnt an die Forschungsarbeiten von Basch & Melchers (2020) und Blacksmith et al. (2016) überprüft und die Histogramme zu der Wichtigkeit der einzelnen Regeln der Gerechtigkeitswahrnehmung von Gilliland (1993) berechnet. Die Hypothese 5 ging davon aus, dass die Teilnehmenden in der virtuellen Stichprobengruppe schlechter von den Beobachtenden bewertet wurden als die Personen in Präsenz. Um diese Hypothese zu prüfen, wurden *t*-Tests mit unabhängigen Stichproben linksseitig auf dem 5%-Niveau berechnet. Der *t*-Test zur Überprüfung der Mittelwertunterschiede zwischen den Kompetenzeinschätzungen der Beobachtenden wurde nicht signifikant ($t_{(98)}$ = -2.45; p = n.s.). Dies bedeutete, dass der Mittelwert der virtuellen Gruppe nicht signifikant schlechter war. Die Betrachtung der Mittelwerte zeigte sogar, dass der Mittelwert der virtuellen Stichprobe *(*$M_{virtuell}$ = 4.12; *CI* [3.99 | 4.43]) größer als der von der in Präsenz Stichprobe war ($M_{Präsenz}$ = 3.75; CI [3.45 | 4.05]). Die Betrachtung der Potenzialaussage zeigte, dass der *t*-Test ebenfalls nicht signifikant wurde. Die Teilnehmenden des virtuellen Workshops wurden besser bewertet, da die Potenzialaussage rekodiert war ($t_{(98)}$ = 1.93; p = n.s.; $M_{virtuell}$ = 1.98; *CI* [1.76 | 2.21]; $M_{Präsenz}$ = 2.32; *CI* [2.05 | 2.59]). Da somit die Bewerbenden in der virtuellen Darbietungsart besser bewertet wurden, wurden beide Tests noch einmal rechtsseitig berechnet, um zu prüfen, ob die Mittelwertunterschiede signifikant sind. Die Tests zeigten, dass die Teilnehmenden des virtuellen Workshops sowohl in der Kompetenzeinschätzung ($t_{(98)}$ = -2.45; p <.01) als auch in der Potenzialaussage ($t_{(98)}$ = 1.93; p <.05) signifikant besser bewertet wurden.

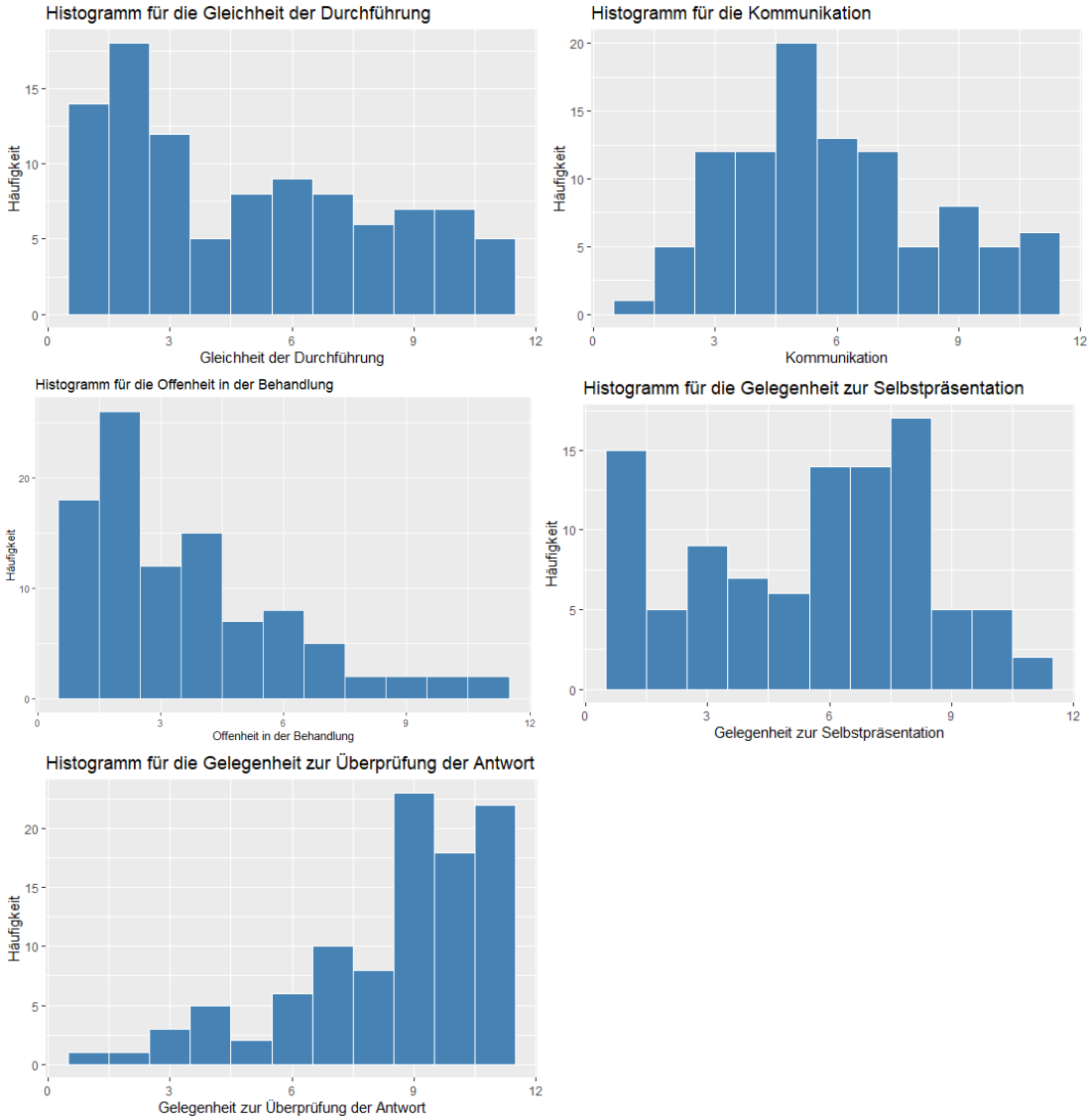
Die Hypothese 6 besagte, dass die Teilnehmenden der virtuellen Stichprobengruppe die Gerechtigkeit der Dialogübung schlechter bewerten als die im Workshop in Präsenz. Der *t*-Test wurde nicht signifikant. Die Betrachtung der Mittelwerte zeigte, dass sich diese kaum unterschieden ($t_{(98)}$ = -.045; p = n.s.; $M_{virtuell}$ = 3.40; *CI* [3.31 | 3.49]; $M_{Präsenz}$ = 3.40; *CI* [3.33 | 3.47]). Diese Hypothese wurde nicht bestätigt, da die Gerechtigkeitswahrnehmung der Teilnehmenden beider Gruppen nicht signifikant verschieden war.

Die Auswertung der Histogramme zur Wichtigkeit der Gerechtigkeitsregeln nach Gilliland (1993) zeigte auf, dass einige Regeln öfter als wichtig bewertet wurden als andere. Einige Regeln wurden eher im mittleren Bereich bewertet (s. Abbildung 15).

Abbildung 15

Verteilungen der Antworten zu der Frage: Was ist Ihnen in Auswahlverfahren wichtig?





Anmerkung. Die Histogramme sind wie folgt angeordnet: 1. Reihe Angemessenheit der Fragen, Berufsbezogenheit (Inhalt); 2. Reihe Behandlung der Bewerbenden, Berufsbezogenheit (Vorhersage); 3. Reihe: Feedback/ Ergebnisrückmeldung, Gleichheit der Durchführung; 4. Reihe: Informationen zum Auswahlverfahren, Kommunikation; 5. Reihe Offenheit in der Behandlung, Gelegenheit zur Selbstpräsentation; 6. Reihe Überprüfung der Antwort

Die Histogramme zeigten, dass vor allem die Regeln Behandlung der Bewerbenden, Offenheit in der Behandlung und Gleichheit in der Durchführung oft niedrig bewertet wurden und somit wichtig zu sein schienen. Demgegenüber wurden die Regeln Gelegenheit zur Überprüfung der Antwort, Feedback/ Ergebnisrückmeldung, Berufsbezogenheit (Inhalt) und Berufsbezogenheit (Vorhersage) hoch bewertet und scheinen daher den Bewerbenden für die Gerechtigkeitswahrnehmung im Auswahlprozess nicht so wichtig zu sein. In den

Histogrammen zu den Regeln Gelegenheit zur Selbstpräsentation, Information zum Auswahlverfahren, Kommunikation und Angemessenheit der Fragen wurden keine eindeutigen Tendenzen gefunden.

Diskussion

Diese Studie beschäftigte sich mit der Fragestellung, ob Dialogübungen im virtuellen Raum eine vergleichbare prädiktive Validität besitzen wie Dialogübungen in Präsenz und ob sich der Zusammenhang zwischen der Gerechtigkeitswahrnehmung mit den Aspekten Enthusiasmus, Durchsetzungsfähigkeit, Offenheit und Intellekt in den beiden Darbietungsarten unterscheidet. Es wurde davon ausgegangen, dass die virtuell durchgeführte Dialogübung vor allem durch ihre örtliche Unabhängigkeit Chancen gegenüber der Dialogübung vor Ort besitzen kann (Fellner, 2019).

