



UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Reine Formalität?
**Subjektive Leistungsbeurteilungen und geschlechterspezifische
Unterschiede in tariflichen Leistungszulagen**

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
– Doctor rerum politicarum –
vorgelegte Dissertation
von

Yilmaz Özdemir-Rose, M.Sc.

geboren (Özdemir) am 19. Januar 1987 in Holzminden

2017

Erstgutachter: Prof. Dr. Martin Schneider

Zweitgutachterin: Prof. Dr. Anja Iseke

Vorwort

Diese Dissertation ist das Ergebnis eines intensiven Forschungsprozesses, der nicht nur zu meiner wissenschaftlichen, sondern auch persönlichen Weiterentwicklung beigetragen hat. Während der Bearbeitung der Dissertation habe ich zahlreiche Personen kennengelernt, die mich durch den Prozess begleitet und auf unterschiedliche Art und Weise unterstützt haben. An dieser Stelle möchte ich mich bei den Personen bedanken, die zum Gelingen dieser Dissertation beigetragen haben.

An erster Stelle danke ich meinem Doktorvater Prof. Dr. Martin Schneider für seine optimale und lehrreiche Betreuung. Er gab mir die Möglichkeit, eigene Ideen umzusetzen und diese stets durch konstruktive Kritik zu verbessern. Besonders schätze ich an ihm das mir entgegengebrachte Vertrauen hinsichtlich der Forschung und der Lehre sowie die Anregungen zu Teilnahmen an internationalen Konferenzen. Ebenso möchte ich mich bei meiner Zweitgutachterin Prof. Dr. Anja Iseke bedanken, die lehrstuhlübergreifend immer ein offenes Ohr für meine Forschungsansätze hatte und durch ihre hilfreichen Anmerkungen zu einem wissenschaftlichen Mehrwert dieser Dissertation beigetragen hat. Prof. Dr. Bernd Frick und Prof. Dr. Claus-Jochen Haake danke ich für das Mitwirken in der Promotionskommission und die Bereitschaft zur Teilnahme an der Disputation.

Einen großen Dank möchte ich an meine Kolleginnen und Kollegen am Lehrstuhl aussprechen: Christian Manfred Wilke, Dr. John Riach, Katharina Radermacher, Paul Hensen, Simon Eisele und Dr. Yanick Kemayou. Das freundliche Arbeitsklima und die gemeinsamen Teilnahmen an den lehrstuhlinternen Forschungskolloquien haben dazu beigetragen, dass ich gerne auf meine ereignisreiche Promotionszeit zurückblicken werde. Den studentischen Hilfskräften am Lehrstuhl und unserer Sekretärin Sarah Kanand danke ich für ihre unterstützenden Tätigkeiten, die im Zusammenhang mit der Dissertation stehen.

Meinen größten Dank widme ich meiner Ehefrau Anica, die immer für mich da war und mir während der Promotionszeit motivierend als auch fachlich zur Seite stand, sowie meiner Familie und meinen Freunden für ihre Liebe und ihr bedingungsloses Vertrauen.

Vielen Dank Euch allen!

Paderborn, im Dezember 2017

Yilmaz Özdemir-Rose

Abstract

Die vorliegende Arbeit untersucht im Rahmen einer Fallstudie die Auswirkungen von Geschlechtereffekten in Leistungsbeurteilungen auf die variable Vergütung von tariflichen Beschäftigten. Der Fallbetrieb stammt aus der Metall- und Elektroindustrie in Nordrhein-Westfalen und ist aufgrund der Bestimmungen des Entgeltrahmenabkommen-Tarifvertrags dazu verpflichtet, seine ca. 3.500 Beschäftigten jährlich hinsichtlich der individuell erbrachten Leistungen zu evaluieren.

Die empirischen Auswertungen von mehr als 12.000 Beschäftigtenbeurteilungen für die Jahre 2011 bis 2014 zeigen, dass Frauen grundsätzlich von einer Formalisierung der Bewertungen profitieren. Demnach erhalten weibliche Beschäftigte im Rahmen eines formalisierten Bewertungssystems bessere Leistungsbeurteilungen als vergleichbare männliche Beschäftigte. Da das Ergebnis der Beurteilungen an die Vergütung der Beschäftigten gekoppelt ist, bekommen Frauen nicht nur bessere Bewertungen, sondern auch höhere entgeltliche Zulagen als Männer. Im umgekehrten Fall, also wenn Führungskräfte die Leistungen von Frauen in einem weniger formalisierten System mit subjektivem Charakter bewerten, werden weibliche Beschäftigte schlechter evaluiert als vergleichbare Männer. Der Geschlechterunterschied macht ca. 0,2%-Punkte der gezahlten Leistungszulagen aus – dies entspricht im Durchschnitt einer geschlechterspezifischen Einkommensdifferenz in Höhe von etwa 62€ pro Jahr (brutto).

Darüber hinaus zeigt die vorliegende Arbeit auf, dass die beschriebenen Geschlechtereffekte durch weitere organisationale Mechanismen, wie zum Beispiel die Entgeltgruppe der Beschäftigten, bedingt werden. Die Ergebnisse der Studie deuten außerdem darauf hin, dass die tariflichen Leistungsbeurteilungen neben dem Geschlecht auch durch andere Faktoren, wie zum Beispiel Alter oder Betriebszugehörigkeit, beeinflusst werden.

Durchgeführte Experteninterviews bieten ein besseres Verständnis der Fallstudienresultate und der generellen Kritik an Leistungsbeurteilungen. Aus den Interviews geht hervor, dass die tariflichen Leistungsbeurteilungen einen geringen Anreiz zur Leistungssteigerung bieten und zudem tendenziell als subjektiv wahrgenommen werden. Durch die Kombination quantitativer und qualitativer Forschungsmethoden liefert die vorliegende Dissertation wichtige Anhaltspunkte, um das tarifliche Leistungsbeurteilungssystem zukünftig leistungsgerechter zu gestalten.

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
1.1 Leistungsbeurteilungen und Geschlechterunterschiede	1
1.2 Forschungslücke und Fokussierung der Fragestellung	4
1.3 Forschungsbeitrag	6
1.4 Aufbau der Arbeit und Vorgehensweise	8
2 Literaturüberblick zu Leistungsbeurteilungen	11
2.1 Historische Entwicklung	11
2.2 Aktueller Forschungsstand.....	12
2.3 Geschlechterspezifische Effekte in Leistungsbeurteilungen.....	19
2.3.1 Haupteffekte	19
2.3.2 Moderatoreffekte	24
2.4 Forschungslücke.....	33
2.4.1 Formalisierungsgrad	33
2.4.2 Geschlechterdyaden.....	36
3 Theoretische Grundlagen und Hypothesen	40
3.1 Ökonomische Ansätze.....	40
3.2 Sozialpsychologische und soziologische Ansätze.....	43
3.3 Hypothesen.....	48
4 Untersuchungskontext	54
4.1 Tarifliches Leistungsbeurteilungssystem	54
4.1.1 Leistungszulagen in der Vor-ERA-Zeit.....	54
4.1.2 Leistungszulagen nach dem aktuellen ERA	59
4.2 Bedeutung des ERA für die betriebliche Praxis.....	64
4.3 ERA und der Bezug zu Entgeltgleichheit	66
4.4 Beurteilungssystem des Fallstudienunternehmens.....	72
5 Forschungssetting	78
5.1 Mixed-Methods-Ansatz.....	78
5.2 Experteninterviews.....	79

5.3	Quantitative Fallstudie	81
5.3.1	Operationalisierung der Variablen.....	82
5.3.1.1	Abhängige Variable und unabhängige Variablen.....	82
5.3.1.2	Kontrollvariablen	85
5.3.2	Methodische Vorgehensweise	90
6	Ergebnisse und Interpretationen.....	93
6.1	Insiderinformationen zur ERA-Leistungsbeurteilung.....	93
6.2	Deskriptive Statistiken	96
6.2.1	Abhängige Variable und unabhängige Variablen.....	96
6.2.2	Kontrollvariablen.....	102
6.3	Multivariate Analysen	106
6.3.1	Hypothesentests	106
6.3.2	Weitere Einflussgrößen	114
6.3.3	Kausalität der Geschlechtereffekte.....	122
7	Diskussion und Schlussbetrachtung.....	129
7.1	Hauptergebnisse der quantitativen Fallstudie	129
7.2	Ergebnisse der Experteninterviews	132
7.3	Praktische Implikationen.....	134
7.4	Limitationen und Anknüpfungspunkte für weitere Forschung	139
8	Anhang.....	142
	Literaturverzeichnis.....	VI

Abbildungsverzeichnis

Abbildung 1: Mitarbeiterbeurteilungen nach Industriesektoren	2
Abbildung 2: Verbreitung entgeltrelevanter Leistungsbeurteilungen in Deutschland.....	3
Abbildung 3: Geschlechtereffekte bei Leistungsbeurteilungen	5
Abbildung 4: Beschäftigtenstruktur der M+E-Industrie nach Entgeltmethoden	66
Abbildung 5: Geschlecht als Einflussgröße auf die Höhe der Leistungszulagen	71
Abbildung 6: Ablauf einer tariflichen Leistungsbeurteilung	74
Abbildung 7: Verteilungen der prozentualen tariflichen Leistungszulage	96
Abbildung 8: Tatsächliche tarifliche Leistungszulage nach Geschlecht.....	97
Abbildung 9: Punkteverteilungen der Beurteilungsmerkmale nach Geschlecht.....	99
Abbildung 10: ERA-Gehaltsgruppen und Leistungszulagen nach Geschlecht	101
Abbildung 11: Verteilungen des Alters und der Betriebszugehörigkeit	102
Abbildung 12: Geschlechteranteile in den ERA-Gehaltsgruppen	103
Abbildung 13: Einkommensverteilungen (ALLBUS und Fallstudie)	104
Abbildung 14: Interaktion zwischen Geschlecht und organisationaler Hierarchiestufe...	110
Abbildung 15: Betriebszugehörigkeit und tarifliche Leistungszulage.....	117
Abbildung 16: Marginale Effekte der Zuordnung zu dem formalisierten System	124
Abbildung 17: Pauschale und ERA-Leistungszulage für Frauen und Männer.....	135

Tabellenverzeichnis

Tabelle 1: Metaanalysen zu Geschlechtereffekten in Leistungsbeurteilungen	21
Tabelle 2: Effekt des Vorgesetztengeschlechts auf Beurteilungen in Metaanalysen.....	25
Tabelle 3: Geschlechtereffekte in experimentellen Studien über Leistungsbeurteilungen.	30
Tabelle 4: Beurteilungsmerkmale nach TV LB Zeitl vom 7. September 1970	56
Tabelle 5: Beurteilungsmerkmale nach TV LB Ang vom 19. Februar 1975.....	58
Tabelle 6: Beurteilungsmerkmale nach ERA-TV vom 18. Dezember 2003	62
Tabelle 7: Fiktives Beispiel zur Bestimmung der tariflichen Leistungszulage.....	75
Tabelle 8: Informationen zu Experteninterviews und Teilnehmern	80
Tabelle 9: ERA-Monatseckentgelte NRW (2011 bis 2014)	83
Tabelle 10: Beschreibung ausgewählter Entgeltgruppen am Beispiel ERA Thüringen	85
Tabelle 11: Korrelationen zwischen tariflichen Beurteilungsmerkmalen.....	98
Tabelle 12: Mittelwerte der tariflichen Beurteilungsmerkmale nach Geschlecht.....	99
Tabelle 13: Aufbau des Datensatzes nach Geschlecht.....	100
Tabelle 14: Statistiken zu Geschlechterdyaden.....	100
Tabelle 15: Statistiken zur Wahl des Beurteilungssystems.....	105
Tabelle 16: Panelschätzungen zu Geschlechtereffekten in Leistungszulagen (H1).....	108
Tabelle 17: Marginale Effekte im weniger formalisierten Beurteilungssystem (H2).....	111
Tabelle 18: Panelschätzungen zu Geschlechtereffekten in Leistungszulagen (H3).....	113
Tabelle 19: Logit-Schätzung zur Wahl des Formalisierungsgrads	123
Tabelle 20: Kritiken und Änderungswünsche am ERA-Beurteilungssystem	137

Abkürzungsverzeichnis

ALLBUS	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften
BMAS	Bundesministerium für Arbeit und Soziales
BMBF	Bundesministerium für Bildung und Forschung
Destatis	Statistisches Bundesamt
EG	Entgeltgruppe
ERA	Entgeltrahmenabkommen
ERA-ETV	ERA-Einführungstarifvertrag
ERA-TV	Entgeltrahmenabkommen-Tarifvertrag
GRA	Gehaltsrahmenabkommen
IG Metall	Industriegewerkschaft Metall
LMX	Leader-Member-Exchange
LRA	Lohnrahmenabkommen
m	männlich
M+E-Industrie	Metall- und Elektroindustrie
QCA	Qualitative Comparative Analysis
REFA	Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung (ehemals: Reichsausschuss für Arbeitszeitermittlung)
SOEP	Sozio-ökonomisches Panel
TV LB Ang	Tarifvertrag zur Leistungsbeurteilung von Angestellten
TV LB Zeitl	Tarifvertrag zur Leistungsbeurteilung von Zeitlohnarbeitern
w	weiblich

Die aufgelisteten Abkürzungen werden im Singular und Plural verwendet.

1 Einleitung

1.1 Leistungsbeurteilungen und Geschlechterunterschiede

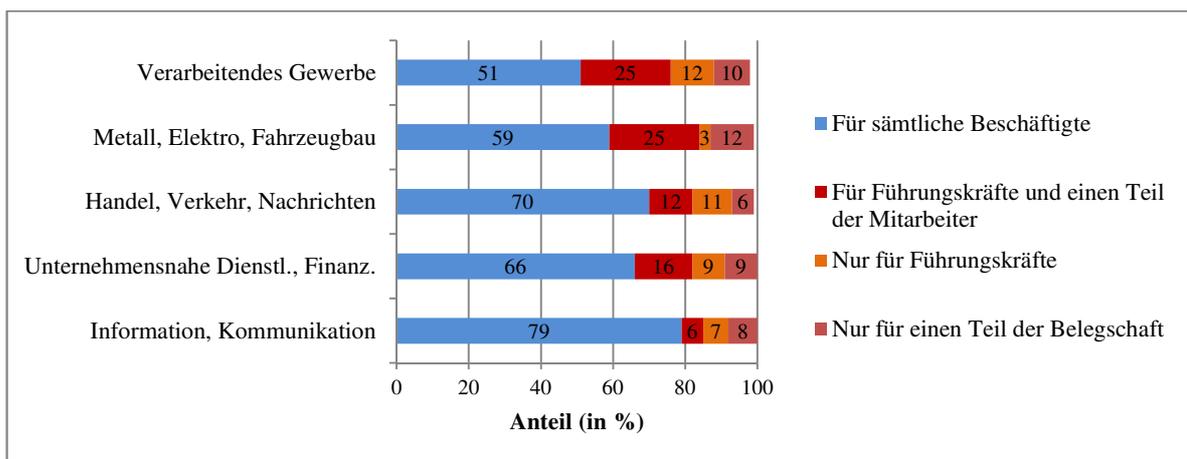
Die Gleichstellung von Geschlechtern in der deutschen Wirtschaft ist ein fortschreitender Prozess. Heutzutage erwerben Frauen in jüngeren Kohorten bereits häufiger einen Hochschulabschluss als Männer (BMBF, 2014) und schließen zu einem früheren Zeitpunkt ihr Studium ab (Destatis, 2013). Doch ungeachtet der Tatsache, dass sich die Bildungsunterschiede zwischen Frauen und Männern verringern und zudem zahlreiche institutionelle Mechanismen existieren, die eine geschlechtsspezifische Diskriminierung am Arbeitsplatz zu verhindern versuchen (für eine Übersicht über internationale Frauenquoten in Aufsichtsräten siehe Terjesen, Aguilera & Lorenz, 2015), ist die Benachteiligung von Frauen in diversen Bereichen des Wirtschaftslebens allgegenwärtig. Eine Folge hiervon ist eine horizontale Segregation auf dem Arbeitsmarkt, wobei Frauen in Berufen mit geringem Einkommen überrepräsentiert und in hochbezahlten Berufen unterrepräsentiert sind (Graves & Powell, 1995). Die Allokation von männlichen und weiblichen Beschäftigten auf unterschiedliche Hierarchieebenen bei gleicher Qualifikation führt ferner zu einer vertikalen Segregation auf dem Arbeitsmarkt, die es Frauen oftmals durch nicht sichtbare oder informelle Barrieren erschwert, Top-Positionen eines Unternehmens zu erreichen (Kreimer, 2009). Neben diesem als *Glass Ceiling* bezeichneten Phänomen sind zudem erhebliche Einkommensdifferenzen hinsichtlich des Geschlechts festzustellen (*Gender Pay Gap*). Beispielsweise lagen im Jahr 2014 der unbereinigte Verdienstunterschied in Deutschland bei 22% und der bereinigte bei 6%. Mit anderen Worten: Arbeitnehmerinnen beziehen durchschnittlich bei vergleichbarer Tätigkeit und Bildungsausstattung im Vergleich zu ihren Kollegen einen um 6% geringeren Bruttostundenverdienst (Destatis, 2017).

Zahlreiche Studien haben sich mit den potentiellen Ursachen für die oben erwähnten Einkommensdifferenzen beschäftigt. Diese Studien zeigen auf, dass die Ungleichbehandlung von weiblichen Beschäftigten in unterschiedlichen Phasen des Karriereverlaufs auftreten kann. Gegenstand dieser Untersuchungen waren zum Beispiel Leistungsbeurteilungen (Bretz, Milkovich & Read, 1992; Koch, D'Mello & Sackett, 2015; Levy & Williams, 2004), Beförderungen (Blau & DeVaro, 2007; Cobb-Clark, 2001; Ransom & Oaxaca, 2005) und Einkommen (Barnett, Baron & Stuart, 2000; Booth, Francesconi & Frank, 2003; Hill, Upadhyay & Beekun, 2015). Darüber hinaus wurde herausgefunden, dass entgeltrelevante Geschlechterunterschiede in Bezug auf Einkommen, Boni und Beförderungs-

raten deutlich höher ausfallen als Differenzen im Rahmen von Leistungsbeurteilungen (Joshi, Son & Roh, 2015). Allerdings bleibt unklar, wie sich geschlechterspezifische Unterschiede in Leistungsbeurteilungen monetär auswirken. Es existiert bis dato keine Untersuchung, die die Größeneffekte von Geschlechterunterschieden in Mitarbeiterbeurteilungen hinsichtlich der Vergütung quantifizieren kann.

Die Analyse von Performanceevaluationen als Erklärungsfaktor für Ungleichbehandlungseffekte ist wichtig, da in Deutschland branchenübergreifend mehr als die Hälfte aller Betriebe (60%) ihre Mitarbeiter mindestens jährlich bewerten. 64% der Unternehmen wenden Leistungsbewertungen als Personalpraktik für sämtliche Mitarbeitergruppen an, d.h. sowohl für Führungskräfte als auch den übrigen Teil der Belegschaft (BMAS, 2013). Abbildung 1 zeigt die Verteilung dieser Praktik in diversen Industriebranchen in Deutschland für unterschiedliche Mitarbeitergruppen und Führungskräfte.

Abbildung 1: Mitarbeiterbeurteilungen nach Industriesektoren



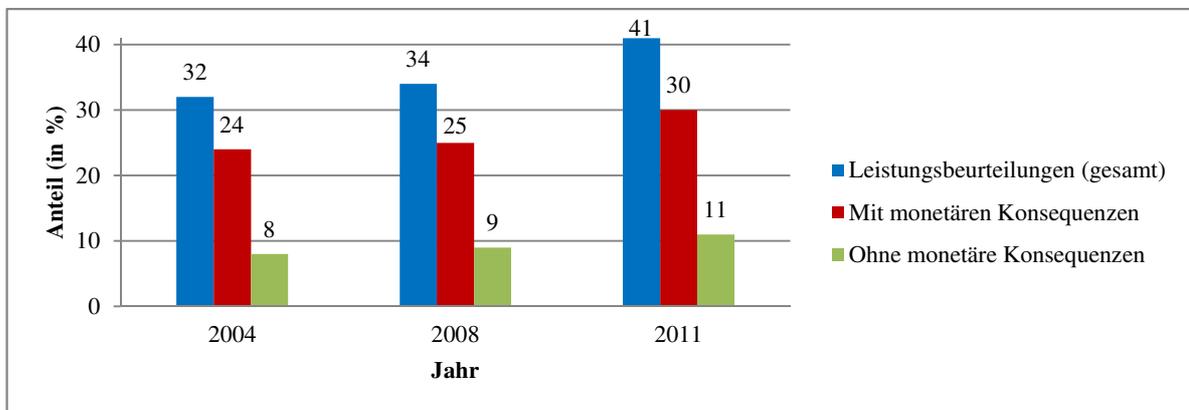
Quelle: Darstellung in Anlehnung an BMAS (2013).

Ein wesentlicher Grund für die weite Verbreitung von Leistungsbeurteilungen in der Wirtschaft ist, dass sie ein fester Bestandteil von Tarifverträgen sind. Beispielsweise gibt das neue Entgeltrahmenabkommen (ERA) der Metall- und Elektroindustrie (M+E-Industrie) vor, dass in neun von 14 Tarifgebieten in Deutschland die Beschäftigten eine entgeltrelevante Leistungszulage erhalten, die durch ein tariflich vorgegebenes Beurteilungsverfahren festgelegt wird (IG Metall, 2015).

Die in der deutschen Wirtschaft praktizierten Mitarbeiterbeurteilungen haben somit eine enorme Tragweite und zudem einen bedeutenden Einfluss auf Gehälter, Leistungszulagen,

Lohnerhöhungen sowie Beförderungsentscheidungen. Erhebungen des Sozio-ökonomischen Panels (SOEP) aus den Jahren 2004, 2008 und 2011 zeigen die Tendenz auf, dass Mitarbeiterbeurteilungen mit monetären Konsequenzen kontinuierlich an Bedeutung gewinnen. Abbildung 2 zeigt die Verbreitung von Leistungsbeurteilungen in deutschen Unternehmen sowie den Anteil an entgeltrelevanten Bewertungen. Während 2004 noch 24% der Mitarbeiterbeurteilungen an eine Bonuszahlung oder Ähnliches gekoppelt waren, ist der Anteil bis 2011 auf 30% gestiegen (Kampkötter, 2017). Es ist festzuhalten, dass Mitarbeiterbeurteilungen in Deutschland nicht nur immer verbreiteter sind, sondern offensichtlich auch die Entgeltrelevanz steigt.

Abbildung 2: Verbreitung entgeltrelevanter Leistungsbeurteilungen in Deutschland



Quelle: Darstellung in Anlehnung an Kampkötter (2017).

Einen Erklärungsansatz für das Vorhandensein von Geschlechterunterschieden im Einkommen von Männern und Frauen können Stereotype sein, die in der Gesellschaft vorhanden sind (O'Reilly, Smith, Deakin & Burchell, 2015). Geschlechterstereotype können bewusst oder unbewusst in der Wahrnehmung von Vorgesetzten hervorgerufen werden, wenn sie die Leistungen von Frauen und Männern evaluieren. Diese Geschlechterstereotype wiederum können – bedingt durch spezifische organisationale Mechanismen in Unternehmen – Auswirkungen auf die Vergütung von Frauen und Männern haben (Abraham, 2017). Falls Führungskräften in Unternehmen Spielraum in Bezug auf Personalpraktiken geboten wird, ist nicht auszuschließen, dass Entscheidungen in Bezug auf Einkommen, Beförderungen und Leistungsbeurteilungen geschlechterspezifisch verzerrt sind (Bielby, 2008). In der Vergangenheit wurden Geschlechtereffekte im Bereich von Leistungsbeurteilungen (Bretz et al., 1992; Koch et al., 2015) und der Formalisierung von Gehaltssystemen (Abraham, 2017; Elvira & Graham, 2002) ausgiebig erforscht. Diese beiden For-

schungsstränge sollen im folgenden Kapitel skizziert werden, um die Fragestellung der vorliegenden Arbeit einzugrenzen.

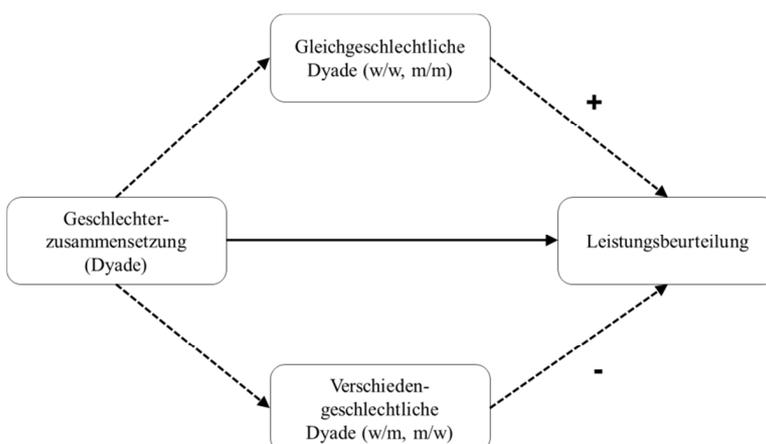
1.2 Forschungslücke und Fokussierung der Fragestellung

Ein Forschungsstrang im Bereich Geschlechtereffekte und Leistungsbewertungen beschäftigt sich mit der Fragestellung, ob der Grad der Formalisierung von Vergütungssystemen in Beziehung steht mit der Ungleichbehandlung zwischen Geschlechtern. In diesem Zusammenhang meint Formalisierung, dass die Bestandteile des Einkommens von Beschäftigten sich an objektiven Kriterien orientieren, wie zum Beispiel die erbrachte Leistung oder die formale Qualifikation. Weniger formalisierte Gehaltssysteme hingegen beinhalten subjektive Bestandteile, die durch den Handlungsspielraum der Führungskraft beeinflusst werden und ungerechtfertigterweise Konsequenzen auf die Vergütung haben können. Wenn Führungskräfte einen großen Entscheidungsspielraum haben, ist anzunehmen, dass Verzerrungen auftreten (Elvira & Graham, 2002). Da ein Großteil der Führungspositionen in Unternehmen von Männern besetzt sind und Männer Stereotype gegenüber Frauen haben können, können weniger formalisierte Gehaltssysteme zu einer größeren Ungleichheit zwischen Geschlechtern beitragen (Reskin, 2000). Jedoch ist auch der umgekehrte Effekt denkbar: Ein größerer Entscheidungsspielraum bietet Führungskräften die Möglichkeit, durch ihre Handlungen Geschlechterdifferenzen in Unternehmen zu verringern (Dencker, 2008). Die Ergebnisse von Studien, die den Zusammenhang zwischen Geschlechterunterschieden und Gehaltssystemen analysieren, sind nicht einheitlich. Tendenziell zeigt sich allerdings, dass formalisiertere Systeme Unterschiede zwischen Männern und Frauen reduzieren (Abraham, 2017). Die vorliegende Arbeit fokussiert sich zwar nicht unmittelbar auf die Formalisierung von Gehaltssystemen und deren Implikationen auf Geschlechterdifferenzen, jedoch auf die Formalisierung von Leistungsbeurteilungssystemen. Das zugrundeliegende Fallstudiensetting lässt direkte Rückschlüsse auf Unterschiede im Gehalt zwischen Männern und Frauen zu, da die Leistungsbewertungen in dem analysierten Unternehmen Auswirkungen auf das Gehalt der Beschäftigten haben.

Weitere Untersuchungen über Geschlechterunterschiede haben sich auf die Fragestellung fokussiert, ob Männer und Frauen mit vergleichbaren demographischen Merkmalen unterschiedlich beurteilt werden. In Metaanalysen, die die Ergebnisse zahlreicher individueller Labor- und Feldstudien berücksichtigen, konnte bestätigt werden, dass das Konfidenzintervall für die durchschnittlich geschätzte Effektgröße in Bezug auf das Geschlecht in der

Regel den Wert 0 einschließt. Dementsprechend kann geschlussfolgert werden, dass Frauen im Mittel nicht systematisch schlechtere Leistungsbeurteilungen erhalten als Männer (Bowen, Swim & Jacobs, 2000; Joshi et al., 2015; Koch et al., 2015; Roth, Purvis & Bobko, 2012). Allerdings berücksichtigen diese Metaanalysen keine dyadischen Effekte, die zwischen den beurteilenden Personen und den beurteilten Individuen wirken. Einzelne Individualstudien jedoch finden, dass Leistungsbewertungen verhältnismäßig schlechter ausfallen, wenn das Geschlecht der Führungskraft und des Beschäftigten unterschiedlich sind (Furnham & Stringfield, 2001; Varma & Stroh, 2001). Die Auswirkungen dieser Geschlechterdyaden zwischen Führungskräften und Beschäftigten führen dazu, dass sich die aufgezeigten Geschlechterunterschiede in der Vergütung von Frauen und Männern niederschlagen, wenn Leistungsbeurteilungen mit dem Entgelt verknüpft sind. Die Einflüsse von Geschlechterdyaden auf die Bewertungen von Beschäftigten scheinen jedoch nicht einheitlich zu sein. Eine weitere Studie kommt beispielsweise zu dem Ergebnis, dass Geschlechterdyaden keine Konsequenzen auf Leistungsbeurteilungen haben (Lyness & Heilman, 2006). Abbildung 3 veranschaulicht den Einfluss von Geschlechterdyaden auf die Leistungsbeurteilung von Beschäftigten. Grundsätzlich sind vier Dyaden zwischen Vorgesetzten und Beschäftigten möglich: Vorgesetzte und Beschäftigte sind weiblich (w/w) oder männlich (m/m), ansonsten verschiedengeschlechtlich, d.h. w/m bzw. m/w. Gleichgeschlechtlichkeit wirkt sich ceteris paribus positiv und Verschiedengeschlechtlichkeit negativ auf die Leistungsbeurteilungen der Beschäftigten aus.

