

2 Theoretischer Hintergrund

Im folgenden Kapitel wird zunächst eine Definition des Begriffs „Führung“ vorgenommen. Im Anschluss werden Aufgaben und Anforderungen der Führung dargestellt und der Begriff Führungserfolg erläutert. Ergänzend wird das Thema Führungspotenzial eingeführt. Eine darauf folgende Darstellung und historische Einordnung der unterschiedlichen Ansätze der Führungsforschung bilden den Rahmen für das methodische Vorgehen der vorliegenden Arbeit. Im zentralen Thema des Kapitels, der Führungsbeurteilung, geht es um das 360°-Verfahren, nachdem über eine Auseinandersetzung mit vorwissenschaftlicher Personbeurteilung auf dieses hingeführt wurde.

2.1 Führung

In der das Kapitel eröffnenden Definition des Begriffs „Führung“ werden die verschiedenen Führungsstil- und Tätigkeitsklassifizierungen berücksichtigt. Anschließend werden das Aufgabenfeld der Führung beschrieben und Führungserfolg sowie Führungspotenzial definiert.

2.1.1 Definition des Begriffs „Führung“

Dem Begriff „Führung“ wurden von verschiedenen Autoren zahlreiche Definitionen zugeordnet. Dabei beziehen sich die folgenden Ausführungen nur auf personale Führung. Indirekte Führung durch schriftlich fixierte Systeme, die den Einfluss einer Führungskraft substituieren, wie in Webers (1921) bürokratischem Organisationsmodell skizziert, wird von der Betrachtung ausgenommen.

Generell wird Führung in der Organisationsforschung als asymmetrischer Prozess einseitiger Einflussnahme vom Vorgesetzten auf den Unterstellten verstanden (Neuberger, 1995; von Rosenstiel, 2001). Weinert (1989) definiert Führung genauer als Einflussnahme auf Unterstellte durch die Interaktion von mindestens zwei Personen, in der über Kommunikationswege angestrebt wird, Ziele zu erreichen. Bezogen auf Führung im Organisationskontext betonen Wunderer und Grunwald (1980), dass die soziale Einflussnahme in einer strukturierten Arbeitssituation mit dem Ziel der Erfüllung gemeinsamer Aufgaben stattfindet. Der Führungsbegriff lässt sich anhand der Betrachtung von zwei Tätigkeitskategorien konkretisieren.

Wunderer (1997) differenziert zwischen personaler und struktureller Führung: Personale Führung ist direkt und vollzieht sich in der Interaktion mit Menschen. Strukturelle, indirekte Führung fokussiert vor allem die Schaffung der Rahmenbedingungen. Dies umfasst z. B. Arbeiten mit Zielsystemen, Festlegen von Verantwortungsbereichen, Delegation und zeitliche Planung von Aufgaben sowie strategische Arbeit.

Tabelle 1: Begriffsunterscheidung und Beschreibung Führung versus Management.

Autoren	Management	Führung
Zaleznik (1977)	„Manager“ Vorgehen sachlich distanziert und analytisch <i>Aufgaben:</i> Prozesssteuerung, Haupttätigkeiten u. a. Ressourcenplanung, Erfolgskontrolle, Maßnahmenumsetzungen	„Leader“ Fokus der Tätigkeit auf Mitarbeitern <i>Aufgaben:</i> Motivieren und Entwickeln der Mitarbeiter, Vermitteln von Sinnhaftigkeit der Geschäftsprozesse und Geschäftsentscheidungen durch Kommunikation
Bennis & Nanus (1985)	<i>Aufgaben:</i> Ausführung und Erfolgs-sicherung von Routinetätigkeiten	<i>Aufgaben:</i> Beschäftigung mit der Zukunft des Unternehmens, veränderungs-orientiertes Handeln
Kotter (1987)	<i>Aufgaben:</i> Formeller, wissenschaftlicher und gegenwartsorientierter Prozess	<i>Aufgaben:</i> Informeller, inspirierender und zukunftsorientierter Prozess

Anmerkung: eigene Zusammenstellung.

Eine ähnliche Kategorisierung findet sich in der Begriffstrennung von Führung und Management wieder. Seit Zaleznik (1977) die Dichotomie *Führer (Leader)* vs. *Manager* einführt, wird in der Literatur eine deutliche Unterscheidung zwischen beiden Begriffen getroffen (Sternberg, 2003). Der Führungsbegriff wird zur Beschreibung der Mitarbeiterführung und insbesondere zur Hervorhebung der interaktionellen Bereiche der Führung genutzt, während im Managementbegriff eher strukturelle und institutionelle Aspekte betont werden (Neuberger, 2002). Verschiedene Beschreibungen beider Begrifflichkeiten finden sich in Tabelle 1.

Die Beschreibung der jeweils typischen Verhaltensweisen von Führern und Managern (hier ausgenommen der Zukunftsbezug in der Führung) ebenso wie die Unterscheidung, die Wunderer (1997) trifft, ist der bekannten Einteilung von Führungsverhaltensweisen in die Stil Kategorien Mitarbeiter- versus Aufgabenorientierung (Likert, 1961) ähnlich. Eine vergleichbare Einteilung wird auch mit den Kategorien sozioemotionales und aufgabenbezogenes Verhalten (Bales & Hare, 1965; Bales & Slater, 1955) vorgenommen. Bales (1950) sowie Bales und Slater (1955) zeigten in Versuchen mit führerlosen Gruppen, dass sich zwei informelle Führer herausbilden, von denen die erste Führungsperson v. a. aufgabenbezogen handelt, während die zweite mehr auf soziale und emotionale Bedürfnisse der Gruppe eingeht (Bales, 1950; Bales & Slater, 1955).

Diese zwei Aspekte der Führung finden sich in den Tätigkeiten wieder, die im folgenden Abschnitt thematisiert werden. Die Beschreibung der Tätigkeiten zeigt, welche Verhaltensweisen unter Führungsverhalten zu verstehen sind und liefert die Grundlage zur Führungsbeurteilung. Die Betrachtung liefert zudem Hinweise, in welchen Situationen und bei welchen Tätigkeiten das Verhalten der Führungskraft durch die verschiedenen sie umgebenden Gruppen beobachtet werden kann.

2.1.2 Aufgaben der Führung

Generell ist die Aufgabe der Führungskraft das Steuern des Austauschs von Informationen, Gütern, Kapital und Dienstleistungen unter Berücksichtigung von und Einflussnahme auf externe Rahmenbedingungen (z. B. ökologischer, ökonomischer und struktureller Art) sowie interne Bereiche, wie das Personalwesen (Grunwald, 1995). Umfassende Beschreibungen von Führungsaufgaben geben Regnet (2003) und Wunderer (2003).

Eine der ersten empirisch begründeten, deskriptiven Klassifizierungen von Führungstätigkeiten stellte Mintzberg (1973) auf. Das Modell basiert auf Beobachtungen des Alltags von Vorstandsmitgliedern in fünf Organisationen und etablierte sich in der Führungsforschung. Mintzberg teilte Führungstätigkeiten in zehn verschiedene Rollen auf und ordnete diese drei Kategorien zu (Abbildung 1). Mehrere Studien konnten Mintzbergs Befunde replizieren (z. B. Bandiera, Guiso, Prat & Sadun, 2011).

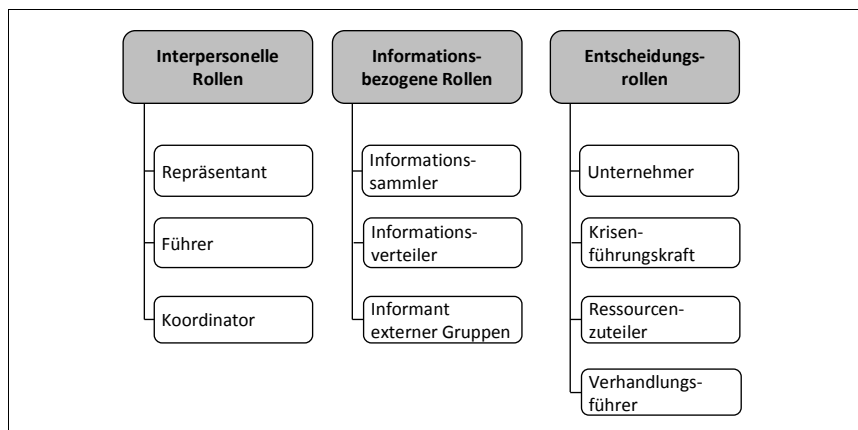


Abbildung 1: Mintzbergs (1973) Klassifizierung der Führungstätigkeiten

Interpersonelle Rollen umfassen das Repräsentieren der Organisation nach innen und außen (Rolle Repräsentant), Personalführungsaufgaben, wie das Einstellen, Anleiten, Entwickeln und Motivieren der Mitarbeiter (Rolle Führer) sowie den Aufbau von Kontakten und Netzwerken (Rolle Koordinator). *Informationsbezogene* Rollen beinhalten das Suchen und Verarbeiten von Information (Rolle Informationssammler), das Informieren der Mitarbeiter und anderer Stellen innerhalb der Organisation (Rolle Informationsverteiler) und das Informieren Dritter über Belange, wie Ergebnisse, Ziele oder Prozesse der Organisation (Rolle Informant externer Gruppen/Sprecher). *Entscheidungsrollen* umfassen alle Bemühungen, den wirtschaftlichen Erfolg der Organisation auszubauen, beispielsweise über die Initiierung von Innovationen (Rolle Unternehmer), das Beseitigen von Störungen (Rolle Krisenführungskraft), die Entscheidungsfindung zur Verteilung von Ressourcen zwischen Personen oder Einheiten (Rolle Ressourcenzuteiler) sowie das Verhandeln von Verträgen oder Plänen mit externen Parteien (Rolle Verhandlungsführer).

Die ausführlichste empirisch begründete Auflistung im Bereich des Führungsverhaltens liefern Yukl, Wall und Lepsinger (1990; zitiert nach Hogan, Gordon & Hogan, 1994). Die

Autoren identifizierten 14 Kategorien von Führungsverhaltensweisen. Die Zusammenstellung enthält im Vergleich zu Mintzbergs Modell keine grundlegend neuen Inhalte, weist aber einen höheren Detaillierungsgrad auf. Sie umfasst die Kategorien Planen und Organisieren, Problemlösen, Klären, Informieren, Überwachen, Motivieren, Beraten, Erkennen/Wahrnehmen, Unterstützen, Konflikte lösen und Teambildung, Netzwerken, Delegieren, Entwicklung und Mentoring sowie Belohnung.

In einer weiteren empirischen Studie befragten Mahoney, Jerdee und Carroll (1965) Führungskräfte zur Verteilung ihrer Arbeitszeit auf unterschiedliche Tätigkeitskategorien. Der Befragung liegt das theoretische Kategorien-Schema PRINCESS (**P**lanning, **R**epresenting, **I**nvestigating, **N**egotiating, **C**oordinating, **E**valuating, **S**upervising, **S**taffing; Mahoney et al., 1965) zugrunde², das bestehenden normativen Modellen stark ähnelt (Neuberger, 2002). Anhand der Ergebnisse unterscheiden die Autoren acht Tätigkeitsprofile. Interessant an den Ergebnissen ist der empirisch erhobene verhältnismäßige Anteil der einzelnen Tätigkeitskategorien an der Gesamttätigkeit bei Führungskräften auf unterschiedlichen Hierarchieebenen (Tabelle 2).

Tabelle 2: Prozentuale Verteilung der acht Tätigkeitsprofile auf drei Hierarchiestufen.

Tätigkeitsprofil	Hierarchiestufe		
	Untere (N = 191)	Mittlere (N = 131)	Höhere (N = 130)
Führer (Mitarbeiterführung im engeren Sinne, v. a. Anleiten und Entwickeln der Mitarbeiter)	51 %	36 %	22 %
Planer (Setzen von Zielen, Rahmenbedingungen und Prozessen)	15 %	18 %	28 %
Generalist (die Arbeitszeit verteilt sich relativ gleichmäßig auf verschiedene Funktionen. Die einzelne Funktion macht maximal 30 % der Arbeitszeit aus)	9 %	10 %	20 %
Informationsverarbeiter (Sammeln und Aufbereiten von Information)	8 %	8 %	6 %
Verhandler (Einkauf, Verkauf, Verhandlungen)	6 %	8 %	3 %
Multispezialist (zwei Aufgabengebiete machen jeweils mehr als 30 % aus)	6 %	8 %	5 %
Koordinator (Informationsaustausch in alle Richtungen)	5 %	7 %	8 %
Beurteiler (Beurteilung/Bewertung von Vorschlägen, Leistung, Personal)	2 %	5 %	8 %

Anmerkung: zitiert nach Mahoney und Kollegen (1965, S. 109).

2 Die Autoren ergänzen das PRINCESS-Schema um die Tätigkeitsprofile des Generalisten und des Multispezialisten (Erläuterungen in Tabelle 2). Die Kategorien Personalausstattung (Staffing) und Repräsentieren (Representing) aus dem PRINCESS-Schema lassen die Autoren aufgrund geringer Häufigkeit (unter 5 % der Arbeitszeit) weg.

Tabelle 3: Häufigkeitsverteilung der Tätigkeitskategorien in drei Gruppen von Führungskräften.

Tätigkeitskategorie (Tätigkeiten)	Gruppe		
	Erfolgs- führungskraft	Leistungs- führungskraft	Erfolgs- und Leistungs- führungskraft
Routinekommunikation (Informationsaustausch)	28 %	44 %	31 %
Traditionelle Führungsfunktionen (Planen/Koordinieren, Entscheiden/Problemlösen, Kontrollieren)	13 %	19 %	34 %
Beziehungspflege (Interagieren mit Externen, Pflegen sozialer Kontakte)	48 %	11 %	20 %
Human Ressource Management (Motivieren/Verstärken, Disziplinieren, Konflikte handhaben, Personal rekrutieren und entwickeln)	11 %	26 %	15 %

Anmerkung: Erfolgsführungskraft: überdurchschnittlicher Karriereerfolg; Leistungsführungskraft: überdurchschnittliche Arbeitsergebnisse; Erfolgs- und Leistungsführungskraft: überdurchschnittliche Leistung und Karriereerfolg. Zitiert nach Luthans (1995).

Diese empirische Untersuchung zeigt auf, dass die Bedeutung einzelner Führungsverhaltensweisen je nach Hierarchieebene unterschiedlich ausfallen kann. So wird die Tätigkeit des Planens mit steigender Hierarchieebene bedeutsamer, das Anleiten und Entwickeln der Mitarbeiter dagegen verlieren an Relevanz. Weiterführende Schlussfolgerungen auf die Erfolgsrelevanz einzelner Führungsverhaltensweisen gestattet Luthans (1995) Untersuchung, auf die im Folgenden eingegangen wird. Diese greifen vor auf die Auseinandersetzung mit der Definition von Führungserfolg in Abschnitt 2.1.3 und vermitteln bereits einen Eindruck von der Vielfältigkeit möglicher zu erhebender Kriterien in der Führungsbeurteilung. Luthans (1995) vergleicht die beobachtete Häufigkeit von Aktivitäten aus vier Tätigkeitskategorien bei Führungskräften, die entweder hinsichtlich Karriereerfolg oder Arbeitsergebnissen als erfolgreich gelten. Tätigkeitskategorien und Ergebnisse sind in Tabelle 3 dargestellt.

Erfolgsführungskräfte verbringen fast die Hälfte ihrer Zeit mit Networking. *Leistungsführungskräfte* hingegen verwenden ihre Zeit hauptsächlich auf Kommunikation und Personalführung. Diese Tätigkeiten führen offenbar zu guten Arbeitsergebnissen, allerdings nicht zu besonderem Karriereerfolg (Luthans, 1995). Aus den Ergebnissen wird deutlich, dass die Erfolgsrelevanz von Tätigkeiten je nach betrachtetem Erfolgskriterium unterschiedlich ausfallen kann.

Tabelle 4 zeigt eine zusammenfassende Übersicht der Aufgaben der Führung, wie sie in bereits zitierten und weiteren Quellen, inklusive deskriptiven Modellen (Yukl et al., 1990; Mahoney et al., 1965; Luthans, 1995; Grunwald, 1995; Mintzberg, 1973; Ulrich, 1970), dargestellt werden. Ergänzt wird die Übersicht abschließend um Erkenntnisse zur Führungstätigkeit aus Beobachtungen des Arbeitsalltags von Führungskräften.

Tabelle 4: Vergleichende Darstellung der von unterschiedlichen Autoren dargestellten Aufgaben der Führung anhand von acht Kategorien.

Kategorie	Mahoney et al. (1965)	Luthans (1995)	Grunwald (1995)	Mintzberg (1973)	Ulrich (1970)	Yukl et al. (1990)
1) Organisation	Typ Planer (Setzen von Zielen, Rahmenbedingungen und Prozessen)	/	Sichern von Arbeitsproduktivität (strukturelle Führung)	Ressourcenzuteilung	/	Planen und Organisieren
2) Informationsverarbeitung	Typ Informationsverarbeiter (Sammeln und Aufbereiten von Information) Typ Koordinator (Informationsaustausch in alle Richtungen fördern)	Routine-Kommunikation	/	Informationsbezogene Rolle	/	Informieren, Erkennen/Wahnehmen und Klären
3) Beurteilen & Kontrollieren	Typ Beurteiler (Bewerten von Vorschlägen, Leistung, Personal)	Kontrollieren	/	/	/	Beraten, Überwachen
4) Entscheiden, Problemlösen	Typ Verhandler (Einkauf, Verkauf, Verhandlungen)	Entscheiden und Problemlösen (Traditionelle Führungsfunktion)	/	Entscheidungsrollen	/	
5) Führung	Typ Führer (Mitarbeiterführung, Anleitung und Entwicklung von Mitarbeitern)	Human Resources Management (Disziplinieren, Bestrafen, Verstärken)	Interaktionelle Führung (Schaffen von Identifikation mit dem Unternehmen, bei den Mitarbeitern, Sichern der Arbeitszufriedenheit) Symbolische Führung (Aufgaben bezüglich Unternehmenskultur und Betriebsklima)	Interpersonelle Rollen – Entscheidungsrollen (Krisenmanagement)	/	Mitarbeiter entwickeln und Mentor sein, Delegieren, Unterstützen, Motivieren; Konflikte lösen und Teambildung
6) Repräsentieren	/	/	/	Interpersonelle Rollen (Repräsentieren)	/	/
7) Netzwerken	/	Beziehungspflege (Soziale Kontakte pflegen, Mikropolitik)	/	/	/	Netzwerken
8) Strategisches Management	/	/	/	Entscheidungsrollen (als Unternehmer agieren)	Strategisches Management	/

Anmerkung: Verglichen werden die Aufgaben nach den Modellen von Yukl et al. (1990), Mahoney et al. (1965), Luthans (1995), Grunwald (1995), Mintzberg (1973) und Ulrich (1970); (eigene Zusammenstellung).

Empirische Studien, die auf Beobachtungen des praktischen Arbeitstags von Führungskräften basieren, ergänzen das Bild der Führungstätigkeit (Grunwald, 1995). Bei den erhobenen Kriterien handelt es sich um Aspekte, wie Länge der Arbeitszyklen (Mintzberg, 1975; Stewart, 1967), Ausmaß an ungeplanten Aktivitäten (Neuberger, 2002) und Arbeitszeiten (Regnet, 2003) sowie Art der genutzten Quellen bei der Suche nach Information (Neuberger, 2002). Das sind die Ergebnisse: Die Arbeitszyklen sind kurz, es besteht ein hohes Maß an ungeplanten Tätigkeiten und die Arbeitszeiten sind generell lang. Informationen werden hauptsächlich aus informellen und unsicheren Quellen bezogen. Kommunikation bildet mit 60-70 % der Gesamtarbeitszeit die Hauptaktivität von Führungskräften (Neuberger, 2002). Dabei finden 7 % der Kontakte der Führungskraft mit dem Vorgesetzten, 16 % mit Kollegen, 48 % mit Mitarbeitern und 28 % mit externen Personen statt (Mintzberg, 1975).