Die Hypothese 1 besagte, dass das Arbeitsverhalten Berufserfolg während des Workshops in der virtuellen Darbietungsart geringer vorhersagt als in Präsenz. Diese Hypothese wurde bestätigt, da der Regressionskoeffizient der Kontrollgruppe größer als bei der virtuellen Gruppe war. Dies bedeutet, dass der Prädiktor Arbeitsverhalten das Kriterium Berufserfolg während des Workshops in Präsenz mehr vorhersagt als in der virtuellen Gruppe. Die Ergebnisse sprechen dafür, dass eine einfache Übersetzung der Übung in eine andere Darbietungsart für die eignungsdiagnostische Methode der Dialogübung nicht funktioniert und stützt somit die Vermutungen von Hertel et al. (2003). Diese Ergebnisse zeigen, dass in der Praxis aufgrund der besseren prädiktiven Validität eher der Einsatz der Dialogübung vor Ort empfehlenswert ist. Sofern Dialogübungen im virtuellen Raum durchgeführt werden müssen, müssen sich Praktiker:innen möglicher Validitätseinschränkungen bewusst sein. Darüber hinaus sollten in einem Auswahlverfahren für eine Position alle Bewerbenden die Dialogübung in der gleichen Darbietungsart durchführen und nicht beispielsweise nur ein Bewerbender aufgrund von einer zu großen Entfernung zum Unternehmen in der virtuellen Darbietungsart. Die Ergebnisse bedeuten für die Forschung, dass es empfehlenswert sein könnte zu betrachten, ob diese Ergebnisse auch in einem realen Verfahren oder einem reinen Laborsetting und nicht in einem Experiment bestehen. Sollte dies der Fall sein, könnte hinterfragt werden, ob sich die Dialogübung in der prädiktiven Validität in allen Berufsgruppen unterscheidet. Darüber hinaus könnte der Inhalt der Dialogübung hinterfragt werden. In dieser Studie wurde ein Konfliktgespräch gewählt. Möglicherweise hängt aber beispielsweise die virtuelle Dialogübung in einem Verkaufsgespräch stärker mit Berufserfolg zusammen. Darüber hinaus könnte auch in der

Wissenschaft hinterfragt werden, inwiefern eine andere Erfassung des Arbeitsverhaltens oder des Berufserfolgs Auswirkungen auf die prädiktive Validität besitzen.

Die Hypothese 2 besagte, dass das Arbeitsverhalten durch die Darbietungsart den tatsächlichen Berufserfolg unterschiedlich vorhersagt. Da die Regressionen des Strukturgleichungsmodells in beiden Darbietungsarten nicht signifikant wurden, konnte diese Hypothese nicht bestätigt werden. Die Hypothese 2 überprüfte die gleichen Zusammenhänge analog zu Hypothese 1 mit der Fremdeinschätzung als ein anderes Kriterium. Ferner bedeuten die nicht signifikanten Regressionsmodelle, dass sie in dieser Stichprobe und diesem Design keine Erklärungskraft durch den Prädiktor Arbeitsverhalten gegenüber dem Kriterium des tatsächlichen Berufserfolgs besitzen. Die Praxis kann anhand der Ergebnisse darauf schließen, dass die Dialogübung in beiden Darbietungsarten keine prädiktive Validität gegenüber der Fremdeinschätzung besitzt. Diesbezüglich sollte untersucht werden, ob die fehlenden Zusammenhänge auf den Inhalt der Dialogübung, die erfassten Kompetenzen oder die Fremdeinschätzung zurückzuführen sind. Dies könnte jedoch daran liegen, dass der Mittelwert bei M =1.63 (SD = .54) lag und somit fast alle Teilnehmenden dahingehend eingeschätzt wurden, dass sie Potenzial für den nächsten Entwicklungsschritt aufweisen. Nachfolgend sollte von Seiten der Wissenschaft angelehnt an die Hypothese 1 untersucht werden, wie das Arbeitsverhalten und der Berufserfolg anders erfasst werden könnten und welche Auswirkungen dies auf die prädiktive Validität hätte.

Die Hypothese 3 überprüfte, ob die vier Aspekte der Offenheit für Erfahrung und Extraversion durch die Darbietungsart das Arbeitsverhalten unterschiedlich vorhersagen. Die Hypothese konnte bestätigt werden, da der Regressionskoeffizient der Gruppe vor Ort größer als das in der virtuellen Gruppe war und somit das Arbeitsverhaltens durch die Aspekte in der Kontrollgruppe mehr vorhergesagt werden konnte. Darüber hinaus wurde die Regression des Strukturgleichungsmodells im Aspekt Intellekt in der Gruppe vor Ort signifikant. Diese Ergebnisse sprechen ebenfalls wie bei Hypothese 1 dafür, dass die Dialogübung vor Ort eine höhere prädiktive Validität zu besitzen scheint als im virtuellen Kontext. Darüber hinaus konnten die beiden Hypothesen das Job Performance Modell nach Tett und Burnett (2003) replizieren, da das Modell davon ausgeht, dass Einflussfaktoren (hier die Darbietungsart) die Zusammenhänge zwischen der Persönlichkeit, dem Arbeitsverhalten und dem Berufserfolg beeinträchtigen. Dies spricht ebenfalls dafür, die Darbietungsarten in der Praxis während eines Auswahlverfahrens nicht zu wechseln. Die Erklärungskraft des Aspekts Intellekt gegenüber des Arbeitsverhaltens bestärkt die Vermutung von Schermuly et al. (2019), dass das Persönlichkeitsmerkmal Offenheit für Erfahrung durch die Digitalisierung immer mehr

an Bedeutung gewinnt und in der Arbeitswelt nicht vernachlässigt werden sollte. Dennoch machen die Ergebnisse der Regressionen zur Vorhersage des Arbeitsverhaltens durch die Aspekte deutlich, dass über den Aspekt Intellekt in der Gruppe vor Ort hinaus keine Zusammenhänge bestehen. Daher muss kritisch betrachtet werden, wie Arbeitsverhalten erfasst wurde und ob neben den Aspekten weitere Variablen bzw. Persönlichkeitsmerkmale in diesem experimentellen Workshop eine Rolle spielen. Aus diesem Grund sollte zum einen betrachtet werden, wie die Zusammenhänge in einem realen Auswahlverfahren sind. Zum anderen sollte in der Forschung hinterfragt werden, ob eine breitere Erfassung der Dimensionen eine größere Aussagekraft besitzen könnte. Es könnte auch sein, dass die drei Dimensionen Neurotizismus, Verträglichkeit und Gewissenhaftigkeit mehr Varianz des Arbeitsverhaltens erklären.

Neben der prädiktiven Validität sollte auch die Gerechtigkeitswahrnehmung der Bewerbenden betrachtet werden, um zu überprüfen, ob die vier Aspekte der Offenheit für Erfahrung und Extraversion durch die Darbietungsart die Gerechtigkeitswahrnehmung unterschiedlich vorhersagen. (Hypothese 4). Die Hypothese 4 konnte nicht bestätigt werden, da keine signifikanten Ergebnisse in beiden Darbietungsarten entstanden, die Regressionskoeffizienten vergleichbar und somit keine Unterschiede erkennbar waren. Dies bedeutet, dass in diesem Experiment die Aspekte in keiner Darbietungsart die Gerechtigkeitswahrnehmung erklären. Somit konnten die Ergebnisse von Moldzio (2014) und van Vianen et al. (2004) nicht repliziert werden. Für die Praxis bedeuten die Ergebnisse, dass beispielsweise eine durchsetzungsstarke Person die Gerechtigkeit der Dialogübung nicht anders einschätzen würde als eine weniger durchsetzungsstarke Person. Dieser Sachverhalt besteht auch unabhängig von der Darbietungsart. Daraus kann für die Praxis abgeleitet werden, dass die Gerechtigkeitswahrnehmung nicht von Persönlichkeitsmerkmalen abzuhängen scheint. Dies würde die Personalauswahl vereinfachen, da sie unabhängig von der Persönlichkeit eines Bewerbers konzipiert werden könnte. Dennoch sollte die Forschung weiterhin überprüfen, ob die Zusammenhänge in einem realen Auswahlverfahren ebenfalls bestehen und welche Zusammenhänge in den anderen Dimensionen bzw. Aspekten zur Gerechtigkeitswahrnehmung existieren. Sollte die Gerechtigkeitswahrnehmung nicht von der Persönlichkeit vorhergesagt werden, ist es dennoch von Bedeutung, zu hinterfragen, ob weitere Faktoren entscheidend sein könnten (zum Beispiel Berufserfahrung etc.). Die Zusammenhänge der Gerechtigkeitswahrnehmung und der Persönlichkeiten sollten jedoch auch betrachtet werden, wenn die Gerechtigkeitswahrnehmung durch andere Konstrukte erfasst wird.

Weitere Analysen

Ergänzend zur Prüfung der Validität von Dialogübungen im virtuellen Raum wurde in der Studie untersucht, ob sich die Ergebnisse von Basch & Melchers (2020) und Blacksmith et al. (2016) auf die Dialogübung im virtuellen Raum übertragen lassen. Dabei wurde überprüft, ob die Versuchspersonen in der virtuellen Darbietungsart schlechter als die in Präsenz von den Beobachtenden bewertet wurden sowie die Teilnehmenden die virtuelle Übung als ungerechter wahrnahmen.