Abbildung 3: Geschlechtereffekte bei Leistungsbeurteilungen



Quelle: Eigene Darstellung.

Die Erkenntnisse aus beiden Forschungssträngen deuten darauf hin, dass Geschlechtereffekte auf indirekte Weise wirken und nur unter bestimmten Gegebenheiten auftreten.

Demzufolge kann beispielweise angenommen werden, dass Geschlechterstereotype sich nicht automatisch in den Leistungsbeurteilungen der Beschäftigten oder im Gehalt widerspiegeln. Das Vorhandensein von Unterschieden zwischen Frauen und Männern hängt vielmehr von indirekten Einflussgrößen ab, wie zum Beispiel organisationalen Praktiken und Gegebenheiten in Unternehmen. Aus diesem Grund ist Abraham (2017) der Auffassung, dass die Forschung sich nicht auf die Frage fokussieren sollte, ob Formalisierung Ungleichheit einschränkt, sondern sich eher damit auseinandersetzen sollte, unter welchen Bedingungen dies geschieht. Reskin (2008) ist der Ansicht, dass sich die Forschung auf kausale Mechanismen und organisationale Praktiken konzentrieren sollte, die in Beziehung mit der Andersbehandlung von Frauen und Männern stehen.

Das vorliegende Fallstudiensetting bietet eine einzigartige und ideale Grundlage für die Analyse von Geschlechtereffekten, die einerseits durch die oben erwähnten Geschlechterdyaden und andererseits durch organisationale Gegebenheiten, insbesondere einer Formalisierung von Beurteilungssystemen, beeinflusst werden. Im Rahmen dieser Dissertation soll daher die folgende übergeordnete Fragestellung beantwortet werden:

Wie wirken sich die Formalisierung eines Beurteilungssystems sowie Geschlechterdyaden auf die Geschlechterunterschiede in Leistungsbeurteilungen und auf die Vergütung von Beschäftigten aus?

1.3 Forschungsbeitrag

Die vorliegende Arbeit liefert mehrere Beiträge zur Aufklärung von geschlechterspezifischen Effekten in Leistungsbeurteilungen und in der Vergütung. An erster Stelle ist zu erwähnen, dass im Rahmen dieser Untersuchung spezifische organisationale Mechanismen eines Unternehmens untersucht werden, um Differenzen in entgeltlichen Leistungszulagen zu erforschen. Die Leistungsdaten der Beschäftigten stammen von einem deutschen Großunternehmen aus der M+E-Industrie, das aufgrund der Bestimmungen des Entgeltrahmenabkommen-Tarifvertrags (ERA-TV) dazu verpflichtet ist, seine Beschäftigten jährlich zu evaluieren und auf Basis dieser Evaluationen Leistungszulagen zu zahlen.¹ Die erwähnten Leistungszulagen sind variabel und werden am Ende des jeweiligen Beurteilungszeitraums

¹ Der ERA-TV wurde eingeführt, um nicht nur eine Gleichbehandlung der ehemals unterschiedenen Beschäftigtengruppen Arbeiter und Angestellte durchzusetzen, sondern auch um eine Entgeltgleichheit zwischen weiblichen und männlichen Beschäftigten zu erreichen (Jochmann-Döll & Ranftl, 2010).

den Grundentgelten der Beschäftigten hinzugerechnet. Der zugrunde liegende Datensatz enthält detaillierte Informationen über die Leistungen und Merkmale der Beschäftigten im Fallbetrieb. Darüber hinaus können die geschätzten Geschlechtereffekte in monetären Beträgen beziffert und in Prozentwerten des Einkommens ausgedrückt werden. Dementsprechend bietet das analysierte Fallstudien-setting nicht nur eine Grundlage zur Untersuchung organisationaler Mechanismen sowie deren Wirkungsweisen auf Geschlechtereffekte in Leistungsbeurteilungen, sondern erstmalig auch die Gelegenheit, die Effektgrößen in monetären Werten auszugeben. Letzteres ist bisher noch in keiner veröffentlichten Studie vorgenommen worden. Dies ist erstaunlich, da Leistungsbeurteilungen Diskriminierungs- als auch Gleichstellungspotenzial bieten und bezüglich dieser Aspekte nicht hinreichend untersucht wurden (Krell, 2011). Im Kontext der M+E-Industrie ist es besonders interessant, Geschlechterunterschiede zu untersuchen, da es sich um eine von Männern dominierte Industrie handelt. Auch im Fallunternehmen ist der Männeranteil verhältnismäßig hoch.

Bei dem skizzierten Forschungsstand kann die vorliegende Studie in mehrfacher Hinsicht einen Beitrag zur Aufklärung von Geschlechtereffekten in Personalbeurteilungen liefern. Da Geschlechtereffekte sich zwischen Branchen und spezifischen Kontexten unterscheiden können, sind Fallstudien, die den besonderen institutionellen Hintergrund berücksichtigen, sinnvoll. Das Fallunternehmen hat mehr als 7.000 Beschäftigte in Deutschland und beurteilt davon jährlich ca. 3.500 im Rahmen des ERA-TV. Zur Erforschung von Geschlechtereffekten wurden die Beurteilungsdaten der Jahre 2011 bis 2014 zur Verfügung gestellt. Bisher wurde in keiner bekannten Studie mit einem Paneldesign, also einer Kombination von Quer- und Längsschnittdaten, gearbeitet. Wenn Effekte häufig statistisch insignifikant sind, kann dies auch daran liegen, dass viele relevante Merkmale unbeobachtet bleiben. Um diese unbeobachtete Heterogenität zu kontrollieren, sind Paneldesigns eine geeignete Methodik.

Einen weiteren Beitrag liefert die vorliegende Arbeit durch die Kombination der quantitativen Ergebnisse zu Geschlechtereffekten mit qualitativen Interviews, die mit acht betrieblichen Experten des Fallunternehmens durchgeführt wurden. Die qualitativen Interviews bieten wichtige Einblicke in die betriebliche Beurteilungspraktik und die organisationalen Mechanismen, die zu Geschlechterunterschieden in Leistungsbewertungen von Frauen und Männern führen. Folglich ist nicht nur der vorliegende Datensatz einzigartig für den Forschungszweck, sondern auch die Kombination zweier Forschungsmethoden, die bisher im

Rahmen der Analyse von Geschlechtereffekten in Leistungsbewertungen noch nicht gemeinsam angewendet wurden. Die Interviews dienen zu guter Letzt auch zur Interpretation der in dieser Arbeit aufgedeckten Wirkungszusammenhänge und stellen eine Grundlage für betriebliche und tarifpolitische Maßnahmen dar, um das Leistungsentgelt von Frauen und Männern gerechter zu gestalten und die Wirkung von Leistungsbewertungen als Anreizinstrument zu optimieren.

1.4 Aufbau der Arbeit und Vorgehensweise

Der Aufbau der Arbeit und die Vorgehensweise zur Beantwortung der Fragestellung gestalten sich wie folgt. In Kapitel 2 wird zunächst der allgemeine Forschungsstand zu Leistungsbeurteilungen skizziert, um einen generellen Überblick darüber zu erhalten, wie sich die Forschung zu diesem Thema historisch entwickelt hat. Das Kapitel soll verdeutlichen, dass im Bereich Leistungsevaluationen nach wie vor intensiv und disziplinübergreifend geforscht wird – zum Beispiel in der Psychologie und im Gebiet Management. In den letzten Jahrzehnten ist feststellbar, dass eine steigende Anzahl an managementnahen Fragestellungen Gegenstand von Forschungsarbeiten über Leistungsbewertungen waren. Im Anschluss an die Skizzierung des allgemeinen Forschungsstands werden in diesem Kapitel Geschlechtereffekte in Leistungsbeurteilungen fokussiert. Da Geschlechtereffekte bisher in mehreren Metaanalysen untersucht wurden und diese die Ergebnisse mehrerer individueller Studien zusammenfassend darstellen, werden in diesem Abschnitt die Erkenntnisse dieser überblicksartigen Untersuchungen detailliert beschrieben. In diesem Zusammenhang wird auf die Geschlechter-Haupteffekte als auch die indirekten Effekte, die durch Moderatoren hervorgerufen werden, eingegangen. Die Beschreibungen berücksichtigter Moderatoren zeigen auf, dass die Faktoren Geschlechterkombinationen zwischen Führungskräften und Beschäftigten als auch organisationale Mechanismen, wie der Formalisierungsgrad, bisher unzureichend erforscht wurden.

Im anschließenden Kapitel 3 werden relevante Theorien beschrieben, die die Entstehung von Geschlechtereffekten in Leistungsbeurteilungen erklären. Der Fokus der Ansätze liegt auf ökonomischen als auch sozialpsychologischen und soziologischen Theorien. Auf Basis theoretischer Argumente und der Berücksichtigung organisationaler Mechanismen werden in diesem Abschnitt testbare Hypothesen formuliert, die mithilfe der quantitativen Fallstudien Daten überprüft werden sollen.

Kapitel 4 skizziert den Untersuchungskontext, indem die Historie des relevanten tarifli-

chen Leistungsbeurteilungssystem beschrieben wird. Die Erörterungen sollen zu einem besseren Verständnis des aktuellen Bewertungssystems und der Beurteilungskriterien beitragen sowie verdeutlichen, weshalb das tarifliche Beurteilungssystem des ERA in direktem Zusammenhang steht mit dem Thema Entgeltgleichheit zwischen Frauen und Männern. Darüber hinaus wird das zugrunde liegende Beurteilungssystem des Fallstudienunternehmens und die Beurteilungspraxis detailliert erörtert.

Das Forschungssetting bildet den Gegenstand von Kapitel 5. Es gliedert sich in einen qualitativen und einen quantitativen Teil. Dem qualitativen Teil sind Experteninterviews zuzuordnen, die durchgeführt worden sind, um näher zu verstehen, wie beispielweise die jährlichen Beurteilungsgespräche verlaufen und was die Experten im Unternehmen von dem angewandten Bewertungssystem halten. Diese Insider, d.h. Führungskräfte und Mitarbeiter sind unter anderem für den Forschungskontext relevant, da sie wichtige Kenntnisse über die tariflich vorgegebene Beurteilungssystematik haben. Die qualitative Studie basiert auf einem standardisierten themenzentrierten Gesprächsleitfaden, der im Anschluss an die empirische Studie entwickelt wurde. Der konzipierte Gesprächsleitfaden basiert auf der Überlegung, dass die gewonnenen Erkenntnisse aus der empirischen Studie sinnvoll um ein qualitatives Forschungselement ergänzt werden, um beispielsweise Kausalmechanismen aufzudecken, die zu statistischen Zusammenhängen führen. Im Rahmen der quantitativen Analyse werden Annahmen zur Formalisierung der beiden im Fallbetrieb praktizierten Bewertungssysteme getroffen sowie die Operationalisierung der in der Studie untersuchten Variablen aufgegriffen. Außerdem wird die methodische Vorgehensweise beschrieben.

Im vorletzten Kapitel 6 werden die Ergebnisse der quantitativen Studie zu Geschlechterunterschieden in Leistungsbeurteilungen vorgestellt. In diesem Zusammenhang wird unter anderem eingegangen auf die deskriptiven Statistiken und multivariaten Analysen zur Wirkungsweise von organisationalen Praktiken auf Geschlechterdifferenzen. Die Ergebnisse werden auch mittels der durchgeführten Experteninterviews interpretiert. Es wird auf weitere Ergebnisse der Untersuchung eingegangen, die nicht im Fokus der Forschungsfrage stehen, aber wichtige Erkenntnisse zur Subjektivität der tariflichen Beurteilungen liefern. Abschließend werden in diesem Abschnitt der Arbeit die aufgezeigten Geschlechtereffekte auf Kausalität überprüft.

Die Diskussion und die Schlussbetrachtung der Forschungsarbeit bilden Kapitel 7. Es

werden die Hauptergebnisse der multivariaten Analysen und die Ergebnisse der Experteninterviews zusammengefasst und deren Implikationen für die betriebliche Beurteilungspraktik diskutiert. In diesem Rahmen werden auch tarifpolitische Empfehlungen zur zukünftigen Gestaltung von Leistungsbeurteilungssystemen gegeben bzw. eine Alternative zum aktuellen Verfahren erläutert. Die Dissertation schließt ab mit den Limitationen der vorliegenden Fallstudie und Ideen für zukünftige Forschungsarbeiten, die sich mit der Fragestellung dieser Arbeit oder weiterführenden auseinandersetzen.

2 Literaturüberblick zu Leistungsbeurteilungen

2.1 Historische Entwicklung

Die geschichtliche Entwicklung der Forschung über Leistungsbeurteilungen lässt sich, wie nachfolgend dargestellt, in vier Phasen untergliedern.

Die erste Phase reicht nach Arvey und Murphy (1998) von 1950 bis 1980 und umfasst Forschungsarbeiten, die sich überwiegend mit der Optimierung von Beurteilungsinstrumenten beschäftigen. Demnach wurden in diesem Zeitraum zahlreiche Studien über Themen veröffentlicht, wie zum Beispiel die Anwendung unterschiedlicher Arten von Ratingskalen, Beurteilungsformaten (absolute versus relative Bewertungen) und Mittel, um möglichst objektive Bewertungsmaße zu entwickeln.

Die zweite Phase der Forschung beginnt nach Arvey und Murphy (1998) in den Anfängen der 1980er Jahre und endet Mitte der 1990er Jahre. In dieser Periode stand nicht länger die Optimierung von Beurteilungsinstrumenten im Fokus, sondern ein besseres Verständnis der kognitiven Prozesse der Beurteilenden. In diesem Zusammenhang haben insbesondere die einflussreichen Forschungsarbeiten von Landy und Farr (1980, 1983) dazu beigetragen, dass das Hauptaugenmerk der Forschung zu Leistungsbeurteilungen auf psychologischen Aspekten lag, die in der Unternehmenspraxis relevant sind, wie beispielsweise kognitive Prozesse der Informationsverarbeitung. Auch der Review von Feldman (1981) hat das Verständnis über Prozesse der sozialen Kognition, die bei Leistungsbeurteilungen relevant sein können, weiter vertieft. Neben den Informationsverarbeitungsprozessen sind zum Ende der zweiten Phase weitere Forschungsfelder in den Fokus gerückt. Beispielsweise weisen Bretz et al. (1992) in ihrer Übersichtsstudie darauf hin, dass in der Zeit zwischen 1985 und 1990 die Forschung über Leistungsbeurteilungen zwar durch kognitive Informationsverarbeitungsprozesse dominiert wurde, jedoch auch Themen, wie der Einfluss demographischer Eigenschaften der am Prozess beteiligten Personen und insbesondere Geschlechtereffekte Aufmerksamkeit erhalten haben.

Der dritte Forschungsabschnitt reicht von ca. Mitte der 1990er Jahre bis in die Anfänge der 2000er Jahre. In dieser Phase lag der Fokus der Untersuchungen zum Thema Leistungsbeurteilungen auf dem (sozialen) Kontext und seine Auswirkung auf den Prozess der Beurteilung (Levy & Williams, 2004). Hiermit sind unter anderem die Rahmenbedingungen, wie das Feedbackumfeld und die -kultur gemeint, die bestimmte Reaktionen bei den

Beschäftigten hervorrufen können, zum Beispiel in Bezug auf die Akzeptanz des Beurteilungssystems, die Zufriedenheit mit dem Bewertungssystem oder die empfundene Gerechtigkeit. Auch die überblicksartigen Studien von Arvey und Murphy (1998) sowie Ferris, Munyon, Basik und Buckley (2008) deuten darauf hin, dass kontextbezogene Variablen der Leistungsbeurteilung den Fokus der dritten Forschungsphase bilden. Im Vergleich zur ersten und zweiten Forschungsphase zeigen diese Übersichtsstudien zudem den Trend auf, dass zum Thema Leistungsbeurteilungen nicht nur in den Gebieten der Psychologie und Sozialpsychologie geforscht wird, sondern auch im Bereich Management: „ ... *research on performance evaluation is alive and well in the psychological and management literatures*“ (Arvey & Murphy, 1998: 162).

Der bis heute andauernde Forschungsabschnitt beginnt ca. Mitte der 2000er Jahre. Diese Phase ist vor allem dadurch gekennzeichnet, dass im Bereich der Leistungsbeurteilungen vermehrt Untersuchungen durchgeführt worden sind, die betriebswirtschaftlichen Fragestellungen nachgehen. Die Phase rückt Themengebiete, wie individuelle und organisationale Konsequenzen von Leistungsbeurteilungen, die Folgen von Verzerrungen durch systematische Beurteilungsfehler, Fairness und Gerechtigkeit, in den Fokus.

Da seit dem Review von Levy und Williams (2004) keine weiteren Literaturübersichten zum Thema Leistungsbeurteilungen veröffentlicht wurden, bezieht sich der nachfolgende Review auf Untersuchungen, die zwischen 2004 bis 2017 durchgeführt worden sind.

2.2 Aktueller Forschungsstand

Die im Rahmen dieser Arbeit erstellte Literaturübersicht, die den aktuellen Stand der Forschung über Leistungsbeurteilungen anreißt, basiert auf einer Recherche der elektronischen Datenbanken EBSCOhost, Taylor & Francis Online und der Suchmaschine für elektronische Zeitschriften Google Scholar. Die daraus entstandene Übersicht umfasst veröffentlichte englischsprachige Artikel in wissenschaftlichen Fachzeitschriften (Journals). Sie bietet einen Einblick in die aktuelle Literatur, die sich dem Thema Leistungsbeurteilungen widmet. Die Literaturübersicht beinhaltet wissenschaftliche Aufsätze aus den Jahren 2004 bis 2017, die im Rahmen eines begutachteten Verfahrens in einer wissenschaftlichen Fachzeitschrift als Diskussions- bzw. Arbeitspapier oder Dissertationsschrift veröffentlicht wurden. Da die Literatur über Leistungsbeurteilungen sich überwiegend auf den angelsächsischen Sprachraum beschränkt, sind im Rahmen der Literaturrecherche die Suchwörter *performance evaluation*, *performance appraisal*, *performance review* und

performance rating verwendet worden, die ähnlich wie die deutschen Begriffe Leistungs- und Mitarbeiterbeurteilung synonym zu verstehen sind.

Die berücksichtigten Artikel des Reviews erfüllen die nachfolgenden Kriterien: 1.) Die Überschriften der Aufsätze und/oder die deklarierten Schlagwörter der Artikel beinhalten die o.g. englischsprachigen Synonyme für die Begriffe Leistungs- bzw. Mitarbeiterbeurteilung. 2.) Die Veröffentlichungen sind empirisch ausgerichtet, d.h., die Autoren erforschen eine oder mehrere spezifische Fragestellung(en) durch die Auswertung eines Datensatzes mithilfe statistischer Analysemethoden², denen beispielsweise regressionsbasierte Verfahren zugrunde liegen. 3.) Die Aufsätze sind über die o.g. Forschungsdatenbanken (EBSCOhost und Taylor & Francis Online) sowie Suchmaschinen für elektronische Zeitschriften (Google Scholar) zugänglich.

Die betrachteten Studien lassen sich grob in sechs Felder eingliedern: 1.) Untersuchungen, die sich mit den Funktionen der Personalpraktik Leistungsbeurteilung und deren Folgen für Mitarbeiter und Unternehmen auseinandersetzen. 2.) Studien über systematische Verzerrungen durch subjektive Faktoren im Beurteilungsprozess. 3.) Fairnessaspekte, die im Zusammenhang mit dem Einsatz von Leistungsbeurteilungen stehen. 4.) Determinanten von Leistungsbeurteilungssystemen. 5.) Verteilungsempfehlungen und 6.) Charakteristiken der am Beurteilungsprozess beteiligten Personen. Im Folgenden wird auf relevante Studien eingegangen, die sich diesen sechs Feldern zuordnen lassen und für die Skizzierung des Forschungsstands von Bedeutung sind.

1.) Individuelle und organisationale Konsequenzen

Nach Kampkötter und Sliwka (2016) haben Leistungsbeurteilungen in der betriebswirtschaftlichen Praxis mehrere Funktionen. Zum Beispiel werden sie von Unternehmen praktiziert, um Anreize zur Leistungssteigerung zu schaffen, indem sie auf Basis der Beurteilungen jährlich einen Bonus an ihre Beschäftigten auszahlen. Des Weiteren werden auf der

² Formaltheoretische Forschungsarbeiten, die keine empirische Analyse von Daten beinhalten und daher nicht Gegenstand des Literaturüberblicks bilden, behandeln diverse Themen in Bezug auf Leistungsbeurteilungen: z.B. Ungleichbehandlung übergewichtiger Beschäftigter (Bento, White & Zacur, 2012), Anreiz- und Selektionseffekte subjektiver Leistungsbeurteilungen (Diaye & Greenan, 2012), den Einfluss von vorläufigem Feedback in Form von Leistungsbeurteilungen auf die Motivation von Mitarbeitern in dynamischen Turnieren (Ederer, 2010) und dessen Auswirkung auf das Anstrengungsniveau von Beschäftigten in Langzeitprojekten (Ray, 2007). Darüber hinaus untersuchen Golman und Bhatia (2012) die Auswirkungen unzureichender Informationen über die Mitarbeiterleistung auf die Tendenzen zur Milde und zur Mitte bei Vorgesetzten. Grund und Przemeczek (2012) fokussieren ebenfalls diese Bewertungsfehler und beschreiben in einem formalen Modell deren Entstehungsgründe.

Grundlage von Beurteilungen Weiterbildungs- und Trainingsmaßnahmen bestimmt sowie Personalentscheidungen, wie zum Beispiel Kündigungen und Beförderungen getroffen.

Diverse Studien, die sich mit den Funktionen und Konsequenzen von Leistungsbeurteilungen beschäftigen, kommen zu dem Ergebnis, dass Mitarbeiterevaluationen einen Einfluss auf die o.g. Aspekte haben. In ihrem Arbeitspapier mit dem Titel „*What do performance appraisals do?*“, das auf Daten einer Fallstudie basiert, zeigen Cappelli und Conyon (2016) exemplarisch auf, dass Leistungsbeurteilungen in Zusammenhang stehen mit Gehaltssteigerungen, Leistungszulagen, Beförderungen und Kündigungen. Auch Frederiksen, Lange und Kriechel (2017) finden in ihrer Studie einen Zusammenhang zwischen Leistungsbeurteilungen und den Kennzahlen Grundgehalt, Leistungszulagen, Beförderungen sowie Kündigungen. In Bezug auf Kennzahlen über die Karriereverläufe von Mitarbeitern bestätigen des Weiteren Smeets, Waldman und Warzynski (2013), dass neben Gehaltssteigerungen auch Beförderungswahrscheinlichkeiten in Beziehung stehen mit Leistungsbeurteilungen.

Neben den individuellen Konsequenzen für Beschäftigte bilden auch die organisationalen Auswirkungen des Einsatzes von Leistungsbeurteilungen einen Fokus der Forschungsarbeiten in diesem Gebiet. Untersuchungsmerkmale sind: die Angleichung individueller Mitarbeiterziele an die Gesamtziele der Organisation und dessen Auswirkungen auf die Unternehmensleistung (Ayers, 2015; Nisar, 2007); der Zusammenhang zwischen der Wahl einer Wettbewerbsstrategie und eines spezifischen Beurteilungssystems (zum Beispiel für Entwicklungs- oder administrative Zwecke) und dessen Einfluss auf die Unternehmensleistung (Bayo-Moriones, Galdon-Sanchez & Martinez-de-Morentin, 2016). Auch die Konsequenzen der Einführung eines Leistungsbeurteilungsinstrumentes auf die betrieblichen Einnahmen (Breuer & Zimmermann, 2010) sowie die Effekte subjektiver Leistungszulagen auf die Produktivität und Profitabilität von Unternehmen (Gibbs, Merchant, van der Stede & Vargus, 2004) waren Untersuchungsgegenstände weiterer Studien. Die zitierten Untersuchungen zeigen, dass individuelle und organisationale Faktoren nicht vernachlässigt werden sollten, wenn Unternehmen Leistungsbewertungen praktizieren.

2.) Verzerrungen durch systematische Beurteilungsfehler

Leistungsbeurteilungen können einen Einfluss auf Kennzahlen haben, die in Verbindung mit den Karrieren von Mitarbeitern oder dem Erfolg von Unternehmen stehen. Allerdings weisen diverse Untersuchungen darauf hin, dass Leistungsbeurteilungen durch eine Viel-

zahl von Faktoren beeinflusst werden, die zu Beurteilungsfehlern führen. Ein Großteil dieser Literatur beschäftigt sich mit Faktoren, die Verzerrungseffekte hervorrufen. Auf diese soll im Folgenden eingegangen werden.

Eine Ursache von Verzerrungen ist der soziale Kontext, in dem Leistungsbeurteilungen durchgeführt werden. Dies betreffend haben Studien zum Beispiel untersucht, wie die Häufigkeit der Wechselbeziehung zwischen beurteilenden und beurteilten Individuen (Saffie-Robertson & Brutus, 2014) oder die Zuneigung zum Vorgesetzten (Sutton, Baldwin, Wood & Hoffman, 2013; Varma, Pichler & Srinivas, 2005) sich auf die Leistungsbeurteilung der Beschäftigten auswirken. Darüber hinaus fokussieren weitere Forschungsarbeiten die Qualität des Vorgesetzten-Beschäftigten-Verhältnisses und dessen Auswirkung auf die Leistungsbeurteilung von Mitarbeitern (Elicker, Levy & Hall, 2006; Pichler, 2012). Auch der Einfluss der sozialen Bindung zum Vorgesetzten, operationalisiert durch die Anzahl der Mitarbeiter in einem Team, bildet den Untersuchungsgegenstand einer Studie (Breuer, Nieken & Sliwka, 2013).

Neben dem sozialen Kontext stellen systematische Urteilsfehler von Vorgesetzten einen weiteren Faktor dar, der Verzerrungen im Rahmen von Personalbeurteilungen hervorrufen kann. In diesem Zusammenhang wurde untersucht, inwieweit objektive Leistungsbeurteilungen durch vorherige subjektive Bewertungen (Woods, 2012) oder subjektive Beurteilungen durch ein zusätzliches objektives Maß beeinflusst werden (Bol & Smith, 2011). Weitere systematische Verzerrungen werden zum Beispiel hervorgerufen durch eine Mittelwertorientierung sowie eine Tendenz zur Milde von Vorgesetzten (Bol, 2011; Moers, 2005), subjektive Bewertungsspielräume von Vorgesetzten (Ahn, Hwang & Kim, 2010), Halo-Effekte (Jennings, Palmer & Thomas, 2004), Leistungstrends bzw. Performance-schwankungen von Mitarbeitern (Lee & Dalal, 2011) sowie Ankereffekte (Thorsteinson, Breier, Atwell, Hamilton & Privette, 2008).

Die zitierten Studien zu Verzerrungseffekten zeigen, dass Leistungsbeurteilungen durch systematische Bewertungsfehler beeinflusst werden können. Dies hat möglicherweise Auswirkungen auf das Fairness- und Gerechtigkeitsempfinden der Mitarbeiter im Unternehmen und könnte dazu führen, dass als subjektiv oder unfair wahrgenommene Leistungsbeurteilungen zum Beispiel das Leistungsniveau der Mitarbeiter negativ beeinflussen oder deren Kündigungswahrscheinlichkeiten erhöhen. Auf diese Aspekte soll im Folgenden eingegangen werden.

3.) Fairness und Gerechtigkeit

Fairnessaspekte, die im Zusammenhang mit der Anwendung von Leistungsbeurteilungen stehen, bilden ein weiteres Forschungsgebiet zum Thema Leistungsbewertungen. Aktuelle Untersuchungen in diesem Bereich zeigen auf, dass als unfair oder ungerecht wahrgenommene Leistungsbeurteilungen einen Einfluss auf den Karriereverlauf von Beschäftigten, die Zufriedenheit mit dem Job oder auch Kündigungsabsichten haben können. Nachfolgend soll auf Forschungsarbeiten über Leistungsbeurteilungen eingegangen werden, die sich mit den Themen Fairness bzw. Gerechtigkeit auseinandersetzen und die Folgen unfairer Bewertungen für Beschäftigte und Unternehmen aufzeigen.

Studien, die die Beziehung zwischen Gerechtigkeit und Zufriedenheit im Kontext von Leistungsbeurteilungen untersuchen, kommen zu dem Ergebnis, dass Gerechtigkeit tendenziell nicht nur die Zufriedenheit der Beschäftigten mit dem Leistungsbeurteilungssystem als solches beeinflusst (Dusterhoff, Cunningham & MacGregor, 2014; Narcisse & Harcourt, 2008; Steensma & Visser, 2007; Sudin, 2011; Thurston & McNall, 2010), sondern auch die Zufriedenheit der Beschäftigten mit ihrem Job (Lau, Wong & Eggleton, 2008; Selvarajan & Cloninger, 2012; Sholihin & Pike, 2009; Umair, Javaid, Amir & Luqman, 2016) als auch die Zufriedenheit mit ihrer Vergütung (Swiercz, Bryan, Eagle, Bizzotto & Renn, 2012).