Für die noch folgende Betrachtung der Urteile verschiedener Personen im Umfeld der Führungskraft über die Führungsleistung sind diese Ergebnisse v. a. deshalb relevant, da sie auf die verhältnismäßige Anzahl an Beobachtungsgelegenheiten des Verhaltens der Führungskraft durch die Beurteiler schließen lassen.

Die Darstellung des Aufgabenfelds in der Führung liefert die Grundlage für Inhalte der Führungsbeurteilung (s. Abschnitt 2.2) sowie für die Definition und Operationalisierung von Führungserfolg. Die Definition des Begriffs Führungserfolg wird im nächsten Abschnitt vorgenommen.

2.1.3 Führungserfolg

Die Definition von Führungserfolg ist in der Forschung heterogen (Neuberger, 2002). Für die Bestimmung von Führungserfolg soll angenommen werden, dass diesem die Erfüllung definierter Anforderungen zugrunde liegt (vgl. Abschnitt 2.1.2 und Abschnitt 2.2). Diese sollen ihrerseits zur Erreichung von für den Unternehmenserfolg relevanten Ergebnissen führen. Als Indikatoren dienen Variablen, wie Gewinn, Produktionsmenge oder die Rate der Kündigungen durch Mitarbeiter (Rosenstiel, 2001). In der Praxis werden die Maße häufig in einem Zielsystem kombiniert. Dies erweist sich als notwendig, da einzelne Variablen nur geringe Zusammenhänge untereinander zeigen (Sackett, Zedeck & Fogli, 1988). Der Erfolg wird in Relation zum Grad der Zielerreichung bestimmt. Ein Kritikpunkt ist der nur eingeschränkte Einfluss, den das Verhalten der Führungskraft auf die Ausprägung genannter Indikatoren hat (Lohaus, 2009).

In der Forschung besteht über das Erfolgskriterium keine Einigkeit (Birkhan, 2013). Überwiegend wird Führungserfolg über Beurteilungsdaten, zum Beispiel über die Leistungsbeurteilung durch den nächsthöheren Vorgesetzten, operationalisiert (Hiller, DeChurch, Murase & Doty, 2011). So nutzten 63 % der Autoren von 406 Validierungsstudien, die in der Fachzeitschrift „Personnel Psychology“ im Zeitraum von 1954 bis 1966 veröffentlicht wurden, Beurteilungsdaten für die Messung von Führungserfolg (Lent, Auerbach & Levin, 1971). Eine Übersicht aller genutzten Maße ist in Tabelle 5 dargestellt.

Beurteilungsdaten sind Gegenstand der Betrachtung in den Abschnitten zu Personwahrnehmung (Abschnitt 2.2.3) und 360°-Beurteilung (Abschnitt 2.2.4). Zwischen Beurteilungsdaten (es handelt sich dabei meist um Leistungs- oder Erfolgsbeurteilung durch

Tabelle 5: Maße für Führungserfolg in der Forschung.

Maße für Führungserfolg	Anteil in den Studien
Beurteilungsdaten	63 %
objektive Leistungsmaße (z. B. Arbeitsstichproben, Produktionszahlen),	17 %
indirekte Maße (z. B. Höhe des Gehalts oder erreichte Position, Fluktuation oder Fehlzeiten der Mitarbeiter)	12 %
Maße einer Restkategorie	8 %

Anmerkung: Daten entnommen aus Lent, Auerbach und Levin (1971).

den nächsthöheren Vorgesetzten; Neuberger, 2002) und objektiven Erfolgsindikatoren bestehen nur niedrige Zusammenhänge (Sackett et al., 1988).

Ein Ansatz, der weniger auf Ergebnisse erfolgreicher Führung als vielmehr auf die Voraussetzungen erfolgreicher Führung achtet, fokussiert auf Merkmale der Person: die Diagnose von Führungspotenzial. Da das Ziel der Bestimmung des Führungspotenzials letztlich die Vorhersage von Führungserfolg betrifft, wird das Konzept im nächsten Abschnitt eingeführt.

2.1.4 Führungspotenzial

Führungspotenzial, in der Literatur auch als Managementpotenzial bezeichnet, ist ein neues Konzept in der Führungsdiagnostik (Sarges, 2000). In der Praxis der Personalauswahl und -entwicklung ist dieses zwar bereits in die Alltagssprache eingeflossen, jedoch wurden nur wenige einschlägige Untersuchungen publiziert. Potenzial beschreibt die theoretische Leistungsfähigkeit einer Person. Konkreter ist das Potenzial einer Person nach Lang-von Wins und Rosenstiel (1998) die Befähigung, Anforderungen eines Aufgabenkomplexes zu bewältigen. Ziel der Messung von Führungspotenzial ist zu bestimmen, ob Voraussetzungen auf Seiten der Person (z. B. kognitive Fähigkeiten) gegeben sind, um die für erfolgreiche Führung benötigten Kompetenzen entwickeln zu können. Mehrheitlich wird in Unternehmen die Erfassung des Potenzials ohne den Einsatz weiterer Verfahren durch eine Einschätzung des Vorgesetzten vorgenommen (Scheinecker & Wallner, 2003)³. Wenn Verfahren zur Führungspotenzialbestimmung in Organisationen eingesetzt werden, so sind diese häufig intuitiv konstruiert worden (Lang-von Wins, 2000). Eine ausführliche Darstellung der Potenzialdiagnostik in Unternehmen geben Lang-von Wins, Triebel, Barth und Sandor (2008).

3 Dies trifft z. B. zu auf 86 % der 135 größten österreichischen Unternehmen (Scheinecker & Wallner, 2003).

2.2 Führungsbeurteilung

Die Führungsbeurteilung ist eine Form der Leistungsbeurteilung. Generell kann zwischen ergebnis- (vgl. Abschnitt 2.1.3), eigenschafts- und verhaltensbezogenen Ansätzen unterschieden werden. Die nächsten Abschnitte bieten einen chronologisch geordneten Überblick über eigenschafts- und verhaltensorientierte Ansätze der Führungsforschung. Für einen ausführlichen Überblick über die Methoden der Leistungsbeurteilung wird auf Schuler (2004a) verwiesen.

2.2.1 Forschungsansätze zur Führungsbeurteilung

In der folgenden Darstellung finden das verhaltenstheoretische Konzept von Hemphill und Coons (1957) in Abschnitt 2.2.1.3 und das eigenschaftstheoretische Konzept besondere Beachtung. Letzteres wird in den Abschnitten 2.2.1.1 und 2.2.1.4 vorgestellt. Für eine umfassende Darstellung der Ansätze wird auf Sarges (2000) verwiesen.

2.2.1.1 Erste eigenschaftstheoretische Ansätze

Bereits im 19. Jahrhundert beschäftigten sich einige Autoren mit der Frage, was gute Führungskräfte von weniger guten und Führungskräfte von Geführten unterscheidet. In der Great-Man-Theorie (Carlyle, 1841; Galton, 1869) wurde davon ausgegangen, dass nur Menschen mit einer spezifischen Persönlichkeitsstruktur zur Führung geeignet sind (Stachle, 1999). Die spätere spezifizierende Betrachtung einzelner Persönlichkeitsmerkmale und ihrer Beziehung zu Indikatoren des Führungserfolgs regte zahlreiche entsprechende Untersuchungen an. Stogdill (1948) und Mann (1959) zogen aus den Ergebnissen ihrer Überblicksarbeiten, die die Befunde jener Studien zusammenfassten allerdings den Schluss, dass es weder Merkmalskonstellationen noch einzelne Merkmale gibt, die Führer von Nicht-Führern zuverlässig unterscheiden konnten (vgl. Schuler & Moser, 2000). Dies führte Ende der 1940er-Jahre dazu, dass der Fokus auf verhaltenstheoretische Ansätze gerichtet wurde.

2.2.1.2 Übersicht über verhaltenstheoretische Ansätze

Die darauf folgende Zeit bis Ende der 1960er-Jahre wurde durch verhaltenstheoretische Ansätze (und bis Anfang der 1980er-Jahre durch situative Ansätze) geprägt, von denen drei besonders hervortraten (Fleishman, Zaccaro & Mumford, 1991). Im Fokus stehen Verhaltensweisen, die als relevant für erfolgreiche Führung gelten (Yukl et al., 1990).

1. Grundlegende Arbeiten aus der verhaltenstheoretischen Führungsforschung entstammen den Untersuchungen der sogenannten Ohio-Schule mit den viel zitierten Autoren Hemphill, Stogdill und Fleishman (vgl. z. B. Fleishman, 1953, 1957; Hemphill, 1950; Hemphill & Coons, 1957). Im Mittelpunkt der Bemühungen stand die Frage, welche Verhaltensmuster oder Führungsstile gute von schlechten Führenden unterscheiden. Um dies zu untersuchen, wurde das wohl einflussreichste Instrument zur Messung von Führungsverhalten erstellt (Neuberger, 2002). In dem Instrument werden zwei Führungsdimensionen unterschieden: Consideration und Initiating Structure. Aufgrund der

Relevanz für die Fragestellung der vorliegenden Arbeit wird der Ansatz in dem sich anschließenden Abschnitt 2.2.1.3 ausführlicher vorgestellt.

2. Zum Verhalten kommt dessen Interaktion mit situativen Komponenten, wie Merkmalen der Arbeit, der Organisation und des externen Umfelds. Diese Interaktion wird in der zweiten Herangehensweise betrachtet. Der bekannteste Vertreter dieses Ansatzes ist Fiedler mit seiner Kontingenztheorie (Fiedler, 1967). Darin geht es um die Kontingenz von Führungsstil und situativer Günstigkeit im Hinblick auf Erfolgskriterien. Fiedler versuchte als Erster, Situationen im Führungskontext systematisch zu klassifizieren. Die grundlegende Annahme eines eindimensionalen Führungsstilkonstrukts, bei dem Mitarbeiter- und Aufgabenorientierung zwei gegensätzliche Pole einer Skala bilden, hat sich jedoch nicht durchgesetzt. Ein weiterer Kontingenzansatz ist das Reifegradmodell von Hersey und Blanchard (1982), nach dem erfolgreiche Führung die Anpassung des Führungsstils an die Situation bzw. Reife (im Sinne von Kompetenzausprägung und Motivation) des Mitarbeiters verlangt. Auch in diesem Modell wird zwischen Mitarbeiter- und Aufgabenorientierung unterschieden, wie sie in Abschnitt 2.1.1 und 2.1.2 beschrieben ist.
3. Ein dritter Weg ist eine transaktionale Betrachtung, in welcher der Austausch zwischen Führer und Geführtem und somit die Geführten selbst mit in den Fokus rücken. Ausgangspunkt transaktionaler Führungstheorien ist die Weg-Ziel-Theorie (House, 1971). Nach dieser Theorie stellt die Führungskraft für den Mitarbeiter attraktive Ziele in Aussicht, deren Erreichung an die Erledigung bestimmter Aufgaben (Weg) durch den Mitarbeiter gebunden ist. Transaktional Führende stellen den Mitarbeitern Bedürfnisbefriedigung unter der Bedingung in Aussicht, dass die gewünschte Leistung erbracht wird. Auf diese Weise steht für den Mitarbeiter nicht mehr der Weg zum Ziel im Fokus, sondern ein Tausch. Persönlichkeitsmerkmale des Führenden werden in den transaktionalen Ansätzen nicht beachtet.

2.2.1.3 Der verhaltenstheoretische Ansatz Consideration und Initiating Structure

Aufgrund der Relevanz für die vorliegende Arbeit soll an dieser Stelle ausführlicher auf das Konzept von zwei voneinander unabhängigen Dimensionen des Führungsverhaltens von Hemphill und Coons (1957) und seine Entstehung eingegangen werden. Im Rahmen einer Fragebogenentwicklung (Leadership Behavior Description Questionnaire: LBDQ, Hemphill & Collins, 1957) wurden 1790 Episodenschilderungen von Berufserfahrenen zu Führungsverhalten gesammelt (Hemphill & Coons, 1957). Diese Episoden wurden inhaltlich analysiert und neun Dimensionen des Führungsverhaltens⁴ aus ihnen abstrahiert. Basierend auf den Episoden wurden 150 Items formuliert und in mehreren Führungsbeurteilungen erprobt. Über Faktorenanalysen wurde eine Struktur von vier Faktoren ermittelt, von denen zwei⁵ weniger varianzstark waren. Eine Reduktion auf 48 Items in weiteren

4 Die neun Dimensionen sind Integration, Initiative, Mitgliedschaft, Repräsentation, Organisation, Domination, Kommunikation, Anerkennung und Leistungs Betonung.

5 „Production emphasis“ (Ergebnisorientierung) und „Sensitivity“ (Sensitivität für Gruppenprozesse).

Tabelle 6: Itembeispiele für die Dimensionen Consideration und Initiating Structure.

Dimension Consideration	Initiating Structure
„In den Gesprächen mit seinen unterstellten Mitarbeitern schafft er eine gelöste Stimmung, so dass sie sich frei und entspannt fühlen“	„Er passt die Arbeitsgebiete genau den Fähigkeiten und Leistungsmöglichkeiten seiner unterstellten Mitarbeiter an“
„Er ist freundlich und man hat leicht Zugang zu ihm“	„Er achtet auf Pünktlichkeit und die Einhaltung von Pausenzeiten“
„Bei wichtigen Entscheidungen holt er erst die Zustimmung seiner unterstellten Mitarbeiter ein“	„Er regt seine unterstellten Mitarbeiter zur Selbstständigkeit an“

Anmerkung: Quelle: Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung (FVVB; Fittkau & Fittkau-Garthe, 1971).

Überarbeitungen des Fragebogens ergab eine Zweidimensionalität des Führungsverhaltens.⁶ Die zwei Dimensionen sind definiert als Consideration und Initiating Structure.

Consideration bedeutet Mitarbeiterorientierung, praktische Besorgtheit und freundliche Rücksichtnahme. Es beschreibt die Aspekte Wärme, Vertrauen, Freundlichkeit und Achtung, die Ermöglichung zweiseitiger Kommunikation sowie Mitsprache. Eine hohe Ausprägung in diesem Merkmal äußert sich über die Anteilnahme der Führungskraft sowie Respekt und Wertschätzung gegenüber den Mitarbeitern, Unterstützung der Mitarbeiter und Sorge um deren Wohlergehen (Bass, 1990a). Initiating Structure beschreibt aufgabenbezogene Organisation und Strukturierung, Planungsinitiative, Aktivierung und Kontrolle. Eine hohe Ausprägung äußert sich im höheren Ausmaß der Bestimmung und Organisation der Aufgaben und Rollen, dem Schaffen von Kommunikationsstrukturen sowie zielorientiertem Handel (Fleishman, 1973). Wie in den in Abschnitt 2.1.1 beschriebenen zwei Aspekten der Führung wird auch in dieser Zweidimensionalität eine Mitarbeiter- von einer Aufgabenorientierung unterschieden (Neuberger, 2002).

Für den deutschen Sprachraum wurde die Version „Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung“ (FVVB; Fittkau-Garthe & Fittkau, 1971) entwickelt. Die Itemmenge wurde hier auf 32 Items reduziert. Auch im Zuge der Konstruktion des FVVB wurden vier Dimensionen identifiziert, die sich auf höherer Ebene jedoch den Dimensionen Consideration und Initiating Structure zuordnen lassen. Genauer ordnen sich „Freundliche Zuwendung“ und „Mitbestimmung, Beteiligung“ Consideration zu, „Mitreißende Aktivität“ und „Kontrolle vs. Laissez-Faire“ Initiating Structure. Nachuntersuchungen am FVVB bestätigten die überlegene Passung der Zweidimensionalität (Allerbeck, 1977; Nachreiner, 1995). Itembeispiele aus dem FVVB, die den beiden Dimensionen Consideration und Initiating Structure von Nachreiner (1978) zugeordnet wurden, sind in Tabelle 6 dargestellt.

Entgegen der Erwartung zeigen sich die beiden Dimensionen miteinander positiv korreliert (Fleishman & Harris, 1962). Weitere Aspekte, die an dem Ansatz der Ohio-Schule kritisch bewertet werden, sind die situationsunabhängige und allgemeine Formulierung

6 Mit Neuberger (2002) soll kritisch erwähnt werden, dass eine Reduktion des Spektrums an Führungsverhaltensweisen durch den gezielten Ausschluss jener Items stattfand, die den beiden Dimensionen nicht zugeordnet werden konnten.

der Items (Implikation ist in Abschnitt 2.2.3.2 zu Moderatoren der Urteilsgüte erläutert) sowie die Heterogenität der Iteminhalte innerhalb der Skalen (Krohne & Hock, 2007).

2.2.1.4 Neue eigenschaftstheoretische Ansätze

Mitte der 1980er-Jahre wandte sich die Führungsforschung erneut den Aspekten der Persönlichkeit zu. Eine der dafür verantwortlichen Tatsachen war die Unmöglichkeit, auftauchende Probleme über die zu der Zeit bestehenden Ansätze zu lösen (Sarges, 2013a; Schuler & Moser, 2000). Die Ergebnisse der verhaltenstheoretischen bzw. situativen Forschung geben Hinweise darauf, wie sich eine Führungskraft verhalten oder die Situation beschaffen sein muss, damit ein Erfolg herbeigeführt werden kann. Für Methoden der Personalentwicklung in Organisationen ist dieser Aspekt von Relevanz, wie in der Einleitung erläutert. Bei der Personalrekrutierung ist jedoch vielmehr die Identifikation von Merkmalen gefragt, anhand derer die Auswahl zwischen Bewerbern getroffen werden kann. Diese Anforderungen in der Wirtschaft trugen dazu bei, dass eigenschaftstheoretische Ansätze in der Führungsforschung wieder eine höhere Bedeutung erlangten.

Kritische Prüfungen der Ergebnisse der Studien von Stogdill (1948) und Mann (1959) öffneten schließlich den Weg für die erneute Auseinandersetzung mit Persönlichkeitsvariablen im Kontext erfolgreicher Führung. Lord, De Vader und Alliger (1986) untersuchten die Ergebnisse von Mann (1959) mit neueren metaanalytischen Herangehensweisen und konnten aufzeigen, dass weit höhere Zusammenhänge zwischen Persönlichkeitseigenschaften und Führungserfolgsindikatoren gefunden werden, wenn die Daten um Verzerrungen durch Stichprobenfehler, mangelnde Zuverlässigkeit und Varianzeinschränkung bereinigt sind.

Im folgenden Abschnitt wird ein aktuell besonders populärer Ansatz in der Führungsforschung dargestellt und damit die Einführung zentraler Konzepte der Führungsforschung abgeschlossen.

2.2.1.5 Transformationale Führung

Im Zuge der erneuten Untersuchung von Persönlichkeitsmerkmalen rückte seit den 1980er-/1990er-Jahren die sogenannte transformationale Führung in den Fokus der Forschung. Aktuell ist es das am häufigsten untersuchte Einzelkonzept in der Führungsforschung (Diaz-Saenz, 2011). Transformationale Führung fokussiert nach Bass und Avolio (1989) auf das Beeinflussen der Bedürfnisse und Ziele der Mitarbeiter. Von transformationalen Führungskräften spricht man, wenn die folgenden vier Komponenten vorliegen (Bass & Avolio, 2004):

1. Der Vorgesetzte bewirkt bei den Mitarbeitern das Bedürfnis, ihn idealisierend zu betrachten sowie sich mit ihm und seinem Vorhaben zu identifizieren („Idealized Influence“). Hierzu gehört zum einen, dass dem Vorgesetzten bestimmte Aspekte zugesprochen werden („Idealized Influence, Attributed“). Solche Aspekte sind zum Beispiel, dass der Vorgesetzte das Interesse der Gruppe gegenüber eigenen Interessen priorisiert und dass er respektiert wird. Zum anderen gehören dazu Verhaltensweisen des Vorgesetzten, wie das Betonen der Wichtigkeit des Sinnzusammenhangs und dem Mitteilen eigener Überzeugungen und Werte („Idealized Influence, Behavioral“).