Hypothese 5 besagte, dass Teilnehmende in der virtuellen Darbietungsart schlechter von den Beobachtenden bewertet wurden als Teilnehmende der Präsenzstichprobe. Die Ergebnisse der *t*-Tests zeigten, dass die Teilnehmenden des virtuellen Workshops sogar signifikant besser bewertet wurden und somit die Hypothese 5 nicht bestätigt werden konnte. Die Ergebnisse replizieren somit die Studie von Basch und Melchers (2020) sowie Blacksmith et al. (2016) nicht. Dies könnte verschiedene Gründe haben. Teilnehmende konnten ihre Fähigkeiten durch die Übung anscheinend gut zeigen. Darüber hinaus könnte die bessere Bewertung von Teilnehmenden in der virtuellen Darbietungsart zeigen, dass die Kompetenzen, die für dieser Studie definiert wurden, in einer Videokonferenz gut erfasst werden können und die Teilnehmenden ihre Fähigkeiten präsentieren können. Außerdem waren die Beobachtenden geschult, wodurch ggf. in der virtuellen Darbietungsart das Verhalten der Teilnehmenden besser beobachtet wurde als bei ungeschulten Beobachtenden. Darüber hinaus handelt es sich um ein Experiment, die Teilnehmenden waren dadurch vielleicht weniger angespannt bzw. aufgeregt als in den Studien von Basch und Melchers (2020) sowie Blacksmith et al. (2016). Die bessere Bewertung in der virtuellen Darbietungsart würde dennoch dafür sprechen, dass vor Beginn der Auswahlverfahren eine Darbietungsart gewählt werden und nicht teilweise zu einer anderen gewechselt werden sollte, da durch die unterschiedlichen Bewertungen der Beobachtenden die Auswahlentscheidungen verzerrt werden könnten. Deshalb sollten Unternehmen auf mögliche Bewertungseffekte aufmerksam gemacht werden. Für die Forschung wäre nachfolgend interessant der Frage nachzugehen, auf welchen Faktoren die unterschiedlichen Bewertungen beruhen. Daraus resultierend könnten ggf. für die Praxis Handlungsempfehlungen entwickelt und überprüft werden, ob die gefundenen Ergebnisse auf ein reales Auswahlverfahren in einer Feldstudie replizierbar sind.

Die Hypothese 6 wurde ebenfalls angelehnt an die Forschung von Basch und Melchers (2020) bzw. Blacksmith et al. (2016) definiert und besagte, dass die Gerechtigkeitswahrnehmung der virtuellen Stichprobe schlechter als die der Stichprobe in

Präsenz ist. Es bestanden in der Gerechtigkeitswahrnehmung keine signifikanten Unterschiede zwischen den verschiedenen Darbietungsarten und somit konnte die Hypothese 6 nicht bestätigt werden. Somit konnten die Ergebnisse von Basch und Melchers (2020) bzw. Blacksmith et al. (2016) nicht auf die Dialogübung repliziert werden. Dafür könnte es folgende Gründe geben. Zum einen fand die Erhebung der Arbeit anders als bei Blacksmith et al. (2016) während der Pandemie statt. Es könnte sein, dass die Bedenken gegenüber virtuellen Verfahren durch zwei Jahre des virtuellen Arbeitens in der Pandemie reduziert wurden. Die Personen könnten durch die virtuelle Arbeit und die vielen Videokonferenzen gelernt haben, wie sie Mimik und Gestik über den Bildschirm vermitteln können. Zum anderen könnte die Studie von Truxillo et al. (2009) ebenfalls eine Erklärung für den nicht signifikanten Unterschied der Gerechtigkeitswahrnehmung gegenüber der Darbietungsart liefern. Sie fanden heraus, dass die Zusammenhänge in Bezug auf die Gerechtigkeitswahrnehmung in Feldstudien stärker waren als in Laborstudien. Dies könnte die Replizierbarkeit der Ergebnisse abschwächen. Daher sollte die Gerechtigkeitswahrnehmung ebenfalls in einer Feldstudie betrachtet werden. Die Ergebnisse liefern jedoch Ansatzpunkte dafür, dass in der Praxis keine Bedenken bezüglich der Gerechtigkeitswahrnehmung im Einsatz von virtuellen Dialogübung bestehen sollten und somit eine ortsunabhängige Möglichkeit der Dialogübung besteht.

Viele Forschungsarbeiten untersuchten, welche Auswahlverfahren als wie stark gerecht wahrgenommen werden (z.B. Basch & Melchers, 2020; Brenner et al., 2016). Es wurde aber nicht hinterfragt, was den Bewerbenden wichtig ist in Bezug auf die Gerechtigkeitswahrnehmung. Dies ist jedoch für die Praxis wichtig, damit Unternehmen wissen, worauf sie bei der Auswahl achten sollten. Die Studie ging einen ersten Schritt, um diese Lücke zu schließen und ließ eine Rangfolge der Gerechtigkeitsregeln von Gilliland (1993) durch die Teilnehmenden generieren. Die Analyse mittels Histogramme war lediglich eine Visualisierungsmöglichkeit der Fragestellung und keine statistische Analyse, kann jedoch einen Hinweis geben, auf dem die nachfolgende Forschung und Praxis aufbauen kann. Es wurde ersichtlich, dass vor allem interpersonelle Aspekte für die Gerechtigkeitswahrnehmung entscheidend sind, da die Regeln „*Behandlung der Bewerbenden*“, „*Offenheit in der Behandlung*“ und „*Gleichheit in der Durchführung*“ oft mit einer hohen Wichtigkeit bewertet wurden. Diese Regeln besagen, dass es den Teilnehmenden wichtig ist, dass sie höflich sowie freundlich behandelt werden und die Beobachtenden ihnen offen gegenübertreten. Außerdem ist den Teilnehmenden wichtig, das Gefühl zu haben, dass alle Bewerbenden gleich behandelt werden. Aus diesem Grund sollten die Unternehmen

darauf achten, dass sich die Beobachtenden ihrer Wirkung bewusst sind und die Beobachtenden zuvor in Trainings geschult werden, beispielsweise auch im Bereich der Kommunikation und Feedbackregeln. Darüber hinaus sollten sich die Beobachtenden vorab fragen, welche Werte in dem Unternehmen wichtig sind und sich diesbezüglich verhalten. Eine schlechte Behandlung könnte zu einer Ablehnung der Stelle und ggf. hohen Kosten führen (Moldzio, 2014), welches durch eine Schulung eventuell vermeidbar gewesen wäre. Überwiegend als weniger wichtig bewertet wurden die Regeln „*Gelegenheit zur Überprüfung der Antwort*“, „*Feedback/ Ergebnisrückmeldung*“, „*Berufsbezogenheit (Inhalt)*“ und „*Berufsbezogenheit (Vorhersage)*“. Dies bedeutet, dass es den Teilnehmenden weniger wichtig ist, dass sie ihre Antwort revidieren können, sie ein ausführliches Feedback erhalten oder die Aufgaben eines Verfahrens einen hohen beruflichen Bezug aufweisen. Der letzte Punkt würde dafürsprechen, dass die Unternehmen gute und valide Auswahlverfahren verwenden können und nicht jede Teilaufgabe eines Auswahlprozesses im kleinsten Detail auf die vakante Stelle abstimmen müssen (z.B. vier oder sechs Mitarbeitende im Team). Dennoch ist wichtig, die für die Stelle benötigten Kompetenzen zu erfassen, um einen Auswahlprozess valide zu gestalten. So muss ein Konfliktgespräch in einem Auswahlverfahren für eine überwiegend im Homeoffice stattfindende Tätigkeit nicht unbedingt als Videokonferenz oder in einem virtuellen Kontext stattfinden, sondern kann auch in Präsenz stattfinden. Obwohl diese Tendenzen erkennbar waren, sollten die Unternehmen darauf achten, alle Regeln von Gilliland (1993) gleichermaßen zu beachten, bis weitere Forschungsarbeiten differenziertere Angaben geliefert haben.

Limitationen

Diese Studie erhob die Daten in einem experimentellen Setting, was dazu führt, dass die Ergebnisse nicht vollständig auf ein Auswahlverfahren replizierbar sind. Dies könnte die Ergebnisse der Gerechtigkeitswahrnehmung abschwächen, da Truxillo et al. (2019) herausfanden, dass die Zusammenhänge in einer Feldstudie stärker als in einer Laborstudie sind. Da die Stichprobe mit $N = 100$ für ein ausführliches Strukturgleichungsmodell zu klein war, wurde mit Mittelwerten als manifeste Variablen gerechnet. Dadurch konnten Messfehlervarianzen durch verschiedene Items nicht berichtet werden, welche aber bestehen könnten (Weiber & Mühlhaus, 2014). Es wird empfohlen, neben der Überprüfung der Zusammenhänge in realen Auswahlprozessen die Stichprobengröße zu erhöhen. Darüber hinaus sollte in der weiterführenden Forschung ebenfalls über einen größeren Stichprobenumfang auf Grund der Power nachgedacht werden. In dieser Studie wurde das

Minimum an erforderlichen Stichprobenumfang gewählt, was jedoch zu einer geringeren Power sowie Aussagekraft der Ergebnisse führt.

Eine weitere Limitation ist die Kriteriumsvariable Potenzialaussage Fremdeinschätzung. Die Variable hatte einen sehr niedrigen Mittelwert und eine geringe Streuung, was dafürspricht, dass alle Teilnehmenden sehr gut von der befragten Person eingeschätzt wurden. Dies könnte die Zusammenhänge verzerrt haben. Die verzerrte Antworttendenz würde bedeuten, dass die Hypothese nicht interpretiert werden kann und somit keine Schlüsse für die Verwendung der Dialogübung in der Praxis geschlossen werden können. Nachfolgend sollte sich die Wissenschaft die Frage stellen, welche Fremdeinschätzungen objektiv und geeignet für die prädiktive Validität sein könnten. Darüber hinaus sollte ggf. der Kontakt zu den Personen der Fremdeinschätzungen gesucht werden, um ihnen die Relevanz der Bewertung genauer zu erklären und somit eventuell eine objektivere Bewertung hervorzurufen. Da die Teilnehmenden die Personen selbst aussuchen konnten, könnte dies ebenfalls zu Verzerrungen der Fremdeinschätzungen geführt haben.