Darüber hinaus stehen ungerechte Leistungsbeurteilungen im Verhältnis zum organisationalen Commitment (Farndale & Kelliher, 2013; Steensma & Visser, 2007) und der Kündigungsabsicht der Mitarbeiter (Brown, Hyatt & Benson, 2010).³ In Bezug auf Letzteres zeigt Poon (2004) jedoch auf, dass ungerechte Leistungsbeurteilungen, die zum Beispiel durch Verzerrungen hervorgerufen werden, in keinem Verhältnis stehen mit der Arbeitszufriedenheit der Beschäftigten und letztlich auch deren Kündigungsabsichten. Abgesehen von den Auswirkungen auf die Kündigungsabsichten der Mitarbeiter können unfaire Leistungsbeurteilungspraktiken auch einen negativen Einfluss auf die Einstellung der Mitarbeiter zu angebotenen Personalentwicklungs- und Weiterbildungsmöglichkeiten in Unternehmen haben (Nurse, 2005).

³ Kuvaas (2006) zeigt einen direkten Zusammenhang zwischen der Zufriedenheit der Belegschaft mit ihren Leistungsbeurteilungen und dem organisationalen Commitment sowie der Kündigungsneigung auf. Auch Kampkötter (2017) fokussiert zwar nicht das Thema Fairness, belegt jedoch im Rahmen seiner Forschungsarbeit über Arbeitszufriedenheit, dass der betriebliche Einsatz formaler Leistungsbeurteilungen positive Auswirkungen hat auf die Zufriedenheit der Mitarbeiter.

Die zitierten Untersuchungen zur Gerechtigkeit bei Leistungsbeurteilungen weisen darauf hin, dass unfaire Beurteilungen die Karriereverläufe von Mitarbeitern negativ beeinflussen können. Forschungsarbeiten in diesem Gebiet setzen sich hauptsächlich mit den Konsequenzen unfairer Beurteilungen auf Faktoren, wie zum Beispiel die Arbeitszufriedenheit oder die Kündigungsneigung der Beschäftigten, auseinander. Allerdings werden in der betrieblichen Praxis Leistungsbeurteilungen nicht von allen Unternehmen eingesetzt. Hierauf deuten nachfolgende Untersuchungen hin, die Determinanten von Leistungsbeurteilungssystemen analysieren.

4.) Determinanten von Beurteilungssystemen

Die Determinanten von Leistungsbeurteilungssystemen lassen sich in individuelle (Geschlecht, Alter, Risikoneigung etc.), arbeitsplatzbezogene (Entgeltgruppe, Übernahme von Verantwortung etc.) und institutionelle Faktoren (Unternehmensgröße, Branche, Gewerkschaftsdichte etc.) einteilen. Die Einflüsse dieser Faktoren werden in Studien häufig gemeinsam betrachtet. Beispielsweise finden Grund und Sliwka (2009) einen positiven Zusammenhang zwischen dem Einsatz von Leistungsbeurteilungen und Faktoren, wie zum Beispiel Geschlecht, Alter sowie Risikobereitschaft von Beschäftigten, Übernahme von Verantwortung, Entgeltgruppe und Unternehmensgröße. Neben diesen Aspekten konnte in weiteren Forschungsarbeiten nachgewiesen werden, dass auch das Angebot an formellem Training und eine Anreizbezahlung einen positiven Einfluss auf den Einsatz von Leistungsbeurteilungen haben (Brown & Heywood, 2005). Weiterbildungsmöglichkeiten und der Einsatz von Gruppenarbeit als arbeitsplatzbezogene Eigenschaften bilden ebenfalls positive Einflussfaktoren (Jirjahn & Poutsma, 2013). Die Betriebszugehörigkeit als individueller Faktor (Addison & Belfield, 2008) sowie die Gewerkschaftsdichte (institutioneller Faktor) hingegen haben einen negativen Einfluss auf den Einsatz von Leistungsbewertungen in Unternehmen (Brown & Heywood, 2005). Auch in der Studie von Jirjahn und Poutsma (2013) wirkt sich der Einfluss von Gewerkschaften negativ auf den Einsatz von Leistungsbeurteilungen in Unternehmen aus. Die Ergebnisse der aufgeführten Studien lassen darauf schließen, dass diverse Faktoren Auswirkungen darauf haben, ob Unternehmen Leistungsbeurteilungen praktizieren und für welche Beschäftigtengruppen sie dies tun.

5.) Verteilungsempfehlungen

Studien zu Verteilungsempfehlungen untersuchen die Auswirkungen von Forced-

Rankings auf die Produktivität von Beschäftigten und auf Sabotageeffekte unter Teammitgliedern (Berger, Harbring & Sliwka, 2013). Darüber hinaus wird analysiert, wie Individuen auf unterschiedliche Formen von Verteilungsempfehlungen reagieren (Blume, Baldwin & Rubin, 2009) und welche Persönlichkeitseigenschaften die Affinität von Beschäftigten zu Verteilungsempfehlungen beeinflussen (Blume, Rubin & Baldwin, 2013). Die zitierten Arbeiten zum Thema Verteilungsempfehlungen zeigen, dass seit dem Beginn der Forschung über Leistungsbeurteilungen Beurteilungsformate nach wie vor ein untersuchungsrelevantes Themengebiet bilden.

6.) Merkmale der am Beurteilungsprozess beteiligten Personen

Neben den oben genannten Aspekten bilden die Merkmale der am Bewertungsprozess beteiligten Personen, also der Vorgesetzten und Beschäftigten, einen weiteren Forschungszweig. Untersuchungsgegenstände sind beispielsweise die Hierarchiestufe des Beurteilten (Eremin, Wolf & Woodard, 2010), das Geschlecht der Führungskräfte (Payne, 2011), der Geschlechteranteil unter den Vorgesetzten (Maas & Torres-González, 2011), deren geschlechterspezifisches Bewertungsverhalten (Breuer, 2011) sowie deren Selbstbild (Brutus, Fletcher & Baldry, 2009; Mishra & Roch, 2013) oder Persönlichkeit (Ogunfowora, Bourdage & Lee, 2010). Die abhängigen Variablen, die diesen zitierten Studien zugrunde liegen, unterscheiden sich und bilden beispielsweise die Streuungen der Leistungsbeurteilungen (Breuer, 2011) oder die Erwartungen der Bewerteten an ihre Bewertungen (Eremin et al., 2010; Maas & Torres-González, 2011).

Fazit zum aktuellen Forschungsstand

Die beschriebenen sechs Forschungsrichtungen zeigen, dass in der betrachteten Periode die Forschung insbesondere durch betriebswirtschaftliche Fragestellungen dominiert wurde. Der Großteil der in der Literaturübersicht zitierten Untersuchungen wurde in Managementzeitschriften veröffentlicht. Diese Entwicklung ist ein Hinweis dafür, dass Leistungsbeurteilungen nicht nur in der psychologischen Forschung relevant sind (siehe Kapitel 2.1), sondern auch vermehrt disziplinübergreifend untersucht werden.

Insbesondere die oben erwähnten Merkmale der am Beurteilungsprozess beteiligten Personen sind in den vergangenen Jahren verstärkt in den Fokus der Forschung über Leistungsbeurteilungen gerückt. Ein wesentlicher Faktor, der in diesem Zusammenhang als Einfluss auf die Bewertungen diskutiert wurde, ist die Beziehung zwischen dem Ge-

schlecht der am Prozess beteiligten Individuen auf die Bewertungen. Auf diese Entwicklung weisen mehrere Metaanalysen hin, die Geschlechtereffekte in Laborexperimenten (Koch et al., 2015) als auch in Feldstudien untersuchen (Joshi et al., 2015; Roth et al., 2012). Vor dem Hintergrund der Fragestellung dieser Arbeit soll im nachfolgenden Kapitel der Forschungsstand zu Geschlechtereffekten in Leistungsbeurteilungen gesondert aufgezeigt werden.

2.3 Geschlechterspezifische Effekte in Leistungsbeurteilungen

Kapitel 2.3.1 gibt zunächst eine Übersicht über Studien, die den Haupteffekt des Geschlechts in Leistungsbeurteilungen analysieren. In diesem Zusammenhang wird auch auf Moderatoren eingegangen, die verstärkend bzw. verringernd auf Geschlechtereffekte wirken. Im Anschluss daran wird in Kapitel 2.3.2 eine Übersicht über weitere Moderatoren gegeben, die im Kontext von Leistungsbeurteilungen ebenfalls von Bedeutung sind und in bisherigen Metaanalysen nicht berücksichtigt wurden. Hierbei handelt es sich um den Formalisierungsgrad von Leistungsbeurteilungen sowie die Geschlechterkomposition zwischen dem Vorgesetzten und dem Beschäftigten.

2.3.1 Haupteffekte

Die Ergebnisse von Metaanalysen zu Geschlechtereffekten basieren auf individuellen Studien, die zum Beispiel Leistungsdaten oder Leistungsbeurteilungen von Frauen und Männern in Bezug auf geschlechterspezifische Verzerrungen untersuchen. Nach Bowen et al. (2000) haben diese Individualstudien mit Daten von Beschäftigten in Unternehmen zu keinem eindeutigen Ergebnis über Geschlechtereffekte in Bewertungen beigetragen. Zum einen existieren Studien, die aufzeigen, dass Männer im Rahmen von Leistungsbeurteilungen vergleichsweise besser beurteilt werden als ihre Kolleginnen (Charles, 1981; Pulakos, White, Oppler & Borman, 1989). Andererseits finden einige Forschungsarbeiten eine positive Beurteilungsverzerrung für Frauen (Mobley, 1982; Pulakos & Wexley, 1983) oder kommen zu dem Ergebnis, dass keine Geschlechterverzerrungen in Leistungsbeurteilungen nachgewiesen werden können (Kirnan, Farley & Geisinger, 1989; Yammarino & Dubinsky, 1988).

Studien über Geschlechterdifferenzen in Leistungsbeurteilungen lassen sich nach Koch et al. (2015) und Roth et al. (2012) zwei unterschiedlichen Forschungsmethoden zuordnen. Dementsprechend existieren Untersuchungen mit Leistungsdaten von tatsächlichen Be-

schäftigten (Feldstudien) oder Experimentaldaten (Laborstudien). In Studien mit experimentellem Design werden Probanden zum Beispiel Leistungsdaten oder Leistungsbeurteilungen einer fiktiven Person präsentiert, wobei das Geschlecht oder weitere soziodemographische Merkmale manipuliert werden. Im Gegensatz dazu werden im Rahmen von Feldstudien Leistungsdaten von Beschäftigten analysiert.

Um eine zusammenfassende und systematische Übersicht über Geschlechtereffekte in Labor- als auch Feldstudien aufzuzeigen und somit ein umfassendes Bild über den Forschungsstand zu erhalten, sollen im Folgenden die Ergebnisse bisher durchgeführter Metaanalysen⁴ aufgezeigt werden. In diesen Metaanalysen werden die Ergebnisse mehrerer Individualstudien (siehe zum Beispiel zitierte Studien oben) über Geschlechtereffekte systematisch zusammengefasst, indem die Effektstärken einzelner Primärstudien zur Berechnung einer durchschnittlichen Effektgröße herangezogen werden. Diese berechnete durchschnittliche Effektgröße wird als d-Wert bezeichnet (Cohen, 1988). Im Kontext von Leistungsbeurteilungen erfasst dieser Wert die Unterschiede in den durchschnittlichen Leistungsbeurteilungen von Frauen und Männern und wird in Einheiten der Standardabweichung angegeben (Bowen et al., 2000; Koch et al., 2015). Zur Berechnung der einzelnen Effektstärken aus Primärstudien nutzen Koch et al. (2015) die folgende Formel, die auf der Maßzahl Cohens d basiert (Cohen, 1988):

$$d = \frac{M_M - M_F}{\sqrt{\frac{(N_M - 1)SA_M^2 + (N_F - 1)SA_F^2}{N_M + N_F - 2}}},$$

wobei M_M und M_F die Bewertungen von Männern bzw. Frauen im Mittel, N_M und N_F die Anzahl an beurteilenden Personen, die Männer bzw. Frauen bewertet haben, und SA_M und SA_F die Standardabweichungen der Bewertungen von Männern und Frauen abbilden. Nach Landy (2010) haben Metaanalysen gegenüber Individualstudien den entscheidenden Vorteil, dass mit einer größeren Zuversicht Schlussfolgerungen über die Ergebnisse gemacht werden können, da auf statistische Artefakte (Verfälschungen von Messergebnissen) kontrolliert wird und der Einfluss möglicher Moderatorvariablen getestet werden kann.

⁴ Von Beschäftigtenbeurteilungen abgesehen wurden Geschlechterdifferenzen auch in anderen Beurteilungskontexten erforscht, wie beispielsweise in Bezug auf die Beurteilung leitender Führungskräfte (Eagly, Karau & Makhijani, 1995; Eagly, Makhijani & Klonsky, 1992) und die Bewertung von Lebensläufen (Davison & Burke, 2000; Olian, Schwab & Haberfeld, 1988). Der Fokus dieser Arbeit liegt jedoch auf Geschlechterdifferenzen in Leistungsbeurteilungen.

Neben der Analyseebene (Individualstudien versus Metaanalysen) bieten auch die Forschungsmethoden (Feld- versus Laborstudien) ihre Vor- und Nachteile. Eine Prämisse von Untersuchungen mit Daten aus Feldstudien ist zum Beispiel, dass tatsächliche und keine fiktiven Leistungsbeurteilungen verwendet werden, um Geschlechterunterschiede zu ermitteln. Dies steigert die externe Validität von Feldstudien, also die Verallgemeinerbarkeit der Ergebnisse (Koch et al., 2015). Andererseits können in Untersuchungen mit Daten von Beschäftigten aus Unternehmen die Ursachen für Geschlechterdifferenzen nicht eindeutig bestimmt werden, da letztlich nicht geklärt werden kann, ob diese Unterschiede durch Verzerrungseffekte oder durch tatsächliche Leistungsunterschiede zwischen Frauen und Männern hervorgerufen werden (Bowen et al., 2000). In experimentellen Laborstudien hingegen können Verzerrungen aufgedeckt werden, da außer der Variable Geschlecht weitere Merkmale, wie die tatsächliche Leistung, konstant gehalten werden (Roth et al., 2012). Nach Johnson und Cochran (2008) sowie Landy (2008) ist jedoch die externe Validität der Ergebnisse fraglich, weil davon auszugehen ist, dass Studierende, die an Experimenten über Geschlechtereffekte in Leistungsbeurteilungen teilnehmen, anders beurteilen als erfahrene Führungskräfte. Mögliche Folgen hiervon sind abweichende Effektstärken in Labor- und Feldstudien (Koch et al., 2015; Roth et al., 2012).

Tabelle 1 zeigt die ermittelten durchschnittlichen gewichteten Effektstärken (d-Werte) von Metaanalysen auf, die Geschlechterunterschiede in Leistungsbeurteilungen untersucht haben.

Tabelle 1: Metaanalysen zu Geschlechtereffekten in Leistungsbeurteilungen

Autoren (Jahr)	Zeitraum	k	n	d	Konfidenzintervall	Design
Bowen et al. (2000)	bis 1998 ⁵	138	./.	-0,01	-0,05 bis 0,04 (95%-Konfidenzintervall)	F
Roth et al. (2012)	1969-2009	61	45.733	-0,11	-0,33 bis 0,12 (80%-Konfidenzintervall)	F
Joshi et al. (2015)	1985-2013	93	95.882	0,04	-0,12 bis 0,20 (95%-Konfidenzintervall)	F
Koch et al. (2015)	1970-2012	25	6.182	0,04	-0,01 bis 0,08 (90%-Konfidenzintervall)	L

Erläuterungen: k=Anzahl an Effektgrößen; n=Anzahl an Beobachtungen; d=durchschnittlich gewichtete Effektgröße (positive d-Werte deuten auf eine Favorisierung von Männern, negative auf eine Favorisierung von Frauen hin); Forschungsdesign der inkludierten Studien: F=Feldstudien, L=Laborstudien.

Die erste Metaanalyse wurde im Jahr 2000 von Bowen et al. durchgeführt. Sie integriert die Ergebnisse mehrerer Individualstudien mit Felddaten, die auch job-relevante Merkma-

⁵ Im Rahmen ihrer Literaturrecherche nutzen die Autoren mehrere elektronische Datenbanken und schließen daher mehrere Veröffentlichungszeiträume ein. Die analysierten Studien wurden bis 1998 veröffentlicht (Bowen et al., 2000).

le, wie zum Beispiel Hierarchiestufe, Arbeitserfahrung und Bildung der zu beurteilenden Person berücksichtigen und daher einen Vergleich der Geschlechterunterschiede in Leistungsbewertungen ermöglichen. Zur Analyse von Geschlechtereffekten haben die Autoren 27 Studien identifizieren können, deren Ergebnisse auf 32 unterschiedlichen Stichproben basieren. Anhand dieser Stichproben berechnen die Autoren insgesamt $k=138$ gewichtete Effektgrößen (d-Werte). Sie finden in ihrer Metaanalyse heraus, dass Geschlechtereffekte in Leistungsbeurteilungen sowohl ökonomisch als auch statistisch eher zu vernachlässigen sind, da die durchschnittlich gewichtete Effektgröße der betrachteten Studien bei einem d-Wert in Höhe von $-0,01$ liegt.⁶ D.h., dass im Unternehmenskontext die Leistungsbewertungen von Frauen vergleichsweise leicht positiv verzerrt sind. Es gilt jedoch zu beachten, dass das 95%-Konfidenzintervall von $-0,05$ bis $0,04$ reicht und die Zahl 0 einschließt. Mit anderen Worten: Unter den berücksichtigten Studien bestehen auch solche, in denen keine Geschlechtereffekte auftreten.

Eine weitere Metaanalyse, die eine Übersicht zu Geschlechtereffekten in Leistungsbeurteilungen und auch Beförderungsbeurteilungen mit Unternehmensdaten bildet, wurde im Jahr 2012 von Roth et al. durchgeführt. In ihrer Untersuchung analysieren die Autoren Leistungsdaten von Beschäftigten aus Studien, die im Zeitraum 1969 bis 2009 veröffentlicht wurden. Hierbei handelt es sich um 61 Studien mit insgesamt $n=45.733$ Teilnehmern (die Anzahl der ermittelten Effektgrößen entspricht ebenfalls $k=61$). Die Ergebnisse der durchgeführten Metaanalyse lassen darauf schließen, dass Frauen im Mittel vergleichsweise besser beurteilt werden als ihre Kollegen, wobei der durchschnittlich korrigierte d-Wert bei $-0,11$ liegt. Dementsprechend finden die Autoren eine leichte Tendenz zu positiven Leistungsbeurteilungen für Frauen im Kontext von Felddaten. Ähnlich wie bei der oben erwähnten Studie von Bowen et al. (2000) weisen auch Roth et al. (2012) darauf hin, dass das ermittelte Konfidenzintervall die Zahl 0 einschließt. Da das angegebene 80%-Konfidenzintervall von $-0,33$ bis $0,12$ reicht, erhalten Männer in einigen Studien vergleichsweise bessere Leistungsbewertungen (positiver d-Wert). Jedoch zeigt die Mehrheit der betrachteten Untersuchungen eine Tendenz zu positiveren Leistungsbeurteilungen bei Frauen (negativer d-Wert) auf. Zusammenfassend betrachtet sind die Geschlechterunterschiede in Leistungsbeurteilungen jedoch eher zu vernachlässigen.

⁶ Nach Cohen (1988) sind Effektstärken bis $d=0,20$ als kleine Effekte definiert, die als vernachlässigbar gelten.

Die aktuellste Metaanalyse mit Leistungsdaten von Beschäftigten aus Unternehmen stammt von Joshi et al. (2015). Neben Geschlechterunterschieden in Leistungsbeurteilungen werden in ihrer Studie auch geschlechterspezifische Differenzen in der Entlohnung (zum Beispiel im Gehalt, in Leistungszulagen und bei Beförderungen) von Beschäftigten fokussiert. Zur Bestimmung der Unterschiede in Leistungsbeurteilungen werden die Ergebnisse aus 73 Studien mit $n=95.882$ Teilnehmern berücksichtigt ($k=93$ Effektstärken). Im Gegensatz zu den obigen Metaanalysen mit Felddaten berechnen Joshi et al. (2015) eine positive durchschnittliche Effektstärke in Höhe von $d=0,04$ mit einem 95%-Konfidenzintervall, das von $-0,12$ bis $0,20$ reicht. Dies bedeutet, dass ceteris paribus Männer im Mittel bessere Leistungsbeurteilungen erhalten als Frauen, wobei – wie in den beiden vorherigen Studien – der Geschlechtereffekt gering ist und nicht auf eine systematische Verzerrung hindeutet. Da die Autoren auch Geschlechtereffekte in Bezug auf die Entlohnung von Beschäftigten betrachten, können die berechneten geschlechterspezifischen Unterschiede in Leistungsbeurteilungen mit denen in der Vergütung von Beschäftigten verglichen werden. Nach Joshi et al. (2015) ist der durchschnittliche Geschlechtereffekt in der Vergütung von Beschäftigten schätzungsweise 14 Mal höher als in Leistungsbeurteilungen.

Die einzige Metaanalyse mit Experimentaldaten wurde im Jahr 2015 von Koch et al. publiziert und beschäftigt sich neben Geschlechterverzerrungen in Leistungsbeurteilungen auch mit geschlechterspezifischen Differenzen im Kontext diverser Personalentscheidungen, zum Beispiel in Bezug auf Einstellungen, Beförderungen und Vergütung. Zur Untersuchung von Geschlechterverzerrungen in Leistungsbeurteilungen identifizieren die Autoren $k=25$ Effektstärken mit $n=6.182$ Beobachtungen. Nach Koch et al. (2015) beträgt der Durchschnitt dieser Effektstärken $d=0,04$, wobei das berechnete 90%-Konfidenzintervall von $-0,01$ bis $0,08$ reicht und – wie bei den oben zitierten Studien auch – die 0 einschließt. Die Ergebnisse dieser Metaanalyse bestätigen, dass Männer in Laborstudien zwar leicht positiv verzerrte Beurteilungen erhalten, jedoch ist der Geschlechterunterschied auch im Kontext experimenteller Settings eher zu vernachlässigen. Demensprechend sind die Ergebnisse aus Feldstudien mit denen aus Laborstudien vergleichbar.

Die Ergebnisse der beschriebenen Metaanalysen zeigen, dass der Haupteffekt des Geschlechts im Rahmen von Leistungsbeurteilungen eher gering ist und dass das Geschlecht der beurteilten Person offensichtlich keine systematischen Verzerrungen verursacht. Dies

bestätigt auch Landy (2010), der einen Literaturüberblick über den Einfluss demographischer Merkmale (zum Beispiel Alter, Geschlecht, Hautfarbe) auf die Leistungsbeurteilung gibt und die Ergebnisse ausgewählter Metaanalysen und aktueller Individualstudien zusammenfasst. In Bezug auf das Thema Geschlechtereffekte bei Leistungsbeurteilungen kommt er zu dem folgenden Fazit: „ ... *my conclusion based on 30 years of research on the question is similar to the conclusion that Jim Farr and I drew in 1980: The evidence of any systematic discrimination in the ratings of working women is scarce and possibly localized*” (Landy, 2010: 241).

Neben dem Geschlechterhaupteffekt sind in Studien auch die Wirkungsweisen potentieller Moderator- und Interaktionsvariablen getestet worden, zum Beispiel: Geschlechterverteilungen in Berufen sowie unter den Vorgesetzten (Koch et al., 2015), Job-Komplexität (Joshi et al., 2015), Industrie bzw. Branche (Roth et al., 2012) und berufsbezogene Geschlechterstereotype (Bowen et al., 2000). Eine Berücksichtigung dieser Moderatoren deutet allerdings darauf hin, dass Geschlechtereffekte in Abhängigkeit der oben genannten Faktoren doch einen Einfluss auf die Leistungsbeurteilungen von Beschäftigten haben können. Im folgenden Kapitel soll daher auf den Einfluss von Moderator- und Interaktionsvariablen eingegangen werden, die verstärkend bzw. vermindern auf Geschlechtereffekte in Leistungsbeurteilungen wirken.

2.3.2 Moderatoreffekte

In bisherigen Studien über Geschlechtereffekte in Leistungsbeurteilungen sind auch die Einflüsse unterschiedlicher Moderatoren getestet worden. Auf die bedeutendsten Einflussfaktoren soll im Folgenden näher eingegangen werden. Diese sind: Geschlecht der beurteilenden Person, Geschlechterstereotype und -verteilungen in Berufen sowie Job-Komplexität.

Geschlecht der beurteilenden Person

Eine bedeutende Einflussgröße auf die Leistungsbeurteilung ist das Geschlecht der Führungskraft, die die Performance der Beschäftigten beurteilt. Forschungsarbeiten haben herausgefunden, dass männliche Vorgesetzte im Vergleich zu weiblichen eher traditionelle Geschlechterstereotype haben (zum Beispiel, dass Frauen passiv und zurückhaltend sind (Massengill & Di Marco, 1979)), sexistisches Verhalten tolerieren (Glick et al., 2000), sich weniger für Geschlechtergleichheit einsetzen (Eagly & Mladinic, 1989; Spence &

Hahn, 1997) und Führungspositionen – gemäß dem Motto „*think manager, think male*“ – eher in Verbindung bringen mit männlichen Attributen (Brenner, Tomkiewicz & Schein, 1989; Koenig, Eagly, Mitchell & Ristikari, 2011; Schein, 2001). Diese Faktoren können möglicherweise dazu beitragen, dass männliche Vorgesetzte sich in ihrem Bewertungsverhalten von weiblichen Führungskräften unterscheiden. Beispielsweise belegen Untersuchungen von Landy und Farr (1980, 1983) und Breuer (2011), dass männliche Führungskräfte anders beurteilen als ihre Kolleginnen. In der letztgenannten Studie von Breuer (2011) zeigt sich, dass männliche Führungskräfte weniger differenziert als Frauen beurteilen.

Tabelle 2 zeigt eine Gegenüberstellung von Metaanalysen, die in ihren Untersuchungen das Geschlecht von Beurteilern als Einflussfaktor auf die Bewertung einbezogen haben. Die aufgelisteten Untersuchungen fokussieren unterschiedliche Beurteilungsarten: zum Beispiel Bewertungen von Verhalten und Kompetenzen (Swim et al., 1989), Bewertungen von Führungskräften (Eagly et al., 1992), Leistungsbeurteilungen von Beschäftigten (Bowen et al., 2000) und Beurteilungen in Bezug auf Leistungen, Kompetenzen oder Beförderungen (Koch et al., 2015). In den erwähnten Metaanalysen mit Labor- als auch Felddaten klassifizieren die Autoren ihre berücksichtigten Primärstudien in Untersuchungen mit nur Beurteilern, nur Beurteilerinnen und Studien mit unterschiedlichen Geschlechtern unter den bewertenden Personen (siehe Spalte Konfidenzintervall für d).

Tabelle 2: Effekt des Vorgesetztengeschlechts auf Beurteilungen in Metaanalysen

Autoren (Jahr)	Gegenstand der Untersuchung	Konfidenzintervall für d	Design
Swim et al. (1989)	Beurteilungen in diversen Dimensionen (zum Beispiel Verhalten und Kompetenzen)	Nur männlich: 0,03 bis 0,09	L
		Gemischt: ./.	
		Nur weiblich: -0,02 bis 0,06	
Eagly et al. (1992)	Bewertungen von Führungskräften	Nur männlich: 0,09 bis 0,21	L
		Gemischt: -0,03 bis 0,05	
		Nur weiblich: -0,04 bis 0,12	
Bowen et al. (2000)	Leistungsbeurteilungen von Beschäftigten	Nur männlich: 0,16 bis 0,45	F
		Gemischt: -0,25 bis -0,09	
		Nur weiblich: ./.	
Koch et al. (2015)	Bewertungen von Beschäftigten in diversen Dimensionen (zum Beispiel Leistung, Kompetenzen, Beförderungen)	Nur männlich: 0,16 bis 0,27	L
		Gemischt: ./.	
		Nur weiblich: -0,02 bis -0,10	

Erläuterungen: Für die ersten drei Metaanalysen in der Tabelle wird das 95%-Konfidenzintervall für die durchschnittliche Effektgröße d angegeben. Die Werte für d in der Untersuchung von Koch et al. (2015) beziehen sich auf das 90%-Konfidenzintervall. Forschungsdesign der inkludierten Studien: F=Feldstudien, L=Laborstudien.

Konsistente Ergebnisse im Hinblick auf das Geschlecht von Beurteilern zeigen sich lediglich für Studien, in denen nur Männer Bewertungen durchgeführt haben. Es bestätigt sich, dass Männer vergleichsweise besser beurteilt werden als Frauen, wenn alle Beurteiler männlich sind. In diesen Fällen schließt das 95%-Konfidenzintervall für die durchschnittlich ermittelte Effektstärke d nicht die 0 ein, wobei die Effektgrößen moderat sind (siehe Werte für die durchschnittliche Effektstärke d). Für gemischtgeschlechtliche Zusammensetzungen unter den bewertenden Individuen oder Studien mit nur Beurteilerinnen zeigen sich keine konsistenten Ergebnisse.

Der Einfluss des Vorgesetztengeschlechts als Moderator im Hinblick auf Leistungsbeurteilungen von Beschäftigten wurde in den Metaanalysen von Bowen et al. (2000) und Koch et al. (2015) berücksichtigt. In Bezug auf Studien mit nur Beurteilerinnen erwähnen die Autoren der erstgenannten Metaanalyse: „*None of the studies in the ... meta-analysis involved all-female raters*“ (Bowen et al., 2000: 2207). Da auch Koch et al. (2015) im Rahmen ihrer Analyse von Leistungsbeurteilungen keine Studie mit ausschließlich Beurteilerinnen berücksichtigen, können sie lediglich Effektstärken für Studien mit nur männlichen Beurteilern und unterschiedlichen Geschlechterverteilungen analysieren. Die Ergebnisse sind ebenfalls moderat und mit denen der oben erwähnten Metaanalysen vergleichbar.