2. Der Vorgesetzte vermittelt den Mitarbeitern eine attraktive Vision und zeigt auf, wie diese Wirklichkeit werden kann („Inspirational Motivation“).
3. Der Vorgesetzte bricht bei den Mitarbeitern alte Sichtweisen auf und ersetzt sie durch neue („Intellectual Stimulation“).
4. Der Vorgesetzte versteht und teilt Bedenken der Mitarbeiter. Er erkennt und befriedigt Entwicklungsbedürfnisse und behandelt jeden Mitarbeiter individuell („Individualized Consideration“).

Zur Messung der Ausprägung transformationaler Führung wurde der Multifactor Leadership Questionnaire (MLQ; Bass & Avolio, 2004) konstruiert.

2.2.1.6 Neue Kontingenzmodelle

Nach Hackman und Wageman (2007) können weder Persönlichkeitsmerkmale noch Situationsvariablen allein den Führungserfolg ausreichend erklären. Deshalb finden sich in neueren Ansätzen (Zaccaro, 2007; Vroom & Jago, 2007; Sternberg, 2007) Modelle, die situationale Variablen als Moderatoren der Beziehung zwischen Persönlichkeitsmerkmalen und Führungserfolgsvariablen einbeziehen. Mittlerweile wurde eine Vielzahl solcher Moderatoren identifiziert, was die Komplexität der Modelle erhöht. Dies gestattet einerseits zwar sehr spezifische Vorhersagen, welche Eigenschaften unter welchen situativen Bedingungen mit Erfolgskriterien assoziiert sind, die gefundenen Zusammenhänge sind jedoch über verschiedene situative Bedingungen nur schwer generalisierbar (Hackman & Wageman, 2007). Ihre Verwendbarkeit für führungsdiagnostische Fragen, die die Messung allgemeiner Führungseignung intendieren (vgl. Abschnitt 2.1.4 und 2.2.1.4), ist dadurch eingeschränkt. In der vorliegenden Arbeit wird daher trotz der Aktualität des Ansatzes auf den Einbezug von Kontingenzmodellen verzichtet.

Nach Darstellung der klassischen Ansätze soll im nächsten Abschnitt ein Konzept ergänzt werden, das in den 1980er-Jahren der Führungsdiagnostik einen neuen Bestandteil hinzufügte, das Konzept der Misserfolgsfaktoren. Dieses Konzept hat aktuell vorrangig unter den Bezeichnungen Managementversagen und Derailment Popularität erlangt (Westermann & Birkhan, 2013).

2.2.2 Das Konzept der Misserfolgsfaktoren

Der allgemeine Fokus der Forschung im Bereich Führungsdiagnostik liegt auf Faktoren, die zum Erfolg der Führungskraft beitragen (Westermann & Birkhan, 2013). Bentz (1985) sowie McCall und Lombardo (1983) fokussierten dagegen nicht mehr nur Merkmale, die dem Erfolg der Führungskraft zuträglich sind. Die Autoren gehen davon aus, dass es auch Merkmale gibt, deren Vorhandensein Führungserfolg⁷ mindern oder verhindern kann. Diese können sowohl eigenschafts- als auch verhaltensbasiert sein. Entsprechende Modelle umfassen sozial unerwünschte Persönlichkeitsaspekte (z. B. Arroganz) oder Verhaltensweisen (z. B. unehrliches Verhalten), unzureichend ausgeprägte Kompetenzen (in die-

⁷ In den empirischen Erhebungen, die zu entsprechenden Sammlungen führten, wird Führungserfolg überwiegend über Karriereerfolg (Aufstieg in der Hierarchie) operationalisiert. Ein negativer Effekt bedeutet entsprechend: kein weiterer hierarchischer Aufstieg, Herabstufung oder Verlust von Führungsverantwortung.

sem Zusammenhang definiert als eine Kombination anwendungsbezogener Fähigkeiten und Fertigkeiten sowie anwendungsbezogenen Wissens, bis hin zu klinischen Auffälligkeiten. Eine aktuelle Sammlung jener Merkmale, die mit dem Versagen in Verbindung gebracht werden, liefert Kanning (2014).

Bentz (1985) identifizierte sieben solcher Aspekte (z. B. die Unfähigkeit zu delegieren oder langsames Lernen) bei Führungskräften, die als Potenzialträger eingestuft worden waren und trotzdem scheiterten (die geplante Führungsstufe nicht erreichten). McCall, Lombardo und Morrison (1988) verglichen 20 erfolgreiche Führungskräfte mit 21, deren Führungskarriere als gescheitert angesehen wurde (s. Fußnote 7). Aspekte, die vermehrt bei den erfolglosen Führungskräften zu finden waren, ordneten sie vier Bereichen zu: (1) Schwierigkeiten mit interpersonellen Beziehungen, (2) Nichterreichen von Geschäftszielen, (3) Unfähigkeit ein Team zu bilden und (4) Unfähigkeit sich an Veränderungen anzupassen. Hierzu muss allerdings kritisch angemerkt werden, dass zwischen Merkmalen der Person und Indikatoren des Misserfolgs teilweise nicht unterschieden wird. Deutlich wird dies am Beispiel des zuvor genannten Themenbereichs „Nichterreichen von Geschäftszielen“.

Foster und Hogan (2006) treffen eine in dieser Hinsicht klarere Unterscheidung und definieren solche Merkmale, die Führungserfolg beeinträchtigen können, als „negative Persönlichkeitsmerkmale“ bzw. „Dark Side“-Charakteristiken. Zur Identifikation solcher Merkmale (z. B. eine zu stark ausgeprägte Vorsicht oder Boshaftigkeit) empfehlen die Autoren Methoden der klinischen Diagnostik. Gemeinsam ist den Ansätzen, dass die Merkmale als eine eigene Kategorie, unabhängig von Führungskompetenzen betrachtet und auch entsprechend in Fragebögen abgebildet werden. Eine aktuelle Bestandsaufnahme zur Forschung in diesem Bereich geben Westermann und Dick (2014).

Obwohl die beschriebenen Ansätze der Führungsforschung bereits zur Entwicklung einer Reihe von Messinstrumenten zur Beurteilung von Führungskräften geführt haben, findet Führungsbeurteilung in Organisationen im Wesentlichen immer noch über eine Einschätzung durch den Vorgesetzten statt⁸. Dabei darf aber nicht vergessen werden, dass es sich bei diesen Einschätzungen um vorwissenschaftliche Personbeurteilung handelt.

2.2.3 Vorwissenschaftliche Personbeurteilung

Personalfachleute und Vorgesetzte, die praktisch die Beurteilung der Führungskraft vornehmen, sind in der Regel nicht in wissenschaftlicher Diagnostik ausgebildet. In den folgenden Abschnitten wird aufgezeigt, welche Implikationen dies für die zu erwartende Urteilsgüte hat (Abschnitt 2.2.3.1) und welche Variablen im Kontext der 360°-Beurteilung die Urteilsgüte moderieren (Abschnitt 2.2.3.2). In zwei weiteren Abschnitten werden der Prozess der Urteilsbildung erläutert (Abschnitt 2.2.3.3) und die Besonderheiten der Beurteilung in Interaktionssituationen herausgestellt (Abschnitt 2.2.3.4). Die Ausarbeitungen sind grundlegender Bedeutung für die in Abschnitt 2.2.4 zur 360°-Beurteilung berichteten Befunde und theoretischen Annahmen. An gegebenen Stellen werden in den nachfolgenden Ausführungen der Bezug zur 360°-Beurteilung hergestellt und bereits erste Ableitungen formuliert.

8 Beispielsweise verwenden laut einer Befragung an deutschen Unternehmen (Hell, Schuler, Boramir & Schaar, 2006) 73,1 % Leistungsbeurteilungen durch den Vorgesetzten für Beförderungsentscheidungen auf unterer Führungsebene.

2.2.3.1 Güte und Fehlerquellen vorwissenschaftlicher Personbeurteilung

Die Güte vorwissenschaftlicher Personbeurteilungen kann durch die Genauigkeit des Urteils oder den Konsens verschiedener Beurteiler bestimmt werden. Die Untersuchung der Genauigkeit erfolgt über die Übereinstimmung eines Urteils mit einem Kriterium (vgl. Kruglanski, 1989). Der Konsens wird darüber bestimmt, in welchem Grade Einschätzer in ihrem Urteil über dieselben Merkmale einer Person übereinstimmen. Auf methodische Einzelheiten in der Bestimmung der Übereinstimmung wird in Abschnitt 2.2.4 eingegangen. Eine differenziertere Behandlung der Berechnung und Bedeutung von Maßen zu Konsens und Selbst-Fremdurteilsübereinstimmung geben z. B. Krohne und Hock (2007). In der sozialpsychologischen Literatur wird die Genauigkeit vorwissenschaftlicher Personbeurteilung unter dem Stichwort „Interpersonale Wahrnehmung“ (vgl. u. a. Jones, 1990; Kenny, 1993) thematisiert. Die Genauigkeit der Beurteilung wird als eher niedrig eingestuft (Cline, 1964; Taft 1955).

Die Genauigkeit von Urteilen kann über Urteilsfehler beeinträchtigt werden. Einen Überblick über die verschiedenen Urteilsfehler, die in Studien zur Personenwahrnehmung (vgl. Tagiuri, 1969) identifiziert worden sind, geben z. B. Strack und Deutsch (2002). Die am häufigsten untersuchten systematischen Fehler in der Personbeurteilung sind der Halo- und der Milde-Effekt (Scullen et al., 2000). Der Halo-Effekt bildet die Hauptursache von Messfehlern in der Beurteilung (Mount et al., 1998b).⁹ Er entsteht dadurch, dass ein Beurteiler bestimmte Eigenschaften als stärker zusammengehörig empfindet und damit eine bestimmte Eigenschaft die Beurteilung anderer Eigenschaften beeinflusst (Krohne & Hock, 2007). In der 360°-Beurteilung wird die Beurteilung einzelner Eigenschaften vor allem durch einen Gesamteindruck von der Leistung einzuschätzender Personen beeinflusst (Mount et al., 1998b). Der Milde-Effekt beschreibt die Tendenz, Personen eher positiv zu beurteilen (Krohne & Hock, 2007). Leistungsbeurteilungen zeigen sich vom Milde-Effekt besonders häufig beeinflusst (Schuler, 2004). Folge sind überhöhte Leistungsbewertungen in der 360°-Beurteilung (Mount et al., 1998b). Eine Quantifizierung des Einflusses der Fehler in der 360°-Beurteilung wird in Abschnitt 2.2.4 aufgeführt.

Fehlertendenzen finden sich sowohl in Selbst- als auch in Fremdurteilen. Selbsturteile unterscheiden sich aber grundlegend von Fremdurteilen. Am „Actor-Observer-Bias“ (Jones & Nisbett, 1972) wird dies deutlich: Beobachtende tendieren dazu, das Verhalten der Zielperson durch deren Eigenschaften zu erklären, während die Ursachen für eigenes Verhalten eher in situativen Merkmalen gesehen werden. Charakteristisch für Selbsturteile ist zudem eine Verzerrung der Einschätzung negativer Merkmale ins Positive („self-bias“; vgl. u. a. Robinson & John, 1997). In der 360°-Beurteilung fallen Selbsteinschätzungen generell positiver aus als Fremdurteile (Harris & Schaubroeck, 1988). Auf einen Vergleich der Güte der Urteile aus verschiedenen Quellen wird in Abschnitt 2.2.4 eingegangen.

Beachtet werden muss, dass mangelnde Zusammenhänge mit einer Kriteriumsvariablen oder mangelnde Beurteilerübereinstimmung nicht in jedem Fall gleichbedeutend mit mangelnder Urteilsgüte sind. Ein niedriger Zusammenhang mit einer Außenvariablen kann unter anderem auch resultieren, wenn Urteil und Kriterium sich auf unterschiedliche Inhalte beziehen. Ein Beispiel wäre die Messung der Genauigkeit eines Urteils zur Planungskompetenz anhand des Zusammenhangs mit der Außenvariable „Verspätung der

9 Die Diskussion um einen „wahren“ Halo-Effekt in der 360°-Beurteilung wird in Abschnitt 4.1.8 geführt.

Fertigstellung eines Projekts in Tagen“. Ein niedriger Zusammenhang trotz genauen Urteils könnte hier darüber erklärt werden, dass die Planung zwar kompetent durchgeführt wurde, aber mangelnde Zielstrebigkeit zu Verspätung führte. (Auf weitere Einschränkungen der Genauigkeitsmessung über Zusammenhänge mit Außenvariablen wird in den Abschnitten 2.1.3 und 2.2.4.3 eingegangen.) Eine niedrige Übereinstimmung kann ähnlich resultieren, wenn Einschätzer zu bewertende Merkmale unterschiedlich interpretieren (fehlendes „shared meaning“; Kenny, 1991). Soll die Planungskompetenz einer Führungskraft bewertet werden, sind verschiedene Interpretationen denkbar: z. B. die Fähigkeit, langfristige Pläne zu entwerfen, oder die Detailliertheit eines Plans. In Abschnitt 2.2.4.2 wird ausgeführt, warum diesem Punkt in der 360°-Beurteilung eine hohe Bedeutung zukommt.

Es bestehen weitere Einwände gegen die Verwendung der Beurteiler-Übereinstimmung zur Evaluierung der Güte vorwissenschaftlicher Urteile (Übersicht in Kenny, 1993).¹⁰ So sagt eine niedrige Übereinstimmung beispielsweise auch nichts darüber aus, ob nur eines der Urteile oder alle betrachteten ungenau sind. In Abschnitt 2.2.4.2 wird eingehender dargelegt, warum die Berechnung der Urteilsübereinstimmung gerade im Kontext von 360°-Verfahren zwar wichtig ist, aber zur Evaluierung der Urteilsgüte von 360°-Urteilen problematisch ist.

Im folgenden Abschnitt werden Moderatoren der Urteilsgüte behandelt. Der Fokus liegt auf Variablen, die relevant in Bezug auf 360°-Verfahren sind. Für eine umfassendere Darstellung von Moderatoren der Urteilsgüte wird auf Krohne und Hock (2007) verwiesen.

2.2.3.2 Moderatoren der Urteilsgüte

Die bislang als wesentlich identifizierten Moderatoren sind nach Funder und West (1993) die Art der Beziehung zwischen Beurteiler und Zielperson (z. B. Maß der Vertrautheit), die Art des zu beurteilenden Merkmals (z. B. globaler versus situationsspezifischer Inhalt), die Art der Information, die erfasst werden soll (z. B. die Informationsmenge) sowie Merkmale der Zielperson (z. B. Extraversion). Taft (1995) beschreibt Merkmale des Beurteilers als ebenfalls relevant (z. B. dessen Alter oder Intelligenz). Tabelle 7 zeigt Variablen, die im Kontext der 360°-Beurteilung eine Rolle spielen.

Für die 360°-Beurteilung sind Moderatorvariablen erst ansatzweise untersucht. Es liegen häufig nur einzelne Befunde vor, die sich zudem nur auf einen Teil der Quellen beziehen. Die stabilsten und relevantesten Moderatorvariablen sind bisherigen Ergebnissen zufolge der Tätigkeitsbereich (s. Tabelle 7, Zeile 7) und die Art des zu beurteilenden Merkmals (s. Tabelle 7, Zeile 1-5 sowie Abschnitt 2.2.4.5).

Ein im Kontext der 360°-Beurteilung nicht direkt beachteter, jedoch als relevant anzunehmender Aspekt ist die Informationsmenge (s. Tabelle 7, letzte Zeile). Auf diese wird im Folgenden näher eingegangen. Mit der Informationsmenge, die dem Beurteiler zur Verfügung steht, steigt die Güte des Urteils (z. B. Borkenau et al., 2004). So wurde eine

¹⁰ Für die ausführliche Kritik an einem globalen Übereinstimmungsmaß siehe Cronbach (1955). Der Autor schlägt das Aufgliedern der Varianz der Beurteilerübereinstimmung in verschiedene Komponenten vor, die eine differenzierte Bestimmung der Urteilsgüte ermöglichen.

Tabelle 7: Moderatoren der Urteilsgüte mit besonderer Relevanz für die 360°-Beurteilung.

Bedeutung in der 360°-Beurteilung		Relevanz für die vorliegende Untersuchung
Art des zu beurteilenden Merkmals		
Global vs. situations-spezifisch	Das Abfragen sehr vieler Merkmale in der 360°-Beurteilung erfordert ein globales Urteil (geringere Urteils-genauigkeit erwartet).	Indirekt für den Kontext von Quellen-unterschieden.
Beobacht-barkeit	Zu beurteilende Merkmale sind in unterschiedlichem Maße beobachtbar. Bei eindeutig beobachtbaren Merk-malen ist die Selbst- und Fremdurteilübereinstimmung höher (Dai, Stiles, Hallenbeck & De Meuse, 2007).	Indirekt für die Analyse der Häufung fehlender Werte in der 360°-Beurteilung und im Kontext von Quellen-unterschieden. Explizit für die Be-trachtungen des Zusammenhangs von eigenschafts- und verhaltensbasierten Variablen.
Bewertungs-bezogenheit	Bewertungsbezogen sind v. a. die sozial unerwünschten Unzulänglichkeitsmerkmale. Bewertungsbezogenheit sollte auf die Übereinstimmung von Selbst- und Fremdurteil wirken (self-bias).	Indirekt für die Interpretation der Faktorenstrukturen und Urteilsgüte.
Verhaltens- vs. eigenschafts-bezogen ¹¹	Urteile zu verhaltensbezogenen Merkmalen stimmen tendenziell stärker überein (Harris & Schaubroeck, 1988).	Hauptsächlich werden verhaltens-bezogene Merkmale erhoben.
Kognitive vs. interpersonelle Fertigkeiten	Urteile zu interpersonellen Aspekten stimmen höher überein (Conway & Huffcutt, 1997; Mount, 1984). Un-terschiede in der Beobachtbarkeit könnten dies erklä-ren.	Beide Aspekte werden erhoben. Be-trachtung im Zusammenhang mit merkmalsspezifischen Quelleneffek-ten.
Kontextmerkmal		
Einsatzzweck des 360°-In-struments	Mitarbeiterurteile stimmen höher überein, wenn die Be-fragung Entwicklungsvorhaben dient (versus adminis-trativen Zwecken ¹² ; Greguras, Robie, Schleicher & Goff, 2003). Selbsturteile zeigen umgekehrt eine höhere Korrelation mit der Zielerreichung bei administrativem Einsatzzweck (Zimmerman, Mount & Goff, 2008).	Einsatzzweck ist Entwicklung.
Tätigkeits-bereich	Der Zusammenhang mit Außenvariablen ist höher bei niedriger Komplexität der Aufgabe der Zielperson (Conway & Huffcutt, 1997).	Die Stichprobe deckt unterschiedliche Komplexitätsgrade ab (Effekt wird ausgeglichen).
	Die Beurteilerübereinstimmung zwischen Quellen variiert zwischen Branchen (Brutus & Fleenor, 1998).	Die Stichprobe beinhaltet Führungs-kräfte aus verschiedenen Branchen.
Bekanntheits-grad	Die Urteilsübereinstimmung ist bei einer Bekanntheits-dauer von fünf Jahren im Vergleich zu einem Jahr höher (Rothstein, 1990). Die Höhe der Zusammenhänge mit Leistungskriterien ist bei der Bekanntheitsdauer von einem und drei Jahren am höchsten, bei drei bis fünf am zweithöchsten, bei weniger als einem Jahr am niedrig-sten (Eichinger & Lombardo, 2004).	Durch Aggregation der Werte inner-halb einer Quelle sollten sich etwaige Effekte ausgleichen.
Informations-menge	Die Güte des Urteils sollte mit der Informationsmenge, die dem Beurteiler zur Verfügung steht, ansteigen (Swann, 1984).	

11 Spezifisch formulierte und verhaltensorientierte Items, die sich auf direkt beobachtbare Merkmale beziehen, führen zu höherer Test-Retest-Reliabilität (Van Velsor & Leslie, 2001).

12 Administrativer Einsatzzweck bedeutet in diesem Zusammenhang die Verwendung der Beurteilung zur Entscheidung über personellen Einsatz, Gehaltsänderung etc.

Anmerkung zu Tabelle 7: Dargestellt sind Variablen, die in der 360°-Beurteilung von besonderer Relevanz sind. Weitere Variablen werden im nachfolgenden Text erläutert (eigene Zusammenstellung).