Sämtliche Zusammenhänge zwischen der Gerechtigkeitswahrnehmung bzw. dem Arbeitsverhalten und der Persönlichkeit wurden ausschließlich in den vier berufsbezogenen Aspekten betrachtet. Dies wich von der Forschung von Tett & Burnett (2003) sowie von Moldzio (2014) ab, welche die übergeordneten Big Five Dimensionen betrachteten. Aus diesem Grund könnte es sein, dass die anderen sechs Aspekten bzw. die fünf übergeordneten Dimensionen einen stärkeren Zusammenhang aufweisen. Dies müsste nachfolgend überprüft werden.

Die Teilnehmenden haben sich freiwillig für diese Studie gemeldet, weshalb die Ergebnisse beeinflusst worden sein und zu einer selektiven Stichprobe geführt haben könnten. Es kann neben der sehr hohen Fremdeinschätzung auch zu einer Verzerrung in den Persönlichkeitsmerkmalen gekommen sein. Ebenso könnte dies zur Folge haben, dass nur die Personen mit einem hohen Wert im Berufserfolg teilgenommen haben. Viele Teilnehmende erklärten der Testleiterin, dass sie schon einmal eine Dialogübung durchliefen und ggf. schon an einer Schulung zur Kommunikationsfähigkeit teilnahmen. Dies könnte die Leistung während der Dialogübung verbessert haben und dann zu einer besseren Bewertung durch die Beobachtenden geführt haben. Aus diesem Grund ist es für die weitere Forschung entscheidend, die Zusammenhänge in einem realen Auswahlverfahren zu betrachten, um Stichprobeneffekte zu minimieren. Eine weitere Limitation könnte auch die schwache Invarianz der Stichprobengruppen sein, welche für einen Varianzunterschied in Bezug auf die Konstanten der latenten Variablen zwischen den beiden Stichprobengruppen (virtuell versus

Präsens) spricht (Werner, 2012). Die Stichprobenunterschiede können die Ergebnisse der Strukturgleichungsmodelle verzerren und somit eine Replizierbarkeit herabsetzen. Aus diesem Grund muss nachfolgend überprüft werden, inwiefern die Stichprobenunterschiede einen Einfluss auf die Ergebnisse hatten. Erst dann kann eine fundierte Aussage über die Interpretierbarkeit der Ergebnisse getroffen werden.

Implikation und weitere Forschung

Die Studie konnte zeigen, dass Unterschiede in der prädiktiven Validität in den beiden Darbietungsarten zu Gunsten der Dialogübung vor Ort bestehen. Dies bedeutet für die Praxis, dass der Einsatz der Dialogübung vor Ort empfehlenswert ist und die Darbietungsarten nicht innerhalb eines Auswahlverfahrens gewechselt werden sollten, da es dadurch zur Verzerrung der Entscheidung kommen könnte. Diese These wird auch von dem Ergebnis, dass die Teilnehmenden in der virtuellen Dialogübung von den Beobachtenden besser bewertet wurden als vor Ort, gestützt. Die Unterschiede waren nicht sehr groß, dennoch sollten sie in weiteren Forschungsarbeiten und in der Praxis berücksichtigt werden. Außerdem sollten noch die Zusammenhänge mit den anderen sechs Aspekten bzw. mit den fünf übergeordneten Big Five Dimensionen betrachtet werden, da dort vielleicht größere Unterschiede in der Vorhersage von dem Arbeitsverhalten entstehen könnten. Darüber hinaus könnte es der Praxis dienen, wenn die Wissenschaft überprüfen würde, ob die Zusammenhänge in allen Dialogübungen bestehen oder beispielsweise in einem Verkaufsgespräch die Unterschiede zwischen den Darbietungsarten in der prädiktiven Validität nicht bestehen.

Diese Studie fand darüber hinaus keine Zusammenhänge zwischen den Aspekten und der Gerechtigkeitswahrnehmung sowie keine Unterschiede in der Gerechtigkeitswahrnehmung zwischen den Gruppen. Dies würde wiederum für den Einsatz der virtuellen Dialogübung sprechen. Diese unterschiedlichen Empfehlungen zeigen, dass die virtuelle Dialogübung weiter untersucht werden sollte, um der Praxis eine genauere Empfehlung zum Einsatz der Dialogübung zu liefern. Darüber hinaus scheinen die Aspekte der Extraversion und Offenheit für Erfahrung nicht mit der Gerechtigkeitswahrnehmung zusammenzuhängen. Dies würde für die Praxis bedeuten, dass in Bezug auf die Gerechtigkeitswahrnehmung die Aspekte nicht relevant sind und nicht berücksichtigt werden müssen. Angelehnt an die Forschungsarbeit von Konradt et al. (2020) würden die Ergebnisse dieser Studie darauf hindeuten, dass die Gerechtigkeitswahrnehmung der Bewerbenden in beiden Darbietungsarten gleichermaßen über den Prozess hinweg abnimmt. Wäre die virtuelle Dialogübung schlechter bewertet worden, hätte dies wahrscheinlich eine stärkere

Abnahme der Gerechtigkeitswahrnehmung innerhalb des Auswahlprozesses zur Folge. Die Zusammenhänge zwischen der virtuellen Dialogübung und der Gerechtigkeitswahrnehmung sollten längsschnittlich über einen Auswahlprozess hinweg untersucht werden. Dennoch sollte die Wissenschaft nachfolgend die Zusammenhänge der Persönlichkeit und der Gerechtigkeitswahrnehmung sowohl in einer Feldstudie als auch anhand der übergeordneten Big Five Dimensionen betrachten.

Die Histogramme zur Wichtigkeit der einzelnen Gerechtigkeitsregeln nach Gilliland (1993) zeigten erste Anhaltspunkte, dass vor allem interpersonale Aspekte für die Gerechtigkeitswahrnehmung wichtig sind. Um gut auf die Ansprüche der Bewerbenden reagieren zu können, sollten vor allem die Beobachtenden sowie Recruiter:innen über gut ausgeprägte Kommunikationsfähigkeiten verfügen und u.a. zu Beobachtungsfehlern qualifiziert werden. Demgegenüber scheint den Teilnehmenden nicht der inhaltliche Berufsbezug der Aufgaben wichtig zu sein. Dies würde für die Praxis bedeuten, dass die Aufgaben nicht bis ins kleinste Detail (z.B. virtueller Bezug im Übungskontext oder nicht) auf die vakante Stelle angepasst werden müssen.

Generelle Diskussion

Die beiden Studien hatten zum Ziel, zwei verschiedene eignungsdiagnostische Methoden zu validieren. Die Entwicklung des berufsbezogenen Fragebogens zu den Aspekten Enthusiasmus, Durchsetzungsfähigkeit, Offenheit und Intellekt von DeYoung et al. (2007) wurde gewählt, da sich die Arbeitswelt stetig weiterentwickelt und durch die Digitalisierung vor allem die Offenheit für Erfahrung bedeutsamer wird (Schermuly et al., 2019). Die Studie zeigte, dass der Fragebogen mit 30 Items reliabel ist und eine faktorielle Validität aufweist. Das nomologische Netzwerk war nicht vollständig gegeben, was jedoch auch auf den Berufsbezug zurückzuführen sein könnte, da ggf. die Aspekte und Big Five Dimensionen doch stärker korrelieren. Dies müsste nachfolgend erforscht werden. Die Kriteriumsvalidität zeigte, dass in verschiedenen Stichprobengruppen (Führungskräfte, Expert:innen, Auszubildende kaufmännisch und technisch) verschiedene Kriterien unterschiedlich stark mit den vier verschiedenen Aspekten zusammenhängen. Dies bedeutet, dass nachfolgend noch weiter die Kriteriumsvalidität in den einzelnen Stichprobengruppen untersucht werden sollte, um spezifische Aussagen pro Stichprobe treffen und der Praxis bessere Einsatzempfehlungen geben zu können. Ersichtlich war jedoch, dass vor allem der Aspekt Intellekt viele Kriterien vorhersagte, was die Bedeutung dieses Aspekts für die aktuelle Arbeitswelt unterstreichen würde (Schermuly et al., 2019). Da die Kriteriumsdaten in der Praxis entwickelt und erhoben wurden, führten sie teilweise zu Verzerrungen und

besaßen eine sehr hohe Fehlervarianz. Dies könnte durch beispielsweise subjektive Einschätzungen der Beobachtenden entstanden sein. Daher sollte in weiteren Studien darauf geachtet werden, dass objektive Kriteriumsdaten (zum Beispiel Gehalt) erhoben und die Kriteriumsvalidität weiter überprüft werden. Insgesamt lässt sich sagen, dass der Fragebogen eine gute Reliabilität aufweist und valide ist. Der Fragebogen weist über die Erfassung des schlussfolgernden Denkens und der Big Five Dimensionen hinaus inkrementelle Validität auf. Die Ergebnisse verdeutlichen, dass es in der Praxis empfehlenswert ist, diesen Fragebogen einzusetzen, da er zu Varianzaufklärung des Merkmals Berufserfolg in dieser Studie beigetragen hat. Dieser Fragebogen ermöglicht es angelehnt an die Forschung von Schermuly et al. (2019) zudem, für die Arbeitswelt relevante Persönlichkeitsmerkmale berufsbezogen zu erfassen. Der Einsatz des Fragebogens scheint in allen Berufsgruppen empfehlenswert zu sein, dies sollte jedoch nachfolgend weiter untersucht werden. Die Ergebnisse der Studie zeigen, dass neben der Forschungsarbeit von Judge et al. (2013) diese Arbeit die Struktur der Aspekte nach DeYoung et al. (2007) replizieren konnte und die Forschung zukünftig eine spezifischere Erfassung der Persönlichkeitsmerkmale durch die Aspekte berücksichtigen sollte. Neben der Validität sollte nachfolgend die Gerechtigkeitswahrnehmung der Bewerbenden in Bezug auf den Fragebogen überprüft werden. Dabei steht die Frage im Vordergrund, ob die Vermutung von Beermann et al. (2013), dass berufsbezogene Fragebögen gerechter wahrgenommen werden, auf diesen Fragebogen repliziert werden kann.