Stereotype und Geschlechterverteilungen in Berufsgruppen

Ein weiterer Moderator, der auf den Zusammenhang zwischen dem Geschlecht und der Leistungsbeurteilung wirkt, sind berufs- und geschlechterspezifische Stereotype und die Verteilung von Geschlechtern in Berufen sowie Branchen. Im Folgenden soll zunächst auf die allgemeine Bedeutung von Stereotypen eingegangen werden. Anschließend wird erläutert, inwieweit Stereotype dazu beitragen können, dass beispielsweise Frauen in männerdominierten Berufen oder Branchen anders beurteilt werden könnten als vergleichbare Kollegen.

Nach Agars (2004) sowie Welle und Heilman (2007) sind Stereotype akzeptierte Vorstellungen über Personen einer bestimmten Gruppe, die zur Folge haben, dass den Mitgliedern dieser Personengruppe durch Kategorisierungsprozesse dieselben Gruppeneigenschaften zugewiesen werden. Diese Gruppenstereotype sind verbunden mit Erwartungen an Verhaltensweisen der Gruppenmitglieder und können Beurteilungsverzerrungen hervorrufen, da nicht die tatsächlichen, sondern die verallgemeinerten Fähigkeiten einer Person beurteilt werden (Koch et al., 2015). Stereotype können automatisch und unbewusst aktiviert wer-

den und sowohl zweckmäßig als auch zutreffend sein (Devine, 1989; Goodwin, Gubin, Fiske & Yzerbyt, 2000; Macrae, Milne & Bodenhausen, 1994).

Insbesondere Geschlechterstereotype werden in der Regel schnell aktiviert, da das Geschlecht ein offensichtliches Gruppenmerkmal darstellt und sich leicht identifizieren lässt (Banaji & Hardin, 1996; Banaji, Hardin & Rothman, 1993; Blair & Banaji, 1996). Beispielsweise sind nach Eagly und Karau (2002) anerkannte Geschlechterstereotype, dass Frauen im Vergleich zu Männern eher Fürsorge für Andere zeigen, da sie unter anderem als hilfsbereit, freundlich und emotional ausdrucksfähig beschrieben werden. Männer hingegen werden allgemein beschrieben als durchsetzungsfähig, da ihr Verhalten als dominant, unabhängig, ehrgeizig und selbstbewusst eingestuft wird. Geschlechterstereotype können nicht nur deskriptiv sein, sondern auch präskriptiv, d.h. sie können vorgeben, wie Frauen oder Männer sein sollen (Eagly, 1987).

Ähnlich wie die oben beschriebenen Geschlechterstereotype wirken berufsspezifische Stereotype (Rollenstereotype). Sie beschreiben, welche Eigenschaften Beschäftigte haben sollten, um spezifische Berufe (Rollen) ausüben zu können (Eagly & Karau, 2002). Beispielsweise sind die Eigenschaften, die man mit einer erfolgreichen Führungsperson in Verbindung bringt Durchsetzungsfähigkeit und emotionale Stärke. Diese Eigenschaften werden erfahrungsgemäß nicht Frauen, sondern eher Männern zugeschrieben (Schein, 2001). Falls ungeachtet dessen eine Führungsposition durch eine Frau besetzt wird, kann es sein, dass eine Diskrepanz aus dem Gruppen- (dem Geschlecht) und Rollenstereotyp (der Führungsposition) resultiert. Die Folge hiervon sind negative Erwartungen in Bezug auf die Leistung der Führungskraft, die wiederum verzerrte Beurteilungen zur Folge haben (Heilman & Eagly, 2008). Studien im Bereich von Leistungsbeurteilungen unterstützen diese Annahme (Lyness & Heilman, 2006).

Abgesehen von den Erläuterungen oben können stereotype Erwartungen an Frauen und Männer in bestimmten Berufen auch durch die geschlechterspezifischen Zusammensetzungen von Berufszweigen beeinflusst werden (Glick, 1991; Gorman, 2005; Heilman, 1983). Dieser Argumentation zufolge werden frauendominierte Berufe (zum Beispiel in den Bereichen Pflege, Erziehung, Reinigung) als adäquat für Frauen und männerdominierte Berufe (zum Beispiel in den Bereichen Technik, Produktion) als angemessen für Männer betrachtet (Joshi et al., 2015). Beispielsweise ist einer Studie von Hausmann und Klei- nert (2014) zufolge, die Daten für den deutschen Arbeitsmarkt auswertet, das Ausmaß der

beruflichen horizontalen Segregation über einen langfristigen Zeitraum betrachtet (1976 bis 2010) hoch und ist über die letzten vier Jahrzehnte kaum zurückgegangen. Zudem kommen die Autoren der Arbeitsmarktstudie zu dem Ergebnis, dass Frauen und Männer mehrheitlich Berufen nachgehen, die überwiegend von ihrem eigenen Geschlecht ausgeübt werden.

Alle dargelegten Argumente, die im Zusammenhang mit Stereotypen stehen, sagen voraus, dass Frauen speziell in frauendominierten Berufen und Männer entsprechend in männerdominierten Berufen vergleichsweise positiver beurteilt werden. Bisher durchgeführte Studien deuten allerdings auf keine einheitlichen Ergebnisse hin.⁷ Beispielsweise untersuchen Bowen et al. (2000) in ihrer Metaanalyse unterschiedliche Effektgrößen in Leistungsbeurteilungen für typische Frauen-, Männer- und gemischtgeschlechtliche Berufe. Die Autoren finden heraus, dass keine systematischen Bewertungsverzerrungen für Berufe festzustellen sind, die mit geschlechterspezifischen Stereotypen assoziiert werden. Das 95%-Konfidenzintervall für typisch männliche, weibliche und geschlechterneutrale Berufe reicht jeweils vom Negativen bis ins Positive und schließt die 0 ein (feminine Berufe: -0,27 bis 0,13; maskuline Berufe: -0,10 bis 0,04 und gemischtgeschlechtliche Berufe: -0,57 bis 0,01).

Die Resultate in der Untersuchung von Koch et al. (2015) zeigen ebenfalls keine systematischen Geschlechtereffekte in Leistungsbeurteilungen und weiteren Beurteilungsdimensionen auf. In ihrer Untersuchung finden die Autoren eine positive, jedoch geringe durchschnittliche Effektgröße für männerdominierte Berufe, d.h., dass Männer leicht positiver beurteilt werden als Frauen ($d=0,13$; 90%-Konfidenzintervall für männerdominierte Berufe: 0,10 bis 0,16). Für frauendominierte ($d=-0,02$) und gemischtgeschlechtliche Berufe ($d=0,02$) hingegen liegen die Werte für d nahezu bei 0. Auch das jeweils berechnete 90%-Konfidenzintervall für diese beiden letztgenannten Kategorien lässt keine eindeutigen Rückschlüsse über den Einfluss von Geschlechteranteilen in Berufsgruppen auf Leistungsbeurteilungen zu, da die untere Grenze ein negatives und die obere Grenze ein posi-

⁷ Auf eine tabellarische Darstellung der Ergebnisse wird an dieser Stelle verzichtet, da sie auf unterschiedlichen Analysemethoden beruhen. Beispielsweise berechnen Bowen et al. (2000) unterschiedliche Konfidenzintervalle (95%) für die durchschnittlichen Effektgrößen d in typischen Frauen-, Männer- und gemischtgeschlechtlichen Berufen. Joshi, Son und Roh (2015) führen hingegen eine Metaregression durch, um den Einfluss des Männeranteils in Berufen auf die Leistungsbeurteilungen zu bestimmen und geben dementsprechend als Ergebnis einen Regressionskoeffizienten aus. Koch et al. (2015) errechnen unterschiedliche Effektgrößen d und Konfidenzintervalle (80% und 90%) für d hinsichtlich Männer-, Frauen- und gemischtgeschlechtlicher Berufe.

tives Vorzeichen für die ermittelten d-Werte (frauendominierte Berufe: -0,08 bis 0,05 und gemischtgeschlechtliche Berufe: -0,02 bis 0,06) hat.

Neben den Untersuchungen von Bowen et al. (2000) und Koch et al. (2015) beschäftigen sich auch Joshi et al. (2015) mit dem Einfluss des Männeranteils in Berufen auf Leistungsbeurteilungen. Um den Einfluss dieses Moderators in den berücksichtigten Feldstudien genauer zu analysieren, führen die Autoren eine Regressionsanalyse auf Basis von Metadaten durch. Die Ergebnisse ihrer Metaregression deuten darauf hin, dass geschlechterspezifische Differenzen in Leistungsbeurteilungen mit dem Anteil an Männern in Berufsgruppen ansteigen ($\beta=1,37$; $p<0,01$). Dieses Resultat lässt jedoch keine eindeutigen Rückschlüsse auf den Einfluss von Geschlechtergruppen zu, da die Geschlechteranteile für die in den Primärstudien berücksichtigten Berufsgruppen approximiert werden und nicht die tatsächlichen Geschlechteranteile abbilden. Aus diesem Grund lässt sich nur vermuten, dass Männer umso besser beurteilt werden, desto höher ihr Anteil in der jeweiligen Berufsgruppe ist.

Tabelle 3 zeigt die ermittelten Effektstärken in den Experimentalstudien, die Koch et al. (2015) in ihrer Metaanalyse über Geschlechterunterschiede in Leistungsbewertungen berücksichtigt haben. Neben den jeweiligen ermittelten Effektstärken sind in der Tabelle auch detaillierte Informationen über die berücksichtigten Primärstudien enthalten. Dies sind: Anzahl an Beobachtungen, Geschlechterverteilung im Beruf, Geschlecht der beurteilenden Person, die Probandengruppe (Studierende, Beschäftigte oder Fachkräfte) und das jeweilige Land, in dem die Untersuchung durchgeführt wurde. In einigen Studien, wie zum Beispiel die von Brief und Wallace (1976), wurden mehrere Effektstärken berechnet, da mehrere Stichproben (beispielsweise Beschäftigte versus Studierende) vorlagen.

Tabelle 3: Geschlechtereffekte in experimentellen Studien über Leistungsbeurteilungen

Autoren (Jahr)	n	d	Verteilung im Beruf	Geschlecht Beurteilende	Teil- nehmer	Land
Bigoness (1976)	60	-0,41	M	m	S	USA
Brief und Wallace (1976), Sample mit Studenten	113	0,06	F	Nein	S	USA
Brief und Wallace (1976), Sample mit Beschäftigten	47	0,06	F	Nein	B	USA
Butler und Skattebo (2004)	96	-0,08	M	Nein	B	USA
Drogosz und Levy (1996)	180	-0,05	F, M, A	Nein	S	USA
Etaugh und Kasley (1981)	368	0,38	A	Ja	S	USA
Frank und Drucker (1977)	55	-0,07	M	Ja	S	USA
Hamner, Kim, Baird und Bigoness (1974)	36	-0,43	M	Ja	S	USA
Heilman und Haynes (2005), Studie 1	60	0,69	A	Nein	S	USA
Heilman und Haynes (2005), Studie 2	61	0,71	A	Nein	S	USA
Heilman und Haynes (2005), Studie 3	90	0,93	A	Nein	S	USA
Hmurovic (2012)	287	-0,03	M	Nein	S	USA
Knez und Enmarker (1998)	80	0,24	F	Nein	S	Schweden
Martell (1991)	202	0,57	M	Nein	B	USA
Mettrick und Cowan (1996)	122	0,39	M	Nein	S	USA
Moore (1984)	69	0,43	M	Nein	B	USA
Norton, Gustafson und Foster (1977)	3.261	-0,04	M	Nein	F	USA
Pazy (1986)	48	0,99	M	Nein	F	Israel
Plake, Murphy-Berman, Derscheid, Gerber, Miller, Speth und Tomes (1987)	85	-0,15	A	Nein	F	USA
Riegelhaupt (1984), Studie 1	68	0,32	F, M, A	Nein	S	USA
Riegelhaupt (1984), Studie 2	304	-0,03	F, M, A	Nein	S	USA
Robbins (1993)	70	0,02	F, M	Nein	S	USA
Schmitt und Lappin (1980)	64	0,07	F	Ja	S	USA
Terborg und Ilgen (1975)	36	0,38	M	m	S	USA
Wood (1999)	320	-0,44	A	Nein	S	USA

Erläuterungen: n=Anzahl an Beobachtungen; d=Effektgröße der Studie. Positive d-Werte deuten auf eine Favorisierung von Männern, negative d-Werte auf eine Favorisierung von Frauen hin. F=frauendominiert, M=männerdominiert, A=ausgewogen. Nein=Effektgrößen wurden auf Basis gemischter oder unbekannter Geschlechtergruppen kalkuliert, Ja=Effektgrößen wurden auf Basis getrennter Geschlechtergruppen berechnet, m=gesamtes Sample setzt sich aus männlichen Bewertern zusammen. S=Studierende, B=Beschäftigte in einem Arbeitsverhältnis, F=Fachkräfte in einem Arbeitsverhältnis.

Auch die Betrachtung der einzelnen Primärstudienresultate lassen keine Rückschlüsse auf systematische Geschlechtereffekte hinsichtlich der Moderatoren Geschlechterverteilungen und Geschlecht des Vorgesetzten zu. Allerdings ist auffällig, dass verhältnismäßig wenige Studien durchgeführt wurden für Berufe, die frauendominiert sind oder in denen die beurteilenden Personen ausschließlich Frauen waren. Zudem ist zu erwähnen, dass in relativ vielen Untersuchungen Studierende als Probanden ausgewählt wurden und dass ein Großteil der Studien in den USA durchgeführt wurde. Nur zwei der 25 aufgelisteten Experimentalstudien fanden in Israel und Schweden statt.

Job-Komplexität

Ein weiterer Moderator, der in direktem Zusammenhang mit Geschlechterdifferenzen in Leistungsbeurteilungen steht, ist die Komplexität von Berufen. Joshi et al. (2015) definieren komplexe Jobs als Berufe, die eine hohe Problemlösefähigkeit, d.h. Ermessensspielräume und Autonomie, bei den auszuführenden Tätigkeit erfordern (siehe auch Auster, 1989).

Einerseits wird argumentiert, dass kognitive Anforderungen an Vorgesetzte besonders hoch sind, wenn sie Beschäftigte mit komplexen Tätigkeiten bezüglich ihrer Leistung evaluieren und vergüten sollen. Demzufolge ist bei solchen Tätigkeiten anzunehmen, dass Benachteiligungen durch Verzerrungen und Stereotype eine Bedeutung haben (Auster, 1989; Fiske, 1998; Fiske, Bersoff, Borgida, Deaux & Heilman, 1991). Zudem ist es wahrscheinlicher, dass Vorgesetzte diskriminieren und Beurteilungen entsprechend verzerren, wenn spezifische Informationen über die Ausführung eines Berufs nicht dokumentiert werden und dadurch schwer nachzuvollziehen sind (Heilman, 1994; Nieva & Gutek, 1980). Darüber hinaus belegen einige Untersuchungen, dass Arbeitgeber zu diskriminierendem Verhalten neigen, zum Beispiel, wenn die zur Verfügung stehenden Informationen über die Leistung der Mitarbeiter begrenzt oder mehrdeutig sind (Gerdes & Garber, 1983; Mobley, 1982).

Im Gegensatz zu dem Argumentationsstrang, dass Geschlechterverzerrungen mit steigender Komplexität des Berufs wahrscheinlicher werden, stehen die Ergebnisse der Metaanalyse von Roth et al. (2012). Bei ihrer Analyse von Feldstudien wird der Moderator Job-Komplexität aus den unterschiedlichen Berufsgruppen, die in den Primärstudien betrachtet werden, abgeleitet. Beispielsweise schätzen die Autoren den Beruf eines LKW-Fahrers als tendenziell weniger komplex, den Beruf eines leitenden Beschäftigten hingegen als kom-

plex ein. Die Ergebnisse der Metaanalyse lassen darauf schließen, dass die Job-Komplexität keinen systematischen Einfluss auf die Leistungsbeurteilung hat: „*There is no clear trend regarding the moderator of job complexity.*“ (Roth et al., 2012: 730) Die durchschnittliche Effektstärke beträgt für weniger komplexe Berufe $d=-0,07$ und für komplexere Berufe $d=-0,13$.

Joshi et al. (2015) zeigen in ihrer Zusammenfassung von Ergebnissen aus Feldstudien die Tendenz auf, dass die Job-Komplexität einen positiven, jedoch geringen Einfluss auf Geschlechterunterschiede in Leistungsbeurteilungen hat ($\beta=0,06$; $p<0,01$). In Ergänzung zu ihrer quantitativen Metaanalyse führen die Autoren eine qualitativ vergleichende Analyse durch (Qualitative Comparative Analysis, kurz QCA), um zu erforschen, ob bestimmte Kombinationen von den in der Metaanalyse berücksichtigten Variablen dazu führen, dass Männer oder Frauen vergleichsweise bessere Leistungsbeurteilungen erhalten. Die QCA deutet auf zwei hinreichende Konfigurationen hin, bei denen Frauen bessere Leistungsbeurteilungen als Männer erhalten. Bei beiden Konfigurationen ist die Bedingung Job-Komplexität abwesend. Zudem zeigt sich, dass ein ausgeglichener Geschlechteranteil im Job und der Anteil an weiblichen Führungskräften in Berufsbranchen (weitere Moderatorvariable, siehe unten) Substitute bilden, da bei Vorhandensein einer Bedingung die jeweils andere abwesend ist. Auf Basis dieser QCA-Befunde schlussfolgern Joshi et al. (2015), dass Frauen vergleichsweise positiver beurteilt werden, wenn der Geschlechteranteil in ihrem Berufsumfeld ausgewogen ist oder die Branche, in der sie tätig sind, durch einen hohen Anteil an Frauen in Top-Positionen gekennzeichnet ist.

Die durchgeführten Untersuchungen hinsichtlich des Moderators Job-Komplexität zeigen keine konsistenten Ergebnisse. Während Roth et al. (2012) keinen systematischen Einfluss auf Geschlechterunterschiede aufzeigen, finden Joshi et al. (2015) einen positiven, jedoch geringen durchschnittlichen Effekt.

Weitere Moderatoren

Im vorangehenden Abschnitt wurde angedeutet, dass neben den erörterten Moderatoren Geschlecht der beurteilenden Person, Stereotype und Geschlechterverteilungen in Berufen sowie Job-Komplexität, weitere Einflussgrößen existieren, die in Studien über Geschlechterunterschiede in Leistungsbeurteilungen berücksichtigt wurden. Hierunter finden sich beispielsweise: Training der Vorgesetzten zur Durchführung akkurater – also nicht verzerrter – Beurteilungen und geschlechterstereotype Leistungsmaße (Bowen et al., 2000);

die Organisationsform (private oder öffentliche Unternehmen) und die Funktion von Leistungsbeurteilungen in Organisationen (Roth et al., 2012); die Art (job-relevante oder nicht relevante) und die Menge an Informationen, die den Vorgesetzten zur Beurteilung von Personen zur Verfügung stehen; die Motivation der Vorgesetzten zur Durchführung akkurater Beurteilungen sowie das Beurteilungsformat, d.h. absolute oder relative Bewertungen (Koch et al., 2015); die durchschnittliche Betriebszugehörigkeit der Beurteilten; die Berufsprestige und der Anteil an Frauen in Managementpositionen in einer Branche (Joshi et al., 2015).

Neben den genannten Moderatoren wird in Studien über Geschlechterdifferenzen in Leistungsbeurteilungen auch der Einfluss des Veröffentlichungszeitraums erforscht, da angenommen wird, dass bei der Analyse von Geschlechtereffekten auch zeitliche Faktoren bedeutend sind. Dies liegt darin begründet, dass sich die Möglichkeiten für Frauen am Arbeitsplatz über den Zeitverlauf geändert und sich dadurch Geschlechterverzerrungen reduziert haben könnten (Bowen et al., 2000; Joshi et al., 2015; Joshi, Jackson & Liao, 2006; Koch et al., 2015; Roth et al., 2012).

Der skizzierte Forschungsstand zu den beschriebenen Moderatoren lässt darauf schließen, dass zwar eine Vielzahl potentieller Moderatoren untersucht wurde, die Wirkungsweisen auf Geschlechtereffekte jedoch unklar bleiben, da sich kein systematisches Bild aufzeigen lässt. Dennoch bedeutet dies nicht, dass die Forschung zu möglichen Moderatoren vernachlässigt werden sollte. Auf die Gründe hierfür wird im folgenden Kapitel 2.4 näher eingegangen.

2.4 Forschungslücke

In Kapitel 2.3.2 wurde auf Moderatoren eingegangen, deren Wirkungsweisen und Effekte auf die Leistungsbeurteilung bisher ausgiebig erforscht wurden und potentielle Gründe dafür liefern, weshalb die Effektgrößen in Primärstudien sich voneinander unterscheiden. Nachfolgend soll auf zwei weitere Einflussfaktoren eingegangen werden, die indirekte Geschlechtereffekte hervorrufen und bislang wenig erforscht wurden. Dies sind zum einen der Formalisierungsgrad von Leistungsbeurteilungssystemen, zum anderen die Geschlechterdyaden aus Vorgesetzten und Beschäftigten.

2.4.1 Formalisierungsgrad

Eine Moderatorvariable, die im Kontext von Leistungsbewertungen bisher kaum erforscht

wurde, jedoch zu indirekten Geschlechtereffekten führen kann, ist der Formalisierungsgrad eines Beurteilungssystems. Im Folgenden wird der Einfluss des Formalisierungsgrads auf Geschlechtereffekte bei Entgeltsystemen zunächst überblicksartig zusammengefasst. Im Anschluss daran soll geklärt werden, inwieweit entgeltrelevante Mitarbeiterbeurteilungen und hieraus resultierende geschlechterspezifische Leistungszulagen durch den Formalisierungsgrad eines Bewertungssystems bedingt werden.

Diverse Studien, die sich mit Gehaltssystemen beschäftigen, sind der Frage nachgegangen, ob Geschlechterungleichheit durch den Grad der Formalisierung eines Systems beeinflusst wird (Abraham, 2017). Formalisierung von Gehaltssystemen kann beispielsweise bedeuten, dass das Verfahren der Entlohnung von Beschäftigten an objektive Faktoren gekoppelt ist, wie beispielsweise die Betriebszugehörigkeit, formelle Qualifikation oder die Leistung. Ein hoch formalisiertes Entgeltsystem ist dadurch gekennzeichnet, dass das Verfahren der Entlohnung und die Regelungen hierzu nicht mehrdeutig sind (Anderson & Tomaskovic-Devey, 1995; Huffman & Velasco, 1997). Weniger formalisierte Entgeltsysteme hingegen bieten den Vorgesetzten Spielraum in Bezug auf die Verteilung des Entgelts an Beschäftigte. Der Spielraum, der den Vorgesetzten hierbei eingeräumt wird, führt jedoch tendenziell zu Bewertungsverzerrungen, von denen in der Regel die männlichen Beschäftigten profitieren (Reskin, 2000; Ridgeway, 1997). In diesem Zusammenhang bestätigen zahlreiche Untersuchungen einen positiven Effekt der Formalisierung von Gehaltssystemen auf die Reduzierung von Geschlechterungleichheit in der Entlohnung (Anderson & Tomaskovic-Devey, 1995; Baron, Hannan, Hsu & Koçak, 2007; Elvira & Graham, 2002; Pfeffer & Cohen, 1984; Reskin & McBrier, 2000; Sutton, Dobbin, Meyer & Scott, 1994). Die Ungleichheit, die durch geschlechterspezifische Verzerrungen auftritt, wird durch Formalisierung verringert, da im Rahmen von Beurteilungen der Entscheidungsspielraum und die Subjektivität von Vorgesetzten eingeschränkt wird (Baron, Davis-Blake & Bielby, 1986; Reskin, 2000).

Im Gegensatz zu dem Forschungsstrang, der eine positive Beziehung zwischen der Formalisierung von Gehaltssystemen und der Geschlechtergleichheit in der Entlohnung aufzeigt, stehen laut Abraham (2017) Untersuchungen, die dieses Verhältnis nicht bestätigen (Castilla, 2008; Dencker, 2008; Kaley, Dobbin & Kelly, 2006; Konrad & Linnehan, 1995). Einige Forschungsarbeiten finden sogar heraus, dass Personalentscheidungen, die auf formalisierten Prozessen basieren, Geschlechterungleichheit am Arbeitsplatz fördern

(Dencker, 2008). Demnach können weniger formalisierte Entgeltsysteme im Vergleich zu hoch formalisierten Systemen dazu beitragen, dass Vorgesetzte ihren Entscheidungsspielraum nutzen, um einer vorherrschenden Geschlechterungleichheit im Unternehmen entgegenzuwirken (Dencker, 2008). Dies ist konsistent mit Befunden über organisationale Bürokratien und formalisierte Routinen, die eine Herstellung von Geschlechtergleichheit am Arbeitsplatz erschweren (Acker, 1990; Kanter, 1993). Obwohl die angesprochenen Studien belegen, dass die Ergebnisse zu Formalisierung von Gehaltssystemen und Geschlechtergleichheit nicht eindeutig sind, bestätigt der überwiegende Teil der Forschungsarbeiten, dass in der Regel eine Formalisierung von Gehaltssystemen zu mehr Geschlechtergleichheit am Arbeitsplatz und im Gehalt führt (Abraham, 2017).

Es wurden bisher zwei Studien durchgeführt, die den Formalisierungsgrad von Leistungszulagen und geschlechterspezifischen Differenzen näher beleuchtet haben. In der Studie von Elvira und Graham (2002), die Daten eines Unternehmens analysieren, werden die jährlichen Leistungszulagen als eine weniger formalisierte Gehaltskomponente betrachtet, da beispielsweise die leistungsbezogenen Kriterien und das Verfahren zur Determinierung der Leistungszulagen von den Vorgesetzten im Unternehmen bestimmt werden. Die Daten ihrer Studie zeigen auf, dass insbesondere in Bezug auf weniger formalisierte Gehaltskomponenten die geschlechterspezifischen Gehaltsunterschiede höher ausfallen als bei Gehaltssystemen mit höherem Formalisierungsgrad (zum Beispiel Grundgehalt, Gehaltssteigerungen). Abraham (2017) analysiert ebenfalls Leistungszulagen und geschlechterspezifische Unterschiede von Beschäftigten eines Fallbetriebs, wobei die Leistungszulagen auf Basis standardisierter Leistungskriterien festgelegt und dementsprechend als formalisierte Gehaltskomponente interpretiert werden. Er findet im Gegensatz zu Elvira und Graham (2002) heraus, dass weibliche Vorgesetzte in vergleichsweise weniger formalisierten Gehaltssystemen ihren Entscheidungsspielraum nutzen, um Geschlechtergleichheit, speziell in unteren Gehaltsgruppen zu fördern.

Die Ergebnisse von Elvira und Graham (2002) als auch Abraham (2017) belegen, dass der Formalisierungsgrad von Gehaltssystemen (wie auch Leistungsbeurteilungen und Leistungszulagen) einen indirekten Faktor darstellt, der Auswirkungen auf Geschlechterunterschiede haben kann. Darüber hinaus bestätigt die Studie von Abraham (2017), dass Geschlechterdyaden zwischen Beschäftigten und Vorgesetzten relevant sind, um Geschlechterunterschiede in Leistungsbeurteilungen zu erklären.

Zusammenfassend lässt sich festhalten, dass die bisherige Forschung zum Moderator Formalisierungsgrad von Leistungsbeurteilungen auf indirekte Geschlechtereffekte hinweist, die nur in spezifischen Kontexten auftreten. Das Vorhandensein von Geschlechterunterschieden in Leistungsbeurteilungen hängt offensichtlich von organisationalen Praktiken und den spezifischen Gegebenheiten in Unternehmen ab. Dieser Argumentation zufolge sollten zukünftige Forschungsarbeiten laut Abraham (2017) nicht den Fokus darauf legen, ob Formalisierung Geschlechterungleichheit reduziert, sondern unter welchen Umständen diese Beziehung bestätigt wird. In diesem Zusammenhang kann die Aufdeckung von Kausalmechanismen helfen, indem beispielsweise die Wirkungsweisen organisationaler Praktiken analysiert werden, um zu verstehen, wie Geschlechterunterschiede im Arbeitskontext entstehen (Reskin, 2008).

2.4.2 Geschlechterdyaden

Neben der Erforschung von Geschlechterhaupteffekten und Moderatoren, wie zum Beispiel dem Einfluss des Vorgesetztingeschlechts oder den Geschlechteranteilen unter den Beurteilten, ist es bedeutsam, dass auch Effekte untersucht werden, die durch Interaktionen demographischer Variablen bedingt werden: *„It would be ... useful to conduct more research on multiple, potentially interacting, demographic variables“* (Roth et al., 2012: 734). In diesem Zusammenhang bilden beispielsweise Geschlechterdyaden einen möglichen Einflussfaktor. Hiermit sind geschlechterspezifische Zusammensetzungen aus Vorgesetzten und Beschäftigten bzw. Interaktionseffekte gemeint, die durch das Zusammenwirken beider Geschlechter hervorgerufen werden und Auswirkungen auf die Leistungsevaluationen von Mitarbeitern haben können. Auch Landy (2010), der eine Übersicht zu demographischen Einflussfaktoren (Alter, Geschlecht, Hautfarbe, Behinderung) auf Leistungsbeurteilungen bietet, weist in seiner Übersichtsstudie unter anderem darauf hin, dass in komplexen Analysen über Stereotype der Einfluss von Geschlechterdyaden nicht unberücksichtigt bleiben sollte: *„ ... I would ... argue that these more complex analyses should be accompanied by more basic analyses of Ratee x Rater gender interaction effects using repeated measures and control variables in the conduct of the research“* (Landy, 2010: 244).