Abnahme des Einflusses von Merkmalen, wie Geschlecht, Alter oder Attraktivität, auf die Einschätzung von Persönlichkeitsvariablen mit zunehmender Menge an Verhaltensinformationen festgestellt (Borkenau et al., 2004). Eine zu große Menge an Informationen kann sich allerdings auch negativ auf die Urteilsgüte auswirken (Kruglanski, 1989). Diese könnte dazu führen, ein Urteil zu früh abzuschließen, ohne neue relevante Information miteinzubeziehen. Ebenso könnte es sein, dass viele Informationen zu sich widersprechenden Hypothesen führen und eine Urteilsbildung damit erschweren. Im Kontext der 360°-Beurteilung variiert die Menge an vorhandener Information durchaus zwischen Einschätzern und vermutlich noch deutlicher (wie in Abschnitt 2.2.4.2 im Hinblick auf Unterschiede in Beobachtungsgelegenheiten ausgeführt wird) zwischen Einschätzern aus verschiedenen Quellen. Kontakthäufigkeit und Bekanntheitsdauer begründen die Unterschiede in diesem Fall, wie zum Thema Quellenunterschiede noch näher erläutert wird. Befunde im Kontext der 360°-Beurteilung (s. Tabelle 7) zeigen die Tendenz einer höheren Genauigkeit mit einer längeren Bekanntheitsdauer (Eichinger & Lombardo, 2004; Rothstein, 1990).

Die in den letzten beiden Abschnitten ausgeführten Aspekte, Moderatoren der Urteilsgüte und Urteilsfehler, gilt es bei der Interpretation von Befunden zur Genauigkeit der 360°-Beurteilung in Abschnitt 2.2.4.3 zu berücksichtigen. Im nächsten Abschnitt werden anhand des Prozesses der Urteilsbildung weitere potenzielle Ursachen von Urteilsunterschieden erläutert.

2.2.3.3 Der Prozess der Urteilsbildung

Der Prozess der Urteilsbildung wird am Modell von Brandstätter (1969), das 2004 von Schuler aktualisiert wurde, veranschaulicht. Hier werden drei Ebenen der Urteilsbildung unterschieden (Abbildung 2).

Auf allen drei Ebenen können Unterschiede in den Urteilen entstehen. Ebene 1 beschreibt, wie das Verhalten der zu beurteilenden Person wahrgenommen wird. Einflussfaktoren auf die Wahrnehmung sind neben der individuellen Aufmerksamkeitsausrichtung auch Kontextfaktoren, wie die allgemeine Arbeitssituation des Beurteilers (Lohaus, 2009). Auf Ebene 2 erfolgt die Eindrucksbildung auf Basis des Wahrgenommenen. Welche Beobachtungen dabei auf welche Weise interpretiert werden, ist wiederum abhängig vom Beurteiler und dessen impliziten Theorien sowie Konzeptualisierungen des zu beurteilenden Sachverhalts. Die Formulierung einer Beurteilung bildet die dritte Ebene ab, auf der Aussagen über die Zielperson getroffen werden. Selbst bei übereinstimmender Wahrnehmung und Interpretation des Beobachteten durch verschiedene Einschätzer können auf dieser Ebene Unterschiede in der Beurteilung entstehen. Bedingt können diese durch bestimmte Erwartungshaltungen, diverse Urteilstendenzen und -fehler (s. Abschnitt 2.2.3.1) oder Motive des Beurteilers sein. Der Einfluss von Motiven des Beurteilers auf die Beurteilung ist bislang in weitaus geringerem Umfang untersucht als jener von Urteilsfehlern (Lohaus, 2009).¹³

13 Einen Überblick über die Motive, die im Kontext der Leistungsbeurteilung zu berücksichtigen sind, gibt Fletcher (2002).

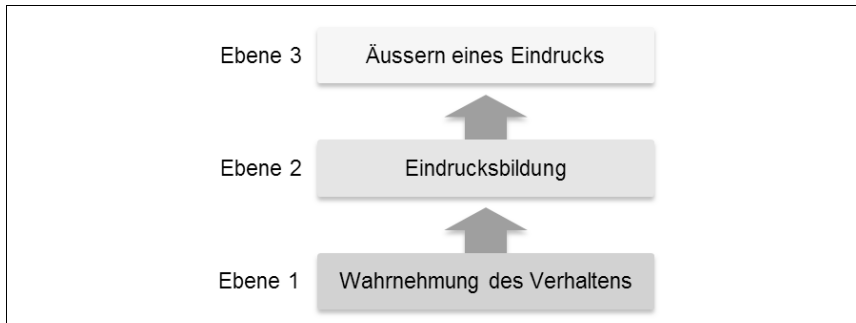


Abbildung 2: Ebenen der Urteilsbildung nach Brandstätter (1969) und Schuler (2004b)

Für die Beantwortung der Fragen der vorliegenden Arbeit ist nicht nur relevant, wie es zu Unterschieden in den Urteilen kommt, sondern insbesondere welche Unterschiede in den Urteilen aus verschiedenen Perspektiven innerhalb der Organisation zu erwarten sind. Hinweise darauf finden sich in der Betrachtung der Eigenheiten von Personbeurteilung unter der Bedingung lebensechter Information: dem Ansatz pragmatischer Zielsetzung des Beurteilers (Swann, 1984) und der Annahme einer Beeinflussung der Urteile durch „verhandelte Identitäten“ (Goffman, 1959). Diese werden im folgenden Abschnitt erläutert.

2.2.3.4 Personbeurteilung als Interaktionsprozess

Auch die Art der Information, lebensecht versus artifiziell, ist eine Moderatorvariable bezüglich der Urteilsgüte (Krohne & Hock, 2007). Die Art der Information, auf der die 360°-Beurteilung basiert, ist lebensecht. Die beiden zentralen Punkte, die Personwahrnehmung im lebensechten Umfeld von jener unter festgelegten Versuchsbedingungen unterscheiden, sind die Aktivität des Beurteilers und die Beurteilungsziele (Krohne & Hock, 2007). Die Personbeurteilung in realen Situationen entspricht einem Interaktionsprozess (Swann, 1984). Da es sich um einen Interaktionsprozess handelt, verfolgen Personen dabei Ziele (Jones, 1990). Swann (1984) ist der Ansicht, die Zielsetzung des Beurteilers zeigt pragmatischen Charakter: Der Beurteiler versucht Informationen vor allem zu jenen Merkmalen der Zielperson zu gewinnen, die für die Erreichung seines Interaktionsziels bedeutsam sind.

Für die 360°-Beurteilung ist anzunehmen, dass Interaktionsziele der Beurteiler (in der Interaktion mit der Führungskraft) innerhalb einer Ebene in der Organisation (z. B. zwischen Mitarbeitern) ähnlicher sind als zwischen verschiedenen Quellen (z. B. Mitarbeiter im Vergleich zu Vorgesetzten). Das gilt entsprechend für den instrumentellen Wert einzelner Informationen bzw. Merkmalskategorien. Ausführlich wird hierzu im nächsten Abschnitt argumentiert. Es kann für einen Mitarbeiter beispielsweise besonders relevant sein, ob die Führungskraft ihn in einer schwierigen Situation unterstützt (Consideration). Für einen Vorgesetzten kann die Information, ob die Führungskraft unternehmerisch handelt sowie Qualität sicherstellt (Initiating Structure) und so den Vorgesetzten in der Erreichung organisationaler Ziele unterstützt, wichtiger sein.

Es resultiert eine höhere Genauigkeit des (zielbezogenen) Urteils bezüglich der Merkmale, die für den Einschätzer bedeutsamer sind (Swann, 1984). Im Kontext der 360°-

Beurteilung sollten diese Aspekte zudem eine höhere Bedeutsamkeit für das Gesamturteil zur Führungsleistung aufweisen (s. Abschnitt 2.2.4.6). Die Genauigkeit zielbezogener Urteile wird zusätzlich dadurch begünstigt, dass der Beurteiler aktiv ist sowie Einfluss nimmt auf die zu beurteilende Person und so auch darauf, welche Information beobachtbar wird.

Weiterführend wird im Konzept der sogenannten verhandelten Identität nach Goffman (1959) beschrieben, dass sich zielbezogene Urteile mehr auf situationsspezifische als auf globale Einschätzungen beziehen. Dem Konzept nach handeln Personen aus, welche Rolle sie in der Interaktion einnehmen. Dabei ist herauszustellen, dass das Verhalten der Zielperson bzw. die angenommene Identität in Abhängigkeit vom Interaktionspartner variieren kann (Swann, 1984). Für die 360°-Beurteilung ist anzunehmen, dass die verhandelten Identitäten der Beurteiler (in der Interaktion mit der Führungskraft) innerhalb einer Ebene in der Organisation (z. B. zwischen Mitarbeitern) ähnlicher sind als zwischen verschiedenen Quellen (z. B. Mitarbeiter im Vergleich zu Vorgesetzten). Begründet werden kann dies über die jeweils ähnlichen Umstände unter denen die Identität verhandelt wird (z. B. die hierarchische Beziehung zur Führungskraft).

Die Wahrnehmung einer Person und damit ihr Urteil über die Zielperson wird stark durch die zwischen Beurteiler und Zielperson verhandelte Identität bestimmt (Goffman, 1959). Dies liegt darin begründet, dass eine Einschätzung der Identität für das Erreichen des Interaktionsziels bedeutsamer ist als die Einschätzung von Eigenschaftsausprägungen. Beispielsweise ist es für den Beurteiler weniger relevant, wie stark generell die Dominanz einer Person ausgeprägt ist; wichtiger ist, ob sich die Person ihm gegenüber dominant oder unterwürfig verhält.

Aus den bisherigen Ausführungen lässt sich schließen, dass Unterschiede zwischen den Urteilen verschiedener Einschätzer zum Verhalten einer Zielperson zu erwarten sind. Diese und ähnliche Überlegungen werden bezogen auf Bedeutung und Nutzen des Einbezugs verschiedener Quellen in die Führungsbeurteilung im Folgenden aufgearbeitet und bewertet. Die noch einzuführenden Ansätze und Annahmen verschiedener Autoren lassen sich in der Mehrzahl letztlich aus den geschilderten Erkenntnissen und Theorien zu vorwissenschaftlicher Personenbeurteilung im Kontext von Organisationen ableiten. Anknüpfend an die in der Einleitung geschilderte Entstehungsgeschichte des 360°-Verfahrens wird in den nächsten Abschnitten der aktuelle Forschungsstand zu Annahmen und Befunden zur Führungsbeurteilung über 360°-Verfahren aufgearbeitet. Zentrale Fragestellungen werden abgeleitet. Für eine detailliertere Beschreibung der historischen Entwicklung der 360°-Beurteilung und ihrer Vorformen wird auf Hedge, Boorman und Birkeland (2001) verwiesen.

2.2.4 Führungsbeurteilung über 360°-Verfahren

Mit 360°-Instrumenten¹⁴ bewerten mehrere Einschätzer Verhalten und Leistung einer Zielperson (meist Führungskraft) in der Regel über standardisierte Fragebogen. Befragt werden vorwiegend vier Quellen (Mount et al., 1998b; siehe Abbildung 3): Kollegen auf gleicher Hierarchieebene (Kollegenurteil; KU), Mitarbeiter der Zielperson (Mitarbeiter-

14 Alternative Begriffe: „360°-Feedback“, „Mehrquellenfeedback“, im angloamerikanischen Sprachraum: „multirater feedback“, „multisource feedback“.

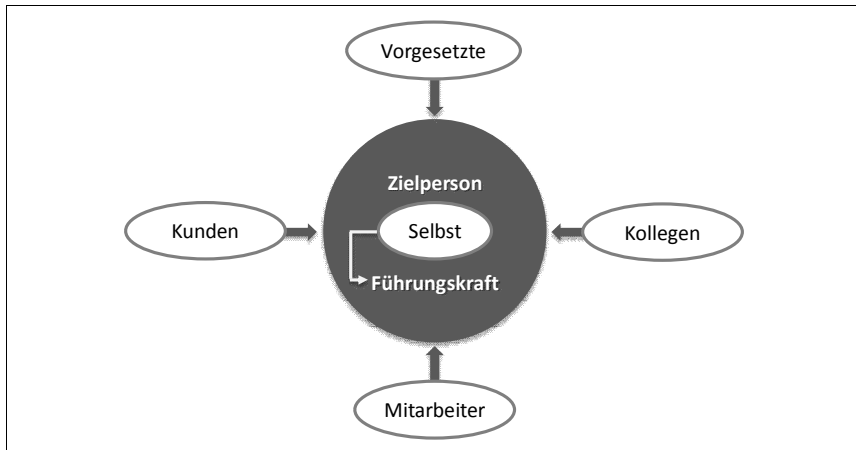


Abbildung 3: Quellen der vollständigen 360°-Beurteilung

urteil; MU), der Vorgesetzte (Vorgesetztenurteil; VU) sowie die Zielperson selbst (Selbsturteil; SU). Für das KU und MU werden üblicherweise jeweils vier bis sechs Personen befragt (London & Smither, 1995). In die Befragung werden seltener außerdem interne oder externe Kunden aufgenommen. Wie in der Einleitung begründet, wird diese Beurteilergruppe aufgrund ihrer relativen Seltenheit in der vorliegenden Arbeit nicht berücksichtigt. Die wesentlichen Ergebnisse sind jedoch im Anhang (Anhang 1) zusammengefasst.

Dem 360°-Verfahren liegt die Annahme zugrunde, dass die Quellen unterschiedliche Perspektiven haben und so zum Teil voneinander unabhängige, sich ergänzende Information beitragen (Borman, 1974, 1991; Bozeman, 1997; Conway & Huffcutt, 1997; Lance et al., 2008; Mount et al., 1998; Tornow, 1993). Nach Waldmann (2003) haben 360°-Beurteilungsinstrumente zwei Hauptaufgaben:

1. interindividuelle Differenzierung zwischen Führungskräften in Bezug auf Führungsmerkmale (Ziel: Vergleich zwischen Führungskräften) und
2. intraindividuelle Differenzierung zwischen den relativen Ausprägungen einzelner Führungsmerkmale der jeweiligen Führungskraft (Ziel: individuelles Profil von Stärken und Schwächen).

Die erste Aufgabe ist entscheidend für administrative Zwecke, wie eine Entscheidung über Lohn oder Beförderung, sowie für Selektionsfragen. Die zweite Aufgabe ist relevant für die Personalentwicklung, da sich aus den Ergebnissen gezielte Förderungsmaßnahmen ableiten lassen.¹⁵ Der Einsatz eines 360°-Verfahrens folgt dabei in der Regel einem gewissen Prozessablauf (vgl. Abschnitt 3.2 zum Prozess der Datenerhebung), der die Rückmeldung der Ergebnisse an die beurteilte Person umfasst (Birkhan & Ringelband, 2013). Die Rückmeldung der Ergebnisse ist fester Bestandteil der Leistungsbeurteilung (Lohaus,

¹⁵ Eine ausführliche Darstellung der Personalentwicklung gibt Sonntag (2006).

2009) und ist bei dem 360°-Verfahren von zentraler Bedeutung (Lepsinger & Lucia, 2009).

Seit den 1990er-Jahren nahm der Einsatz von 360°-Beurteilungen stetig zu (Church & Bracken, 1997; Conway, et al. 2001; Hazucha et al., 1993; van Hooft et al., 2006; London & Beatty, 1993; London & Smither, 1995; Scherm & Sarges, 2002)¹⁶. In der Personalentwicklung ist das 360°-Instrument zur favorisierten Methode geworden (Church & Bracken, 1997). Die Entwicklung der Führungskraft ist das häufigste Einsatzgebiet des Instruments (Scherm, 2013b). Zunehmend (innerhalb der letzten fünfzehn Jahre) wird das Verfahren auch für administrative Zwecke (Entscheidungen über Gehaltsentwicklung, Beförderung oder Kündigung) genutzt (van Hooft et al., 2006; London, 2001; Schleicher & Goff, 2003)¹⁷ und ist mittlerweile ein wesentlicher Bestandteil der Leistungsbeurteilung (Harris & Kuhnert, 2008). Auch wird die 360°-Beurteilung zur Einschätzung des Führungspotenzials eingesetzt (Scheinecker & Wallner, 2003), was in der Literatur auch durchaus Unterstützung findet (z. B. Sarges & Westermann, 2013).¹⁸ Eine ausführliche Beschreibung der Anwendung von 360°-Verfahren in Organisationen geben Scherm und Sarges (2002). Eine der umfassendsten Ratgeber zur Anwendung liefern Lepsinger und Lucia (2009).

Als Beurteilungsinstrument unterscheidet sich das multiperspektivische Verfahren von der klassischen Beurteilung durch den Vorgesetzten vor allem durch den Einbezug weiterer Quellen (Mount et al., 1998b, s. Abbildung 4). Ein anderes Unterscheidungsmerkmal zum VU ist die Anonymität von zumindest KU und MU.¹⁹

Berichtet werden in der Regel Mittelwert und Standardabweichung pro Quelle (Church & Bracken, 1997). Diese sind auf Item- und Dimensionsebene sichtbar gegenübergestellt (London & Smither, 1995). Empirische Befunde zeigen, dass die Aggregation der Einzelwerte die Genauigkeit des Urteils erhöhen kann (Cronbach, Gleser, Nanda & Rajaratnam, 1972). Über die Reduktion von Beobachtereffekten und Messfehlern durch die Aggregation der Einzelwerte pro Quelle wird dabei eine reliablere Messung erreicht (Conway & Huffcutt, 1997; Latham & Wexley, 1994; Scullen et al., 2000; Waldmann, 2003). Die Reliabilität von 360°-Verfahren wird im Folgenden betrachtet.

Als Indikatoren der Reliabilität der Urteile werden in der Regel die interne Konsistenz und die Test-Retest-Reliabilität verwendet. Die psychometrische Qualität von 360°-Instrumenten wird generell kritisch betrachtet (z. B. Mount et al., 1998; Murphy, 2008). Die interne Konsistenz der Skalen ist oft hoch bis sehr hoch (entsprechende Werte berichten z. B. Atwater, Waldman, Atwater und Cartier in 2000; Bradley in 2006 sowie Waldmann in 2003). Unzureichend zeigt sich aber bei vielen Instrumenten die Test-Retest-Reliabilität (Van Velsor & Leslie, 2001).

Über den genauen Fehleranteil der Messungen durch 360°-Verfahren geben Varianzanalysen Aufschlüsse. Wie sich in diversen Studien gezeigt hat, ist der Großteil der Va-

16 Inzwischen, zwischen 2006/2007 und 2009 wurde eine gleichbleibende, moderate Nachfrage auf dem amerikanischen Markt beobachtet (Lepsinger & Lucia, 2009).

17 Die meisten Unternehmen nennen als Grund für die Einführung eines 360°-Instruments, organisationale Entscheidungen glaubwürdiger zu machen (Ewen & Edwards, 2001).

18 Neuerdings wird die 360°-Beurteilung auch zur Evaluation des Erfolgs von Personalentwicklungsmaßnahmen eingesetzt (z. B. Tyson & Ward, 2004).