Die zweite Studie betrachtete die prädiktive Validität und die Gerechtigkeitswahrnehmung einer virtuellen Dialogübung. Die Ergebnisse zeigten, dass die prädiktive Validität in virtuellen Dialogübungen in der Vorhersage von Berufserfolg durch das Arbeitsverhalten geringer ist als in der Gruppe in Präsenz. Darüber hinaus sagt die Persönlichkeit (erfasst durch die vier oben genannten Aspekte) das Arbeitsverhaltens in der Dialogübung vor Ort besser vorher. Diese Ergebnisse verdeutlichen die Aussagen des Job Performance Modell von Tett und Burnett (2003), dass die unterschiedlichen Darbietungsarten Einfluss auf die Zusammenhänge zwischen der Persönlichkeit, dem Arbeitsverhalten und dem Berufserfolg besitzen. Dies würde für die Praxis bedeuten, dass der Einsatz von Dialogübungen vor Ort empfehlenswerter ist als der von virtuellen Dialogübungen. Sollte sich ein Unternehmen dennoch für die virtuelle Dialogübung entscheiden, dann muss es sich der Validitätseinschränkungen bewusst sein. Diese Studie stellt nur einen kleinen Teil der Zusammenhänge in einem experimentellen Setting dar. Aus diesem Grund sollten virtuelle Dialogübungen in Bezug auf eine andere Erfassung der

Persönlichkeit, dem Arbeitsverhalten und Berufserfolg erforscht werden. Darüber hinaus sollten auch andere Dialogübungskontexte (Verkaufsgespräch, Überzeugungsgespräch etc.) betrachtet werden. Interessant für die Praxis könnte die Frage sein, ob die Zusammenhänge in allen Stichproben bzw. Berufsgruppen bestehen oder die virtuelle Dialogübung in manchen Berufsgruppen (zum Beispiel Führungskräfte) valider ist.

Neben der prädiktiven Validität wurde auch die Gerechtigkeitswahrnehmung der Bewerbenden in der Dialogübung betrachtet. Die Ergebnisse fanden keinen Unterschied in der Vorhersage der Gerechtigkeitswahrnehmung der Bewerbenden durch die vier Aspekte der Extraversion und Offenheit für Erfahrung in den beiden Darbietungsarten. Daraus kann für die Praxis abgeleitet werden, dass die Gerechtigkeitswahrnehmung nicht von den vier Persönlichkeitsaspekten abhängt. Dies würde die Personalauswahl vereinfachen, da sie in Bezug auf die Gerechtigkeitswahrnehmung unabhängig von der Persönlichkeit eines Bewerbers konzipiert werden könnte. Dennoch sollte die Forschung weiterhin überprüfen, ob die Zusammenhänge in einem realen Auswahlverfahren ebenfalls bestehen und welche Zusammenhänge in den anderen Persönlichkeitsdimensionen bzw. Aspekten zur Gerechtigkeitswahrnehmung bestehen. Darüber hinaus war ersichtlich, dass die Gerechtigkeitswahrnehmung in beiden Dialogübungen nicht signifikant verschieden war, was für den Einsatz von virtuellen Dialogübungen in Bezug auf die Gerechtigkeitswahrnehmung sprechen würde. Angelehnt an die Forschung von Konradt et al. (2020) sollte die Gerechtigkeitswahrnehmung in beiden Dialogübungen über den Prozess hinweg nachfolgend betrachtet werden, um eine Aussage über die unterschiedliche Abnahme der Gerechtigkeitswahrnehmung in der virtuellen Dialogübung und der Dialogübung in Präsenz treffen zu können. Darüber hinaus sollte auch betrachtet werden, ob die Zusammenhänge der Gerechtigkeitswahrnehmung in Feldstudien nach der Forschungsarbeit von Truxillo et al. (2009) stärker ausfallen. Diese Arbeit konnte erste Muster in der Wichtigkeit einzelner Gerechtigkeitsregeln (Gilliland, 1993) für die Teilnehmenden finden. Aufbauend auf diese Erkenntnisse sollte in weiteren Forschungsarbeiten das Thema weitergehend betrachtet werden, um für Unternehmen bzw. Praktiker:innen weitere Handlungsempfehlungen erarbeiten zu können. Diese Handlungsanweisungen könnten die Praktiker:innen dahingehend unterstützen, zu beurteilen, auf welche Gerechtigkeitsregeln zumeist in der Konzipierung eines Verfahrens geachtet werden sollte und welche vernachlässigt werden könnten.

Literaturverzeichnis

Allport, F. H. & Allport, G. W. (1921). Personality Traits: Their Classification and Measurement. *The Journal of Abnormal Psychology and Social Psychology*, *16*(1), 1–60. https://doi.org/10.1037/h0069790

Allport, G. W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), 1–171. https://doi.org/10.1037/h0093360

Anderson, N. R., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive metaanalysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, *18*, 291–304. https://doi.org/10.1111/j.1468-2389.2010.00512.x

Armoneit, C. (2019). *Trendstudie zur Nutzung, Validität, Praktikabilität und Akzeptanz psychologischer Personalauswahlverfahren in Deutschland* [Masterarbeit, Hochschule für Angewandte Psychologie FHNW]. https://doi.org/10.26041/fhnw-1722

Ashton, M. C., & Lee, K. (2002). Six independent factors of personality variation: A response to Saucier. *European Journal of Personality*, *16*(1), 63-75. https://doi.org/10.1002/per.433

Backhaus, K., Erichson, B., & Weiber, R. (2015). *Fortgeschrittene Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (3. Aufl.). Springer Gabler.

Bagozzi, R. P. (1981). Evaluating structural equation models with unobservable variables and measurement error: A comment. *Journal of Marketing Research*, *18*, 375–381. https://doi.org/10.1177/002224378101800312

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94. https://doi.org/10.1007/BF02723327

Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Basch, J. M. & Melchers, K. G. (2020). Technologie-mediierte Einstellungsinterviews: Ein Überblick über Befunde und offene Fragen. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, *51*(1), 71–79. https://doi.org/10.1007/s11612-020-00497-y

Basch, J. M., Melchers, K. G., Kurz, A., Krieger, M. & Miller, L. (2020). It takes more than a good camera: Which factors contribute to differences between face-to-face interviews and videoconference interviews regarding performance ratings and interviewee perceptions? *Journal of Business and Psychology*, *36*(5), 921–940. https://doi.org/10.1007/s10869-020-09714-3

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: development of the selection procedural justice scale (SPJS). *Personnel Psychology*, 54, 387-420.

Beauducel, A. & Leue, A. (2014). *Psychologische Diagnostik* (2. Aufl.). Hogrefe Verlag.

Beermann, D., Kersting, M., Stegt, S. & Zimmerhofer, A. (2013). Vorurteile und Urteile zur Akzeptanz von Persönlichkeitsfragebogen. *PersonalQuarterly. 65*(1), 41-45.

Bender, R. & Lange, S. (2001). Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology. 54*(4). 343-349. https://doi.org/10.1016/S0895-4356(00)00314-0

Bierhoff, H-W. (2021). *Kommunikation – Dorsch - Lexikon der Psychologie.* https://dorsch.hogrefe.com/stichwort/kommunikation#search=aa5c8a2abf23382e8c2d0e3313fe3367&offset=1

Blacksmith, N., Willford, J. & Behrend, T. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions*, *2*(1), 12–20. https://doi.org/10.25035/pad.2016.002

Blickle, G. (2014). Personalauswahl. In Nerning, F. W., Blickle, G. & Schaper, N. (Hrsg.), *Arbeits- und Organisationspsychologie* (3. Auflage, S. 241-318). Springer-Verlag. https://doi.org/10.1007/978-3-662-56666-4_17

Borkenau, P. & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und Mc Crae* (2. Aufl.). Hogrefe.

Böge, M. (2016). *Determinanten und Auswirkungen von Gerechtigkeitswahrnehmungen im Verlauf eines Personalauswahlprozesses für Ausbildungsplätze* [Dissertation, Christian-Albrechts-Universität zu Kiel]. https://macau.uni-kiel.de/receive/diss_mods_00020295

Brandt, H. (2020). Explorative Faktorenanalyse (EFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 575-614). Springer-Verlag.

Brandt, H. & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 39 – 66). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_3

Brenner, F., Ortner, T., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. *Frontiers in Psychology*. *https://doi.org/10.3389/fpsyg.2016.00863*

Brislin, R. W. (1980). Translation and content analysis of oral and written material. *Handbook of cross-cultural psychology, 2*(2)*,* 349-444.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. Aufl.). Guilford Press.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.)*.* Pearson*.*

Bühner, M. (2021). *Einführung in die Test-und Fragebogenkonstruktion* (4. Aufl.). Pearson.