Darüber hinaus ist die Analyse von Geschlechterdyaden wichtig, da aus der Untersuchung der oben erörterten Haupteffekte nur Vermutungen darüber angestellt werden können, welche spezifischen Geschlechterkombinationen zwischen Beschäftigten und Vorgesetz-

ten zu unterschiedlichen Leistungsbeurteilungen führen. Ein näherer Blick auf Kategorisierungsprozesse, die möglicherweise bei der Durchführung von Bewertungen wirken, kann Aufschluss über den Einfluss von Geschlechterdyaden bringen. Nach Varma und Stroh (2001) haben Kategorisierungsprozesse zur Folge, dass Beschäftigte bessere Ergebnisse in ihren Beurteilungen erhalten, eine geringe Fluktuationsrate aufweisen und zufriedener sind mit ihrem Verhältnis zur Führungskraft, wenn diese dasselbe Geschlecht hat. Das Geschlecht als soziodemographisches Merkmal stellt für Führungskräfte eine beobachtbare Eigenschaft dar, die die Wahrnehmung von Ähnlichkeit beeinflusst (Graves & Powell, 1995). Durch eine höhere wahrgenommene Ähnlichkeit können sich beispielsweise interpersonelle Vorlieben entwickeln. Diese haben wiederum zur Folge, dass die Interaktionen zwischen gleichgeschlechtlichen Vorgesetzten und Beschäftigten zunehmen und infolge dessen Personalentscheidungen positiv verzerrt werden (Graves & Powell, 1995, 1996; Powell & Graves, 2003). Folglich können im Rahmen von Personalbeurteilungen Vorgesetzte diejenigen Beschäftigten besser beurteilen, die ihnen hinsichtlich sichtbarer demographischer Merkmale – wie dem Geschlecht – ähnlich sind.

Die dargelegten Argumente lassen darauf schließen, dass im Rahmen von Leistungsbeurteilungen spezifische Geschlechterkombinationen zwischen Beschäftigten und Vorgesetzten zu positiven bzw. negativen Verzerrungen führen können. Vor diesem Hintergrund ist es erstaunlich, dass im Rahmen bisher durchgeführter Metaanalysen über Geschlechtereffekte in Leistungsbeurteilungen nicht die Interaktion zwischen den Geschlechtern der Beschäftigten und Vorgesetzten als potentieller Moderator berücksichtigt wurde. Ein möglicher Grund hierfür ist die geringe Anzahl an Primärstudien, die für diesen Effekt kontrollieren. Eine dieser wenigen Studien, die sich insbesondere mit dem Zusammenhang zwischen wahrgenommener Ähnlichkeit, Geschlechterdyaden und Leistungsbewertungen von Mitarbeitern beschäftigt, wurde 1982 von Mobley durchgeführt. Die Untersuchung kommt zu dem Ergebnis, dass Vorgesetzte Beschäftigte ihres eigenen Geschlechts nicht besser beurteilen, da der Einfluss spezifischer Geschlechterkombinationen auf die Leistungsbeurteilung statistisch nicht signifikant ist ($F=0,01$; $p>0,05$). Weitere Studien von Peters, O'Connor, Weekley, Pooyan, Frank und Erenkrantz aus dem Jahr 1984 ($F=3,20$; $p>0,05$), Pulakos et al. aus dem Jahr 1989 ($F=0,28$; $p>0,05$) und Lyness und Heilman aus 2006 ($\beta=-0,06$; $p>0,05$) bestätigen diese Erkenntnis.

Den Befunden der oben zitierten Forschungsarbeiten, die keinen statistisch signifikanten

Einfluss von Geschlechterdyaden auf Leistungsbewertungen von Beschäftigten aufzeigen, stehen die Resultate der zwei nachfolgend erörterten Untersuchungen von Varma und Stroh (2001) sowie Furnham und Stringfield (2001) entgegen. Die erstgenannte Studie analysiert Umfragedaten aus vier unterschiedlichen US-amerikanischen Elektronikunternehmen und kommt zu dem Ergebnis, dass Vorgesetzte gleichgeschlechtlichen Beschäftigten gegenüber eine höhere Zuneigung ausdrücken als Beschäftigten mit anderem Geschlecht. Dies wiederum führt zu positiv verzerrten Mitarbeiterbewertungen. Der Studie zufolge haben Geschlechterdyaden einen statistisch signifikanten Einfluss auf Leistungsbeurteilungen ($F=81,91$; $p<0,01$), wobei diese Beziehung durch die zuvor genannten Eigenschaften mediiert wird (Varma & Stroh, 2001).

Eine weitere Forschungsarbeit, die den Einfluss von Geschlechterdyaden auf Leistungsbeurteilungen von Beschäftigten näher beleuchtet, nutzt Daten einer neuseeländischen Fluggesellschaft (Furnham & Stringfield, 2001). Die Autoren werten hierfür Beurteilungsformulare von Beschäftigten aus dem Jahr 1997 aus und finden heraus, dass der Interaktionseffekt aus dem Geschlecht des Vorgesetzten und des Beschäftigten einen statistisch signifikanten Einfluss auf die Beurteilung der Mitarbeiter hat ($F=5,78$; $p<0,05$). Zudem merken die Autoren an, dass die Daten in ihrer Untersuchung darauf hindeuten, dass weibliche Vorgesetzte vergleichsweise strikter sind bei der Evaluation der Leistung von Männern: *„The significant interaction is perhaps the most interesting of the results. The results indicate that female, not male managers (raters) appear to differentiate in terms of the sex of their reports, being more strict.“* (Furnham & Stringfield, 2001: 286) Den Autoren zufolge kann dieser Befund auch damit zusammenhängen, dass Frauen bei der Leistungsbewertung von Männern andere Maßstäbe heranziehen als für Frauen.

In Bezug auf den Moderator Geschlechterdyaden lässt sich zusammenfassend sagen, dass sich durch die Skizzierung des Forschungsstands kein systematisches Bild des Einflusses von Geschlechterinteraktionen auf die Leistungsbeurteilung aufzeigen lässt. Die Befunde der zitierten individuellen Studien sind unterschiedlich, während vier der sechs Studien mit Geschlechterdyaden keinen statistisch signifikanten Einfluss aufzeigen (Lyness & Heilman, 2006; Mobley, 1982; Peters et al., 1984; Pulakos et al., 1989), liefern die beiden Untersuchungen von Varma und Stroh (2001) sowie Furnham und Stringfield (2001) Gründe dafür, dass Geschlechterdyaden als Moderator in Studien über geschlechterspezifische Differenzen in Leistungsbeurteilungen nicht unberücksichtigt bleiben sollten. Mög-

liche Erklärungsansätze für die unterschiedlichen Ergebnisse sind zum Beispiel die Branchen, in denen die Studien durchgeführt wurden. Dementsprechend ist in einer frauendominierten Branche eher zu erwarten, dass Frauen als Vorgesetzte ihre Mitarbeiterinnen vergleichsweise besser beurteilen.

Levy und Williams (2004), die sich in ihrem Übersichtsartikel auf die Studien mit Geschlechterdyaden von Varma und Stroh (2001) sowie Furnham und Stringfield (2001) beziehen, merken an: „*Clearly additional research is needed more fully to examine any potential rating bias due to gender.*“ (Levy & Williams, 2004: 893) Die in den Metaanalysen berücksichtigten Primärstudien beinhalten bislang kaum Geschlechterdyaden zwischen Beschäftigten und Vorgesetzten und setzen ihren Fokus, aufgrund fehlender Informationen, lediglich auf die Haupteffekte des Geschlechts. Auf Basis detaillierter Personaldaten (unter anderem zum Geschlecht des Beurteilenden sowie Beurteilten und deren Interaktion), trägt die vorliegende Dissertation dazu bei, diese Forschungslücke zu schließen.

Im nachfolgenden Kapitel soll der theoretische Rahmen für die Wirkungsweisen von Geschlechtereffekten in Leistungsbeurteilungen dargelegt werden. In diesem Zusammenhang soll auch auf organisationale Mechanismen eingegangen werden, die moderierend auf Geschlechtereffekte wirken. Zunächst werden interdisziplinäre theoretische Konzepte aus den Bereichen Ökonomie (Kapitel 3.1), Sozialpsychologie sowie Soziologie (Kapitel 3.2) erörtert, um das Entstehen und die Wirkungsweisen von Geschlechterdifferenzen bei Mitarbeiterbeurteilungen zu erklären. Auf Grundlage der theoretischen Argumente werden im nächsten Schritt überprüfbare Hypothesen hergeleitet.

3 Theoretische Grundlagen und Hypothesen

3.1 Ökonomische Ansätze

Zur Aufklärung von Ungleichbehandlung können unterschiedliche Disziplinen herangezogen werden. Im Vergleich zu theoretischen Ansätzen aus den Bereichen der Sozialpsychologie oder Soziologie gehen ökonomische Theorien nicht von möglichen Verhaltensmustern aus, die die unterschiedliche Behandlung von Individuen oder Gruppen erklären, sondern von einem ökonomischen Kalkül. Diesbezüglich werden in der Literatur zwei grundlegende theoretische Konzepte unterschieden: die präferenzbasierte Diskriminierung (*taste-based discrimination*) nach Becker (1971) sowie die statistische Diskriminierung (*statistical discrimination*) nach Arrow (1971) und Phelps (1972). Auf die erwähnten Theorien wird im Folgenden näher eingegangen.

Präferenzbasierte Diskriminierung

Von präferenzbasierter Diskriminierung ist nach Becker (1971) allgemein die Rede, wenn Arbeitgeber, Arbeitskollegen/innen oder Kunden⁸ Vorlieben bzw. Abneigungen gegenüber Individuen einer bestimmten Religionszugehörigkeit, eines ethnischen Hintergrunds, einer sozialen Klasse oder – wie im Fokus dieser Arbeit – eines bestimmten Geschlechts haben. Durch die Vermeidung von Interaktionen mit diesen Personengruppen entstehen für einen Arbeitgeber unter Umständen nicht-pekuniäre Kosten, die für ihn einen negativen Nutzen darstellen. Mit anderen Worten: Da Individuen Präferenzen gegenüber Personen mit bestimmten demographischen Merkmalen haben, nehmen sie bewusst Kosten in Form von entgangenen Gewinnen oder Einkommenseinbußen in Kauf, um nicht mit Personen zusammenarbeiten zu müssen, gegenüber denen sie abgeneigt sind. Diskriminierende Arbeitgeber orientieren sich demnach nicht an der Produktivität von Beschäftigten, sondern an ihrer Präferenz bzw. Abneigung gegenüber diesen. Eine Konsequenz hiervon ist, dass Frauen mit einem Verdienstabschlag rechnen müssen, da ihre Produktivität systematisch gering eingeschätzt wird und ein Arbeitgeber sich einen negativen Nutzen aus der Beschäftigung verspricht. Die Nachfrage nach Arbeitskräften, die zur nicht diskriminierten Gruppe gehören, steigt, sodass letztlich männliche Beschäftigte – die bevorzugte Gruppe – im Vergleich zu weiblichen Beschäftigten höhere Einkommen beziehen. Diesen Lohnauf-

⁸ Im Folgenden wird lediglich die Ungleichbehandlung durch Arbeitgeber näher betrachtet, da im Kontext von Leistungsbeurteilungen in der Regel die Vorgesetzten als Beauftragte des Arbeitgebers die Leistungen der Beschäftigten bewerten.

schlag für männliche Beschäftigte definiert Becker wie folgt: $\pi_i(1 + dc_{ie})$, wobei π_i den Lohnsatz darstellt, den ein Arbeitgeber i bezahlt und dc_{ie} den Diskriminierungskoeffizienten, eine Maßzahl für das Ausmaß an Diskriminierung dieses Arbeitgebers. Da nun die Produktionskosten für diskriminierende Arbeitgeber steigen, können sich Konkurrenten einen Kostenvorteil verschaffen, indem sie Frauen beschäftigen, die unter ihrer tatsächlichen Produktivität entlohnt werden. Die Folge ist, dass nicht-diskriminierende Unternehmen ihren Marktanteil kontinuierlich vergrößern und diskriminierende Unternehmen langfristig vom Markt verdrängen. Nach Becker (1971) verbleiben am Ende dieses Prozesses nur noch diejenigen Unternehmen am Markt, die in geringem Ausmaß bzw. gar nicht diskriminieren.

In Bezug auf den vorliegenden Kontext der Mitarbeiterbeurteilungen ist eine mögliche Diskriminierung durch den Arbeitgeber bzw. den direkten Vorgesetzten von zentraler Bedeutung. In einem Unternehmen mit vielschichtigen Führungsstrukturen hat ein Vorgesetzter verglichen mit einem Arbeitgeber üblicherweise keine Residualansprüche auf Gewinne aus der Unternehmung (Prendergast, 1999). Es ist prinzipiell denkbar, dass Arbeitgeber die Leistungszulagen, die sich aus Leistungsbeurteilungen ergeben, systematisch auf einem niedrigen Niveau halten, um die Lohnkosten zu senken und die Residualansprüche zu maximieren (Bol, 2011). Jedoch können auch direkte Vorgesetzte, die in der Regel für den Arbeitgeber die Beurteilungen der Beschäftigten durchführen, individuelle Interessen und Präferenzen haben, die einen Einfluss auf die Beurteilungen ausüben (Bol, 2011). „*Principals (i.e., managers) are predicted ... to take their own preferences into account when they appraise performance*“ (Bol, 2011: 1552). Auch Jirjahn und Poutsma (2013) sind der Ansicht: „*Superiors' prejudices and preferences toward subordinates may enter performance appraisal.*“ (Jirjahn & Poutsma, 2013: 809). Nach Becker (1971) können Vorgesetzte aufgrund von Vorurteilen Abneigungen gegenüber weiblichen Beschäftigten entwickeln und diese schlechter beurteilen als sonst vergleichbare männliche Beschäftigte. Mögliche Folgen hiervon sind nicht nur Einkommensdifferenzen, die sich durch verzerrte Personalbeurteilung ergeben, sondern auch die weiter oben angesprochene vertikale Segregation des Arbeitsmarktes. Wenn beispielsweise Frauen bei ähnlicher Leistung systematisch schlechtere Evaluationen bekommen als ihre Kollegen, lässt sich annehmen, dass Frauen in Unternehmen geringere Beförderungschancen aufweisen und überwiegend in den unteren Gehaltsgruppen beschäftigt werden.

Statistische Diskriminierung

Im Gegensatz zu den Ausführungen von Becker (1971) argumentiert Arrow (1971), dass Diskriminierung auch langfristig bestehen kann, wenn durch Informationsdefizite die tatsächliche Produktivität verschiedener demographischer Gruppen unterschiedlich eingeschätzt wird. In diesem Fall ist von statistischer Diskriminierung die Rede. Im theoretischen Modell der statistischen Diskriminierung, entwickelt von Arrow (1971) und Phelps (1972), wird angenommen, dass Unternehmen aufgrund unvollständiger Informationen über den Output der Beschäftigten eine Einschätzung über die Produktivität von Mitgliedern einer Gruppe auf Basis soziodemographischer Merkmale (zum Beispiel Geschlecht, Nationalität, Religionszugehörigkeit, soziale Herkunft, Alter) vornehmen. Diesbezüglich müssen nach Arrow (1971) drei Grundbedingungen erfüllt sein. Zum einen sollten Arbeitgeber in der Lage sein, zu vertretbaren Kosten zwischen zwei Gruppen von Beschäftigten zu unterscheiden, die sich durch beobachtbare Merkmale, wie zum Beispiel das Geschlecht oder die Nationalität, voneinander differenzieren. Zum anderen sollte die tatsächliche Produktivität eines potentiellen Beschäftigten ex ante unbekannt und die Feststellung der Produktivität mit hohen Investitionskosten verbunden sein. Eine weitere Voraussetzung ist, dass der Arbeitgeber subjektiv einschätzen kann, wie sich die zu erwartende durchschnittliche Arbeitsproduktivität für Mitglieder einer bestimmten Gruppe unterscheidet. Die Einschätzung über den erwarteten Output hängt wiederum ab von den Erfahrungen des Arbeitgebers mit statistischen Daten sowie soziologisch akzeptierten Einstellungen zu diesen Gruppen.

Wenn beispielsweise Arbeitgeber davon ausgehen, dass Frauen als Gruppe eine geringere Produktivität als Männer aufweisen, da sie in der Regel eine geringere Beschäftigungsstabilität haben, kann statistische Diskriminierung die Folge sein. Da sich Mitarbeiterbeurteilungen jedoch auf die Vergangenheit beziehen (üblicherweise auf einen Zeitraum von einem Jahr), ist davon auszugehen, dass die Produktivität von Beschäftigten für die Beurteilung der Leistung zumindest besser eingeschätzt werden kann als für Entscheidungen, die sich auf die Zukunft beziehen (zum Beispiel Einstellungen oder Beförderungen). Vorgesetzte haben in der Regel in einem definierten Beobachtungszeitraum die Möglichkeit, den Output der zu beurteilenden Person zu beobachten und einzuschätzen und bauen daher Informationsasymmetrien ab, die zwischen Vorgesetzten und Beschäftigten bestehen. Informationsasymmetrien bleiben jedoch weiterhin bestehen, da Vorgesetzte die Performance ihrer Beschäftigten nicht kontinuierlich beobachten können und somit den Output

für spezifische Leistungsdimensionen (zum Beispiel Arbeitseinsatz, Zusammenarbeit, Beweglichkeit) einschätzen müssen. Es ist daher theoretisch zu vermuten, dass die Produktivität von Frauen im Vergleich zu Männern unterschätzt wird und eine statistische Diskriminierung im Hinblick auf das Merkmal Geschlecht vorliegen kann.

Gegenüberstellung der ökonomischen Ansätze

Die ökonomische Theorie der präferenzbasierten Diskriminierung impliziert ein Vorhandensein von Vorlieben eines Vorgesetzten gegenüber Beschäftigten eines bestimmten Geschlechts. Im Rahmen der statistischen Diskriminierung wird hingegen postuliert, dass Vorgesetzte unvollständige Informationen über die Produktivität ihrer Beschäftigten verfügen und sich daher an repräsentativen soziodemographischen Merkmalen wie dem Geschlecht von Beschäftigten orientieren, um zum Beispiel die Produktivität von Arbeitnehmerinnen (besser) einschätzen zu können. Da Frauen als Gruppe eine geringere Beschäftigungstabilität als Männer aufweisen, wird angenommen, dass Vorgesetzte die Produktivität von Frauen systematisch geringer einschätzen als die von Männern – die Folge ist eine Vergütung, die unterhalb der Produktivität liegt.

Sowohl die ökonomische Theorie der präferenzbasierten als auch statistischen Diskriminierung sagen voraus, dass Frauen schlechtere Leistungsbeurteilungen erhalten, wenn sie von männlichen Führungskräften beurteilt werden. Diese theoretischen Voraussagen sind konsistent mit den nachfolgend erörterten sozialpsychologischen und soziologischen Ansätzen und deuten darauf hin, dass Ungleichbehandlungen auf Basis des Geschlechts insbesondere bei Marktunvollkommenheiten auftreten.

3.2 Sozialpsychologische und soziologische Ansätze

Die präferenzbasierte und die statistische Diskriminierung (Kapitel 3.1) erklären nicht, unter welchen Bedingungen Abneigungen bzw. Bevorzungen gegenüber Individuen eines bestimmten Geschlechts entstehen. Zudem bleibt ungeklärt, inwieweit das Auftreten von Ungleichbehandlung durch unterschiedliche dyadische Beziehungen zwischen Beschäftigten sowie Vorgesetzten beeinflusst wird. Es bleibt zum Beispiel offen, ob weibliche Vorgesetzte auch entsprechende Präferenzen für Frauen und Abneigungen gegenüber männlichen Beschäftigten haben können.

Ein Großteil der nachfolgenden theoretischen Ansätze basiert auf der gemeinsamen Annahme, dass Individuen andere Individuen aufgrund soziodemographischer Merkmale

kategorisieren. Diese Kategorisierungen wiederum helfen Individuen sich in komplexen Sozialwelten zurechtzufinden, d.h. andere Menschen zu verstehen, deren Verhalten vorauszusagen und abzuschätzen, wie man auf sie reagiert (Ashforth & Mael, 1989; Stangor & Lange, 1994; Tajfel, 1981; Taylor, Fiske, Etcoff & Ruderman, 1978). Merkmale, die offensichtlich sind, wie zum Beispiel das Geschlecht einer zu beurteilenden Person, führen zu automatischen Kategorisierungen (Fiske & Neuberg, 1994; Ito & Urland, 2003; Norton, Vandello & Darley, 2004). Kategorisierungsprozesse, die auf geschlechtsspezifischen Merkmalen basieren, schaffen Raum für Ungleichbehandlung. Im Folgenden werden – analog zu den ökonomischen Theorien in Kapitel 3.1 – aus sozialpsychologischer als auch soziologischer Perspektive mögliche Gründe für Geschlechtereffekte bei Leistungsbeurteilungen näher beleuchtet.

Stereotype

Kategorisierungen, die auf Grundlage offensichtlicher Merkmale, wie des Geschlechts, stattfinden, können Geschlechterstereotype hervorrufen (Stangor & Lange, 1994). Im Allgemeinen werden Stereotype definiert als „... *beliefs about the characteristics, attributes, and behaviors of members of certain groups*“ (Hilton & Hippel, 1996: 240). Untersuchungen zu allgemeinen Geschlechterstereotypen bestätigen, dass Frauen zum Beispiel als sanftmütig, nett und gefühlsvoll, wohingegen Männer als dominant, kompetent sowie kompetitiv angesehen werden (Carli & Eagly, 1999; Deaux & Kite, 1993; Williams & Best, 1990). Neben diesen allgemeinen Stereotypen können Personen auch Stereotype gegenüber spezifischen Untergruppen (zum Beispiel Hausfrauen versus Karrierefrauen) oder gegenüber Menschen, die eine Kombination mehrerer soziodemographischer Merkmale aufweisen, entwickeln (Deaux, 1995; Fiske, 1998; Stangor, Lynch, Duan & Glas, 1992). Stereotype führen in der Regel dazu, dass Personen andere Personen in Abhängigkeit ihrer sozialen Gruppenzugehörigkeit entweder als Mitglieder ihrer In- oder Outgroup betrachten. Gemäß diesem Ansatz werden beispielsweise Mitarbeiterinnen von gleichgeschlechtlichen Vorgesetzten als Ingroup-Mitglieder betrachtet, wohingegen männliche Beschäftigte zur Outgroup gehören. Stereotypisierungsprozesse führen dazu, dass den Mitgliedern der Ingroup positive Attribute zugeschrieben werden, wie beispielsweise eine hohe Kompetenz oder Einsatzbereitschaft. Im Gegensatz dazu werden die Mitglieder der Outgroup in den erwähnten Kompetenzdimensionen vergleichsweise schlechter evaluiert. Stereotype können generell vielfältig sein, zum Beispiel werden Karrierefrauen als fachlich kompetent, jedoch nicht als sozial kompetent angesehen (Fiske, Cuddy, Glick & Xu,

2002; Rudman & Glick, 1999). Unabhängig davon, welche Stereotype über Individuen vorherrschen, verarbeiten Vorgesetzte im Rahmen von Personalentscheidungen Informationen über Qualifikation, Leistung oder das zukünftige Potenzial, die konsistent mit einem spezifischen Stereotyp sein können (Fiske, 1998; Ilgen & Youtz, 1986; Stangor & Lange, 1994). Untersuchungen zeigen, dass Vorgesetzte gezielt nach stereotypkonformen Informationen suchen und sich eher mit diesen befassen (Biernat & Ma, 2005; Fiske, 1998). Daraus folgt im Umkehrschluss, dass Informationen, die sich nicht mit einem spezifischen Stereotyp decken, weniger Beachtung geschenkt wird.

Stereotype sind auch im Rahmen von Mitarbeiterbeurteilungen bedeutend, wenn beispielsweise verzerrte Beurteilungen das Resultat vorherrschender Geschlechterstereotype sind. Dieser Argumentation zufolge ist es denkbar, dass Frauen aufgrund von Geschlechterstereotypen bei gleicher Leistung und Qualifikation wie ihre Kollegen *ceteris paribus* schlechtere Beurteilungen erhalten.

Soziale Identität und Selbstkategorisierungs-Theorie

Weitere theoretische Erklärungsansätze, die auf der Argumentation von In- und Outgroups aufbauen und Geschlechterunterschiede in Leistungsbeurteilungen beschreiben können, sind die Theorie der Sozialen Identität (Ashforth & Mael, 1989; Capozza & Brown, 2000; Tajfel & Turner, 1986) und die Selbstkategorisierungs-Theorie nach Turner (1985). Beide Ansätze postulieren, dass Individuen in Abhängigkeit von ihrer sozialen Schicht bzw. Kategorie soziale Identitäten entwickeln. Sie versuchen ein positives Selbstbild zu erhalten, indem sie sich selbst und andere Individuen in soziale Kategorien einordnen und positiv verzerrte Vergleiche zwischen Mitgliedern der In- und Outgroup durchführen. Demzufolge identifizieren sich Menschen grundsätzlich eher mit Personen, die dieselbe Gruppenzugehörigkeit haben, wie zum Beispiel dasselbe Geschlecht (Brewer & Miller, 1984; Williams & O'Reilly, 1997). Es kann sogar angenommen werden, dass positiv verzerrte Vergleiche stattfinden, selbst wenn keine Interaktionen zwischen Gruppenmitgliedern stattfinden (Tsui, Egan & O'Reilly, 1992).

Individuen bevorzugen eher gleichgeschlechtliche Personen, wenn das relevante Kategorisierungsmerkmal, in diesem Fall das Geschlecht, eine bedeutende Komponente des Selbstbilds darstellt (Powell & Butterfield, 1994). Die Bedeutung dieser Komponente kann verstärkt werden durch persönliche Faktoren, wie zum Beispiel Erfahrungen mit Geschlechterungleichbehandlung und situative Faktoren, wie zum Beispiel einem geringen

Frauenanteil in einem Unternehmen oder in einer Abteilung (Wharton, 1992). Darüber hinaus ist der soziale Status, der mit dem Geschlecht eines Individuums in Verbindung steht, ebenfalls relevant bei der Bestimmung des Selbstbilds. Beispielsweise können Frauen, die in der Regel Sozialgruppen mit niedrigem Status angehören, ein positives Selbstbild erhalten, wenn sie sich von ihrer Ingroup distanzieren und sich gleichzeitig mit Mitgliedern der Outgroup (Männern) identifizieren, die einen höheren sozialen Status haben (Ely, 1994; Graves & Powell, 1995; Tajfel & Turner, 1986).

Die obigen Ausführungen zur Theorie der Sozialen Identität und Selbstkategorisierungstheorie lassen darauf schließen, dass im Kontext von Mitarbeiterbeurteilungen Vorgesetzte gleichgeschlechtliche Beschäftigte bevorzugt behandeln bzw. bei gleicher Leistung und gleicher Qualifikation Beschäftigte des anderen Geschlechts schlechter evaluieren. Diese Ungleichbehandlung ist selbst dann anzunehmen, wenn Vorgesetzte und Beschäftigte kaum bzw. nicht miteinander interagieren. Dies ist im Kontext von Mitarbeiterbeurteilungen eher unwahrscheinlich. Darüber hinaus lässt sich festhalten, dass Vorgesetzte mit einem geringen Sozialstatus andersgeschlechtliche Beschäftigte im Rahmen von Mitarbeiterbeurteilungen favorisieren können, um ihr positives Selbstbild aufrechtzuerhalten.

Similarity-Attraction-Theorie

Die Similarity-Attraction-Theorie (Byrne, 1971; Byrne & Neuman, 1992) besagt, dass Menschen sich zu anderen Menschen hingezogen fühlen, je öfter sie mit diesen interagieren und sie bevorzugt behandeln, wenn sie ihnen, zum Beispiel im Hinblick auf soziodemographische Merkmale, ähnlich erscheinen. Das Geschlecht als soziodemographisches Merkmal stellt dabei einen Faktor dar, der die Wahrnehmung der Ähnlichkeit beeinflusst (Graves & Powell, 1995). Der Theorie zufolge ist davon auszugehen, dass Gleichgeschlechtlichkeit zwischen Vorgesetzten und Beschäftigten zu einer Zunahme der wahrgenommenen Ähnlichkeit führt, zum Beispiel in Bezug auf Einstellungen und Werten, und sich interpersonelle Vorlieben entwickeln. Diese haben wiederum zur Folge, dass die Interaktionen zwischen gleichgeschlechtlichen Vorgesetzten und Beschäftigten zunehmen und letztlich Personalentscheidungen positiv verzerrt werden (Graves & Powell, 1995, 1996; Powell & Graves, 2003). Dementsprechend könnten auch im Rahmen von Personalbeurteilungen Vorgesetzte Beschäftigte bevorzugt behandeln, die ihnen hinsichtlich sichtbarer demographischer Merkmale – wie dem Geschlecht – ähnlich sind.