19 Vor- und Nachteile anonymer Beurteilung beschreiben z. B. Antonioni und Woehr (2001).

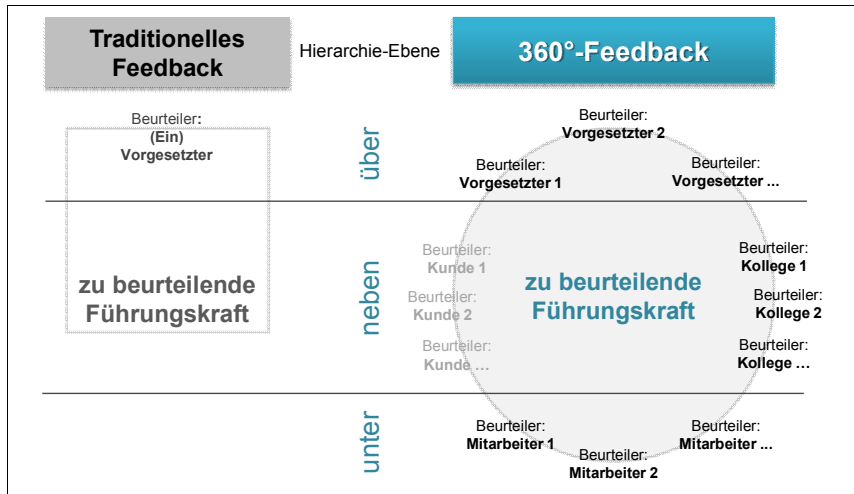


Abbildung 4: 360°-Beurteilung mit Hierarchieebenen im Vergleich zum traditionellen Bewertungsverfahren (das Kundenurteil ist aufgrund seiner relativen Seltenheit grau unterlegt)

rianz in 360°-Urteilen (trotz Urteilsaggregation; Fletcher, Baldry & Cunningham-Snell, 1998) auf Attribute der Einschätzer, Beobachtereffekte, Einschätzer- und Eingeschätzter-Interaktionseffekte sowie Fehler zurückzuführen (Greguras et al., 2003; Mount et al., 1998b; Scullen et al., 2000; Trost, 2001; Waldmann, 2003). Scullen und Kollegen (2000) beziffern die Varianz, die auf die tatsächlichen Unterschiede im Verhalten der Zielperson zurückgeht für das MU auf 14 %, für das VU auf 27 bis 32 % und für das KU auf 27 bis 29 %.²⁰ Waldmann (2003) kommt zu einem ähnlichen Ergebnis (33 % der Varianz im KU und MU).

Ein Teil der übrigen Varianz entstammt dem sogenannten Quelleneffekt (Scullen et al., 2000). Quelleneffekte beschreiben einen Einfluss der Quelle auf das Urteil. Im nächsten Abschnitt wird der Quelleneffekt näher erläutert. Hierzu wird zunächst auf die Interrater-Reliabilität als zentrales Maß zur Bestimmung von Quelleneffekten eingegangen (vgl. auch Abschnitt 2.2.3.2). Die Interrater-Reliabilität erfasst die relative Konsistenz der Urteile unterschiedlicher Einschätzer und damit die Vergleichbarkeit der relativen Einstufungen (LaBreton & Senter, 2008).

2.2.4.1 Urteilsübereinstimmung und Quelleneffekte

Metaanalytisch wurde für die Urteile innerhalb von Quellen eine niedrige Interrater-Reliabilität (anhand jeweils zweier Urteile²¹) ermittelt: .37 für das KU, .30 für das MU und .50

20 Die Ergebnisse basieren auf der Analyse jeweils eines Urteils. Die Mittelung mehrerer Urteile innerhalb einer Quelle führt zu einer Erhöhung des wahren Anteils (Scullen et al., 2000).

21 Durch die Aggregation von mehr als zwei Einschätzern in der 360°-Beurteilung resultieren höhere Interrater-Reliabilitäten. Conway und Huffcutt (1997) schätzen über die Spearman-Brown-Formel, dass mit drei

Tabelle 8: Übersicht über die metaanalytisch ermittelte Interrater-Reliabilität innerhalb und zwischen Quellen.

Quelle	VU			KU			MU			SU		
	1	2	3	1	2	3	1	2	3	1	2	3
Studie												
VU	.50	.52	/									
KU	.34	/	.48	.37	.42	/						
MU	.22	/	/	.22	/	/	.30	/	/			
SU	.22	/	.22	.19	/	.24	.14	/	/	/	/	/

Anmerkung: Abgebildet ist die Interrater-Reliabilität innerhalb und zwischen Quellen beim Vergleich des Urteils je zweier Einschätzer. Zahlen entnommen: 1: Metaanalyse von Conway und Huffcutt (1997); 2: Viswesvaran et al., (1996); 3: Metaanalyse von Harris und Schaubroeck (1988), Werte mit korrigiertem Stichprobenfehler.²²

für das VU (Conway & Huffcutt, 1997). Bedeutsam für die Fragestellung der vorliegenden Arbeit ist, dass die Interrater-Reliabilität der Urteile innerhalb einer Quelle höher ausfällt als jene zwischen Quellen (Conway, 1996; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; LeBreton et al., 2003; Mabe & West, 1982). Dies ist neben den varianzanalytischen Befunden ein zusätzlicher – und in der Literatur der zentrale – Hinweis auf Quelleneffekte. In Tabelle 8 sind die Interrater-Reliabilitäten im Vergleich aufgeführt.

Quelleneffekte werden auch über Multi-Trait-Multi-Method-Analysen sowie über konfirmatorische Faktorenanalysen bestätigt, indem sich ein signifikanter Methodeneffekt der Quelle ergibt. Einen Überblick über die relevanten Befunde geben Lance und Kollegen (2008). Anzumerken ist, dass die Betrachtung der Quellen als unterschiedliche Methoden in diesem Kontext allerdings fraglich ist, da die Messung anhand der gleichen Methode (üblicherweise der gleichen Skalen) erfolgt.

Nach den Berechnungen von Scullen und Kollegen (2000) klären Quelleneffekte im VU 11 %, im MU 15 % und im KU 0 % der Varianz auf. Waldmann (2003) bestätigt die ungefähre Höhe der Quelleneffekte im MU (13 %) und identifiziert ergänzend 13 % im KU.²³ Effekte der Quelle sind somit zwar geringer als die zuvor berichteten individuellen Beurteilereffekte, für das MU sind sie allerdings ebenso hoch wie der erwähnte Anteil wahrer Varianz, der das Verhalten der Führungskraft abbildet. Die Effekte stellen somit einen wichtigen Einflussfaktor dar, welcher zugleich der Grund ist, überhaupt unterschiedliche Quellen zu befragen (Borman, 1997; Murphy & Cleveland, 1991).

Vorgesetzten, fünf Kollegen oder sechs Mitarbeitern eine Interrater-Reliabilität von jeweils .70 pro Quelle erreicht werden kann.

- 22 Um weitere Artefakte (z. B. Reliabilitätsmängel innerhalb der Quellen) korrigiert, betragen die Werte .62 für VU und KU, .36 für KU und SU sowie .35 für VU und SU (Harris & Schaubroeck, 1988).
- 23 Weitere Werte veröffentlichten Conway und Huffcutt (1997). Sie schätzten die Höhe der Quelleneffekte, indem sie die Interrater-Reliabilität zwischen Quellen durch jene innerhalb einer Quelle teilten. Hiernach sind 43 % der Varianz zwischen MU und VU, 34 % zwischen MU und KU und 21 % der Varianz zwischen VU und KU auf Perspektivunterschiede zurückzuführen. Nach Waldmann (2003) führt die Methode jedoch zu einer Überschätzung.

Um die Unterschiede in den Urteilen verschiedener Quellen interpretieren zu können, müssen zunächst zwei Aspekte bekannt sein: die Ursache der Effekte (wodurch kommt es zu Quelleneffekten?), auf die bereits im vorherigen Kapitel generell eingegangen wurde, und der Ausdruck der Effekte (worin genau unterscheiden sich die Urteile verschiedener Quellen/was wird über die verschiedenen Quellen gemessen?). Um die Fragen zu beantworten, werden in den folgenden Abschnitten bestehende Annahmen herausgearbeitet, strukturiert und empirischen Befunden gegenübergestellt.

2.2.4.2 Erklärung von Quelleneffekten

Harris und Schaubroeck (1988) finden in der Literatur zwei Erklärungsansätze zu Perspektivitätseffekten, die auf alle vier Quellen Bezug nehmen. Der erste Ansatz erklärt Quelleneffekte durch Unterschiede zwischen Hierarchieebenen. Innerhalb dieses Ansatzes wird entweder vermutet, dass Einschätzer auf verschiedenen Hierarchieebenen einzelne Aspekte der Führungsleistung unterschiedlich gewichten, was zu abweichenden Gesamturteilen führt (siehe Klimoski & London, 1974; Zammuto, London & Rowland, 1982). Oder es wird vermutet, dass sich Einschätzer auf verschiedenen Hierarchieebenen nicht nur hinsichtlich der Gewichtung einzelner Aspekte der Führungsleistung unterscheiden, sondern auch hinsichtlich deren Definition (siehe Borman, 1974; Harris & Schaubroeck, 1988; Landy et al., 1976). Daher bestünden nicht nur Unterschiede im Gesamturteil, sondern auch in den Urteilen pro Dimension. Der zweite Ansatz führt die Quelleneffekte auf Unterschiede in Beobachtungsmöglichkeiten zurück. Dieser wird weiter unten erläutert.

Die Erklärung der Unterschiede nur über die Hierarchieebene ist fragwürdig. Sie impliziert die Unterscheidung von nur drei Perspektiven (1. Mitarbeiter, 2. die Führungskraft selbst und Kollegen und 3. der Vorgesetzte). Da alle vier Quellen sich im Urteil voneinander unterscheiden, bieten Hierarchieeffekte keine ausreichende Erklärung. Die Übereinstimmung der Urteile von Einschätzern innerhalb einer Hierarchieebene (SU und KU) ist zudem nicht höher als jene von Einschätzern verschiedener Hierarchieebenen (siehe Übersicht über die Beurteilerübereinstimmung in Tabelle 8). Da die Konkordanz zwischen Fremdbeurteilungen jedoch generell höher ist als jene zwischen Selbst- und Fremdbeurteilung (Atwater & Yammarino, 1992; Mabe & West, 1982), ist die Aussagekraft dieses Vergleichs zwar gering. Ein eindeutigerer metaanalytischer Vergleich der Urteilsübereinstimmung zwischen Selbst- und Kollegeneinschätzung mit jener zwischen SU und VU (Harris & Schaubroeck, 1988) kommt aber zum gleichen Ergebnis. Die Hierarchieebene erklärt Quelleneffekte unzureichend.

Borman (1979) berichtet drei mögliche Gründe für Quelleneffekte. Wie auch Harris und Schaubroeck (1988) beschreibt der Autor, dass durch die Quellen entweder zwar die gleichen Aspekte beobachtet, diese aber unterschiedlich gewichtet werden, oder dass aus der jeweiligen Perspektive (Kollege, Vorgesetzter, Mitarbeiter) partiell unterschiedliche Verhaltensweisen sichtbar werden (Unterschiede in den Beobachtungsgelegenheiten). Als weiteren Grund fügt Borman Unterschiede in der Aufmerksamkeitsausrichtung hinzu: die Quellen beobachten womöglich die gleichen Verhaltensweisen, achten jedoch verstärkt auf jeweils unterschiedliche Aspekte der Leistung.

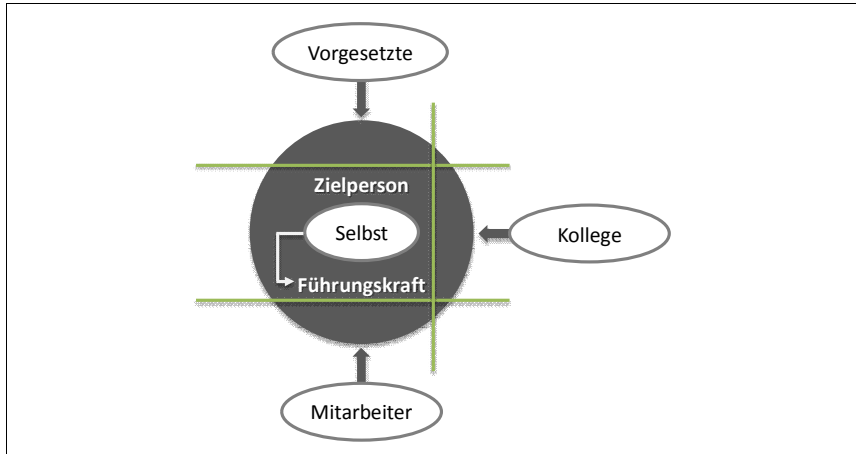


Abbildung 5: Die unterschiedlichen Perspektiven der 360°-Quellen

Unterschiede in den Beobachtungsgelegenheiten bedeuten, dass je nach Perspektive unterschiedliche Teilaspekte des Verhaltens der Zielperson sichtbar sind. Folglich kann für eine Quelle ein anderer Ausschnitt der fraglichen Verhaltensweise beobachtet werden. So ist beispielsweise nur für den Mitarbeiter beobachtbar, ob die Führungskraft Fehlverhalten der Mitarbeiter anspricht und mit dem Mitarbeiter gemeinsam unter Berücksichtigung seiner Sichtweise erarbeitet, wie dies in Zukunft zu verhindern ist (der Mitarbeiter beurteilt hier den Aspekt Einbezug der Sichtweise des Mitarbeiters). *Unterschiede in der Aufmerksamkeitsausrichtung* bedeuten dagegen, dass Einschätzer aus verschiedenen Quellen bestimmte Verhaltensaspekte unterschiedlich stark beachten. So würde der Vorgesetzte, wenn er eine ähnliche Situation beobachtete, möglicherweise nur darauf achten, ob das Fehlverhalten angesprochen wurde (Beurteilen eines Aspekts der Konfrontationsfähigkeit), während Kollegen darauf achten würden, ob die Lösungsfindung unter Einbezug des Anderen passiert (Beurteilen eines Aspekts der Teamorientierung).

Das Vorliegen von Unterschieden in Beobachtungsgelegenheiten und der Aufmerksamkeitsausrichtung zwischen Einschätzern aus unterschiedlichen Quellen impliziert, dass bestimmtes Verhalten jeweils von einer Quelle besser beurteilt werden kann als von einer anderen. Die Vermutungen zu Perspektivunterschieden und resultierenden Unterschieden im Urteilsverhalten, die auf Basis der Annahmen und Befunde zur vorwissenschaftlichen Personbeurteilung formuliert wurden, werden hierdurch gestützt. Abbildung 5 illustriert die Perspektivunterschiede.

Im Folgenden sollen die genannten Untergliederungen der Annahmen differenziert und erweitert werden. Insgesamt liegt eine Vielzahl von Ansätzen und Einzelannahmen zu diesem Thema in der Literatur vor. Diese lassen sich in fünf Ansätze gliedern.

1. Der Effekt von Quellenunterschieden geht auf Unterschiede in der *Konzeptualisierung von Leistung* zurück (Campbell & Lee, 1988; Thornton, 1980). Der Effekt der Unterschiede in der Konzeptionalisierung äußert sich sowohl im Gesamtergebnis als auch in den einzelnen Dimensionen, da Leistung verschieden definiert wird (Borman, 1974;

Harris & Schaubroeck, 1988; Landy et al., 1976). Auch das Verwenden unterschiedlicher Standards in der Bewertung der Führungskraft je nach Quelle kann eine Rolle spielen (Mount et al., 1998b).

Betrachtet man implizite Theorien zur Führungseignung, werden Unterschiede in der Konzeptualisierung gefunden (Nachweis von Unterschieden zwischen Mitarbeitern und Vorgesetzten: Neubauer, 2000). Unterschiede in der Konzeptualisierung implizieren Unterschiede in der Wahrnehmung und Interpretation beobachteter Verhaltensweisen wie auch in ihrer Gruppierung und Bewertung. Dies greift vor auf die unter Punkt 4 geschilderte Vermutung Pennys (2003), dass Unterschiede im Urteil der Quellen etwas über deren Wahrnehmung aussagen. An dem eingeführten Beispiel würde sich dies wie folgt gestalten. Kollegen betrachten das Verhalten (Ansprechen von Problemen und Einbezug Anderer bei Lösungsfindung) möglicherweise unter dem Aspekt der Teamorientierung. Den Einbezug Anderer bei der Lösungsfindung rechnen sie der Führungskraft als Zeichen der Teamorientierung an und geben ein entsprechend gutes Urteil. Der Einbezug durch die Führungskraft könnte für Mitarbeiter dagegen ein Indikator für Mitarbeiterorientierung sein. Von den Mitarbeitern erhielt die Führungskraft folglich ein gutes Urteil diesbezüglich. Der Vorgesetzte wiederum könnte das gleiche Verhalten als einen Indikator der Unselbstständigkeit in der Lösungsfindung interpretieren und die Beobachtung in ein entsprechend niedrigeres Urteil der Selbstständigkeit einfließen lassen.

2. Quelleneffekte gehen auf Unterschiede in den *Beobachtungsmöglichkeiten* und damit auf das Vorliegen unterschiedlicher *Informationen* zurück („Distinctiveness“, Conway & Huffcutt, 1997; Ansatz der „ökologischen Perspektive“, Lance, Baxter & Mahan, 2006; Lance, Teachout & Donnelly, 1992; Latham & Wexley, 1982; Lawler, 1967; „Perspektivität“, Mount et al., 1998b; Murphy & Cleveland, 1995; „Stelleneffekte“, Waldmann, 2003).

Annahme ist, dass unterschiedliche Einschätzer nur teilweise überlappende Bereiche der Verhaltensweisen der Zielperson beobachten (Lance et al., 1992). Argumente, die diese Annahme stützen, ziehen die Hierarchieebene als Erklärung heran oder formulieren alternativ Einzelannahmen zu Unterschieden in den Beobachtungsgelegenheiten einzelner Quellen. Diese gründen häufig auf der Interaktionshäufigkeit zwischen Führungskraft und Quelle.²⁴

3. Quellenunterschiede gehen auf Unterschiede in der *Aufmerksamkeitsausrichtung* zurück. Der Effekt ist auf Ebene 1 der Urteilsbildung nach dem Modell von Brandstätter (s. Abbildung 2) anzuordnen, dem Wahrnehmen von Verhalten. Der Ansatz rollenspezifischer Effekte schreibt diese den Unterschieden in den betrieblichen Rollen der Einschätzer zu (Bozeman, 1997; Humphrey, 1985). Ein ähnlicher Ansatz beschreibt Unterschiede in der Aufmerksamkeitsausrichtung sowie in der *Gewichtung* einzelner Aspekte des Führungsverhaltens und führt diese auf einen Unterschied in der Beziehung zwischen Einschätzer und Führungskraft zurück. Danach würden Vorgesetzte in ihrer Be-

24 Ein Beispiel ist die Annahme einer besonders niedrigen Urteilsübereinstimmung zwischen Vorgesetzten und Kollegen, da Kollegen mehr Gelegenheiten zur Beobachtung der einzuschätzenden Person haben als Vorgesetzte (Latham & Wexley, 1982; Lawler, 1967). Diese Annahme ist empirisch nicht bestätigt (vgl. Tabelle 8). Es ist allerdings anzumerken, dass die Untersuchung des Unterschieds zwischen MU und VU möglicherweise lohnenswerter wäre, da nach Mintzberg (1975) Mitarbeiter noch deutlich mehr Gelegenheiten zur Beobachtung der Führungskraft haben als Kollegen (vgl. Abschnitt 2.1.2).

ziehung zur Führungskraft hauptsächlich auf Arbeitsergebnisse achten, während für Mitarbeiter zwischenmenschliche Aspekte entscheidend in der Beziehung zur Führungskraft sind (Bradley, 2004). Der Effekt von Quellenunterschieden äußert sich hier im Gesamtergebnis, indem Dimensionen unterschiedlich gewichtet werden (Klimoski & London, 1974; Zammuto, London & Rowland, 1982). Der Effekt ist auf Ebene 2 der Urteilsbildung nach dem Modell von Brandstätter anzuordnen, der Eindrucksbildung.

Dem rollenbasierten Ansatz liegt die Annahme zugrunde, dass Wahrnehmung und Aufmerksamkeit der Einschätzer über betriebliche Rollen der einzelnen Positionen gesteuert werden (Bozeman, 1997; Humphrey, 1985). Die Einschätzer fokussieren in ihrer Beobachtung auf das, was sie gemäß ihrer Rolle als am bedeutsamsten einstufen. Dies könnte z. B. für Kollegen Hilfsbereitschaft sein und für den Vorgesetzten Zuverlässigkeit. Bewertet würden somit jeweils unterschiedliche Aspekte der beruflichen Leistung (Bozeman, 1997).