Bundesregierung. (2022). *Corona-Schutz am Arbeitsplatz: Das sind die aktuellen Regeln.* https://www.bundesregierung.de/breg-de/themen/coronavirus/infektionsschutz-arbeitsplatz-1983894#:~:text=Arbeitgeber%20m%C3%BCssen%20bei%20B%C3%BCroarbeiten%20oder,soweit%20ihrerseits%20keine%20Gr%C3%BCnde%20entgegenstehen.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Cattell, R. B. (1946). *The description and measurement of personality.* World Book.

Chin, W.W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Hrsg.), *Modern methods for business research* (1.Aufl.*,* S. 295-336). Jossey-Bass.

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Does the medium matter? No. *Journal of Research in Personality*, *40*(4), 359–376. https://doi.org/10.1016/j.jrp.2005.01.006

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Lawrence Erlbaum.

Costa, Paul T.; McCrae, Robert R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I.

Cronbach, L. J., & Gleser, C. G. (1957). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press. https://doi.org/10.1002/bimj.19660080311

Cropanzano, R., Rupp, D. E., Mohler C. J. & Schminke, M. (2001). Three roads to organizational justice. *Research in Personnel and Human Resources Management,* 20, 1-113. http://dx.doi.org/10.1016/s0742-7301(01)20001-2

Daft, R.L., & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571. https://doi.org/10.1287/mnsc.32.5.554

Day, D. V., & Bedeian, A. G. (1991). Predicting job performance across organizations: The 1nteraction of work orientation and psychological climate. *Journal of Management, 17*(3), 589–600. https://doi.org/10.1177/014920639101700304

Dennis, A. & Valacich, J. (1999). Rethinking media richness: towards a theory of media synchronicity. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. https://doi.org/10.1109/hicss.1999.772701

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/ intellect: cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality*, *73*(4), 825–858. https://doi.org/10.1111/j.1467-6494.2005.00330.x

DeYoung, C. G., Quilty, L. C. & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

DeYoung, C. G., Quilty, L. C., Peterson, J. B. & Gray, J. R. (2013). Openness to Experience, Intellect, and Cognitive Ability. *Journal of Personality Assessment*, *96*(1), 46–52. https://doi.org/10.1080/00223891.2013.806327

Ellwart, T., Jaster, C., & Peiffer, H. (2018). MotivSORT. Entwicklung eines Instruments zum Screening individueller Motiverfüllung im Handwerk. *ZPID (Leibniz Institute for Psychology Information)*. https://doi.org/10.23668/psycharchives.916

Eysenck, H. J. (1967). *The biological basis of personality*. Transaction Publishers.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.

Fellner, K. (2019). *Moderne Personalauswahl: Renommierte Experten über Trends, neue Technologien, Chancen und Risiken in der Eignungsdiagnostik* (1. Aufl.). Springer. https://doi.org/10.1007/978-3-658-25897-9

Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C. & Frink, D. D. (2005). Development and validation of the political skill inventory. *Journal of Management*, *31*(1), 126–152. https://doi.org/10.1177/0149206304271386

Gäde, J. C., Schermelleh-Engel, K. & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 615–660). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_24

Gäde, J. C., Schermelleh-Engel, K. & Werner, C. S. (2020). Klassische Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 307-368). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_14

Geister, S. & Rastetter, D. (2009). Aktueller Stand zum Thema Online-Tests. In H. Steiner (Hrsg.), *Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (1. Aufl.,S. 3–16). Springer Medizin Verlag. https://doi.org/10.1007/978-3-540-78919-2_1

Gentil, A. (2019). *May the situation be with you: Is agreeableness a major predictor for leader potential in different situations?* EAWOP-Conference. Turin, 2019.

George, D. & Mallery, P. (2002). *SPSS for Windows step by step: A simple guide and reference*. 11.0 upgrade. Allyn & Bacon.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, *18*(4), 694–734. https://doi.org/10.5465/amr.1993.9402210155

Goldberg, L. R. (1990). An alternative „description of personality“: The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Häcker, H. O. (2016). *Fähigkeit – Dorsch - Lexikon der Psychologie*. https://dorsch.hogrefe.com/stichwort/faehigkeit

Häcker, H. & Stapf, K. H. (1998). *Dorsch Psychologisches Wörterbuch* (13. Aufl.). Verlag Hans Huber.

Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta□analysis. *Personnel Psychology*, *57*(3), 639–683. https://doi.org/10.1111/j.1744-6570.2004.00003.x

Hermann, T. (1976). *Lehrbuch der empirischen Persönlichkeitsforschung*. Hogrefe.

Hertel, G., Konradt, U. & Orlikowski, B. (2003). Ziele und Strategien von E-Assessment aus Sicht der psychologischen Personalauswahl. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment: Rekrutierung, Auswahl und Beurteilung von Personal im Inter-und Intranet* (1. Aufl., S. 31 - 44). Hogrefe Verlag.

Hertel, G., Schroer, J., Batinic, B., & Naumann, S. (2008). Do shy people prefer to send e-mail?: Personality effects on communication media preferences in threatening and nonthreatening situations. *Social Psychology, 39*(4), 231–243. https://doi.org/10.1027/1864-9335.39.4.231

Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1-28. https://doi.org/10.18637/jss.v042.i08

Höft, S. & Obermann, C. (2010). Der Praxiseinsatz von Assessment Centern im deutschsprachigen Raum: Eine zeitliche Verlaufsanalyse basierend auf den Anwenderbefragungen des Arbeitskreises Assessment Center e.V. von 2001 und 2008. *Wirtschaftspsychologie, 12*(2), 5-16.

Hossiep, R. & Krüger, C. (2012). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – 6 Faktoren* (1. Aufl.). Hogrefe.

Hossiep, R. & Paschen, M. (1998). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung* (1. Aufl.). Hogrefe.

Hossiep, R., Schecke, J. & Weiß, S. (2015). Zum Einsatz von persönlichkeitsorientierten Fragebogen: Eine Erhebung unter den 580 größten deutschen Unternehmen. *Psychologische Rundschau*, *66*(2), 127–129. https://doi.org/10.1026/0033-3042/a000235

IBM Corp. (2016). *IBM SPSS Statistics for Windows*, Version 24.0. IBM Corp.

Isenschmid, J. (2013). *Führen – In der Einfachheit liegt die Stärke*. Springer Gabler. https://doi.org/10.1007/978-3-658-00617-4_2

Janczyk, M. & Pfister, R. (2020). Fehlertypen, Effektstärken und Power. In M. Janczyk, & R. Pfister (Hrsg.), *Inferenzstatistik verstehen* (3. Aufl., S. 81-97). Springer Spektrum. https://doi.org/10.1007/978-3-662-59909-9_7

Jang, K. L., Livesley, W. J., Angleitner, A., Reimann, R., & Vernon, P. A. (2002). Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and Individual Differences*, *33*(1), 83–101. https://doi.org/10.1016/S0191-8869(01)00137-4

Johns G. (1999). A multi-level theory of self-serving behavior in and by organizations. *Research in Organizational Behavior*, *21*, 1–38.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M. & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling* (R package version 0.5-6.) [Computer Software]. https://CRAN.R-project.org/package=semTools

Joubert, T. & Kriek, H. J. (2009). Psychometric comparison of paper-and-pencil and online personality assessment in a selection setting. *SA Journal of Industrial Psychology*, *35*(1), 78-88. https://hdl.handle.net/10520/EJC89184

Judge, T. A., Cable, D. M., Boudreau, J. W. & Bretz, R. D. (1995). An empirical investigation of the predictors of executive career success. *Personnel Psychology*, *48*(3), 485–519. https://doi.org/10.1111/j.1744-6570.1995.tb01767.x

Judge, T. A., Higgins, C. A., Thoresen, C. J. & Barrick, M. R. (1999). The big five personality traits, general mental ability and career success across the life span. *Personnel Psychology*, *52*(3), 621–652. https://doi.org/10.1111/j.1744-6570.1999.tb00174.x

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S. & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *The Journal of Applied Psychology*, *98*(6), 75–925. https://doi.org/10.1037/a0033901

Kanning, U. P. (2022). Forschung und Praxis in der Personalpsychologie – Plädoyer für eine evidenzbasierte Personalarbeit. *Report Psychologie*, *47*(4), 5-8. http://www.report-psychologie.de/

Kelava, A. & Moosbrugger, H. (2020). Deskriptivstatistische Itemanalyse und Testwertbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 143-158). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_7

Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, *42*(2), 61-75.

Kersting, M. (2006). Stand, Herausforderungen und Perspektiven der Managementdiagnostik. *Personalführung*, *10*(1), 16-27.

Kersting, M. (2015). Eignungsdiagnostik in Zeiten des Personalmangels: Bewährte Praxis und neue Akzente. In L. Gooßens, M. Kersting & S. Koch (Hrsg.). *Auf die richtigen Mitarbeiter kommt es an - Eignungsdiagnostik und ihre Anwendung in der Sparkassen-Finanzgruppe* (1. Aufl., S. 21 – 37). Deutscher Sparkassen Verlag.

Kersting, M. (2018). Was König Bewerber denkt. Zur Akzeptanz von Personalauswahlverfahren. *Personalmagazin*, *3*(1), 26-29.

Kersting, M. & Ziegler, M. (2020). Same same but different. Eignungsdiagnostik auf Distanz. *Personalmagazin*, *8*(1), 34-40.