Leader-Member-Exchange-Theorie

Der Leader-Member-Exchange-Ansatz (Graen & Scandura, 1987; Graen, 1976; Graen & Cashman, 1975), kurz LMX, baut auf den theoretischen Überlegungen zur Similarity-Attraction-Theorie auf und geht ebenfalls davon aus, dass Vorgesetzte ihre Beschäftigten nicht gleich behandeln und somit bei Personalbeurteilungen nicht in gleicher Weise evaluieren. Da Führungskräfte nur begrenzte Ressourcen haben und ihr Handeln zeitlichen Restriktionen unterliegt, kategorisieren sie ihre Beschäftigten, wie bei den theoretischen Ausführungen oben, in Mitglieder ihrer In- und Outgroup. Die Mitglieder der Ingroup erhalten von ihrem Vorgesetzten im Vergleich zu den Beschäftigten der Outgroup mehr Zeit, Zuneigung und Vertrauen. Die Kategorisierung der Mitglieder hat gemäß der LMX-Theorie letztlich zur Folge, dass die Mitglieder der Ingroup bessere Ergebnisse in Leistungsbeurteilungen erhalten, zufriedener sind mit ihrem Verhältnis zum Vorgesetzten und geringere Fluktuationsraten aufweisen (Varma & Stroh, 2001). Verschiedene Untersuchungen, die den LMX-Ansatz empirisch getestet haben, zeigen, dass Führungskräfte ihre Beschäftigten anhand demographischer Merkmale, die nicht mit der Mitarbeiterleistung in Verbindung stehen, in In- und Outgroup-Mitglieder kategorisieren (Dansereau, Graen & Haga, 1975; Dienesch & Liden, 1986; Graen, Liden & Hoel, 1982). Der Qualität der Beziehung zwischen Vorgesetzten und Beschäftigten kommt eine besondere Bedeutung zu, da sie auch einen Einfluss auf die Leistungsbeurteilung ausübt (Vecchio & Gobdel, 1984).

Gegenstand diverser Forschungsarbeiten bildet die Identifikation von Faktoren, die die Qualität von LMX-Beziehungen determinieren (Varma & Stroh, 2001). Beispielsweise kommen Liden und Mitchell (1988) zu dem Schluss, dass neben individuellen demographischen Merkmalen von Beschäftigten (zum Beispiel Geschlecht, Herkunft) und dem Arbeitsverhalten auch die Präferenz der Vorgesetzten für einzelne Beschäftigte von Bedeutung ist, um die Qualität einer LMX-Beziehung zu bestimmen. Die Präferenz für Beschäftigte wiederum kann nach Varma und Stroh (2001) maßgeblich durch das Geschlecht erklärt werden, da das Geschlecht eine Kategorisierung nach In- und Outgroup-Mitgliedern ermöglicht. Forschungsarbeiten von Turban und Jones (1988) sowie Wayne, Shore und Liden (1997) deuten beispielsweise darauf hin, dass Gemeinsamkeiten hinsichtlich soziodemographischer Merkmale das Ähnlichkeitsempfinden und somit die Präferenzen eines Vorgesetzten für einen Beschäftigten beeinflussen. Dies wiederum hat laut Varma und Stroh (2001) Auswirkungen auf die Qualität der Beziehung, die sich zwischen Vorgesetzten und Beschäftigten entwickelt.

Gegenüberstellung der sozialpsychologischen und soziologischen Ansätze

Es lässt sich festhalten, dass sozialpsychologische als auch soziologische Erklärungsansätze auf Basis von In- und Outgroup-Mechanismen argumentieren, um Geschlechterdifferenzen zu erklären. Im Zusammenhang mit der Bildung von Stereotypen kann beispielsweise angenommen werden, dass männliche Vorgesetzte die Fähigkeiten von Outgroup-Mitgliedern – in diesem Fall die Fähigkeiten der weiblichen Beschäftigten – als gering einstufen und folglich anders beurteilen. Die Theorie der Sozialen Identität und die Similarity-Attraction-Theorie gehen ebenfalls von einer Favorisierung der Ingroup-Mitglieder aus und unterscheiden sich in ihren Erklärungsmustern nur geringfügig. In der Regel kann davon ausgegangen werden, dass die soziale Gruppe aus Individuen besteht, die das Geschlecht des Vorgesetzten haben. Auch im Fall der LMX-Theorie, die auf den Überlegungen zur Similarity-Attraction-Theorie aufbaut, wird auf Basis des beschriebenen Zweikategorien-Schemas argumentiert. Demnach bauen Vorgesetzte ein besseres Verhältnis zu Beschäftigten auf, die zum Beispiel dasselbe Geschlecht haben und behandeln diese bevorzugt, wenn sie Personalentscheidungen treffen.

Beide Perspektiven, sowohl die ökonomische als auch die sozialpsychologische sowie soziologische, zeigen Nachteile für Frauen im Berufsleben. Diese Benachteiligungen entstehen insbesondere dann, wenn Unvollkommenheiten am Arbeitsmarkt vorhanden sind und wenn Führungspositionen überwiegend durch Männer besetzt werden. Eine Folge hiervon wäre, dass Frauen systematisch schlechtere Leistungsbeurteilungen erhalten als vergleichbare Kollegen. Diese theoretische Annahme kann jedoch anhand zahlreicher empirischer Studien nicht bestätigt werden (siehe Kapitel 2.3.1). Allerdings kann die Tendenz aufgezeigt werden, dass bestimmte Faktoren dazu führen, dass Geschlechtereffekte einen systematischen Einfluss auf die Leistungsbeurteilung von Beschäftigten ausüben können (siehe Kapitel 2.3.2). Im anschließenden Kapitel 3.3 werden auf der Grundlage theoretischer Überlegungen zu den Wirkungsweisen von organisationalen Mechanismen auf Geschlechtereffekte in Leistungsbeurteilungen testbare Hypothesen hergeleitet.

3.3 Hypothesen

Reskin (2008) argumentiert, dass die Untersuchung der Wirkungsweisen organisationaler Mechanismen bedeutsam ist, um ein besseres Verständnis darüber zu erhalten, wie zum Beispiel vorhandene Stereotype zu geschlechterspezifischen Unterschieden in Leistungsbeurteilungen führen. Demzufolge können Geschlechtereffekte in Bewertungen kontext-

spezifisch sein und durch organisationale Faktoren beeinflusst werden.

Der Grad der Formalisierung eines Leistungsbeurteilungssystems bildet einen moderierenden Faktor, der auf Geschlechtereffekte in Leistungsbeurteilungen wirken kann. In dieser Dissertation wird ein Leistungsbeurteilungssystem als formalisiert verstanden, wenn Vorgesetzten nur ein geringer oder kein Spielraum geboten wird, um Einfluss auf entgeltrelevante Leistungsbeurteilungen zu nehmen. Ist in diesem Zusammenhang gar kein Bewertungsspielraum im Rahmen von Beurteilungen gegeben, können auch Stereotype oder Präferenzen gegenüber Beschäftigten keine Auswirkungen auf deren Evaluationen und Gehälter ausüben. In dieser Form von Beurteilungssystemen ist erwartungsgemäß auch eine statistische Diskriminierung von Beschäftigten auszuschließen. Im Gegensatz dazu ist bei Beurteilungssystemen, die einen geringen Bewertungsspielraum bieten, zum Beispiel in Bezug auf das Verfahren der Leistungsbeurteilung oder den Bewertungsdimensionen, zu erwarten, dass Vorgesetzte sich mit individuellen Informationen über die Beschäftigten befassen. Diese individuellen Informationen wiederum können einerseits die Wahrnehmung von Stereotypen über Frauen beeinflussen (Fiske, 1998; Hilton & Hippel, 1996) oder andererseits Stereotype beseitigen, die beispielsweise im Rahmen von Leistungsbeurteilungen zu einer statistischen Diskriminierung von Frauen geführt haben.

Es lässt sich festhalten, dass der Einsatz formalisierter Gehaltssysteme, und somit auch entgeltrelevanter Leistungsbeurteilungen, zu einer geringeren Geschlechterungleichheit im Gehalt führen kann (Reskin, 2000). Dieser Wirkungszusammenhang wurde in einigen Studien über die Formalisierung von Gehaltssystemen und dessen Auswirkungen auf geschlechterspezifische Effekte in Frage gestellt (Abraham, 2017), da der Einsatz weniger formalisierter Systeme auch dazu führen kann, dass weibliche Vorgesetzte durch den gegebenen Bewertungsspielraum Mitarbeiterinnen, insbesondere die in den unteren Gehaltsgruppen beschäftigt sind, vergleichsweise besser evaluieren, um einer vorherrschenden Geschlechterungleichheit in Unternehmen entgegenzuwirken (siehe Argumentation in Kapitel 2.4.1). Eine Gegenüberstellung der Studien über unterschiedliche Formalisierungsgrade in Gehaltssystemen zeigt jedoch die Tendenz auf, dass eine Formalisierung zu einer Neutralisierung von Geschlechtereffekten führt. Hinweise dafür bieten zahlreiche empirische Studien, die auf eine positive Beziehung zwischen Formalisierung und einer geschlechtergerechten Vergütung hindeuten (Baron et al., 1986; Elvira & Graham, 2002; Pfeffer & Cohen, 1984; Reskin & McBrier, 2000; Sutton et al., 1994). Beispielweise un-

tersuchen Elvira und Graham (2002) neben Geschlechterunterschieden in Leistungsbeurteilungen zwei weitere Gehaltskomponenten, die sich in Bezug auf den Formalisierungsgrad unterscheiden. In ihrer Studie bilden formalisiertere Gehaltssysteme Gehaltserhöhungen und Grundgehälter, Leistungszulagen hingegen werden in dem untersuchten Fallunternehmen auf Basis weniger formalisierter Bestimmungen vergeben. Die Autoren finden heraus, dass die Geschlechterdifferenzen für Frauen und Männer am geringsten sind, wenn die Beschäftigten im Rahmen formalisierter Gehaltssysteme vergütet werden. Baron et al. (2007) zeigen mit Daten von Start-ups auf, dass neu gegründeten Unternehmen mit formalisierten Beschäftigungspraktiken eher weibliche Fachkräfte einstellen als Start-ups, die auf formalisierte Methoden verzichten. Die Beschäftigung eines in Vollzeit für Personalbelange Verantwortlichen wird von den Autoren als Signal für Formalisierung in einem Start-up definiert. Der Formalisierungsgrad von Beschäftigungspraktiken wird ebenfalls in der Studie von Reskin und McBrier (2000), der Daten von 506 Unternehmen zugrunde liegen, fokussiert. In dieser Studie bestätigt sich, dass die Anzahl formalisierter Personalpraktiken (das sind zum Beispiel schriftlich festgehaltene Regelungen zu Stellenprofilen, Arbeitsverträgen oder Einstellungen und Entlassungen) mit einer Reduzierung des Männeranteils auf Führungspositionen einhergeht.

Im Gegensatz zu den oben genannten Studien existieren auch solche, die den positiven Zusammenhang zwischen Formalisierung und Geschlechtergleichheit nicht belegen können (Castilla, 2008; Dencker, 2008; Huffman & Velasco, 1997; Konrad & Linnehan, 1995). Wie auch Elvira und Graham (2002) kommt Abraham (2017) jedoch zu der Folgerung, dass eine Formalisierung in der Regel zu Geschlechtergleichheit führt: „ ... *the preponderance of empirical evidence supports the argument that more formalized pay systems serve to reduce gender pay inequality ...* “ (Abraham, 2017: 31). Da auch Leistungsbeurteilungen sowie darauf basierende Zulagen eine entgeltrelevante Vergütungspraktik bilden und zu einer Geschlechterungleichheit am Arbeitsplatz führen können, lautet die erste Hypothese dieser Forschungsarbeit:

Hypothese 1: Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem formalisierten Bewertungssystem kleiner als in einem weniger formalisierten Bewertungssystem.

Abhängig vom Formalisierungsgrad des Leistungsbeurteilungssystems kann auch die organisationale Hierarchiestufe die Beziehung zwischen Geschlecht und Leistungsbeurtei-

lung beeinflussen. Experimentelle Studien über Geschlechterrollen zeigen, dass weibliche Führungskräfte in männerdominierten Bereichen vergleichsweise schlechter evaluiert werden als ihre Kollegen, da sie nicht solche Verhaltenseigenschaften aufweisen, die mit Führungspositionen assoziiert werden (Eagly & Karau, 2002). Doch nicht nur in Führungspositionen können Mitarbeiterinnen in von Männern dominierten Bereichen Benachteiligungen erfahren. In diesem Zusammenhang argumentieren Joshi et al. (2015), dass einerseits Führungskräfte die Leistung ihrer weiblichen Beschäftigten vergleichsweise schlechter evaluieren, da ihre Kenntnisse und Fähigkeiten nicht denen entsprechen, die für die Ausübung einer mit männlichen Attributen assoziierten Tätigkeit erforderlich sind. Andererseits können nicht stereotypkonforme Arbeitsweisen zu einer Änderung des Status quo in Unternehmen führen. Eigenschaften, die für den Erfolg auf hohen Hierarchiepositionen erforderlich sind, werden in der Regel eher mit männlichen Attributen in Verbindung gebracht (Eagly, 1987; Heilman, Block & Martell, 1995; Rudman & Glick, 1999). Im Gegensatz dazu werden Berufe mit geringem Status tendenziell mit weiblichen Attributen und einer niedrigen Position in der organisationalen Hierarchie assoziiert (Baron & Newman, 1990; Glick, 1991). Um den beschriebenen Status quo beizubehalten, können männliche Führungskräfte dazu neigen, die Leistungen ihrer Mitarbeiterinnen systematisch schlechter zu beurteilen, um zum Beispiel Frauen den Aufstieg in der organisationalen Hierarchie zu verhindern (Fiske et al., 1991; Rudman, Moss-Racusin, Phelan & Nauts, 2012). Dass Leistungsbeurteilungen auch beförderungsrelevant sein können, wurde im Literaturteil der vorliegenden Dissertation aufgezeigt. Der oben dargelegten Argumentation zufolge ist daher nicht auszuschließen, dass Frauen insbesondere in männerdominierten Bereichen und in Positionen, die mit einem vergleichsweise niedrigen Status einhergehen, schlechter evaluiert werden als ihre Kollegen.

Abraham (2017) argumentiert, dass benachteiligte Mitarbeitergruppen insbesondere in unteren Gehaltsgruppen gefördert werden, wenn Führungskräften durch Bewertungsspielräume die Möglichkeit hierzu gegeben wird. Demzufolge ist anzunehmen, dass in weniger formalisierten Systemen mit steigender Hierarchiestufe die Geschlechterungleichheit höher ist als in formalisierten Bewertungssystemen. Es kann jedoch auch angenommen werden, dass der Eintritt des umgekehrten Effekts wahrscheinlicher ist, insbesondere wenn individuelle Informationen über Beschäftigte die Wahrnehmung von Vorgesetzten und somit auch die Leistungsbeurteilung beeinflussen. Eine Frau in einer Führungsposition, d.h. hohen Hierarchiestufe, könnte dementsprechend besser beurteilt werden als eine Frau

auf einer niedrigeren Gehaltstufe, da sie die Karriereleiter bereits erfolgreich aufgestiegen ist. Demnach haben Frauen in Führungspositionen vermutlich nicht nur eine Signalwirkung auf andere Frauen im Unternehmen (Joshi et al., 2015), sondern auch auf männliche Vorgesetzte, die Beurteilungen durchführen. Das Signal einer Statuserrungenschaft kann als individuelle Information dazu beitragen, dass Geschlechterstereotype abgeschwächt werden und eine statistische Diskriminierung weniger wahrscheinlicher ist. Die dargelegten Argumente führen zu der folgenden Hypothese:

Hypothese 2: Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem weniger formalisierten Bewertungssystem auf höheren organisationalen Hierarchiestufen kleiner als auf niedrigeren.

Die Argumentation zu Hypothese 2 verdeutlicht, dass in Bezug auf den Formalisierungsgrad eines Beurteilungssystems auch das Geschlecht der Führungskraft zu berücksichtigen ist, um die Wirkungsweisen organisationaler Mechanismen auf Geschlechterunterschiede in Leistungsbeurteilungen näher zu untersuchen. Sowohl die Theorie der Sozialen Identität als auch die LMX-Theorie bieten Erklärungsansätze dafür, dass männliche als auch weibliche Führungskräfte die Leistungen der Beschäftigten anders beurteilen. Den erwähnten Theorien zufolge kategorisieren Führungskräfte, bewusst oder unbewusst, Beschäftigte auf Basis ihres Geschlechts in Mitglieder ihrer Ingroup, die positiver beurteilt werden und in Mitglieder ihrer Outgroup, die negativer evaluiert werden. Das Auftreten eines solchen Kategorisierungsprozesses ist wahrscheinlich, da das Geschlecht ein offensichtliches und für Führungskräfte beobachtbares Merkmal der Beschäftigten bildet (Ridgeway, 1991) und Individuen dazu neigen, auf Basis von Kategorisierungsprozessen Personen desselben Geschlechts bevorzugt zu behandeln (Elliott & Smith, 2004; Gorman, 2005). Hinzu kommt, dass insbesondere Führungskräfte durch Kategorisierungsprozesse beeinflusst werden (Goodwin, Operario & Fiske, 1998), da sie ein Interesse daran haben könnten, den Status quo im Unternehmen beizubehalten bzw. zu ändern (Jost & Banaji, 1994). Folglich ist aufgrund der dargelegten Argumente nicht auszuschließen, dass beispielsweise im Rahmen eines weniger formalisierten Beurteilungssystems, das einen vergleichsweise hohen Bewertungsspielraum bietet, weibliche Führungskräfte Frauen besser bewerten und männliche Vorgesetzte Männer positiver beurteilen. Abgesehen vom Formalisierungsgrad des Beurteilungssystems zeigen Studien über die Wirkung von Geschlechterdyaden, die auf Geschlechterkategorisierungen basieren, dass weibliche Vorgesetzte dazu tendieren, ihre

Mitarbeiterinnen besser zu beurteilen als ihre männlichen Beschäftigten (Furnham & Stringfield, 2001; Varma & Stroh, 2001).

Um den angesprochenen Status quo entgegenzuwirken, könnten Frauen ihre Stellung als Führungskraft im Unternehmen dazu nutzen, um andere Frauen bevorzugt zu behandeln und ihnen Karrierechancen durch bessere Leistungsbeurteilungen zu ermöglichen oder um ihnen höhere Leistungszulagen auszuzahlen. Diese Annahme ist umso wahrscheinlicher, wenn den weiblichen Führungskräften durch ein weniger formalisiertes Beurteilungssystem Spielraum geboten wird und ein Unternehmen aus einer männerdominierten Industrie stammt. Zum Beispiel lässt sich der Geschlechterunterschied in der Vergütung von Männern und Frauen dadurch erklären, dass Männer in Führungspositionen überrepräsentiert sind und den Status quo nutzen, um gezielt Männer bei Gehaltsentscheidungen zu bevorzugen und Frauen zu benachteiligen (Reskin, 2000). Aufgrund der dargelegten Argumente ist jedoch auch anzunehmen, dass auch weibliche Führungskräfte insbesondere in einem männerdominierten Umfeld gezielt Frauen bevorzugen, wenn ihnen die Möglichkeit hierzu gegeben wird. Daher lautet die Hypothese in Bezug auf die Wirkungsweise von Geschlechterdyaden auf die Leistungsbeurteilung von Beschäftigten:

Hypothese 3: Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem weniger formalisierten Bewertungssystem vergleichsweise kleiner, wenn sie von weiblichen und nicht von männlichen Vorgesetzten beurteilt werden.

Nachdem die theoretischen Grundlagen erörtert und die für den Untersuchungskontext relevanten Hypothesen hergeleitet wurden, soll in Kapitel 4 auf den Kontext dieser Forschungsarbeit eingegangen werden. In diesem Zusammenhang werden zunächst die historische Entwicklung und die Bedeutung tariflicher Leistungsbeurteilungen erörtert. Anschließend werden spezifische Merkmale der M+E-Industrie in Nordrhein-Westfalen (NRW) aufgezeigt, die zur Erklärung von Geschlechtereffekten in Leistungsbeurteilungen relevant sind und das Fallstudienunternehmen sowie sein Beurteilungssystem vorgestellt.

4 Untersuchungskontext

4.1 Tarifliches Leistungsbeurteilungssystem

Die Forschungshypothesen, die in Kapitel 3.3 hergeleitet wurden, sollen anhand von Daten eines Fallstudienunternehmens getestet werden. Da dieses Unternehmen Mitglied im Arbeitgeberverband Gesamtmetall ist, tritt es mit der Industriegewerkschaft Metall (IG Metall) in Verhandlungen und unterliegt den Bestimmungen des ERA-TV für die M+E-Industrie in NRW, der Ende 2003 verabschiedet wurde. Die Beschäftigten erhalten im betrachteten Unternehmen ein Zeitentgelt, d.h., dass den Beschäftigten für ihre Zeit am Arbeitsplatz ein Grundeinkommen gezahlt wird. Die tariflichen Bestimmungen legen fest, dass Beschäftigte im Zeitentgelt je nach Output eine Leistungszulage erhalten (siehe § 10 ERA-TV). Die Höhe dieser gezahlten Leistungszulage bemisst sich nach dem Ergebnis der Beurteilung der Leistung. Während beispielsweise bei einer Zielvereinbarung messbare Bezugsgrößen herangezogen werden (zum Beispiel die über einen Zeitraum anzufertigende Stückzahl an Elektronikkomponenten), wird bei der individuellen Leistungsbeurteilung auf Merkmale zurückgegriffen, deren Erfüllung vom Vorgesetzten subjektiv beurteilt wird.

Um besser zu verstehen, wie das tarifliche Beurteilungssystem im Allgemeinen und insbesondere die tariflichen Bewertungskriterien entstanden sind, soll nachfolgend auf die historische Entwicklung eingegangen werden. Diese lässt sich in eine Vor-ERA-Periode und eine ERA-Periode einteilen.

4.1.1 Leistungszulagen in der Vor-ERA-Zeit

Im Jahr 1970 wurde im Rahmen von Tarifverhandlungen zwischen den beteiligten Tarifparteien vereinbart, dass ein zweiter Teil des damals gültigen Lohnrahmenabkommens (LRA) mit Bestimmungen über eine Leistungszulage für Zeitlohnarbeiter eingeführt wird. Da für die tarifvertraglichen Verhandlungen nur eine begrenzte Zeit zur Verfügung stand, war es nicht möglich, ein Verfahren zur Leistungsbeurteilung von Zeitlohnarbeitern zu bestimmen. Die Tarifparteien verpflichteten sich jedoch zu Verhandlungen über ein solches Verfahren und zu einer Einigung bis Ende 1970. Im September 1970 – also kurz vor Ablauf der genannten Frist – führten die Verhandlungen letztlich doch zu einem Ergebnis: dem Tarifvertrag zur Leistungsbeurteilung von Zeitlohnarbeitern (TV LB Zeitl). Dieser Tarifvertrag war eine Ergänzung zum damals geltenden LRA und konnte nur in Verbin-

derung mit den tariflichen Bestimmungen dieses LRA angewendet werden (Böhner & Ziepke, 1972).

Das ursprüngliche Ziel der Einführung einer tariflichen Leistungsbeurteilung und der damit einhergehenden Leistungszulage war die Schaffung von Leistungsanreizen für Zeitlohnarbeiter. Demzufolge sollte den Arbeitern die Möglichkeit gegeben werden, ihre Entlohnung durch das individuelle Leistungsverhalten direkt zu beeinflussen, um letztlich ihre Arbeitsanstrengung zu erhöhen. Neben diesem Aspekt wurde mit der Einführung angestrebt, dass eine differenzierte Beurteilung des Leistungsstands erfolgt und dieser über den Zeitverlauf beobachtet wird. Zudem sollten die Leistungszulagen das Betriebsklima verbessern, da die damaligen Verdienstunterschiede zwischen den Zeitlohnarbeitern und den im Akkord- sowie im Prämienlohn beschäftigten Arbeitnehmern zu Unzufriedenheit führte. In diesem Zusammenhang sollte die Zahlung von Leistungszulagen einen Lohnausgleich zwischen den erwähnten Mitarbeitergruppen sicherstellen. Darüber hinaus wurde angenommen, dass die Leistungsbeurteilung eine ordnende Funktion übernahm, da sie für die Rekrutierung und Weiterbildung des Personals verwendet werden konnte (Böhner & Ziepke, 1972).

Die damaligen rechtlichen Bestimmungen für die Zahlung einer Leistungszulage für Zeitlohnarbeiter waren in § 9 Nr. 4 des Lohnrahmenabkommens für die Eisen-, Metall- und Elektroindustrie NRW aus 1970 zu finden. Den Regelungen zufolge erhielten Zeitlohnarbeiter eine leistungsabhängige Zulage, die vom Arbeitgeber in Prozent auszuweisen war. Diese Leistungszulagen mussten im betrieblichen Durchschnitt in den Lohngruppen 1 bis 10 mindestens 13% der tariflichen Lohnsumme betragen. Der § 1 Nr. 1 TV LB Zeitl besagte, dass für jeden Zeitlohnarbeiter eine Beurteilungskarte anzulegen ist, auf der das Ergebnis und der Zeitraum, auf den sich die Leistungsbeurteilung bezieht, vermerkt wird. Darüber hinaus sollte die Beurteilung durch einen Beauftragten des Arbeitgebers durchgeführt und von diesem unterzeichnet werden. Die spezifischen Beurteilungsmerkmale wurden in § 2 TV LB Zeitl erwähnt. Den dort aufgeführten Bestimmungen zufolge hing die Bewertung der persönlichen Leistung von den in Tabelle 4 aufgeführten Beurteilungsmerkmalen ab.

Tabelle 4: Beurteilungsmerkmale nach TV LB Zeitl vom 7. September 1970

Beurteilungsmerkmale	Beschreibung
Arbeitsergebnis	Leistungsverhalten durch Intensität und Wirksamkeit
Arbeitsausführung	Güteleistung durch Einhaltung der vorgeschriebenen Qualität (Ausschuss) Arbeitsmethoden und Fertigungsvorschriften
Arbeitseinsatz	Einsatz für verschiedene Arbeitsaufgaben
Arbeitsorgfalt	Sachgemäße Behandlung der Betriebsmittel, sparsamer Verbrauch von Werkzeugen, Energie, Werk- und Hilfsstoffen Beachtung der Sicherheitsvorschriften

Quelle: Eigene Darstellung in Anlehnung an § 2 TV LB Zeitl.

Böhner und Ziepke (1972) weisen darauf hin, dass die Tarifparteien sich bewusst auf nur vier Beurteilungsmerkmale festgelegt hatten, die durch die erklärenden Worte (siehe Spalte Beschreibung in Tabelle 4) eine Legaldefinition erhielten. Die Beurteilungsmerkmale, wie auch viele weitere Begriffe des LRA stehen im engsten Zusammenhang mit der Lehre des Verbands für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung (REFA).⁹

In § 3 TV LB Zeitl waren die übergeordneten Beurteilungsmerkmale und -stufen mit der anzuwendenden Punkteskala aufgelistet. Die erwähnten Evaluationsstufen bezogen sich stets auf das jeweilige Merkmal und reichten beispielsweise im Fall des zu bewertenden Arbeitsergebnisses von a) Leistungsverhalten umständlich, b) Leistungsverhalten noch nicht ganz zweckmäßig, c) Leistungsverhalten zweckmäßig, d) Leistungsverhalten sehr zweckmäßig, bis hin zu e) Leistungsverhalten vorzüglich. Ein Zeitlohnarbeiter konnte im besten Fall – wenn er die Leistungsanforderungen voll erfüllt hatte – mit 8 Punkten bewertet werden, im schlechtesten Fall mit 0 Punkten. Die Abstufung der Punkteskala erfolgte entsprechend der Bewertungsstufen in Zwei-Punkte-Schritten. Dementsprechend hatte die Bewertungsskala 5 Ausprägungen, die von den beteiligten Tarifparteien aus Gründen der Praktikabilität für die Unternehmen bewusst in dieser Form festgelegt wurden (Böhner & Ziepke, 1972). Die Gesamtpunktzahl, die ein Zeitlohnarbeiter erhielt, wurde nach § 4 Nr. 2 und § 9 Nr. 4 Abs. 1 TV LB Zeitl vom Unternehmen im Rahmen einer schriftlichen Mit-

⁹ REFA ist die älteste Organisation ihrer Art in Deutschland. Sie befasst sich unter anderem mit Verfahren der Arbeitsbewertung, den Methoden und Kriterien und stimmt diese mit den Tarifparteien ab (www.refa-consulting.de).

teilung an den Arbeiter in Prozent ausgewiesen. Die relevanten Umrechnungsfaktoren wurden tarifvertraglich nicht vorgegeben (siehe § 4 Nr. 1 TV LB Zeitl) und mussten daher von den Betrieben selbst bestimmt und bekanntgegeben werden.