In den meisten 360°-Instrumenten sind die Fragen, die den unterschiedlichen Quellen vorgelegt werden, zwar identisch (Waldmann, 2003) und spezifisch im Bezug auf einzelne Leistungsaspekte. Ein solches Item könnte lauten „Er/sie bezieht seine/ihre Mitarbeiter in Entscheidungen mit ein“. Effekte der Rolle, wie sie von Bozeman (1997) bezeichnet werden, sind allerdings dennoch eine zu erwartende Einflussgröße, da Forschung zur Informationsverarbeitung darauf hinweist, dass Bewertungen vielmehr basierend auf vorhandenen globalen Urteilen zu der Person als auf Basis konkreter Beobachtungen vorgenommen werden (Feldman, 1981).

Nach Bradley (2004) sowie Bradley, Allen, Hamilton, Filgo (2006) setzten Einschätzer verschiedener Quellen unterschiedliche Schwerpunkte in der Beurteilung der Führungskraft. Für Vorgesetzte haben nach Bradley (2004) beispielsweise Ergebnisse, die die Führungskraft (gemessen an Unternehmenszielen) erzielt, eine höhere Bedeutung als zwischenmenschliche Aspekte im Verhalten, während für Mitarbeiter der umgekehrte Fall angenommen wird. Eine vertiefende Betrachtung dieser Annahme und Referenz auf empirische Befunde erfolgt im Abschnitt zu merkmalspezifischen Quelleneffekten, 2.2.4.5. Ein unterschiedliches Gewichten einzelner Dimensionen der Führungsleistung in der Beurteilung durch die Quellen wird von weiteren Autoren angenommen (z. B. Pulakos, Schmitt & Chan, 1996; Tsui & Ohlott, 1988).

4. Quellenunterschiede gehen auf Unterschiede in der *Interpretation der Itemformulierung* zwischen den Einschätzern zurück (Penny, 2003).

Penny (2003) nimmt Unterschiede in der Interpretation von Itemformulierungen an, die er auf (mit der Hierarchieebene) variierende Komplexität der Umgebung des Einschätzers zurückführt. Penny (2003) stellte beispielsweise fest, dass ein Item zu Konfliktbereitschaft („Accepts conflicts as inevitable and does not shy away from them“) von Mitarbeitern höher als von Vorgesetzten beurteilt wurde. Eine mögliche Erklärung wäre, dass Mitarbeiter hierbei den Aspekt beurteilten, wie die Führungskraft mit bestehenden Konflikten umgeht, während der Vorgesetzte der Führungskraft hier zusätzlich beurteilt, inwiefern Entscheidungen getroffen werden, die Konflikte auslösen können. Diese Annahme ist im Wesentlichen eine Auslegung der Problematik fehlenden „shared meanings“ (s. Abschnitt 2.2.3.1). Penny folgert, dass eine 360°-Rückmeldung Auskünfte darüber geben kann, wie die einzelnen Quellen die Führungsaufgabe verstehen und bewerten. Die Schlussfolgerung des Autors ist eine wesentliche: Gleiche Verhaltensweisen können unterschiedlich inter-

pretiert werden. Die Annahme, dass das Urteil etwas darüber aussagt, wie die Quelle die Führungsaufgabe versteht, ist bereits in dem Ansatz von beispielsweise Harris und Schaubroeck (1988) impliziert, nach dem Quellen die Leistungsdimensionen unterschiedlich definieren bzw. unterschiedliche Konzeptualisierungen der Führungsdimensionen haben.

5. Der Effekt von Quellenunterschieden geht auf *tatsächliche Unterschiede im Verhalten* der Führungskraft gegenüber verschiedenen Quellen zurück (Penny, 2003).

Grundsätzlich ist die Annahme von tatsächlichen Unterschieden im Verhalten der Führungskraft als Ursache der Quellenunterschiede durchaus vertretbar²⁵, wie in Abschnitt 2.2.3.4 aus dem Konzept der verhandelten Identität abgeleitet wurde. Die verschiedenen Quellen interagieren mit der Führungskraft in je andersartigen Situationen und in einem anderen Kontext, was verschiedene Verhaltensweisen der Führungskraft hervorrufen könnte. Es liegen unterschiedliche Interaktionsziele vor und die Führungskraft interagiert mit den Einschätzern in unterschiedlichen Rollen (Lance et al., 2008; Tsui 1984).

Die beschriebenen Ansätze lassen sich, sortiert nach Erklärungsausgangspunkt, auf fünf das Wesentliche zusammenfassende Erklärungen reduzieren. Die dargestellten Gründe (Tabelle 9) bieten die Basis für die Ableitung der Hypothesen der vorliegenden Arbeit.

In der vorliegenden Arbeit wird davon ausgegangen, dass Quelleneffekte über mehrere Aspekte bewirkt werden: Es bestehen tatsächliche Unterschiede im Verhalten der Führungskraft den Personen der unterschiedlichen Quellen gegenüber. Innerhalb des Verhaltensspektrums der Führungskraft werden den Quellen zudem teilweise unterschiedliche Ausschnitte sichtbar. Zusätzlich lenken die Quellen ihre Aufmerksamkeit auf zum Teil unterschiedliche Aspekte des Verhaltens, die in der Folge nicht nur genauer beurteilt werden, sondern auch ein entsprechend höheres Gewicht in der Beurteilung einnehmen. Führungsleistung als solche wird außerdem je nach Perspektive unterschiedlich definiert, Verhalten unterschiedlich kategorisiert und an unterschiedlichen Maßstäben bewertet.

Die Ansätze stützen die Vermutung, dass verschiedene Aspekte des Führungsverhaltens durch die einzelnen Quellen jeweils unterschiedlich gut beurteilt werden können. Die nächsten Abschnitte stellen entsprechende Vermutungen und relevante empirische Befunde zu den Zusammenhängen des VU, KU und MU mit externen Variablen dar.

25 Penny (2003) argumentiert, indem er aus der von Hershey und Blanchard (1969) angenommenen Notwendigkeit der Anpassung des Führungsstils an die Situation (den Reifegrad) des Mitarbeiters ableitet, dass Unterschiede in der Beurteilung der Zielperson zwischen den Beurteilergruppen als Zeichen für ein unterschiedliches und damit der Situation angemessenes Verhalten gewertet werden können. Obwohl die Vermutung zutreffend erscheint, die Anwendung der Annahmen von Hershey und Blanchard (1969) ist kritisch zu betrachten, da diese sich nur auf den Reifegrad der Mitarbeiter bezogen. Unterschiede in der Beurteilung der Führungskraft durch die einzelnen Mitarbeiter könnten so jedoch erklärt werden, wie London und Smitther (1995) vermuten.

Tabelle 9: Übersicht über die fünf Hauptgründe für Quellenunterschiede aus der Literatur.

Gründe	Abhängig von	Autoren	Bezeichnung
1 Unterschiede in Beobachtungsmöglichkeiten (den Quellen liegen jeweils unterschiedliche Teile der Information vor)	Beobachtungsmöglichkeiten (Quelle)	Conway & Huffcutt (1997); Lance et al. (2006); Lance et al. (1992); Latham & Wexley (1982); Lawler (1967); Mount et al. (1998b); Murphy & Cleveland (1995); Waldmann (2003)	Perspektivunterschiede; Distinktheit; Perspektivität; Stelleneffekte
2 Gleiche Information liegt vor, es wird auf unterschiedliche Ausschnitte geachtet (Unterschiede in der Aufmerksamkeitsausrichtung ; anzuordnen auf Ebene 1 der Urteilsbildung nach dem Modell von Brandstätter: Wahrnehmen von Verhalten)	Rolle	Bozeman (1997); Humphrey (1985)	Rolleneffekte
3 Gleiche Information liegt vor, Konzeptionen, Definitionen von Leistung sind unterschiedlich	3a) Hierarchieebene 3b) Perspektive (Quelle)	Borman (1974); Landy et al. (1976); Harris & Schaubroeck (1988) Campbell & Lee (1988); Thornton (1980)	
4 Gleiche Information liegt vor, einzelne Aspekte werden unterschiedlich gewichtet (anzuordnen auf Ebene 2 der Urteilsbildung nach dem Modell von Brandstätter: Eindrucksbildung)	4a) Hierarchieebene 4b) Beziehung	Klimoski & London (1974); Zammuto et al. (1982) Bradley (2004); Bradley et al. (2006) Pulakos et al. (1996); Tsui & Ohlott (1988)	Hierarchieeffekte Beziehungseffekte
Tatsächliche Unterschiede im Verhalten liegen vor	Verhalten der Zielperson	Lance et al. (2008); Penny (2003)	Wahre Unterschiede
Unterschiedliche Umgebungs-komplexität führt zu unterschiedlicher Interpretation der Items	Komplexität der Umgebung	Penny (2003)	Interpretation der Itemformulierung
Gleiche Information aber unterschiedliche Bewertungsstandards liegen vor	Perspektive (Quelle)	Mount et al. (1998b)	

Anmerkung: Graue Schrift = theoretische Annahmen, die ergänzend zu den fünf Gründen betrachtet werden können (eigene Zusammenstellung).

2.2.4.3 Unterschiede zwischen den Urteilen der verschiedenen Quellen

Im Kontext der Führungsbeurteilung befasst sich die Forschung nur am Rande damit, was die Kollegen, Mitarbeiter und die Führungskraft selbst zur Beurteilung tatsächlich beitragen sollen. In der Literatur beschriebene Gründe zum Einbezug der einzelnen Quellen in die Beurteilung der Führungskraft sind stattdessen vornehmlich aus allgemeinen Problemstellungen der Organisation abgeleitet, ohne Bezug auf die Erhöhung der Genauigkeit der Beurteilung zu nehmen. Dies gilt zum Beispiel für folgende Beweggründe: über den Einbezug von Mitarbeitern der Führungskraft eine stärkere Beteiligung der Mitarbeiter zu verwirklichen (McGregor, 1957, 1959), diese zu motivieren (Lawler, 1967), die Beurteilung weniger autoritär erscheinen zu lassen (Meyer, 1991) und die Akzeptanz der Beurteilung durch die Führungskraft zu erhöhen (Bassett & Meyer, 1968; Meyer, Kay & French, 1965). Um einschätzen zu können, inwiefern der Einbezug einzelner Quellen die Genauigkeit der Beurteilung der Führungskraft erhöhen könnte, wird nachfolgend dargestellt, welche Information durch welche Quelle erfasst werden kann. Die Quellen werden zudem hinsichtlich der Genauigkeit ihres Urteils verglichen.

2.2.4.3.1 Rangreihe der Urteilsgenauigkeit der Quellen

Es existieren bisher nur wenige Studien zur Urteilsgenauigkeit der einzelnen Quellen (van Hooft et al., 2006). Die Befundlage ist lückenhaft (Übersichtsanalyse u. a. bei Atwater, Waldman & Brett, 2002). Gewählte Variablen umfassen Indikatoren der Führungsleistung (z. B. Produktivität oder Leistungsbeurteilung), Persönlichkeitseigenschaften und kognitive Fähigkeiten.

Generell werden Zusammenhänge von 360°-Urteilen und objektiven Außenvariablen nur in geringem Ausmaß festgestellt und sind nur selten substanziell (Atkins & Wood, 2002). Hazucha und Kollegen (1993) beispielsweise berichten von Zusammenhängen der 360°-Urteile mit Gehaltsentwicklung und Beförderungen, die die Höhe von .23 (für das SU, .12 für VU, KU und MU) respektive .19 (für VU, KU und MU, .09 für das SU) nicht überschreiten. Auch die Verwendung von subjektiven Außenvariablen führt zu keinen deutlich höheren Zusammenhängen. Maylett (2005) fand keine signifikanten Korrelationen der 360°-Beurteilung mit der Leistungseinschätzung durch den Vorgesetzten, Ostroff, Atwater und Feinberg (2004) fanden niedrige Korrelationen von KU und MU ($r = .22$ bzw. $r = .32$). Weitere Studien, über die Smither und Kollegen (2005) eine Übersicht geben, bestätigen die insgesamt schwache Assoziation mit Außenvariablen.

Die Befundlage zur Höhe der Assoziation der Urteile der vier Quellen mit Außenvariablen lässt nach Eichinger und Lombardo (2004) eine Rangfolge der Genauigkeit der Urteile pro Quelle über verschiedene Kriterien und Merkmale hinweg erkennen: Die höchste Genauigkeit besitzt danach das VU (Lombardo & Eichinger, 2003; Zimmermann et al., 2008), gefolgt vom KU, gefolgt vom MU (Conway & Huffcutt, 1997). Für das SU wurden bisher die wenigsten Hinweise auf Genauigkeit gefunden (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Lombardo & Eichinger, 2003).²⁶

Im ersten Kapitel wurden Erwartungen geschildert, die mit dem Einbezug der jeweiligen Quelle in die Beurteilung der Führungskraft verbunden sind. Diese werden in den

26 Die Rangfolge der Zusammenhangshöhe mit Außenvariablen ist der Rangfolge der Kontakthäufigkeit entgegengesetzt (siehe Abschnitt 2.1.2; Mintzberg, 1975). Die höheren Zusammenhänge des VU scheinen nicht Folge häufigerer Beobachtungsgelegenheiten zu sein.

folgenden Abschnitten ergänzt und empirischen Befunden zu Zusammenhängen mit Außenvariablen, Differenziertheit des Urteils und Urteilsübereinstimmung gegenübergestellt. Die Besonderheiten der einzelnen Quellen werden zunächst für das VU beschrieben, es folgen das KU und das MU. Das SU wird in Abschnitt 2.2.4.7 gesondert betrachtet. Ein Vergleich der Urteilsgenauigkeit bezogen auf die Messung einzelner Merkmale erfolgt im Abschnitt zu merkmalspezifischen Quelleneffekten (Abschnitt 2.2.4.5).

2.2.4.3.2 Das Vorgesetztenurteil

Die berichtete Rangfolge der Urteilsgüte (bezogen auf die Höhe der Zusammenhänge mit Außenvariablen) bestätigt traditionelle Annahmen, dass der Vorgesetzte die beste Gesamtübersicht über die Leistung der Führungskraft hat (vgl. Bradley, 2004). In der Einleitung wurden die Kritikpunkte an der Vorgesetztenbeurteilung bereits dargelegt. Im Folgenden wird daher nur auf jene Gründe fokussiert, die erklären, warum das Urteil des Vorgesetzten als denen der übrigen Quellen überlegen gilt.

Da Leistungsbeurteilung zu den Aufgaben des Vorgesetzten gehört, hat dieser in der Regel die größte Erfahrung und Übung darin (Antonioni & Park, 2001). Erfahrene Vorgesetzte sollten auch über einen gut ausgebildeten Vergleichsmaßstab verfügen (Borman, 1991). Häufig hat der Vorgesetzte außerdem die Position der zu beurteilenden Führungskraft selbst innegehabt und kennt damit die Verantwortlichkeiten der Stelle (Mount et al., 1998b). Es ist zudem anzunehmen, dass der Vorgesetzte gezielt Situationen aufsucht, die diagnostisch wertvoll sind sowie sich Rückmeldungen zu der Führungskraft von Anderen einholt. Die Gewissenhaftigkeit, mit der die Beurteilung vorgenommen wird, sollte beim Vorgesetzten besonders hoch sein, da er für die Ergebnisse seiner Beurteilung am ehesten in die Verantwortung genommen wird (zumal die Einschätzung des Vorgesetzten nicht anonym ist).

Die Annahme überlegener Genauigkeit der Vorgesetztenbeurteilung wird über eine Reihe weiterer empirischer Befunde gestützt.²⁷ Das VU zeigt die höchste Beurteilerübereinstimmung (Conway & Huffcutt, 1997; Mount et al., 1998b) und den geringsten Anteil an Antworttendenzen (Scullen et al., 2000). Es zeigt die im Vergleich zu den anderen Quellen höchsten Zusammenhänge mit externen Variablen, wie z. B. Persönlichkeitsmerkmalen (van Hoof et al., 2006), dem Gesamtwert eines Assessment Centers (Atkins & Wood, 2002) und der Zielerreichung (Zimmermann et al., 2008).

2.2.4.3.3 Das Kollegenurteil

Da Kollegen sich überwiegend ebenfalls in Führungsfunktionen befinden, kann von einer gewissen Geübtheit in der Leistungsbeurteilung ausgegangen werden. Der Einbezug von Kollegen verspricht deshalb einen zusätzlichen Beitrag zur Beurteilung durch den Vorgesetzten, da Kollegen häufiger als der Vorgesetzte mit der Führungskraft interagieren (Mintzberg, 1975) und damit mehr Beobachtungsgelegenheiten haben. Auf diese Weise beobachten sie auch jene Situationen, die dem Vorgesetzten nicht zugänglich sind (Lawler, 1967); beispielsweise erleben Kollegen durch die Zusammenarbeit mit der Führungskraft deren Verhalten im Team direkt (Lepsinger & Lucia, 2009; Waldmann, 2003). Kol-

²⁷ Zu erwähnen ist, dass das VU meist aus nur einem Urteil besteht. Da das KU und MU in der Regel ein Mittelwert mehrerer Einzelurteile darstellt, mag es sein, dass die Güte des VU im Verhältnis unterschätzt wird. (Dies gilt ebenso für das SU.)

legen können zudem Führungskraft-Mitarbeiter-Interaktionen beobachten und sollten diese besonders gut beurteilen können, da sie mit den Ansprüchen der Stelle an die Führungskraft vertraut sind (Bradley, 2004). Auch Borman (1991) verweist auf die besonders große Bandbreite an Verhaltensweisen, die Kollegen beobachten können. Dies mache es für die Führungskraft schwieriger, schwache Leistungen zu verbergen. Die Führungskraft teilt Informationen über Persönliches oder Fehler möglicherweise auch eher mit Kollegen als mit dem Vorgesetzten. Gerade dies impliziert allerdings einen möglichen Rollenkonflikt zwischen objektivem Beurteiler und Vertrautem, den Krohne und Hock (2007) herausstellen. Denkbar ist jedoch auch eine mögliche intendierte Urteilsverfälschung durch Kollegen, da diese mitunter in einem Konkurrenzverhältnis zur Zielperson stehen.

Studien im militärischen Umfeld in den 1950er-Jahren zeigten optimistisch stimmende Befunde zu Reliabilität und Validität des KU (siehe z. B. Hollander, 1954). Eine Übersicht über die relevanten Studien zu Reliabilität und Validität des KU in den 1950er-Jahren geben Hedge, Borman und Birkeland (2001). Die derzeitigen Befunde aus dem Wirtschaftskontext platzieren das KU in seiner Genauigkeit insgesamt hinter dem VU (Eichinger & Lombardo, 2004), obwohl vereinzelt eine vergleichbare Vorhersagekraft ermittelt wurde (siehe z. B. Lance et al., 1992).

2.2.4.3.4 Das Mitarbeiterurteil

In einer weniger konfliktreichen Rolle befinden sich Mitarbeiter. Da Mitarbeiter die Wirkung von Führungsverhalten direkt erleben (Arnold & Kelsey, 1995), verfügen sie bei der Beurteilung über eine besondere Perspektive (Conway & Huffcutt, 1997; Murphy & Cleveland, 1995). Dies zeigt sich auch in dem im Vergleich zu den übrigen Quellen größten Effekt der Perspektive in der Varianz des Urteils (Scullen et al., 2000). Mitarbeiter sind die einzigen Einschätzer, welche die Führungskraft in ihrer Rolle als direkten Vorgesetzten beurteilen (vgl. Bozeman, 1997). Viele Führungsverhaltensweisen, insbesondere auch Führungsfertigkeiten (Borman, 1991), werden nur in der direkten Interaktion mit den Mitarbeitern sichtbar und in der Regel nur von diesen beobachtet (Lawler, 1967). Basierend auf dem Ansatz der Rolleneffekte sind Mitarbeiter darüber hinaus die einzigen Einschätzer, die die Führungskraft in ihrer Rolle als direkten Vorgesetzten beurteilen.²⁸ Möglicherweise ist dies eine Erklärung für den Befund einer besonders niedrigen Urteilsübereinstimmung des MU mit KU und VU von Conway und Huffcutt (1997). Keine andere Quelle interagiert so häufig mit der Führungskraft wie die Mitarbeiter (Mintzberg, 1975).