Kersting, M. (2021). Digitale Transformation. Trends und Herausforderungen in der Eignungsdiagnostik. *Personalführung*, *6*(1), 15-21.

Kirbach, C., Montel, C., Oenning, S. & Wottawa, H. (2004). *Recruiting und Assessment im Internet. Werkzeuge für eine optimierte Personalauswahl und Potenzialerkennung*. Vandenhoeck & Ruprecht.

König, C. J., Klehe, U. C., Berchtold, M. & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, *18*(1), 17-27. https://doi.org/10.1111/j.1468-2389.2010.00485.x

Konradt, U., Garbers, Y., Erdogan, B. & Bauer, T. (2016). Patterns of change in fairness perceptions during the hiring process. *International Journal of Selection and Assessment*, *24*(3), 246-259. https://doi.org/10.1111/ijsa.12144

Konradt, U., Oldeweme, M., Krys, S. & Otte, K-P. (2020). A meta-analysis of change in applicants' perceptions of fairness. *International Journal of Selection and Assessment*, *28*(4), 365-382. https://doi.org/10.1111/ijsa.12305

Konradt, U. & Sarges, W. (2003). Einleitung. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment: Rekrutierung, Auswahl und Beurteilung von Personal im Inter-und Intranet* (1. Aufl., S. 5 - 15). Hogrefe Verlag.

Lawrence, A. D., Quist, J. S., & O'Connell, M. S. (2009). *Unproctored internet testing: Examining the impact of test environment*. In 24th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testkonstruktion* (6. Aufl.). Beltz.

Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R* (2. Aufl.). Hogrefe.

Liepmann, D., Beauducel,A. Brocke, B. & Nettelnstroth, W. (2012). *Intelligenz-Struktur-Test – Screening* (1. Aufl.). Hogrefe.

Lounsbury, J. W., Loveland, J. M., Sundstrom, E. D., Gibson, L. W., Drost, A. W. & Hamrick, F. L. (2003). An investigation of personality traits in relation to career satisfaction. *Journal of Career Assessment*, *11*(3), 287–307. https://doi.org/10.1177/1069072703254501

Meade, A. W., Michels, L. C. & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, *10*(2), 322-345. https://doi.org/10.1177/1094428106289393

Moldzio, T. (2014). *Akzeptanz von Auswahlverfahren aus Bewerbersicht: Einflussfaktoren und Auswirkungen*. Masuhr.

Moldzio, T., Böge, M. &Wedemeyer, P.S. (in Vorbereitung). *Arbeitsbezogene Verträglichkeitsskalen (AVS)*. Hogrefe.

Moldzio, T., Peiffer, H., Dreier, K., Gergovska, T., Reiner, A. & Felfe, J. (2019). *Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen* (1. Aufl.). Hogrefe.

Moldzio, T., Peiffer, H., Wedemeyer, P. S., & Gentil, A. (2021). Differentiated measurement of conscientiousness and emotional stability in an occupational context–greater effort or greater benefit? *European Journal of Work and Organizational Psychology*, 1-14. https://doi.org/10.1080/1359432X.2020.1866066

Moldzio, T., Wedemeyer, P.S. & Böge, M. (in Vorbereitung). *Business Big 5 (BB5)*. Hogrefe.

Moosbrugger, H. & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 15-38). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_2

Murphy, K. R.; Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*(4), 343–357. https://doi.org/10.1207/s15327043hup1804_2

Ng, T. W.H., Eby, Lillian T., Sorensen, K. L. & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, *58*(2), 367–408. https://doi.org/10.1111/j.1744-6570.2005.00515.x

Oh, I.-S. (2009). *The five factor model of personality and job performance in east Asia: A cross-cultural validity generalization study* [Dissertation, University of Iowa].

Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae* (revidierte Fassung). Hogrefe.

Ott, M., Ulfert, A. & Kersting, M. (2017). „Online-Assessments“ und „Self-Assessments“ in der Eignungsdiagnostik. In: D.E. Krause (Hrsg.), *Personalauswahl* (1. Aufl., S. 215-242). Springer Fachmedien. https://doi.org/10.1007/978-3-658-14567-5_10

Petri, P., Bianucci, D. & Kersting, M. (2019). *Alles online – Akzeptanzurteile von Bewerber*innen zum Einsatz unbeaufsichtigter Eignungstests*. Tagung der Fachgruppe „Arbeits-, Organisations- und Wirtschaftspsychologie“, Braunschweig, September 2019.

Randolph, J. J., Falbe, K., Manuel, A. K. & Balloun, J. L. (2014). A step-by step guide to propensity score matching in R. *Practical Assessment, Research & Evaluation*, *19*(18), S. 1-6. https://doi.org/10.7275/n3pv-tx27

Rehrl, M., Harteis, C. & Gruber, H. (2006). Potentialanalysen in der Personalentwicklung: Ein kritischer Diskurs, *German Journal of Human Resource Management: Zeitschrift für Personalforschung*, *20*(2), 185-191. https://doi.org/10.1177/239700220602000207

Revelle, W. (2022). *psych: Procedures for personality and psychological research (2.2.5)* [Computer Software]. https://CRAN.R-project.org/package=psych Version = 2.2.5.

Rosseel, Y. (2012). *lavaan: An R package for structural equation modeling 48*(2), 1–36. [Computer Software]. https://CRAN.R-project.org/package=psych Version = 2.2.5.

Rothstein, M. G., & Jelley, R. B. (2003). The challenge of aggregating studies of personality. In K. R. Murphy (Hrsg.), *Validity generalization: A critical review* (1. Aufl., S. 223–262). Lawrence Erlbaum Associates Publishers.

RStudio Team (2020). *RStudio: Integrated development environment for R*. RStudio, PBC. [Computer Software]. http://www.rstudio.com/

Ryan, A. M., & Ployhart, R. E. (2000). Applicants’ perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*(3), 565-606. https://doi.org/10.1016/S0149-2063(00)00041-6

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Salgado, J. F. (1997). The five factor model of personality and job performance in the european community. *The Journal of Applied Psychology*, *82*(1), 30–43. https://doi.org/10.1037/0021-9010.82.1.30

Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B. & Berges, A. (2014). Validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work an Organizational Psychology*, *24*(3), 325-349. https://doi.org/10.1080/1359432x.2014.903241

Schermelleh-Engel, K., Geiser, C. & Burns, L. (2020). Multitrait-Multimethod-Analysen (MTMM-Analysen). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 661-686). Springer-Verlag. https://doi.org/10.1007/978-3-662-61532-4_25

Schermuly, C. C., Arlt, R. & Geissler, C. (2019). Welche Kompetenzen erfordert agiles Arbeiten? *Human Resources Manager*, *19*(4), 72-74.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262-274. https://doi.org/10.1037/0033-2909.124.2.262

Schmithüsen, F. & Krampen, G. (2015). Persönlichkeitspsychologie. In F. Schmithüsen (Hrsg.), *Lernskript Psychologie: Die Grundfächer kompakt* (1. Aufl., S. 287-314). Springer. https://doi.org/10.1007/978-3-662-44941-7

Schmitt, M. (2019). *Gerechtigkeit, Gerechtigkeitsprinzip – Dorsch - Lexikon der Psychologie*. https://dorsch.hogrefe.com/stichwort/gerechtigkeit-gerechtigkeitsprinzip

Schönbrodt, F. D. & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schuler, H. (2000). *Psychologische Personalauswahl* (3.Aufl.). Verlag für Angewandte Psychologie.

Schuler, H. & Marcus, B. (2006). Biografieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 189-226). Hogrefe.

Schuler, H., Frier, D. & Kauffmann, M. (1993). *Personalauswahl im europäischen Vergleich*. Hogrefe.

Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen - Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, *6*(2), 60-70. https://doi.org/10.1026/1617-6391.6.2.60

Schuler, H. & Höft, S. (2006): Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 101-144). Hogrefe.

Schulz, C., Schuler, H. & Stehle, W. (1985). Die Verwendung eignungsdiagnostischer Methoden in deutschen Unternehmen. In H. Schuler & W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation* (1. Aufl., S. 126 - 132). Hogrefe.

Schyns, B. & von Collani, G. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology*, *11(*2), 219-241. https://doi.org/10.1080/13594320244000148

Seibert, S. E., Kraimer, M. L. (2001). The five-factor model of personality and career success. *Journal of Vocational Behavior*, *58*(1), 1–21. https://doi.org/10.1006/jvbe.2000.1757

Stange, K. (2013). *Angewandte Statistik: Erster Teil Eindimensionale Probleme*. Springer-Verlag.

Stone, C. A. & Tang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. *Practical Assessment, Research & Evaluation*, *18*(13), 1-12. https://doi.org/10.7275/qkqa-9k50

Tett, R. P. & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *The Journal of Applied Psychology*, *88*(3), 500–517. https://doi.org/10.1037/0021-9010.88.3.500

Tett, R. P., Jackson, D. N. & Rothstein (1991). Personality measures as predictors of job performance: A meta-Analytic review. *Personnel Psychology, 44*(4), 703–742. https://doi.org/10.1111/j.1744-6570.1991.tb00696.x

Truxillo, D. M., Bodner, T. B., Bertolino, M., Bauer, T. N., & Yonce, C. (2009). Effects of explanations on applicant reactions: A meta-analytic review. *International Journal of Selection and Assessment*, *17*(4), 346–361. https://doi.org/10.1111/j.1468-2389.2009.00478.x

Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: development of the selection procedural justice scale (SPJS). *Personnel Psychology*, *54*(2), 387-419. https://doi.org/10.1111/j.1744-6570.2001.tb00097.x

Urban, D., & Mayerl, J. (2013). *Strukturgleichungsmodellierung: Ein Ratgeber für die Praxis*. Springer-Verlag.