Im Gegensatz zu den Zeitlohnarbeitern wurde Böhner und Ziepke (1975) zufolge der Angestelltenbereich für tarifvertragliche Bestimmungen über eine Leistungszulage zunächst nicht berücksichtigt, jedoch erfuhren die Angestellten durch die Einführung eines neuen Gehaltsrahmenabkommens (GRA) im Oktober 1970 erhebliche Verbesserungen im Hinblick auf ihr Gehalt. Neben dieser Entwicklung hielt die IG Metall an der Einführung einer Leistungsbeurteilung für den Angestelltenbereich fest. So wurde seitens der Gewerkschaft von einer Kündigung des GRA abgesehen, falls die Arbeitgeberseite sich dazu bereit erklärte, auch für die Gruppe der Angestellten ein Verfahren zur Leistungsbeurteilung einzuführen. Die Konsequenz hiervon war, dass nur über wenige Änderungen des GRA verhandelt und das Abkommen seitens der IG Metall nicht gekündigt wurde. Letztendlich signalisierte die Arbeitgeberseite ihr Einverständnis, da zwischenzeitlich in nahezu allen Tarifgebieten der Metallindustrie Leistungszulagen für Angestellte eingeführt worden waren. Aufgrund der Tatsache, dass die Leistung von Akkord- oder Prämienlohnarbeitern nicht in derselben Weise wie die der Angestellten messbar ist, wurde trotz heftiger Kritik sowohl seitens der Arbeitgeber als auch Arbeitnehmer ein Tarifvertrag zur Leistungsbeurteilung von Angestellten beschlossen. Dieser Tarifvertrag wurde vor allem von kleineren und mittelständischen Unternehmen als zu impraktikabel und kompliziert wahrgenommen. Die Arbeitnehmer kritisierten die sich ergebenden Konsequenzen, die beispielweise dazu führten, dass nachgewiesene Leistungssteigerungen nicht mit entsprechend höheren Leistungszulagen einhergingen. Dies konnte realisiert werden, da der Arbeitgeber die Faktoren zur Umrechnung der erreichten Punkte in Prozent für spezifische Lohngruppen selbst bestimmen und bekanntgeben konnte (Böhner & Ziepke, 1975).

Abgesehen von dem Verfahren zur Ermittlung der Leistungszulage wurde der TV LB Zeitl als Vorbild für die Entwicklung eines adäquaten Tarifvertrags für den Angestelltenbereich verwendet: der Tarifvertrag zur Leistungsbeurteilung von Angestellten (TV LB Ang). Ein positiver Aspekt dieser Vorgehensweise war, dass die betroffenen Betriebe in der Industrie sich nicht auf ein neues Beurteilungsverfahren einstellen mussten. Ein Ziel der Implementierung dieses Leistungsbeurteilungsverfahrens war, wie auch bei den Zeitlohnarbeitern, dass die Höhe der Leistungszulagen durch den Output der Angestellten bestimmt und nach

einem tariflich festgelegten Beurteilungssystem determiniert wird. Darüber hinaus sollte eine leistungsgerechtere Entlohnung sichergestellt, der Führungsstil sowie die Beziehungen zu den Mitarbeitern durch ein modernes Beurteilungswesen verbessert und die Personalplanung sowie die Weiterbildung der Mitarbeiter optimiert werden (Böhner & Ziepke, 1975).

Mit der Einführung eines neuen GRA der Eisen-, Metall-, Elektro- und Zentralheizungsindustrie NRW vom 19. Februar 1975 wurde letztlich auch die Anwendung von Leistungsbeurteilungen im Angestelltenbereich tarifvertraglich festgehalten. Nach Ziepke (1975) hat dieses GRA neben anderen Regelungen insbesondere neue Bestimmungen über Leistungszulagen erfahren, auf die im Folgenden näher eingegangen wird.

Der TV LB Ang vom 19. Februar 1975 stellte nach Böhner und Ziepke (1975) eine Ergänzung zu § 5 des relevanten GRA dar und war nur in Verbindung mit diesem Paragraphen zu verstehen. In § 5 Nr. 1 GRA war geregelt, dass Angestellte bezüglich ihrer Leistung eine entsprechende Zulage erhielten, deren Berechnung sich nach dem TV LB Ang richtete. Die allgemeinen tariflichen Bestimmungen der Leistungsbeurteilungen waren in § 1 TV LB Ang zu finden. Dieser Paragraph besagte, dass für jeden Angestellten eine Beurteilungskarte anzulegen war, auf der das Ergebnis der vergangenen Beurteilungsperiode vermerkt wurde (siehe Nr. 1). Die Beurteilung konnte einmal im Jahr von einem Beauftragten des Arbeitgebers durchgeführt werden und war von diesem zu unterzeichnen (siehe Nr. 1 und Nr. 5). Informationen zu den einzelnen Beurteilungsmerkmalen konnten § 2 TV LB Ang entnommen werden, der die individuelle Leistung von den in Tabelle 5 aufgelisteten Beurteilungsmerkmalen abhängig machte.

Tabelle 5: Beurteilungsmerkmale nach TV LB Ang vom 19. Februar 1975

Beurteilungsmerkmal	Beschreibung
Anwendung der Kenntnisse	Sorgfalt, Genauigkeit, Zuverlässigkeit
Arbeitseinsatz	Intensität, Wirksamkeit, Selbstständigkeit, Kostenbewusstsein
Arbeitsverhalten bei unterschiedlichen Arbeitssituationen	Überblick, Beweglichkeit, Setzen von Prioritäten
Zusammenarbeit	Informationsaustausch, Überzeugungsfähigkeit, Zusammenarbeit zur gemeinsamen Lösung von Aufgaben

Quelle: Eigene Darstellung in Anlehnung an § 2 TV LB Ang.

Der darauffolgende § 3 TV LB Ang enthielt Informationen zum eigentlichen Verfahren der Beurteilung. Dort waren noch einmal die Beurteilungsmerkmale sowie die Beurteilungsstufen und die vom Arbeitgeber bzw. dem Beauftragten des Arbeitgebers zu vergebenden Punkte festgehalten. Die Beurteilungsstufen reichten von a) Genügt den Anforderungen nicht immer, b) Genügt den Anforderungen fast immer, c) Genügt den Anforderungen in vollem Umfang, d) Übertrifft die Anforderungen, bis hin zu e) Übertrifft die Anforderungen in besonderem Umfang. Angestellte erhielten in der schlechtesten Beurteilungsstufe a) 0 Punkte und in der besten e) 8 Punkte. Dementsprechend war die Beurteilungsskala in Zwei-Punkte-Schritte unterteilt. Die ermittelten Punkte wurden anschließend verwendet, um die Leistungszulage in Prozent auszuweisen. Die rechtlichen Grundlagen hierzu fanden sich in § 4 TV LB Ang. Nr. 1 dieses Paragraphen besagte, dass die erreichte Punktzahl mit dem Faktor 0,25 zu multiplizieren war, um die Leistungszulage eines Angestellten zu berechnen. Diese Leistungszulage war nach Nr. 2 desselben Paragraphen den Angestellten in Prozent schriftlich mitzuteilen und auszuweisen.

Seit den 1970er Jahren fanden in den darauffolgenden drei Jahrzehnten keine signifikanten Änderungen der tariflichen Bestimmung zur Leistungsbeurteilung statt. Erst am 18. Dezember 2003 wurde ein neues ERA für NRW abgeschlossen, das überarbeitete Bestimmungen zur tariflichen Leistungsbeurteilung von Arbeitern und Angestellten enthält, wobei seitdem keine Unterscheidung mehr zwischen diesen beiden Gruppen vorgenommen wird.

4.1.2 Leistungszulagen nach dem aktuellen ERA

In den Jahren 1978 bzw. 1991 kündigte die IG Metall NRW die in Kapitel 4.1.1 vorgestellten LRA und GRA, um über einen gemeinsamen Entgeltrahmenabkommen-Tarifvertrag, dem ERA-TV, zu verhandeln (Ohl, 2011). Eine Absichtserklärung, in der die Gewerkschaft festhielt, dass Verhandlungen über ein neues ERA geführt werden sollten, wurde bereits im Jahr 1980 formuliert. Das ERA ersetzt nicht nur die Bestimmungen der gekündigten LRA und GRA, sondern beinhaltet darüber hinaus umfassende Regelungen zur Leistungsbeurteilung und Zahlung von Leistungszulagen und ersetzt daher auch die Tarifverträge zur Leistungsbeurteilung von Zeitlohnarbeitern und Angestellten (TV LB Zeitl und TV LB Ang). Die damals geltenden Tarifverträge umfassten insgesamt 20 Paragraphen zuzüglich der sieben Paraphen der jeweiligen Tarifverträge zur Leistungsbeurteilung von Zeitlohnarbeitern und Angestellten. Im Vergleich dazu enthält das ERA lediglich

12 Paragraphen, wobei bei der Schaffung des Tarifvertrags keine Verschlinkung der tariflichen Bestimmungen im Fokus stand, sondern eine Anpassung an die aktuellen wirtschaftlichen Gegebenheiten (Metall NRW, 2004).

Seit dem Inkrafttreten des ERA am 1. März 2004 (siehe § 12 ERA-TV) erhalten Beschäftigte im Zeitentgelt neben den tariflich festgelegten Monatsgrundentgelten eine Leistungszulage, deren Bestimmungen in § 10 ERA-TV zu finden sind. Diese kann individuell in einer Spannbreite zwischen 0% und 20% variieren und muss nach § 6 Nr. 3 ERA-TV im betrieblichen Durchschnitt mindestens 10% der tariflichen Monatsgrundentgeltsummen aller Beschäftigten im Zeitentgelt betragen. Demzufolge verändert sich durch die Einführung des ERA theoretisch auch das Verhältnis zwischen den gezahlten Monatsgrundentgelten und den Leistungszulagen (Metall NRW, 2006). Wie im vorangehenden Kapitel 4.1.1 erläutert, existierten in den bisherigen Tarifverträgen unterschiedliche Bestimmungen zu den gezahlten Leistungszulagen von Zeitlohnarbeitern und Angestellten. Während die Leistungszulagen der Zeitlohnarbeiter zuletzt im Durchschnitt bei ca. 16% lagen, erhielten die Angestellten hingegen im Mittel ca. 4% an leistungsabhängigen Zahlungen (Metall NRW, 2007a), wobei im Angestelltenbereich keine tarifliche Vorgabe in Bezug auf ein durchschnittliches Zulagenvolumen existierte. Die Differenzen in den Leistungszulagen für Zeitlohnarbeiter und Angestellte versuchte man durch die Einführung des ERA zu kompensieren. Die einheitliche Behandlung der zwei Mitarbeitergruppen hatte zur Folge, dass sich der Anteil der Leistungszulagen bei den ehemaligen Arbeitern im Mittel von 16% auf 10% verringerte und bei den Angestellten von ca. 4% auf 10% erhöhte. Der tarifliche Entgeltanspruch – d.h. das Grundentgelt plus die Leistungszulage – eines Beschäftigten blieb jedoch konstant, da der Anteil der in den Tariftabellen enthaltenen Grundentgelte sich für die ehemaligen Arbeiter erhöhte bzw. für die Angestellten verringerte (Metall NRW, 2006). Die beschriebene Angleichung wurde im Angestelltenbereich mit der Tarifierhöhung im Jahr 2004 und der Einführung einer festen ERA-Leistungszulage herangeführt, die bis zur betrieblichen Umsetzung des ERA wie ein fixer Gehaltsbestandteil behandelt wurde, zum Beispiel im Hinblick auf die Berechnungen der durchschnittlichen als auch individuellen Leistungszulagen. Die nachfolgenden Tarifierhöhungen flossen ebenfalls in diese fixe ERA-Leistungszulage ein, die letztlich in der im ERA festgelegten betriebsdurchschnittlichen Vorgabe von ca. 10% variabler Leistungszulage aufgeht. Das Ergebnis dieses Angleichungsprozesses war, dass die ehemaligen Angestellten zukünftig keine Benachteiligung erfuhren (Metall NRW, 2007a).

Den betroffenen Betrieben wurde seit dem Inkrafttreten des ERA am 1. März 2004 Zeit gegeben, sich auf die betriebliche Einführung des ERA und speziell dem neuen Verfahren der tariflichen Leistungsbeurteilung vorzubereiten. Die tariflichen Regelungen zum Einführungszeitraum sind in § 2 ERA-Einführungstarifvertrag (ERA-ETV) zu finden. In § 2 Nr. 2 wird festgehalten, dass die Einführungsphase vier Jahre umfasst. Diese konnte durch die Zustimmung der Tarifvertragsparteien um 12 Monate verlängert werden, falls betroffene Unternehmen mehr Zeit zur Umsetzung des ERA benötigten. Demzufolge ist das ERA spätestens seit dem 1. März 2009 – auch als Einführungsstichtag bezeichnet – grundsätzlich für alle Betriebe verbindlich. Eine Einführung des ERA während der Vorbereitungsphase war laut Nr. 1 desselben Paragraphen nur möglich, falls eine Zustimmung der Tarifparteien (Metall NRW und IG Metall) vorlag.

Seit der stichtagsbezogenen Einführung des ERA sind Unternehmen dazu verpflichtet, ihre tariflich beschäftigten Mitarbeiter im Zeitengelt mit dem neuen einheitlichen Verfahren zu bewerten und auf Basis dessen deren Leistungszulagen zu bestimmen. Der betroffene Beschäftigtenkreis setzt sich zusammen aus den ehemaligen Zeitlohnarbeitern sowie Angestellten, für die der ERA-TV Anwendung findet.¹⁰ Die Grundlage des ERA-Beurteilungssystems bilden im Wesentlichen die Bestimmungen für die ehemaligen Angestellten, zum Beispiel in Bezug auf die Beurteilungsmerkmale, Bewertungskriterien oder Evaluationsstufen (Metall NRW, 2007b). Die ERA-Beurteilungsmerkmale und Bewertungsstufen, die für die Evaluation der persönlichen Leistung zu berücksichtigen sind, werden in § 10 Nr. 8 ERA-TV aufgelistet. Den dort beschriebenen Regelungen zufolge sind zur Bewertung der individuellen Leistung die in Tabelle 6 aufgezeigten Beurteilungsmerkmale ausschlaggebend.

¹⁰ Zu den übrigen Beschäftigtenkreisen zählen Mitarbeiter, die ein Leistungs- (Prämien- oder Akkordlohn) oder ein Zielentgelt erhalten (im Rahmen einer Zielvereinbarung).

Tabelle 6: Beurteilungsmerkmale nach ERA-TV vom 18. Dezember 2003

Beurteilungsmerkmal	Beschreibung
Anwendung der Kenntnisse	Sorgfalt, Genauigkeit und Zuverlässigkeit
Arbeitseinsatz	Intensität, Wirksamkeit, Selbstständigkeit, Kostenbewusstsein, sachgemäße Behandlung der Betriebsmittel
Beweglichkeit	Überblick, Setzen von Prioritäten, Arbeitsverhalten bei verschiedenen Arbeitssituationen
Zusammenarbeit/Führungsverhalten	Informationsaustausch, Überzeugungsfähigkeit, aufgabenorientierte Zusammenarbeit

Quelle: Eigene Darstellung in Anlehnung an § 10 Nr. 8 ERA-TV.

Analog zum TV LB Ang können die zu beurteilenden Mitarbeiter gemäß ERA-TV minimal 0 und maximal 8 Punkte pro Bewertungskriterium erhalten – auch die Abstufung erfolgt in Zwei-Punkte-Schritten. Des Weiteren gelten dieselben fünf Bewertungsstufen, die von a) genügt den Leistungsanforderungen nicht immer bis hin zu e) übertrifft die Leistungsanforderungen in besonderem Umfang reichen.

Bei näherer Betrachtung der Kriterien für die Leistungsbeurteilung der ehemaligen Angestellten und der neuen ERA-Bewertungskriterien fällt auf, dass diese sich nur marginal voneinander unterscheiden. Dies ist darauf zurückzuführen, dass die beteiligten Tarifparteien sich im Rahmen der ERA-Verhandlungen darauf geeinigt hatten, die neuen ERA-Beurteilungskriterien sowie -stufen an das Beurteilungsverfahren der ehemaligen Angestellten anzugleichen. Die Bewertungskriterien und -merkmale für den Arbeiterbereich blieben bei dieser Angleichung unberücksichtigt, da die Beurteilungssystematik nach dem ERA prinzipiell auch für die Gruppe der ehemaligen Zeitlohnarbeiter anwendbar ist und keine Benachteiligung dieser Beschäftigtengruppe darstellt. Dies ist damit zu begründen, dass auch Zusammenhänge zwischen den ERA-Kriterien und den Kriterien zur Leistungsbeurteilung der ehemaligen Zeitlohnarbeiter bestehen. Beispielsweise stehen die Bewertungsdimensionen Anwendung der Kenntnisse und Fertigkeiten (Leistungsbeurteilung gemäß dem ERA-TV) und Arbeitsausführung (Leistungsbeurteilung entsprechend dem TV LB Zeitl) in unmittelbarem Zusammenhang zueinander (Metall NRW, 2007b).

Tarifliche Rahmenbedingungen für die Leistungsbeurteilung nach ERA-TV

Nach § 10 Nr. 1 ERA-TV erhalten Beschäftigte im Zeitgelt nach Ablauf ihrer Probezeit zusätzlich zu ihrem tariflich bestimmten Monatsgrundentgelt eine variable Leistungszulage. Zur Ermittlung dieser Leistungszulage hat der Arbeitgeber einmal jährlich die Leistung

bzw. das Leistungsverhalten der Beschäftigten zu beurteilen, wobei eine Bewertung in der Regel spätestens bis 18 Monate nach der letzten zu erfolgen hat (siehe § 10 Nr. 7 Abs. 1 ERA-TV). Die Ergebnisse für den Bewertungszeitraum sind vom Arbeitgeber bzw. dessen Beauftragten auf einer Beurteilungskarte festzuhalten und zu unterzeichnen. Die Beurteilungskarten können auch in elektronischer Form angefertigt und archiviert werden (siehe § 10 Nr. 3 ERA-TV). Da keine tariflichen Vorgaben hinsichtlich der Gestaltung einer typischen Beurteilungskarte existieren, bleibt es den Unternehmen selbst überlassen, wie sie diese gestalten. Um jedoch sicherzustellen, dass eine Leistungsbeurteilung durchgeführt und eine Leistungszulage bestimmt wird, sollte die Beurteilungskarte notwendige Mindestinformationen enthalten. Diese sind zum Beispiel der Vor- und Zuname der zu beurteilenden Person, dessen Entgeltgruppe, die Organisationseinheit, das Datum der Bewertung und die in der Beurteilung erreichten Gesamtpunkte (Metall NRW, 2008).

Das Beurteilungssystem gemäß dem ERA-TV umfasst vier Bewertungskriterien mit fünf Beurteilungsstufen. Da ein Mitarbeiter pro Beurteilungsdimension minimal 0 und maximal 8 Punkte erhalten kann, entspricht die Gesamtpunktzahl nach Addition der Punkte pro Beurteilungskriterium im schlechtesten Fall 0 und im besten Fall 32 Bewertungspunkten. Gemäß § 10 Nr. 10 Abs. 4 ERA-TV beträgt die Gesamtsumme der von den Betrieben zu zahlenden Leistungszulagen ca. 10% der tariflichen Grundentgelte der im Zeitentgelt Beschäftigten. Hieraus ergibt sich, dass für einen durchschnittlich leistenden Beschäftigten mit 16 Punkten in der Leistungsbeurteilung die individuelle Leistungszulage 10% betragen muss. Demzufolge liegt der tarifliche Faktor bei 0,625% (10%/16), der dem Geldwert pro Beurteilungspunkt entspricht und für die Ermittlung der prozentualen Zulagen anzuwenden ist.¹¹

Gesamtbetrieblich liegt das Zulagenvolumen für die Beschäftigten im Zeitentgelt mit Leistungszulage in der tariflichen Bandbreite zwischen 9% und 11% (Metall NRW, 2008). Fällt das Zulagenvolumen unter oder über diese tariflich definierte Bandbreite, sind von den Betrieben folgende Anpassungsmechanismen zu berücksichtigen. Falls die betrieblich ermittelte Gesamtsumme der Leistungszulagen unter 9% liegt, sind die Betriebe zur Korrektur der Summe auf 10% verpflichtet, indem sie den tariflich definierten Faktor anheben (Muss-Korrektur). Liegt das Zulagenvolumen oberhalb der tariflich vorgegebenen Bandbreite, d.h. 11%, können die Betriebe mittels einer Senkung des tariflichen Faktors die

¹¹ Auf die Bestimmung der individuellen Leistungszulagen wird in Kapitel 4.4 näher eingegangen.

Gesamtsumme der Leistungszulagen auf 10% korrigieren (Kann-Korrektur). Der betriebliche Korrekturfaktor im Fall einer Unter- bzw. Überschreitung lässt sich nach Metall NRW (2009) anhand folgender Formel berechnen¹²:

$$\text{Korrekturfaktor} = \frac{9\% \text{ bzw. } 11\%}{\text{Betriebliche Gesamtsumme der ermittelten Leistungszulagen in \%}} * 0,625.$$

Für die Leistungsbeurteilung der Beschäftigten sind ausschließlich die von den Tarifparteien festgelegten Beurteilungsmerkmale relevant (siehe Tabelle 6). Der Arbeitgeber kann allerdings aufgrund freiwilliger Betriebsvereinbarungen ein abweichendes Beurteilungsverfahren mit arbeitsbezogenen Beurteilungsmerkmalen anwenden (siehe § 10 Nr. 12 ERA-TV) und die Gewichtung der Kriterien anpassen (siehe § 10 Nr. 13 ERA-TV).¹³

Die oben erwähnten Bestimmungen des ERA sind für einen Großteil der Beschäftigten in der M+E-Industrie in NRW als auch für die Beschäftigten der übrigen Tarifgebiete relevant. Nachfolgend soll auf die Bedeutung der ERA-Leistungsbeurteilung eingegangen werden. In diesem Zusammenhang werden Statistiken zur Nutzung der Bewertungen in der unternehmerischen Praxis präsentiert.

4.2 Bedeutung des ERA für die betriebliche Praxis

Deutschlandweit gibt es ca. 2,4 Mio. Beschäftigte, die in den 14 Tarifgebieten der M+E-Industrie arbeiten (siehe Anhang 1).¹⁴ Der Großteil dieser Beschäftigten erhält ein festes Entgelt für die Arbeitszeit im Unternehmen und zusätzlich eine Leistungszulage, die durch die Bewertung relevanter Leistungskriterien bestimmt wird. *„Demzufolge stellt die Methode ‚Leistungsbeurteilung‘ das häufigste Bindeglied zwischen der Leistung bzw. dem Leistungsverhalten der betreffenden Beschäftigten und dem dafür vorgesehenen Entgelt dar.“* (Metall NRW, 2008) Für tarifvertraglich Beschäftigte in NRW ist die oben erwähnte Systematik des ERA-TV für die M+E-Industrie vom 18. Dezember 2003 anzuwenden.

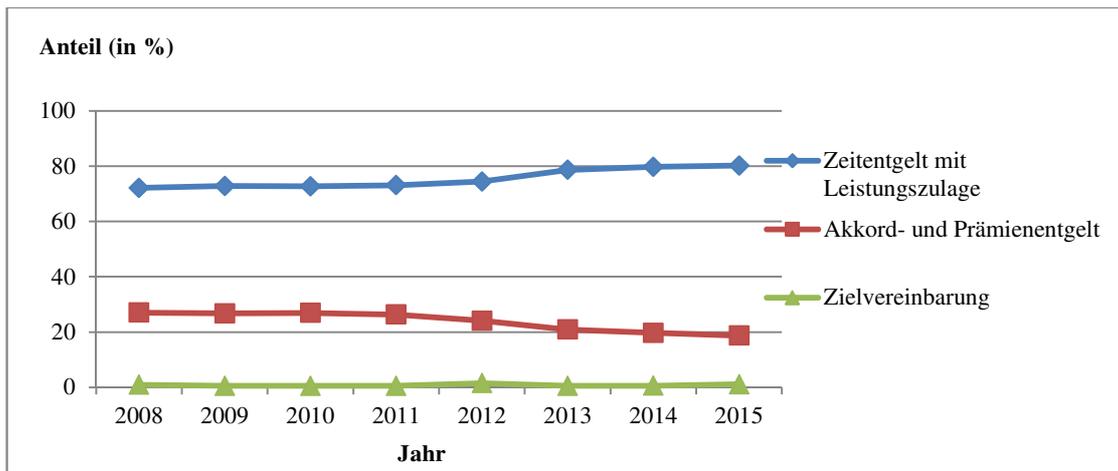
¹² Die Neuberechnung der individuellen Leistungszulagen erfolgt durch die Multiplikation der erreichten Gesamtpunkte in der Leistungszulage mit dem Korrekturfaktor.

¹³ Ist ein Beschäftigter mit seiner Leistungsbeurteilung nicht einverstanden, kann er diese innerhalb von zwei Wochen nach Bekanntgabe des Beurteilungsergebnisses beanstanden (siehe § 10 Nr. 14 Abs. 1 und 2 ERA-TV). Die Abwicklung der Beanstandung erfolgt durch eine paritätische Kommission, die sich aus jeweils zwei Vertretern des Betriebsrats und durch den Arbeitgeber bestimmten Betriebsangehörigen zusammensetzt (siehe § 10 Nr. 14 Abs. 3 und 4 ERA-TV). Weitere Bestimmungen zur ERA-Leistungsbeurteilung finden sich in § 6 ERA-TV (Allgemeine Bestimmungen zum Leistungsentgelt), § 10 ERA-TV (Zeitentgelt und Leistungszulage) sowie § 2 ERA-ETV (Einführungszeitraum) und § 3 ERA-ETV (Voraussetzungen zur Einführung des ERA).

¹⁴ Laut der IG Metall sind es allein in NRW 700.000 Beschäftigte (IG Metall, 2017).

Auch in den übrigen Tarifgebieten ist die Beurteilungspraxis nach dem ERA ausschlaggebend. Gemäß der IG Metall wird die Methode *Beurteilen* von Unternehmen in allen 14 Tarifgebieten der M+E-Industrie angewendet, wobei sich das durchschnittliche Leistungsentgelt bzw. die Leistungszulage der Beschäftigten in den einzelnen Tarifgebieten unterscheidet (siehe Anhang 2). In neun von 14 Tarifgebieten geben die tariflichen Bestimmungen des ERA-TV vor, dass der variable Anteil des Entgelts im betrieblichen Mittel mindestens 10% betragen muss (IG Metall, 2015).

Der Verband der Metall- und Elektroindustrie Nordrhein-Westfalen e.V. geht davon aus, dass ca. 75% der tarifvertraglich Beschäftigten in NRW ein Zeitentgelt mit Leistungszulage erhalten (Metall NRW, 2008). In der M+E-Industrie Baden-Württembergs sind es laut einer schriftlichen Befragung zufolge ca. 60% der Beschäftigten, die auf Basis einer Beurteilung ein Leistungsentgelt beziehen. Werden darüber hinaus Kombinationsmodelle, zum Beispiel eine Leistungsbeurteilung und eine weitere Entgeltmethode, wie eine Zielvereinbarung, berücksichtigt, steigt der Anteil auf ca. 70% (Bahnmüller & Hoppe, 2016a). Die Bedeutung der Leistungszulage nimmt nicht nur in NRW und in Baden-Württemberg zu, sondern im gesamten Bundesgebiet. Laut einer Statistik der Arbeitgeberverbände der M+E-Industrie ist der Anteil an Beschäftigten mit einer Leistungsbeurteilung seit der bundesweiten Einführungsphase des ERA Ende 2008 um ca. 8%-Punkte auf 80% in 2014 gestiegen. Der Anteil der Beschäftigten im Akkord-/Prämienentgelt bzw. mit Zielvereinbarungsentgelt betrug in 2014 nur noch 20% bzw. 1% (Gesamtmetall, 2016). Abbildung 4 zeigt diese bundesweiten Anteile der Beschäftigten in der M+E-Industrie nach entsprechenden Entgeltmethoden auf Basis einer ERA-Verdiensterhebung seit 2008.

Abbildung 4: Beschäftigtenstruktur der M+E-Industrie nach Entgeltmethoden

Quelle: Eigene Darstellung basierend auf Daten von Gesamtmetall (2016).

4.3 ERA und der Bezug zu Entgeltgleichheit

Die Einführung des aktuell gültigen ERA verfolgte nicht nur das Ziel einer Gleichbehandlung zwischen den damaligen Arbeitern und Angestellten, sondern auch die Durchsetzung einer Entgeltgleichheit von Frauen und Männern (Jochmann-Döll & Ranftl, 2010). In diesem Kapitel werden Defizite des ERA aufgedeckt, die dennoch eine Ungleichbehandlung zwischen den Geschlechtern ermöglichen.

Status quo und Gründe für den Gender Wage Gap

Nach Boll, Leppin, Rossen und Wolf (2016) beträgt der unbereinigte Gehaltsunterschied zwischen Frauen und Männern EU-weit 15,3% und in Deutschland sogar 22,2%. Die in Deutschland vorhandene Gehaltsdifferenz kann einerseits zu 14,5% durch unterschiedliche Merkmale von Beschäftigten erklärt werden, wie zum Beispiel Beruf, Alter, Betriebszugehörigkeit, Bildungsstand und andererseits durch Unternehmenseigenschaften, wie beispielsweise Branche oder Unternehmensgröße. Der unerklärte Anteil beträgt dementsprechend 7,7% und lässt sich auf eine unterschiedliche Vergütung von Frauen und Männern und weitere Faktoren zurückführen, die nicht beobachtbar sind (Boll et al., 2016). Diese Einkommenskluft, die auch als *Gender Pay Gap* bezeichnet wird, hat sich in den vergangenen Jahren kaum verändert (Destatis, 2017). Verdienststatistiken für ausgewählte Branchen, die teilweise in den Geltungsbereich des ERA fallen, deuten ebenfalls auf das Vorhandensein eines *Gender Pay Gap* in der deutschen M+E-Industrie hin, der in 2006 zwischen 16% in der Branche Fahrzeugbau und 31% im Wirtschaftszweig der Herstellung von Büromaschinen, Datenverarbeitungsgeräten und -einrichtungen, Elektrotechnik,

Feinmechanik und Optik lag (Jochmann-Döll & Ranftl, 2010).