Gerade für das MU wäre demnach eine höhere Genauigkeit zu erwarten. Im Gegensatz zu Vorgesetzten und Kollegen sind Mitarbeiter in der Beurteilung jedoch weniger geübt und urteilen möglicherweise daher weniger differenziert (Scullen, Mount & Judge, 2003). Eine Ausnahme bildet der Befund einer im Vergleich zu anderen Quellen höheren Urteilsdifferenziertheit in der Beurteilung von Items, die Verhaltensweisen beschreiben, welche die Mitarbeiter in besonderem Maße betreffen (Maurer et al., 1998). Eine Metaanalyse (Conway et al., 2001) zeigt einige bedeutsame Zusammenhänge mit objektiven Variablen (untersucht wurden Variablen, wie der Anstieg in Verkaufszahlen, Bruttogewinn und die Reaktionszeit auf Kundenanfragen), diese fallen aber im Vergleich zu VU und KU schwächer aus. Insgesamt zeigt das MU einen größeren Anteil Einschätzer-

28 Für einen Überblick über den Einsatz des MU außerhalb der 360°-Beurteilung wird auf Kiefer (2000) verwiesen.

Effekte im Urteil als VU und KU (66 - 79 %; Scullen et al., 2000). Auch zeigt das MU im Vergleich zu VU und KU die niedrigste Urteilsübereinstimmung der Einzelurteile²⁹ (Conway & Huffcutt, 1997). Die Ausnahme einer höheren Beurteilerübereinstimmung im MU verglichen mit dem KU zeigte Waldmann (2003) bezüglich der Beurteilung von Misserfolgsmerkmalen im Verhalten der Führungskraft (z. B. mangelndes Durchhaltevermögen der Führungskraft und fehlende Verhaltensänderung nach dem Erhalt von Feedback).

Abschließend soll darauf hingewiesen werden, dass gerade die empirischen Studien zum Zusammenhang von 360°-Urteilen mit Außenkriterien durch methodische Schwachpunkte gekennzeichnet sind. Solche Schwachpunkte betreffen Konfundierungen zwischen der abhängigen Variablen und dem VU (z. B. bei Zimmermann et al., 2008: Die abhängige Variable Zielerreichung wurde ebenfalls durch den Vorgesetzten bewertet) faktoriell nicht (z. B. bei Hooft et al., 2006 und Atkins & Wood, 2002) oder nur anhand des VU (Zimmermann et al., 2008) geprüfte Skalen, Erhebungsinstrumente mit geringem Umfang (14 Items bei Hooft et al., 2006) oder wenigen Items pro Skala sowie unzureichender interner Konsistenz (drei bis sechs Items, Cronbachs Alpha ab .49 bei Atkins & Wood, 2002) sowie kleine Stichproben ($N = 63$ bei Atkins & Wood, 2002). Festzuhalten gilt jedoch, dass Urteile aller Fremdeinschätzergruppen bedeutsame Zusammenhänge mit Außenvariablen aufweisen und diese zwischen den Quellen stark variieren (Conway & Huffcutt, 1997). Der nächste Abschnitt ergänzt die empirischen Befunde zur Güte der Urteile um Nachweise zusätzlicher Varianzaufklärung in einzelnen Kriteriumsvariablen durch KU, MU und SU in Kombination mit dem VU.

2.2.4.4 Inkrementelle Varianzaufklärung durch die Urteile einzelner Quellen

Der Beitrag einzelner Quellen in der Varianzaufklärung in Führungserfolgsvariablen über das VU hinaus ist eine der wichtigsten Fragen bezüglich der Nutzung von 360°-Instrumenten (Borman, 1974). Untersucht ist sie jedoch kaum. Die erste Untersuchung von MU und KU in Kombination mit dem VU führten Conway und Kollegen (2001) nach eigener Aussage³⁰ durch. In einer hierarchischen Regression der Werte einer metaanalytisch erstellten Korrelationsmatrix (Datenbasis sind Studien, die die Zusammenhänge zwischen den Urteilen der drei Quellen und Außenvariablen berichten) zeigte das *R-Quadrat* einen Anstieg um .03 nach Hinzunahme des KU (zum VU) auf .15. Der gleiche Anstieg resultierte nach Hinzunahme des MU zum VU. Wurden MU oder KU als jeweils dritte Variable hinzugenommen, resultierte ein weiterer Anstieg um .02.

Van Hooft und Kollegen (2006) berichten die folgenden Ergebnisse einer Regression zur zusätzlichen Varianzaufklärung durch KU und SU aus eigener Studie. Der Anteil aufgekklärter Varianz stieg bei sieben von zehn Persönlichkeitsvariablen an, wenn SU und KU zusätzlich zum VU in die Regressionsgleichung aufgenommen wurden. Die entsprechende Skala aus dem KU zeigte für drei Persönlichkeitsvariablen ein signifikantes Beta-Gewicht von bis zu .21. Die entsprechende Skala aus dem SU zeigte für fünf Persönlich-

29 Wie bereits erwähnt, ist dies möglicherweise eine Konsequenz tatsächlicher Unterschiede im Verhalten der Führungskraft gegenüber einzelnen Mitarbeitern (London & Smither, 1995; s. Abschnitt 2.2.4.2).

30 Nach Borman (1997) existiert eine frühere Studie zu inkrementeller Validität von Quellen unterschiedlicher Hierarchieebenen: Buckner (1959) fand, dass die Kombination des Urteils des direkten mit dem des nächsthöheren Vorgesetzten einen größeren Anstieg in der Korrelation mit Leistungskriterien (im Vergleich zu den Einzelurteilen) zeigte, wenn beide Urteile miteinander niedriger korreliert waren.

keitsvariablen ein signifikantes Beta-Gewicht von bis zu .30. Kritisch anzumerken ist, dass der Basisprädiktor (VU) nur mit zwei dieser Persönlichkeitsvariablen signifikant assoziiert ist. Eine weitere Studie, die diesem Bereich zuzuordnen ist, führten Atkins und Wood (2002) durch. Sie berichten eine um .10 höhere Korrelation des Mittelwerts aus VU, KU und MU ($r = .39, p < .05$) verglichen mit dem VU ($r = .29, p < .01$). Dieser Unterschied kann allerdings auch auf eine höhere Reliabilität des über die Urteile beider Quellen aggregierten Werts zurückgehen.

Im nächsten Abschnitt soll betrachtet werden, ob die Rangfolge der Zusammenhangshöhe mit Außenvariablen und der Ausprägung anderer Indikatoren der Güte der Urteile unabhängig vom gemessenen Merkmal bestehen. Ausgangsüberlegungen dazu wurden bereits eingeführt (Abschnitte 2.2.3.4 und 2.2.4.2). Die verschiedenen Quellen nehmen das Verhalten der Führungskraft unterschiedlich wahr und bewerten es unterschiedlich. Quellenunterschiede sollten daher auf der Ebene einzelner Merkmale noch deutlicher sichtbar werden. Im folgenden Abschnitt werden theoretische Annahmen sowie empirische Befunde zu dieser Frage aufgeführt.

2.2.4.5 Quellenunterschiede in der Genauigkeit der Messung einzelner Merkmale

Quellenunterschiede hinsichtlich der Genauigkeit der Messung einzelner Merkmale sind bisher wenig untersucht. Hypothesen, die in diesem Bereich formuliert werden, nehmen kaum Bezug auf existierende Theorien und beschränken sich häufig auf Einzelannahmen zu einer bestimmten Quelle oder einem bestimmten Merkmal. Zimmermann und Kollegen (2008) äußern, dass es weder empirische noch theoretische Gründe gäbe, Hypothesen zu spezifischen Unterschieden zwischen Quellen zu formulieren. Die im Folgenden beschriebenen theoretischen Annahmen und die Befundlage (basierend auf sieben relevanten Studien) legen insgesamt jedoch nahe, dass sich die Rangfolge in der Genauigkeit des Urteils je nach Merkmal unterscheidet und Merkmale ein jeweils unterschiedlich starkes Gewicht in der Gesamtbeurteilung aufweisen – die Quellen messen gemäß ihrer Perspektive (s. Abschnitte 2.2.3.4 und 2.2.4.2) Führungsleistung unterschiedlich.

Studien zu merkmalspezifischen Unterschieden der Urteilsgüte von einzelnen Quellen nutzen im Wesentlichen fünf Methoden, die in Tabelle 10 dargestellt sind. In vier der Studien wurden die erhobenen Führungskompetenzen/-verhaltensweisen Kategorien zugeordnet, die das in Abschnitt 2.2.1.3 eingeführte Konzept Consideration und Initiating Structure reflektieren (s. Tabelle 10). Das Konzept wird auch in der vorliegenden Arbeit verwendet. Ein Vorteil besteht darin, dass Consideration und Initiating Structure Gruppen von Verhaltensweisen beschreiben, die jeweils vermehrt Interaktionen mit bestimmten Personengruppen beinhalten. So bezieht sich Consideration beispielsweise vornehmlich auf Interaktionen mit Mitarbeitern. Unterschiede in den Urteilen zwischen den Quellen sind bereits aus diesem Grund zu erwarten. Im Folgenden werden Annahmen sowie empirische Befunde zu merkmalspezifischen Quellenunterschieden für das VU, anschließend für das MU und zum Schluss für das KU gegenübergestellt.

Tabelle 10: Übersicht über Studien zu merkmalspezifischen Quellenunterschieden in der Urteilsgröße.

Studie	Methode (was verglichen wird)	Merkmalszuordnung auf höherer Ebene	Hypothesengeleitet
Atkins & Wood (2002)	Zusammenhangshöhe der Urteile mit Außenvariablen	Keine (Kompetenzdimensionen)	Nein
Bradley (2004)	Relative Höhe der Faktorenladung verschiedener Dimensionen*	Ergebnis- versus Beziehungsaspekte**	Ja
Conway & Huffcutt (1997)	Interrater-Reliabilität (metaanalytisch)***	Kognitiv versus interpersonell****	Ja
Maurer und Kollegen (1998)	Antwortverhalten hinsichtlich Diskriminanz	Keine (Itemebene)	Nein
Viswesvaran und Kollegen (1996)	Interrater-Reliabilität	Kognitiv versus interpersonell	Nein
Waldmann (2003)	Beurteilerübereinstimmung	Keine (Itemebene)	Nein
Zimmermann und Kollegen (2008)	Zusammenhangshöhe der Urteile mit Außenvariablen	Initiating Structure versus Consideration	Nein

Anmerkung: *Aus der relativen Höhe der Faktorenladung verschiedener Merkmale in einer konfirmatorischen Faktorenanalyse zweiter Ordnung wird geschlossen, welche Merkmale pro Quelle einen besonders großen Einfluss auf die Gesamtbewertung haben. **Ergebnisaspekte umfassen die von der Führungskraft erreichten Arbeitsergebnisse, Beziehungsaspekte umfassen Merkmale sozialer Kompetenz und die Qualität der zwischenmenschlichen Beziehungen der Führungskraft zu den Einschätzern. ***Aus der Interrater-Reliabilität sowie aus der Beurteilerübereinstimmung wird auf die Genauigkeit des Urteils geschlossen. ****Kognitive Dimensionen umfassen Aspekte, wie visionäres Denken oder Beurteilen, interpersonelle Dimensionen Aspekte, wie Beziehungsfähigkeiten oder Consideration; (eigene Zusammenstellung).

Das Urteil des Vorgesetzten sollte vermehrt auf Initiating Structure Aspekten basieren und diese exakter beurteilen können als Consideration Aspekte (Bradley, 2004; Conway & Huffcutt, 1997; Conway et al., 2001). Insbesondere sollte der Vorgesetzte auf das Erbringen von Ergebnissen achten und diesen ein besonderes Gewicht in der Beurteilung zusprechen (Bradley, 2004; Conway et al., 2001; Zamuto, London & Rowland, 1982) sowie kognitive Dimensionen genauer beurteilen als interpersonelle (Conway & Huffcutt, 1997). Hinsichtlich der entsprechenden Unterschiede in der Faktorenladung wurde die Annahme empirisch bestätigt: Eine Faktorenanalyse zweiter Ordnung erbrachte eine höhere Faktorenladung ergebnisbezogener Skalen verglichen mit beziehungsbezogenen Skalen im VU (Bradley, 2004). Hinsichtlich der Interrater-Reliabilität wurde die Annahme tendenziell bestätigt, die Reliabilität des VU für kognitive Dimensionen war mit .47 um .03 höher als für interpersonelle .44 (Conway & Huffcutt, 1997). Ein den Erwartungen entgegengesetzter Befund zeigte sich im Zusammenhang mit der vom Vorgesetzten eingeschätzten Zielerreichung (Zimmermann et al., 2008): Das VU zu Consideration korrelierte mit Kriterienvariablen zu .42 in einem für administrative Fragen und zu .30 in einem zur

Entwicklung eingesetzten 360°-Instrument³¹, während Initiating Structure nur zu .20 bzw. -.09 korrelierte.

Bezüglich des MU wird der umgekehrte Zusammenhang angenommen. Mitarbeiter sollen zwischenmenschliches Verhalten häufiger beobachten können als Vorgesetzte (Murphy & Cleveland, 1995) sowie stärker auf den Beziehungsaspekt als auf den Ergebnisaspekt fokussieren (Bradley, 2004) – insbesondere auf unterstützendes Verhalten (Conway et al., 2001). Das MU sollte besser als alle anderen Quellen zur Messung des mitarbeiterbezogenen Führungsstils geeignet sein (Waldmann, 2003). Empirisch ist die Befundlage nicht eindeutig. Die Annahme wurde bestätigt bezüglich höherer Interrater-Reliabilität im MU (.34 bzw. .25) bei der Bewertung interpersoneller versus kognitiver Dimensionen (Conway & Huffcutt, 1997). Zwei weitere unterstützende Befunde finden sich in der Literatur. Das MU sagte Assessment Center Ergebnisse nur zu den Kompetenzen „Personen coachen“ und „Personen entwickeln“ über die entsprechenden 360°-Skalen besser vorher als die anderen Quellen (Atkins & Wood, 2002) und zeigte verglichen zum KU im Antwortverhalten eine höhere Diskriminierung der Leistung der Führungskräfte bezogen auf ein Item zum Ausmaß des Zeigens von Anerkennung für Teamleistungen (Maurer et al., 1998). Nicht bestätigt wurde die Annahme bezüglich entsprechender Unterschiede in der Höhe der Faktorenladung von Skalen mit zwischenmenschlichen versus ergebnisbezogenen Inhalten im MU (Bradley, 2004).

Ähnliche Annahmen wie für das MU finden sich auch bezüglich des KU. Hinsichtlich der Kollegen wird angenommen, dass sie der Leistung der Führungskraft weniger Beachtung schenken als den zwischenmenschlichen Aspekten (Conway & Huffcutt, 1997; Murphy & Cleveland, 1995), stärker auf Zusammenarbeit und Teamorientierung achten (Sculen et al., 2003; Lawler, 1967) und diese genauer beurteilen (Waldmann, 2003) bzw. hauptsächlich Umgänglichkeit und Hilfsbereitschaft bewerten (Conway et al., 2001). Generell sollten Kollegen (und Mitarbeiter) zwischenmenschliches Verhalten häufiger beobachten und besser beurteilen können als Vorgesetzte (Conway & Huffcutt, 1997; Murphy & Cleveland, 1995). Nur ein Autor vertritt eine entgegengesetzte Annahme (Kollegen fokussierten mehr auf Ergebnisse als auf Beziehungsaspekte; Bradley, 2004), die sich aber empirisch nicht bestätigte (Bradley, 2004). Gestützt werden die Annahmen der übrigen Autoren durch den Befund, dass interpersonelle und Kommunikationskompetenzen von Kollegen reliabler eingeschätzt werden als von Vorgesetzten (Viswesvaran et al., 1996) und die Beurteilerübereinstimmung eines Items zur Teamorientierung im KU höher ausfiel als im MU (Waldmann, 2003)³². Ebenso zeigen sich metaanalytisch ermittelte Unterschiede in der Zusammenhangshöhe mit dem Persönlichkeitsmerkmal Verträglichkeit unterstützend: Das Urteil von Kollegen und Mitarbeitern korreliert mit Verträglichkeit zu .15 bzw. .14, während für das VU Nullkorrelationen berichtet werden (Conway et al., 2001). Nicht bestätigt wird die Annahme durch den metaanalytischen Befund, dass Kollegen kognitive Dimensionen reliabler einschätzen als interpersonelle Dimensionen (.43 versus .34; Conway & Huffcutt, 1997).

31 Die Befragung im administrativen Kontext wurde zeitlich näher zur Erhebung der Zielerreichung eingesetzt und enthielt mehr Einschätzungen pro Person. Dies kann den Unterschied zwischen beiden Instrumenten in der Zusammenhangshöhe erklären.

32 Die Beurteilerübereinstimmung wurde auf Itemebene berechnet.

Die fehlende Eindeutigkeit der Befundlage zum KU lässt sich über folgende Verhältnisse erklären. Da Kollegen häufiger (Mintzberg, 1975) und auf Teamebene³³ mit der Führungskraft zusammenarbeiten als der Vorgesetzte, können Kollegen interpersonelle Aspekte vermutlich besser beurteilen als Vorgesetzte. Kollegen arbeiten aber weniger häufig direkt mit der Führungskraft zusammen als Mitarbeiter (Mintzberg, 1975) und zudem fokussiert auf Teamsituationen. So können sie vermutlich nur jenen begrenzten Aspekt des Consideration besonders genau beurteilen, der sich auf die Zusammenarbeit im Team bezieht. Consideration insgesamt können Mitarbeiter daher wahrscheinlich besser beurteilen als Kollegen. Kollegen haben bezüglich kognitiver und aufgabenbezogener Aspekte wiederum mehr Einsichten und Kenntnisse als die Mitarbeiter, jedoch weniger als der Vorgesetzte.

Die insgesamt zwar lückenhafte und heterogene Datenlage unterstützt zusammenfassend die folgenden, überwiegend vertretenen theoretischen Annahmen:

- Vorgesetzte beobachten und achten verstärkt auf leistungs-/ergebnisrelevantes Verhalten sowie auf Ergebnisse und fokussieren mehr auf kognitive Fähigkeiten und Aspekte des „Initiating Structure“.
- Mitarbeiter beobachten und achten verstärkt auf das mitarbeiterbezogene Verhalten der Führungskraft. Sie fokussieren auf den Umgang/die zwischenmenschlichen Beziehungen/die soziale Kompetenz und Mitarbeiterorientierung der Führungskraft bzw. Aspekte des „Consideration“.
- Kollegen scheinen tendenziell mehr auf zwischenmenschliches Verhalten im Team (Aspekt des „Consideration“) zu achten.

Entsprechend werden die folgenden Unterschiede in der merkmalspezifischen Urteilsgüte der Quellen angenommen, die in Abbildung 6 illustriert sind:

- Vorgesetzte beurteilen Initiating Structure genauer als Mitarbeiter oder Kollegen.
- Kollegen beurteilen Initiating Structure weniger genau als der Vorgesetzte, aber genauer als die Mitarbeiter.
- Mitarbeiter beurteilen Consideration genauer als Vorgesetzte oder Kollegen.
- Kollegen beurteilen Consideration genauer als der Vorgesetzte, aber weniger genau als Mitarbeiter.

Die schwache Befundlage hinsichtlich der Unterschiede zwischen Quellen in der Höhe der Zusammenhänge einzelner Merkmale mit Außenvariablen wird vermutlich durch drei einschränkende Aspekte mit bedingt. Neben der bereits geäußerten Kritik ist erstens die unzureichend geprüfte Skalenstruktur der Instrumente zu nennen. Die Höhe der Assoziation von Urteilen zu bestimmten Merkmalen mit Außenvariablen ist zweitens auch durch generelle Unterschiede in der Urteilsgenauigkeit verschiedener Quellen beeinflusst (s. Abschnitt 2.2.4.3.1). Wie zu den Aufgaben der Führung (Abschnitt 2.1.2) dargestellt, variiert drittens die Beziehung zwischen verschiedenen Führungsmerkmalen und Füh-

33 Die Annahme einer besonders akkuraten Beurteilung von Teamfähigkeit oder -orientierung durch Kollegen kann in der vorliegenden Arbeit nicht überprüft werden, da Aspekte des Teamverhaltens faktoriell keine separate Skala ergeben.