Van Aarde, N., Meiring, D. & Wiernik, B. M. (2017). The validity of the big five personality traits for job performance: Meta-analyses of south african studies. *Int J Select Assess*, *25*(3), 223–239. https://doi.org/10.1111/ijsa.12175

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*(2), 89-99. https://doi.org/10.1027/1016-9040.1.2.89

Van Vianen, A. E. M., Taris, R., Scholten, E. & Schinkel, S. (2004). Perceived fairness in personnel selection: determinants and outcomes in different stages of the assessment procedure. *International Journal of Selection and Assessment*, *12*(1-2), 149-159. https://doi.org/10.1111/j.0965-075X.2004.00270.x

Wedemeyer, P.S. & Moldzio, T. (in Vorbereitung). *Arbeitsbezogene Extraversions- und Offenheitsskalen (AEOS)*. Hogrefe.

Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Springer Gabler. https://doi.org/10.1007/978-3-642-35012-2

Weisberg, Y. J.; DeYoung, C. G. & Hirsh J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in Psychology*, *2*, 178. https://doi.org/10.3389/fpsyg.2011.00178

Wentura, D. & Pospeschill, M. (2015). *Multivariate Datenanalyse: Eine kompakte Einführung* (1. Aufl.), Springer.

Werner, C. (2012). *Parametertests und Modellvergleiche in Strukturgleichungsmodellen*. http://www.psychologie.uzh.ch/fachrichtungen/methoden/team/christina werner/sem/parametertests_modellvergleiche.pdf

West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In: R. H. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (1. Aufl., S. 56–75). Sage Publications.

West, S. G.; Taylor, A. B.; Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Hrsg.), *Handbook of structural equation modelling* (1. Aufl., S. 209–231). Guilford Press.

Westhoff, K., Hagemeister, C., Kersting, M., Lang, F., Moosbrugger, H., Reimann, G. & Stemmler, G. (2010). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3. Aufl.). Pabst.

Ziegler, M., & Bühner, M. (2012). *Grundlagen der psychologischen Diagnostik*. Springer-Verlag.

Anhang

Dialogübung

Stellen Sie sich vor, Sie leiten ein fünfköpfiges Projektteam in dem Unternehmen „DigaCom“. Sie haben für das Projekt die Verantwortung übertragen bekommen. Ihr Projektteam wurde vor drei Monaten gegründet und ist für die Umsetzung der Digitalisierung im Unternehmen zuständig.
Frau Müller ist ein Teil Ihres Projektteams und Sie kennen sie bereits aus einer vorherigen Zusammenarbeit als leistungsstarke Kollegin.
Nach den ersten vier Wochen der Zusammenarbeit stehen Sie Frau Müller jedoch recht kritisch gegenüber. Sie beobachten, dass ihr Einsatz bei der Arbeit nachgelassen hat. So haben Sie bemerkt, dass Frau Müller wiederholt zu den Teambesprechungen zu spät gekommen ist und abwesend wirkte.
Zudem scheint sich Frau Müller im Team nicht sehr kooperativ zu verhalten. Sie gibt nur wenig Informationen preis, woraufhin sich schon zwei Kollegen (m/w/d) aus dem Projektteam bei Ihnen beschwert haben.
Vor diesem Hintergrund führen Sie nun unmittelbar ein erstes Gespräch mit Frau Müller, um ihr Engagement zu steigern. Aufgrund Ihres engen Terminplans können Sie sich für dieses Gespräch lediglich ca. zehn Minuten Zeit nehmen.

Sie haben nun 15 Minuten Zeit, sich auf das Gespräch vorzubereiten.

Einschätzung der Beobachtenden

Code / Name der Teilnehmerin/ des Teilnehmers:

	Trifft überhaupt nicht zu	Trifft nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft zu	Trifft vollständig zu
Kommunikationsfähigkeit: Die Teilnehmerin/ der Teilnehmer …						
ist in der Lage, ihre/ seine Ideen, Meinung und Entscheidungen gezielt an ihr/ sein Gegenüber zu vermitteln	☐	☐	☐	☐	☐	☐
setzt Mimik, Gestik und die Stimme überzeugend ein.	☐	☐	☐	☐	☐	☐
drückt sich klar und verständlich aus.	☐	☐	☐	☐	☐	☐
hört aktiv zu und lässt den Gesprächspartner ausreden.	☐	☐	☐	☐	☐	☐
Gesamteinschätzung: Kommunikationsfähigkeit	☐	☐	☐	☐	☐	☐
Konfliktfähigkeit: Die Teilnehmerin/ der Teilnehmer …						
ist in der Lage, ihre/ seine eigenen Ziele für einen gemeinsamen Kompromiss hintenanzustellen.	☐	☐	☐	☐	☐	☐
zeigt sich unvoreingenommen gegenüber anderen Meinungen.	☐	☐	☐	☐	☐	☐
erkennt Konfliktpunkte und versucht diese zu lösen.	☐	☐	☐	☐	☐	☐
scheut sich nicht, Konflikte einzugehen.	☐	☐	☐	☐	☐	☐
Gesamteinschätzung: Konfliktfähigkeit	☐	☐	☐	☐	☐	☐
Durchsetzungsfähigkeit: Die Teilnehmerin/ der Teilnehmer…						
ist in der Lage, ihr/ sein Gegenüber mit Argumenten von ihrer/ seiner Meinung zu überzeugen.	☐	☐	☐	☐	☐	☐
ist auch in der Lage andere Meinungen anzunehmen und ihr/ sein Gegenüber nicht zu überreden.	☐	☐	☐	☐	☐	☐
teilt ihre/ seine Meinung auch bei eventuellen Widerständen mit.	☐	☐	☐	☐	☐	☐
gewinnt andere Personen für ihre/ seine Ideen.	☐	☐	☐	☐	☐	☐
Gesamteinschätzung: Durchsetzungsfähigkeit	☐	☐	☐	☐	☐	☐

Potenzialaussage: Die Teilnehmerin/ der Teilnehmer…	
hat das Potenzial für weitere berufliche Entwicklungs-schritte und benötigt keine begleitenden Fördermaßnahmen.	☐
hat das Potenzial für weitere berufliche Entwicklungs-schritte, benötigt jedoch noch Fördermaßnahmen.	☐
weist zum aktuellen Zeitpunkt noch nicht das Potenzial für weitere berufliche Entwicklungsschritte auf, das Potenzial könnte sich durch Fördermaßnahmen entwickeln.	☐
weist nicht das Potenzial für weitere berufliche Entwicklungsschritte auf.	☐

Profilbogen

Der unten angeführte Profilbogen wurde aufgrund des Copyrights der Unternehmensberatung vereinfacht ohne Merkmalsbeschreibungen dargestellt.

Fragebogenergebnisse im Rahmen des Workshops

Code: Datum:

Im Folgenden ist jeweils eine kurze Beschreibung der erfassten Merkmalsbereiche vorangestellt. Es erfolgt eine Darstellung der Ergebnisse mittels einer Eingruppierung in fünf Kategorien, die sich an der Systematik der Normalverteilung orientiert. Es gibt einen großen mittleren Bereich mit nach außen hin schmaler werdenden Kategorien. Die Ergebnisse werden eingruppiert in:

weit unterdurchschnittlich	**<<∅**	<∅	∅	>∅	>>∅
unterdurchschnittlich	<<∅	**<∅**	∅	>∅	>>∅
durchschnittlich	<<∅	<∅	**∅**	>∅	>>∅
überdurchschnittlich	<<∅	<∅	∅	**>∅**	>>∅
weit überdurchschnittlich	<<∅	<∅	∅	>∅	**>>∅**

Der Unternehmensberatung liegen Daten von über 12.000 Kandidaten aus eignungsdiagnostischen Untersuchungen vor. Als Referenzgruppen für dieses Verfahren dienen Daten von bis zu 2.500 Experten.

NEO-Fünf-Faktoren Inventar (NEO-FFI)

Das NEO-FFI ist ein Persönlichkeitsfragebogen, in dem man seine Ausprägungen in fünf wichtigen Eigenschaften beschreibt. Dabei handelt es sich um zeitlich überdauernde, stabile Merkmale, die unabhängig von kulturellen und sozialen Hintergründen sind.

- *Ausgeglichenheit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Extraversion:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Offenheit für Erfahrung:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Verträglichkeit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Gewissenhaftigkeit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS) und Arbeitsbezogene Verträglichkeitsskalen (AVS)

Die ABGS und AVS sind in einem Fragebogen zusammengeführte, arbeitsbezogene Selbstbeschreibungsinventare zur differenzierten Erfassung der Persönlichkeitsmerkmale Belastbarkeit, Gewissenhaftigkeit und Verträglichkeit. Die Aspekte Soziale Belastbarkeit und Dauerbelastbarkeit, Fleiß und Ordnung sowie Einfühlungsvermögen und Bescheidenheit sind für die meisten Berufsgruppen von Bedeutung.

- *Soziale Belastbarkeit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Dauerbelastbarkeit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Fleiß:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Ordnung:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Einfühlungsvermögen:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

- *Bescheidenheit:* Beschreibung

<<∅	<∅	∅	>∅	>>∅

Self-efficacy Scale (SE-kurz)

Die Skala zur Erfassung der Selbstwirksamkeit ist eine 6-Item-Skala, die in die ABGS integriert wurde.

- *Selbstwirksamkeit: Beschreibung*

<<∅	<∅	∅	>∅	>>∅