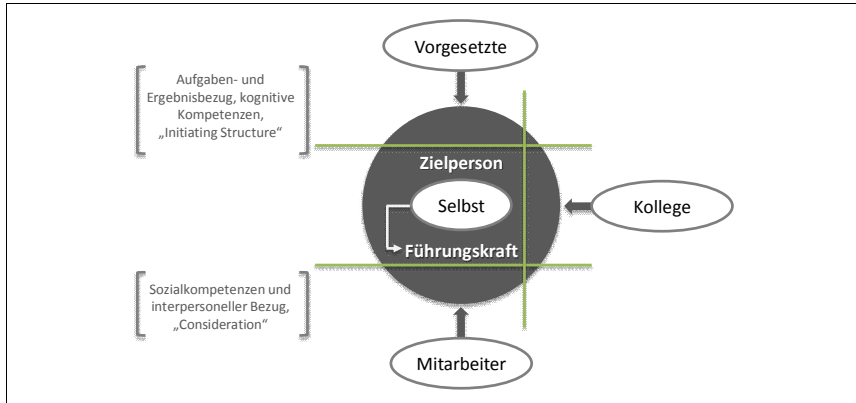


Abbildung 6: Annahmen zu merkmalspezifischen Quelleneffekten

rungserfolgsvariablen in Abhängigkeit der gewählten Außenvariablen bereits unabhängig von der Urteilsquelle. Initiating-Structure-bezogene Merkmale sagen so eher Ergebnisse oder Leistungskennwerte vorher, Consideration-bezogene Merkmale eher Variablen der Mitarbeiterzufriedenheit (Judge et al., 2004).

Zusammenfassend wurde in den letzten drei Abschnitten aufgezeigt, dass Unterschiede zwischen Quellen bezogen auf die generelle und merkmalspezifische Urteilsgenauigkeit feststellbar sind und einzelne Quellen in Kombination mit dem VU zusätzliche Varianz in Führungserfolgsvariablen aufklären. Auf eine weitere, grundlegendere Frage wird im nächsten Abschnitt eingegangen: Wird über die verschiedenen Quellen überhaupt das Gleiche gemessen?

2.2.4.6 Quellenunterschiede in der faktoriellen Struktur der Fragebogen

Den Erklärungen von Quelleneffekten (Abschnitt 2.2.4.2) zufolge wird Führungsleistung aus verschiedenen Perspektiven unterschiedlich interpretiert, gewichtet und kategorisiert. Es wurde herausgearbeitet, dass die Quellen teilweise Unterschiedliches zu messen scheinen (Abschnitt 2.2.4.5). Dies müsste über Abweichungen zwischen den Faktorenstrukturen der Urteile³⁴ oder dem Antwortverhalten hinsichtlich einzelner

34 Erläuterungen zur Methode des Vergleichs der Faktorenstrukturen: Cheung (1999) beschreibt sieben Kennwerte der konfirmatorischen Faktorenanalyse, in denen sich Unterschiede in den Urteilen zwischen Quellen abbilden können. Er untergliedert die Kennwerte in zwei Kategorien: erstens Parameter, die Auskunft über *Unterschiede in der Konzeptualisierung* der Wahrnehmung von Führungsleistung geben (die Anzahl der Faktoren, deren Itemzusammensetzung und Unterschiede in den Faktorladungen); zweitens Aspekte der *Vergleichbarkeit der Bewertung* der Führungsleistung zwischen Quellen (psychometrische Aspekte der Größe des zufälligen Messfehlers, Unterschiede in der latenten Faktorvariabilität, Unterschiede im latenten Dimensionsmittelwert und Korrelationen der Dimensionswerte pro Quelle).

Tabelle 11: Übersicht über Studien zu Quellenunterschieden auf Item- und Faktorenebene.

Studie	Methode	Quellen	Befund
Bradley, 2004	CFA zweiter Ordnung, Vergleich der Faktorenladungen	VU, KU, MU	Unterschiede in Faktorenladungen (s. Abschnitt 2.2.4.5)
Cheung, 1999	CFA an Instrument mit 10 Dimensionen	VU, SU	Höhe der Fehlervarianz variiert für 4* der 10 Dimensionen (Modell nach Mintzberg, s. Abschnitt 2.1.2)
Facteau & Craig, 2001	CFA an Instrument mit 10 Dimensionen (mit z. T. nur drei Items)	VU, KU, MU, SU	Unterschiede in Fehlerkovarianzen**; Quellenunterschiede auf Itemebene einer Skala („Motivieren Anderer“)
Lance & Bennett, 1997	EFA an Instrument zu interpersonellen Fähigkeiten	VU, KU, SU	Unterschiede in den Itemladungen zwischen VU, KU und SU
Maurer und Kollegen, 1998	CFA an eindimensionalem Instrument (Messung der Fähigkeit, Teams zu bilden)	KU, MU	Keine Unterschiede in Modellpassung
Oppler & Sager, 1992	CFA	VU, KU	Signifikante Unterschiede***
Pulakos und Kollegen 1996	Strukturgleichungsmodell	VU, KU	Signifikante Unterschiede
Scullen et al., 2003	CFA an 2 mehrdimensionalen 360°-Instrumenten, 2 Stichproben	VU, KU, MU, SU	Gleiche Faktorenstruktur für alle Quellen bestätigt

Anmerkung: CFA: konfirmatorische Faktorenanalyse, EFA: exploratorische Faktorenanalyse, VU = Vorgesetztenurteil, KU = Kollegenurteil, MU = Mitarbeiterurteil, SU = Selbsturteil. *Niedrigere Fehlervarianzen im SU als im VU bei Dimensionen Informant und Repräsentant, höhere als im VU bei Krisenmanager und Verhandlungsführer. **Eine Fehlerkovarianz im SU und MU war im KU und VU nicht enthalten. ***Am KU faktoriell bestimmte Skalen zeigten konfirmatorisch eine höhere Passung für das Urteil anderer Kollegen als für das VU; (eigene Zusammenstellung).

Items³⁵ sichtbar werden (Facteau & Craig, 2001). Nur wenige Studien wurden zu diesem Thema publiziert (Smither, London & Reilly, 2005). Tabelle 11 listet eine Übersicht über die acht wesentlichen Untersuchungen.

Die Befundlage ist uneinheitlich. Eine Reihe von Studien (Cheung, 1999; Scullen, Facteau & Craig, 2001; Maurer et al., 1998; Mount & Judges, 2003) fanden keine oder nur geringfügige Quellenunterschiede und nehmen Messäquivalenz zwischen den untersuchten Quellen an. Einschränkend ist zu erwähnen, dass teilweise nur einzelne Quellen oder

35 Erläuterung zu der Item-Response-Theorie (Rasch, 1960): Das Latent-Trait-Modell schließt aus dem Antwortverhalten einer Person auf zugrunde liegende Eigenschaften der Person und die Itemschwierigkeit. Bzgl. 360°-Instrumenten werden die dem Antwortverhalten verschiedener Quellen zugrunde liegenden mathematischen Funktionen (Item-Response Functions) verglichen.

eindimensionale Instrumente betrachtet worden sind und Skalen oft eine geringe Itemzahl aufweisen (z. B. Skalen mit 3 Items bei Fecteau & Graig, 2001). Scullen und Kollegen (2003) weisen zudem darauf hin, dass das konfirmatorische Bestätigen der gleichen Faktorenstruktur für unterschiedliche Quellen keine Aussage darüber erlaubt, ob die einzelnen Dimensionen von den Quellen unterschiedlich gewichtet werden oder ob es Unterschiede in der Definition von Leistung gibt. Andere Studien berichten signifikante Quellenunterschiede auf Faktoren- und Itemebene (Lance & Bennett, 1997; Oppler & Sager, 1992; Pulakos et al., 1996; Thornton, 1980³⁶).

Die Befundlage erlaubt den Schluss, dass die Anwendung gleicher Skalen auf alle Quellen, ohne diese zumindest auf Passung für die einzelne Quelle zu prüfen, vermutlich zu einer niedrigen Genauigkeit der Messung für einzelne Quellen führt. Dies könnte beispielsweise erklären, warum die auf Itemebene festgestellte Differenziertheit des MU nicht in einer erhöhten Genauigkeit der Urteile auf Dimensionsebene resultiert. Da die Skalen nicht am MU faktoriell geprüft oder gar entwickelt wurden, könnte es sein, dass Items den Skalen nicht gemäß dem Muster ihrer Faktorenladungen zugeordnet wurden. Aufgrund der angenommenen Besonderheit der Mitarbeiterperspektive (Conway & Huffcutt, 1997; Murphy & Cleveland, 1995; Scullen et al., 2000) kann gerade hier ein einzigartiges Ladungsmuster der Items angenommen werden.

Bevor das Kapitel zur 360°-Beurteilung mit einer Zusammenfassung abgeschlossen wird, werden im nächsten Abschnitt wesentliche Punkte zum SU beschrieben. Von einem direkten Vergleich zwischen Selbst- und Fremdurteilen ist abzusehen. In Abschnitt 2.2.3 zur Urteilsbildung wurde auf grundlegende Unterschiede aufmerksam gemacht. Da nahezu alle 360°-Rückmeldesysteme eine Selbsteinschätzung beinhalten (Eichinger & Lombardo, 2003), ist es jedoch notwendig für die Zielsetzung der vorliegenden Arbeit, das SU mit in den Vergleich aufzunehmen. Gerade für das SU liegen zudem Hinweise auf die Spezifität auf Item- und Faktorebene vor, wie im Anschluss aufgeführt. Diese lassen einen Gewinn in der Genauigkeit des 360°-Urteils unter der Verwendung quellenspezifischer Skalen erwarten.

2.2.4.7 Bedeutung der Selbsteinschätzung

Das Selbsturteil der Führungskraft sollte – im Vergleich zum Urteil anderer Einschätzer – auf einer vollständigeren Informationsbasis beruhen, da die Führungskraft selbst als Einzige ihr Verhalten in allen Situationen erlebt (Murphy, Cleveland & Mohler, 2001). Die übrigen Quellen können nur Ausschnitte des Verhaltens der Führungskraft beobachten (vgl. z. B. Lawler, 1967).

Eine Erhöhung der Genauigkeit der Beurteilung über den Einbezug des Selbsturteils scheint man sich in Veröffentlichungen zu diesem Thema jedoch nicht zu versprechen. Vornehmlich existieren Zweifel an der Fähigkeit oder Bereitschaft von Führungskräften zu einer genauen Beurteilung des eigenen Verhaltens (z. B. Harris & Schaubroeck, 1988). Die Aufnahme des SU in die Führungsbeurteilung wurde vornehmlich mit der resultierenden erhöhten Akzeptanz der Beurteilung begründet. Spätere Literatur hebt hervor, dass die Selbstbeurteilung der Führungskraft ermöglicht, mehr aus den Ergebnissen der 360°-Beurteilung zu lernen (z. B. Antonioni, 1996). Dies geschehe über Gegenüberstellung von Selbst- und Fremdwahrnehmung.

36 Thornton berichtet Unterschiede in den Faktorenstrukturen zwischen SU und VU.

Im Vergleich zu den Übereinstimmungen von Fremdeinschätzungen untereinander zeigt die Selbsteinschätzung durchgehend geringere Übereinstimmungen mit Fremdurteilen (Conway & Hunnicut, 1997; Mabe & West, 1982). Selbsturteile fallen in 360°-Beurteilungen zudem positiver aus als Fremdurteile (Harris & Schaubroeck, 1988). Überwiegend zeigt das SU eine niedrigere Vorhersagekraft als die anderen Quellen (Mount, Barrick & Strauss, 1994), in der Regel³⁷ finden sich keine Zusammenhänge mit objektiven Führungserfolgskriterien (Lombardo & Eichinger, 2003). Auch Zusammenhänge mit (ebenfalls auf Selbsteinschätzung beruhenden) Persönlichkeitsvariablen konnten von van Hooft und Kollegen (2006) nicht nachgewiesen werden.

Einige Autoren sprechen der Selbstbeurteilung jedoch eine höhere Differenziertheit zu: Die Varianz der Werte über die Dimensionen hinweg ist im SU höher als in den Urteilen anderer Quellen und es besteht die geringste Assoziation zwischen den Werten einzelner Dimensionen (Cheung, 1999; Scullen et al., 2000; Waldmann, 2003). Tatsächlich liegt ein geringerer Halo-Effekt vor im Vergleich zu den Fremdeinschätzungen (Thornton, 1980). Die insgesamt niedrigeren Zusammenhänge des SU mit Außenvariablen sind vermutlich durch den relativ hohen Anteil (höher als der Anteil in VU und KU) von Antworttendenzen im Urteil (Scullen et al., 2000) mit bedingt. Wie beim MU kann ein Beitrag auch in den ungeprüften Skalen vermutet werden. Im nächsten Abschnitt werden die Erkenntnisse zur 360°-Beurteilung zusammengefasst und die Fragestellung abgeleitet.

2.2.4.8 Zusammenfassende Ableitungen und Fragestellung zum 360°-Ansatz

Der Abschnitt zur 360°-Beurteilung beschrieb Unterschiede zwischen Quellen hinsichtlich der Genauigkeit der Beurteilung einzelner Merkmale. Es wurde zudem gezeigt, dass Quelleneffekte auf Item- und Faktorenebene reflektiert sind. Entsprechend kann die Nicht-Berücksichtigung von Quelleneffekten in der Konstruktion der 360°-Instrumente zum Einsatz von Skalen führen, die erhobene Information ungenau oder falsch abbilden.

Eine faktorielle Bestimmung von Skalen für die einzelnen Quellen könnte gewährleisten, durch genauere Urteile höhere Zusammenhänge mit Außenvariablen zu erreichen. Der Fokus würde dadurch zudem stärker auf den bisher wenig beachteten Nutzen von 360°-Instrumenten gerichtet, eine Einsicht in die Wahrnehmung und Konzeptualisierung von Führungsleistung durch die verschiedenen Quellen zu ermöglichen (Penny, 2003). In der vorliegenden Arbeit sollen daher die folgenden drei zentralen Forschungsfragen erstmals mit pro Quelle faktoriell bestimmten Skalen untersucht werden:

1. Welche spezifischen Dimensionen können für die einzelnen Quellen faktoriell gesichert werden und sind diese zwischen den Quellen vergleichbar?
2. Trägt die einzelne Quelle zusätzliche, einzigartige Informationen zur Führungsbeurteilung bei?
3. Hängt die Genauigkeit der Messung einzelner Merkmale von der Quelle ab? Wenn ja, welches Merkmal kann durch welche Quelle am genauesten gemessen werden?

37 Die Studie von Fleenor (1996), in der das SU einige Assessment-Center-Dimensionen besser vorhersagte als das VU, bildet beispielsweise eine Ausnahme.

2.3 Zielsetzung und methodische Ansatzpunkte

Das übergeordnete Ziel der vorliegenden Arbeit ist die Erhöhung der Genauigkeit der Führungsbeurteilung. Es wird geprüft, ob die Kombination verschiedener Quellen und der Einbezug von eigenschaftsbasierten Variablen diesen Zweck erfüllen. Vier Ziele leiten sich ab.

Das erste Ziel ist die quellenspezifische faktorielle Bestimmung von Skalen für das 360°-Instrument. Da davon ausgegangen wird, dass teilweise Unterschiedliches gemessen wird, ist ein exploratorisches Vorgehen in der Bestimmung der Faktorenstruktur der Einschätzung durch die einzelnen Quellen zu wählen. Exploratorischen Verfahren wird häufig vorgeworfen, aufgrund des Mangels von Überprüfungsmöglichkeiten der statistischen Signifikanz unzureichend objektiv zu sein, weshalb sie in diesem Zusammenhang selten eingesetzt werden und wenig Beachtung finden (Cheung, 1999). Durch eine konfirmatorische Überprüfung der exploratorisch ermittelten Modelle an einer weiteren unabhängigen Stichprobe sowie einer Gegenüberstellung verschiedener Modelle wird dem Einwand begegnet. In diesem Zuge wird ebenso die Möglichkeit einer einheitlichen Struktur für alle Quellen geprüft.

Die Untersuchung merkmalspezifischer Quelleneffekte auf die Höhe des Zusammenhangs mit Außenvariablen bildet das zweite Ziel. Es wird davon ausgegangen, dass die Genauigkeit eines Urteils einer Quelle auch vom zu beurteilenden Merkmal abhängt. Methodische Mängel in bisherigen Studien wurden herausgestellt und dienen als Ansatzpunkt der vorliegenden Arbeit. Hierzu werden die Fragestellungen erstmals anhand von faktoriell begründeten, quellenspezifischen Skalen untersucht. Zudem werden verschiedene Kriteriumsvariablen betrachtet sowie Konfundierungen zwischen Prädiktor- und Kriteriumsvariable vermieden. Um die angenommene Rangfolge der Urteile bezüglich der Zusammenhangshöhe mit Außenvariablen in Abhängigkeit des Merkmals empirisch zu überprüfen, werden die erhobenen Merkmale in die in Abschnitt 2.2.1.3 eingeführten Kategorien Consideration und Initiating Structure (Hemphill & Coons, 1957) gegliedert. Das Konzept findet aktuell in der Führungsforschung Verwendung (Judge, Piccolo & Ilies, 2004) und es liegt eine breite empirische Basis zum Zusammenhang der Dimensionen mit Führungserfolgsmerkmalen vor (Judge et al., 2004).

Das dritte Ziel ist es zunächst zu untersuchen, ob signifikante Zusammenhänge zwischen Indikatoren des Führungserfolgs und KU, MU sowie SU bestehen und weiterführend zu prüfen, ob diese Quellen in Kombination mit dem VU inkrementelle Validität aufweisen. Die faktorielle Bestimmung und Prüfung der Skalenstruktur pro Quelle blieb in bisherigen Untersuchungen gänzlich unberücksichtigt, was zu einer insgesamt mangelhaften Prüfung der Annahme führte.

Das vierte Ziel ist die Beantwortung der Frage, ob 360°-Variablen nennenswerte Zusammenhänge mit Indikatoren des Führungspotenzials aufweisen. Die Anwendbarkeit des 360°-Verfahrens für Fragen der Potenzialbeurteilung wird zwar empfohlen (z. B. Sarges, 2005), wurde bislang aber nicht untersucht.

2.4 Theoretische Erwartungen

Die zentralen theoretischen Erwartungen werden in zwei Kategorien formuliert: 1. Quellenunterschiede, 2. Inkrementelle Varianzaufklärung in Außenvariablen durch KU, MU und SU in Kombination mit dem VU. Die Inhalte der Kategorien werden im Folgenden genannt und in Abschnitt 3.7 über die Ableitung von Hypothesen konkretisiert. Die theoretischen Erwartungen sind:

1. Es bestehen Quelleneffekte in der 360°-Beurteilung.
 - Unterschiedliche Quellen konzeptualisieren die Dimensionalität der Führungsleistung uneinheitlich und einzelne Items der 360°-Beurteilung stehen über verschiedene Quellen hinweg in unterschiedlicher Beziehung zu den zugrunde liegenden Faktoren (vgl. Abschnitt 2.2.4.6).
 - Es werden durch quellenspezifische Skalen höhere Zusammenhänge mit Indikatoren des Führungserfolgs erreicht als in bisherigen Studien. Auch mit Führungspotenzialvariablen bestehen signifikante Assoziationen.
 - Es gibt merkmalspezifische Unterschiede in der Urteilsgenauigkeit zwischen Quellen (vgl. Abschnitt 2.2.4.5 und 2.2.4.8).
2. Das Mehrquellenprinzip ist der traditionellen Vorgesetztenbeurteilung in der Vorhersage von Führungserfolgs- und Führungspotenzialvariablen überlegen: Die Urteile von Kollegen, Mitarbeitern und das Selbsturteil der Führungskraft sind signifikant mit Indikatoren des Führungserfolgs assoziiert und es kann ein inkrementeller Anteil der Varianz in den Außenvariablen in Kombination mit dem VU aufgeklärt werden.

Die Führungskraft aus unterschiedlichen Blickwinkeln
Führungsbeurteilung über quellenspezifische
360°-Skalen

Puckett, S.

2016, XIII, 178 S. 11 Abb., Softcover

ISBN: 978-3-658-13798-4