Für die Beständigkeit der oben erwähnten Einkommenslücke in Deutschland lassen sich nach Boll et al. (2016) mehrere Gründe anführen. Eine indirekte Ursache ist beispielsweise die Gegebenheit, dass Männer in Berufen überrepräsentiert sind, in denen bei gleicher Qualifizierung die Gehälter vergleichsweise höher sind, insbesondere im produzierenden Gewerbe. Darüber hinaus haben Frauen aufgrund familiärer Verpflichtungen häufiger Erwerbsunterbrechungen oder gehen Teilzeittätigkeiten nach, die tendenziell mit geringeren Stundenlöhnen vergütet werden. Zu der bestehenden Einkommenslücke trägt auch das etwas geringere Bildungsniveau von Frauen bei (Boll et al., 2016). Direkte Ursachen für die geschlechterspezifischen Einkommensunterschiede sind hingegen nach Jochmann-Döll & Ranftl (2010) auf diskriminierendes Verhalten von Arbeitgebern zurückzuführen, die durch eine ungleiche Bezahlung von Frauen und Männern für identische oder gleichwertige Tätigkeiten dem Prinzip der Entgeltgerechtigkeit widersprechen.

Diskriminierungspotenzial der Tätigkeitsbewertungen nach dem ERA-TV

Das neue ERA geht auf Entgeltgerechtigkeit zwischen Geschlechtern ein, indem die oben erwähnte Wertigkeit einer Tätigkeit im Rahmen des tarifvertraglich geregelten Arbeitsbewertungsverfahrens neu bestimmt wurde (detaillierte Informationen zum Punktebewertungsverfahren und zur Entgelteingruppierung finden sich in § 3 ERA-TV für die M+E-Industrie in NRW). Die auf Basis der Arbeitsbewertung festgelegte Wertigkeit einer Tätigkeit legt fest, dass Frauen und Männer für die Ausführung der evaluierten Arbeit nicht unterschiedlich vergütet werden dürfen (Jochmann-Döll & Ranftl, 2010). Diverse Forschungsarbeiten weisen jedoch auf Diskriminierungspotenziale hin, die sich durch das Verfahren der Arbeitsbewertung und die Anwendung in der Praxis ergeben können (Krell & Winter, 2011; Ranftl, Buchinger, Gschwandtner & Meggeneder, 2002). Beispielsweise werden bestimmte Tätigkeiten mit Eigenschaften von Frauen in Verbindung gebracht, d.h. stereotypisiert und als geringwertig interpretiert. Diese Stereotypisierung von „Frauenarbeit“ wird durch die Methodik und Anwendung der Arbeitsbewertung und weitere Instrumente von Gehaltssystemen beeinflusst und reproduziert (Chicha, 2006). Untersuchungen mit näheren Informationen über die monetären Folgen der geschlechterspezifischen Auf- und Abwertungen liegen jedoch nicht vor.

Durch die Einführung des ERA und der damit verbundenen Neubewertung von Tätigkeiten haben sich zwangsläufig Änderungen in den Gehältern von Frauen und Männern erge-

ben. Eine nähere Betrachtung der geschlechterbezogenen Verteilung von Auf- und Abwertungen von Tätigkeiten kann Hinweise darüber liefern, ob beispielsweise Frauen vor der Einführung des ERA ein höheres Einkommen bezogen haben und dementsprechend zu den sogenannten Überschreiterinnen gehören. Jochmann-Döll und Ranftl (2010) zeigen im Rahmen einer Befragung von vier Fallbetrieben in unterschiedlichen Tarifgebieten der M+E-Industrie auf, dass Frauen in der Tat überproportional vertreten sind in der Gruppe der Personen, die nach der ERA-Einführung ein geringeres Entgelt erhalten. Mit anderen Worten: den Statistiken der Befragung zufolge waren sie vergleichsweise häufiger von Abwertungen ihrer Tätigkeiten betroffen als männliche Beschäftigte. Die von Bahnmüller und Schmidt (2009) durchgeführte Begleitstudie in dem Tarifbezirk Baden-Württemberg bestätigt die Tendenz, dass Frauen als Überschreiterinnen vergleichsweise deutlich überrepräsentiert sind.

Diskriminierungspotenzial der Leistungsbeurteilungen nach dem ERA-TV

Neben dem Gehalt, das im Rahmen des ERA durch das Verfahren der Arbeitsbewertung und der entsprechenden Eingruppierung festgelegt wird, bieten auch weitere Gehaltsbestandteile, wie die Leistungszulage, Diskriminierungspotenziale, die zu einer unterschiedlichen geschlechterspezifischen Vergütung beitragen können. Wie in Kapitel 4.2 dargestellt, bilden die Beschäftigten im Zeitentgelt mit Leistungszulage die größte Beschäftigtengruppe in der M+E-Industrie. Da die Beschäftigten im Zeitentgelt nach § 10 ERA-TV NRW auf Basis einer Leistungsbeurteilung eine entgeltliche Zulage erhalten, ist fraglich, ob die Leistungszulage als fester Bestandteil der tariflichen Vergütung ebenfalls zur oben diskutierten Lohnlücke zwischen Frauen und Männern beiträgt, die auch in Wirtschaftszweigen im Geltungsbereich des ERA vorhanden ist.

Dass Leistungsbeurteilungen Diskriminierungspotenzial bieten, zeigen unter anderem die Forschungsarbeiten von Krell (2011) und Schreyögg (2011). In der Diskussion zu potentiellen Diskriminierungsursachen, die im Rahmen der Leistungsevaluation auftreten und zu einer ungleichen Beurteilung von Männern und Frauen führen können, erwähnt Krell (2011) klassische Beurteilungsfehler, wie beispielsweise den „Similar-to-me-Effekt“, „Hierarchie-Effekt“ oder „Kleber-Effekt“. Bei der zuerst genannten Verzerrungsart beurteilen Führungskräfte Beschäftigte desselben Geschlechts vergleichsweise positiver (siehe auch Argumentation zur theoretischen Wirkung von Geschlechterdyaden in Kapitel 3.2). Da jedoch die überwiegende Mehrheit der Führungskräfte männlich ist, hat

dieser Effekt eher negative Folgen für Frauen im Unternehmen. Der „Hierarchie-Effekt“ hingegen beschreibt das Phänomen, dass Beschäftigte in höheren Entgeltgruppen systematisch besser bewertet werden. Von dieser Verzerrungsform profitieren jedoch ebenfalls hauptsächlich männliche Personen im Unternehmen, da sie im Vergleich zu Frauen mehrheitlich hierarchisch höher positioniert sind. Ähnliche Wirkungsweisen zeigen sich auch beim „Kleber-Effekt“. In diesem Zusammenhang werden Frauen schlechter beurteilt, wenn sie seit einem längeren Zeitraum – ungerechtfertigt – nicht befördert worden sind (Krell, 2011).

Die oben dargelegten Argumente lassen vermuten, dass Frauen insbesondere in männerdominierten Branchen hinsichtlich Leistungsbeurteilungen nachteilig behandelt werden können. Die M+E-Industrie bildet einen solchen traditionell männerdominierten Industriezweig, da der bundesweite Anteil an Frauen seit mehreren Jahren konstant bei ca. 21% liegt (Gesamtmetall, 2016). Dieser Anteil wird in den Büro- und Verwaltungsberufen überschritten und in den technischen Facharbeiter- und Ingenieursberufen unterschritten. Da jedoch speziell in diesem Industriezweig Führungskräfte tendenziell aus Ingenieurs- und Technikbereichen rekrutiert werden, bilden Männer den Großteil der Führungskräfte. Der Frauenanteil in den obersten fünf Führungsebenen der Branche beträgt daher im Mittel nur 10% (Gesamtmetall, 2011). Da die M+E-Industrie traditionell männerdominiert ist, kann somit angenommen werden, dass klassische Beurteilungsverzerrungen, die in Verbindung stehen mit Geschlechterstereotypen, auch bei der tariflich geregelten ERA-Leistungsbeurteilung relevant sind und Auswirkungen auf die Gehaltszulagen von Frauen haben.

Neben den beschriebenen Bewertungsverzerrungen können Beurteilungskriterien, Verfahren und auch Beschreibungshilfen für Bewertungsmerkmale potentielle Gründe für diskriminierende Leistungsbeurteilungen liefern. Beispielsweise leitet Krell (2011) auf der Grundlage der Anforderungen für eine geschlechterneutrale Arbeitsbewertung (siehe Krell & Winter, 2011) Kriterien für eine Diskriminierungsprüfung von Leistungsbeurteilungen ab. Die Anforderungen beinhalten unter anderem, dass Beurteilungssysteme objektive Kriterien enthalten müssen, die zur einheitlichen Bewertung von Frauen und Männern herangezogen werden. Explizit erwähnt wird auch eine diskriminierungsfreie Auslegung der Evaluationskriterien. Allgemein sind Evaluationen von merkmalsbasierten Kriterien, wie beispielsweise Belastbarkeit und Durchsetzungsvermögen, kritisch zu sehen, da durch

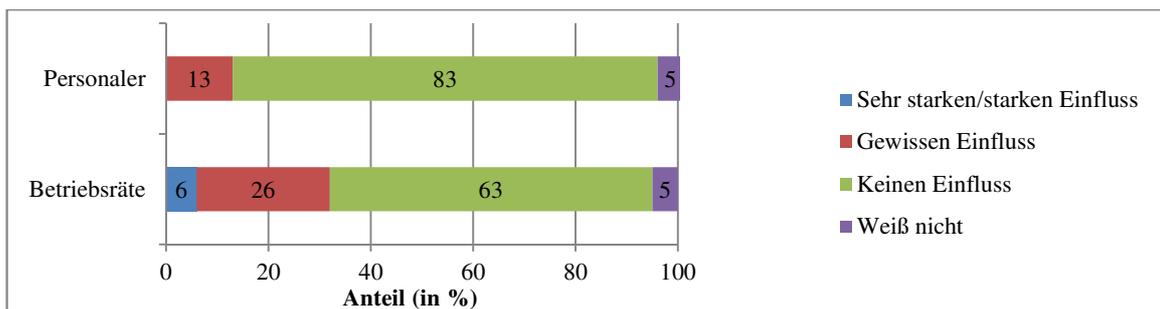
die Beurteilung dieser Dimensionen unbewusst Geschlechterstereotype und somit Bewertungsverzerrungen hervorgerufen werden können (Niedersächsisches Ministerium für Soziales, Frauen, Familie und Gesundheit, 2005; Tondorf & Jochmann-Döll, 2005). Auch in Bezug auf die im ERA-TV aufgelisteten verhaltensorientierten Beurteilungsmerkmale (zum Beispiel Arbeitseinsatz oder Beweglichkeit) ist anzunehmen, dass in der praktischen Anwendung des Beurteilungssystems Geschlechterstereotype eine Bedeutung haben können, da beispielweise hinsichtlich des Merkmals Beweglichkeit unter anderem „ ... *die Belastbarkeit bei der Übertragung anderer Tätigkeiten/Arbeitsaufgaben* ... “ zu beurteilen ist (Metall NRW, 2007b: Kapitel 12, S. 20)¹⁵. Im Hinblick auf die Evaluation des Kriteriums Zusammenarbeit/Führungsverhalten wird den Führungskräften empfohlen, die Überzeugungsfähigkeit zu evaluieren. Für die Beurteilung der Überzeugungsfähigkeit wiederum kann das „ ... *konstruktive Durchsetzungsvermögen gegenüber anderen* ... “ als Bewertungsgrundlage herangezogen werden (Metall NRW, 2007b: Kapitel 12, S. 20). Die Beschreibungen der Beurteilungsmerkmale deuten darauf hin, dass sie Diskriminierungspotenzial bieten und folglich zu geschlechterspezifischen Leistungsbeurteilungen und Vergütungen führen können, falls das Ergebnis der Beurteilungen einer Gehaltskomponente entspricht, wie es im tariflichen System des ERA bei Beschäftigten im Zeitentgelt der Fall ist.

Ein weiterer Faktor, der die Ungleichbehandlung von Geschlechtern bei tariflichen Leistungsbeurteilungen und somit auch das Gehalt von Männern und Frauen beeinflussen könnte, bildet die oben angesprochene Objektivität der Merkmalskriterien. Kratzer und Nies (2009) kommen im Rahmen ihrer Befragung von Fallunternehmen in unterschiedlichen Branchen und Tarifbezirken zu dem Ergebnis, dass die im ERA enthaltenen Beurteilungskriterien sowohl von betrieblichen Experten als auch Beschäftigten mehrheitlich nicht als objektiv, sondern subjektiv wahrgenommen werden. Als mögliche Gründe für Unterschiede in der entgeltlichen Leistungszulage identifizieren die Autoren Faktoren, wie Seniorität, Sympathie und Änderungen in der Zuordnung zu Entgeltgruppen. Bahn Müller und Hoppe (2016b) gehen im Rahmen ihrer Befragung von Betriebsräten und Personalverantwortlichen in der M+E-Industrie unter anderem der Frage nach, inwieweit das Geschlecht eine Einflussgröße auf die Höhe und Verteilung der Leistungszulagen hat. Aus Sicht der befragten Betriebsparteien steht das Geschlecht tendenziell nicht in einem Ver-

¹⁵ In dieser und der nachfolgenden Quellenangabe werden die Kapitelnummern explizit angegeben, da der zitierte Leitfaden für Führungskräfte für jedes Kapitel separat fortlaufende Seitenzahlen enthält.

hältnis mit dem Beurteilungsergebnis. Diesem Zusammenhang stimmen 83% der befragten Personalverantwortlichen und 63% der Betriebsratsmitglieder zu. Jedoch sind 32% der Betriebsratsmitglieder der Ansicht, dass das Geschlecht einen sehr starken/starken (6%) bzw. gewissen (26%) Einfluss auf die Bewertungen hat. Bei den befragten Personalverantwortlichen stimmen 13% der Aussage zu, dass Geschlechtereffekte einen gewissen Einfluss auf die tariflichen Beurteilungen ausüben (siehe Abbildung 5).

Abbildung 5: Geschlecht als Einflussgröße auf die Höhe der Leistungszulagen



Quelle: Eigene Darstellung basierend auf Befragungsdaten von Bahnmüller und Hoppe (2016b).

Die zuvor dargelegten Argumente führen zu der Annahme, dass die ERA-Bewertungen und somit auch die gezahlten Leistungszulagen Diskriminierungspotenzial hinsichtlich des Geschlechts bieten. Es kann zumindest nicht ausgeschlossen werden, dass sowohl das Verfahren der Beurteilung als auch die tariflich vorgegebenen Kriterien zu einer ungleichen Evaluation von Männern auch Frauen führen können. Da Bahnmüller und Hoppe (2016b) in ihrer oben zitierten Umfrage Betriebsräte und Personalverantwortliche befragt haben, ist darüber hinaus nicht geklärt, inwieweit aus Sicht der Beschäftigten – also die von den Zulagenzahlungen Betroffenen – das Geschlecht einen Einflussfaktor auf die Bewertung ihrer Leistung darstellt. Bis dato hat sich keine publizierte Studie mit den monetären Konsequenzen potentieller Geschlechtereffekte beschäftigt. Dies ist erstaunlich, da beispielsweise auch Krell (2011) in ihrem Aufsatz über das Diskriminierungs- und Gleichstellungspotenzial von Leistungsbeurteilungen in ihrer Schlussbemerkung anführt: „Leistungsbeurteilungen enthalten sowohl Diskriminierungspotenzial als auch Gleichstellungspotenzial. Beide Aspekte sind ... noch immer nicht hinreichend bekannt, beachtet und bearbeitet“ (Krell, 2011: 272).

Die vorliegende Fallstudie soll auch einen Beitrag zur Aufklärung des zu Beginn des Kapitels diskutierten *Gender Pay Gaps* beitragen, der indirekt auch in der M+E-Industrie

vorhanden ist, da Frauen zu den sogenannten Überschreiterinnen gehören, die vor der Einführung des ERA verhältnismäßig höhere Einkommen bezogen haben. Da die tariflichen Leistungsbeurteilungen gemäß ERA entgeltrelevant sind und – wie zuvor erörtert – Diskriminierungspotenzial bieten, ist fraglich, inwieweit eine mögliche Geschlechterungleichbehandlung bei den Leistungsbewertungen zur Aufklärung der vorhandenen Gehaltskluft beiträgt, die auch in der M+E-Industrie vorhanden ist. In Deutschland beträgt der bereinigte *Gender Pay Gap* ca. 8%, d.h., dass Frauen bei vergleichbarer Bildung, Berufserfahrung etc. 8% weniger verdienen als Männer. In bisherigen Studien (siehe Kapitel 2.3.1) konnten festgestellte Geschlechtereffekte nicht in monetären Größenordnungen wiedergegeben werden. Da der vorliegenden Studie unter anderem approximierete Gehaltsdaten von Beschäftigten mit tariflichen Leistungsbeurteilungen vorliegen, soll neben den hergeleiteten Forschungshypothesen zu Geschlechtereinflüssen und dem Formalisierungsgrad von Beurteilungen (siehe Kapitel 3.3) ein Beitrag zur Aufklärung der Entgeltgleichheit zwischen Männern und Frauen im tariflichen Umfeld des ERA gegeben werden.

4.4 Beurteilungssystem des Fallstudienunternehmens

Der Fallbetrieb ist mit seinem Produktangebot an Elektrotechnik und Automatisierung in diversen Branchen vertreten. Ein breites Produktangebot an Elektronikkomponenten trägt dazu bei, dass das Unternehmen Kunden aus diversen Branchen bedient, unter anderem Fahrzeugbau, Elektrizitätsversorgung, Maschinen- und Anlagenbau sowie Verkehrstechnik. Der Fallbetrieb produziert weltweit an mehreren Standorten und besitzt Vertriebsgesellschaften in zahlreichen Ländern. Das Unternehmen beschäftigt mehr als 14.000 Mitarbeiter, von denen ca. die Hälfte aus Deutschland stammt und setzt mit seinen Produkten schätzungsweise 2 Mrd. € um (Stand 2016). Der Fallbetrieb gliederte sich zum Zeitpunkt der Datenerhebung in vier Unternehmensbereiche, die neben Deutschland auch internationale Märkte bedienen.

Vergütung der Beschäftigten

Da der Fallbetrieb dazu verpflichtet ist, die tariflichen Regelungen des ERA im Unternehmen umzusetzen, beinhaltet das Gehaltssystem der Beschäftigten sowohl formalisierte als auch weniger formalisierte Bestandteile. Die Belegschaft erhält ein Grundentgelt, das mit höherer Entgeltgruppenzugehörigkeit (in der M+E-Industrie in NRW existieren 14 tariflich vorgegebene Entgeltgruppen) steigt. Bei der Entlohnung seiner Beschäftigten orientiert sich das Fallunternehmen an den vom Arbeitgeberverband der M+E-Industrie in

NRW vorgegebenen Grundentgelten.¹⁶ Diese Gehaltskomponente wird gemäß dem tariflichen Arbeitsbewertungsverfahren (siehe § 3 ERA-TV für nähere Informationen zum Punktebewertungsverfahren) festgelegt und orientiert sich an den empfohlenen Grundentgelten für die entsprechenden Entgeltgruppen. Der fixe Lohnbestandteil ist somit tariflich formalisiert und bietet den Führungskräften keine Spielräume bei der Entlohnung ihrer Beschäftigten. Über das Grundentgelt hinaus zahlt das Fallunternehmen seinen Beschäftigten weitere Zulagen, die den Führungskräften – wie nachfolgend beschrieben – unterschiedliche Spielräume bei der Festlegung bieten.

Tarifliche Leistungszulage

Die Kriterien zur Bestimmung der individuellen Leistungszulagen orientieren sich im Fallstudienunternehmen an den vier tariflich vorgeschlagenen Beurteilungsmerkmalen (siehe Tabelle 6 in Kapitel 4.1.2), die jede Führungskraft jährlich im Einzelgespräch mit dem Beschäftigten zu evaluieren hat. Diese sind:

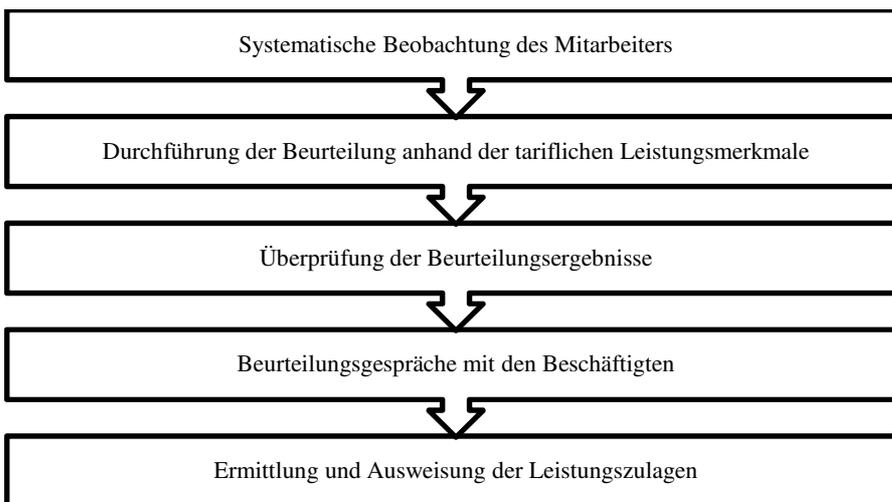
- 1) *Anwendung der Kenntnisse und Fertigkeiten* (Sorgfalt, Genauigkeit und Zuverlässigkeit),
- 2) *Arbeitseinsatz* (Intensität, Wirksamkeit, Selbstständigkeit, Kostenbewusstsein und sachgemäße Behandlung),
- 3) *Beweglichkeit* (Überblick, Setzen von Prioritäten, Verhalten bei verschiedenen Arbeitssituationen)
- 4) und *Zusammenarbeit/Führungsverhalten* (Informationsaustausch, Überzeugungs-fähigkeit, aufgabenorientierte Zusammenarbeit).

Ein Beschäftigter kann in den vier oben genannten tariflichen Beurteilungsdimensionen minimal 0 Punkte und maximal 8 Punkte erhalten, wobei 0 Punkte für eine besonders schlechte und 8 Punkte für eine besonders gute Leistung steht. Die Bewertungsskala ist in Zwei-Punkte-Schritte unterteilt. Neben dem Beurteilungsergebnis werden im Rahmen einer Evaluation unter anderem folgende Angaben über den Beurteilten festgehalten: Perso-

¹⁶ In NRW werden vom Verband der Metall- und Elektroindustrie Nordrhein-Westfalen e.V. jährlich ERA-Entgelttabellen veröffentlicht. Diese Entgelttabellen stellen für die Unternehmen in NRW eine Orientierung für die zu zahlenden Gehälter in den entsprechenden Entgeltgruppen dar. In der Entgelttabelle für das Jahr 2016/2017 beträgt das monatliche Bruttoeinkommen für die niedrigste Entgeltgruppe EG 1 2.294,00€ (bzw. 2.523,40€ mit einer durchschnittlichen Leistungszulage in Höhe von 10% des Bruttoeinkommens) und für die höchste Entgeltgruppe EG 14 5.904,50€ (bzw. 6.494,95€ mit einer durchschnittlichen Leistungszulage in Höhe von 10% des Bruttoeinkommens). Die zitierte Entgelttabelle ist auf der Seite des Verbands verfügbar: www.metall.nrw.

nalnummer, Entgeltgruppe, Organisationseinheit, Datum der Leistungsbeurteilung sowie die erreichten Gesamtpunkte. Das Dokument ist am Ende eines Beurteilungsgesprächs sowohl vom Vorgesetzten als auch von der beurteilten Person zu unterzeichnen. Die tarifliche Leistungszulage wird auf der ersten Seite des Beurteilungsbogens des Fallunternehmens festgehalten.¹⁷ Abbildung 6 stellt den Ablauf des tariflichen Beurteilungsverfahrens vor, das sich in fünf Schritte gliedert und im Fallunternehmen praktiziert wird.

Abbildung 6: Ablauf einer tariflichen Leistungsbeurteilung



Quelle: In Anlehnung an Metall NRW (2008).

Laut den tariflichen Bestimmungen soll die Höhe der durchschnittlichen Leistungszulagen für alle Beschäftigten bei ca. 10% liegen. Übertragen auf den einzelnen Mitarbeiter des Fallbetriebs bedeutet dies, dass bei einer mittelmäßigen Performance mit 16 Bewertungspunkten die individuelle Leistungszulage 10% des Bruttoeinkommens betragen sollte. Dementsprechend lässt sich eine tarifliche Größe von 0,625% ($10\%/16$) ermitteln, die dem Geldwert pro Beurteilungspunkt entspricht (siehe Kapitel 4.1.2). Ein durchschnittlich leistender Beschäftigter erhält dieser tariflich bestimmten Größe zufolge eine Leistungszulage in Höhe von 10% ($16 \cdot 0,625\%$). Ein außerordentlich gut leistender Beschäftigter hingegen bekommt eine Zulage in Höhe von 20% ($32 \cdot 0,625\%$). Tabelle 7 zeigt beispielhaft die Bestimmung einer tariflichen Leistungszulage für einen fiktiven, durchschnittlich leistenden Beschäftigten.

¹⁷ Aus Gründen der Anonymisierung wird der Beurteilungsbogen in dieser Arbeit nicht abgebildet.

Tabelle 7: Fiktives Beispiel zur Bestimmung der tariflichen Leistungszulage

Tarifliches Kriterium	Erreichte Punkte
1) Anwendung der Kenntnisse und Fertigkeiten	2
2) Arbeitseinsatz	5
3) Beweglichkeit	3
4) Zusammenarbeit/Führungsverhalten	6
Summe:	16
Leistungszulage in Prozent:	$16 * 0,625 = 10\%$

Erläuterungen: Fiktives Beispiel in Anlehnung an § 10 ERA-TV. Maximal sind pro Kriterium 8 Punkte möglich.

Nachdem die individuelle tarifliche Leistungszulage in Prozent ermittelt wurde, wird das tarifliche Monatsgrundentgelt (brutto), das bestimmt wird durch die Entgeltgruppe eines Beschäftigten, mit der individuellen tariflichen Leistungszulage multipliziert. Ein Beschäftigter, der beispielsweise in der Entgeltgruppe (kurz EG) 14 eingestuft ist (monatliches Bruttogehalt in Höhe von ca. 5.904,50€ in 2016/2017) und ihm aufgrund seiner Leistungsbeurteilung eine Zulage in Höhe von 10% ($16 * 0,625\%$) zusteht, erhält demzufolge eine individuelle Leistungszulage in Höhe von monatlich 590,45€ (brutto).

Außertarifliche Zulage

Zusätzlich zu der oben beschriebenen tariflichen Leistungszulage erhalten die Beschäftigten des Fallunternehmens einen außertariflichen Bonus. Dafür werden die Beschäftigten nicht nur bezüglich der tariflich festgelegten Beurteilungskriterien bewertet, sondern auch bezüglich ihres Arbeitsergebnisses im abgelaufenen Jahr. Im Fallstudienunternehmen werden die tariflichen sowie außertariflichen Zulagen addiert und an den Beschäftigten ausgezahlt. Bei der Evaluation des Arbeitsergebnisses können die Beschäftigten zwischen zwei Bewertungssystemen entscheiden: Bei einem *Zielvereinbarungsrating* bewertet die Führungskraft, inwieweit vorab festgelegte Ziele im abgelaufenen Jahr erfüllt wurden. Alternativ kann die *Arbeitsquantität* (Menge, Termin, Umfang des Arbeitsergebnisses) beurteilt werden.

Um die erzielten Punkte der außertariflichen Zulagen mit den Punkten der tariflichen Merkmalskriterien vergleichbar zu machen, werden diese ebenfalls in Prozentwerte umgerechnet. Hierfür wird bei der Wahl des *Zielvereinbarungsratings* die erreichte Summe der

Zielerreichungsfaktoren durch 4 dividiert. Die Zielvereinbarung wird auf der zweiten Seite des Beurteilungsbogens explizit festgehalten. Dort werden bis zu drei Ziele und jeweils zwei Kriterien bestimmt. Eine beispielhafte Zielbeschreibung für einen Beschäftigten in der Produktion könnte lauten: Die Ausschussquote und Nachbearbeitungshäufigkeit sollen in der Produktion gesenkt werden. Das erste Kriterium beinhaltet die Senkung der Ausschussquote von 2,5% auf 2,3%. Das zweite Kriterium legt fest, dass die Nachbearbeitungshäufigkeit gesenkt wird und nennt keine expliziten Kennzahlen. Diese beispielhaften Ziele können an einem Zwischen- und/oder Endtermin erfüllt werden. Die Gewichtung der Ziele muss in Summe 9 Punkte betragen. Die Gewichtung wird anschließend mit der Anzahl der erreichten Zielkriterien multipliziert und ergibt letztlich den Zielerreichungsfaktor, der wiederum durch den Wert 4 dividiert wird. In dem oben genannten Beispiel könnten die Senkung der Ausschussquote und der Nachbearbeitungshäufigkeit in der Produktion in Summe mit 9 Punkten gewichtet werden, falls keine weiteren Ziele bestimmt wurden. Bei Erfüllung beider Zielkriterien beträgt der Zielerreichungsfaktor 18 (9*2). Da dieser Wert zur Ermittlung der außertariflichen Zulage durch 4 dividiert wird, beträgt die Zulage auf Basis des Zielvereinbarungsratings im Beispiel 4,5% des Bruttoeinkommens.

Hätte ein Beschäftigter sich der Beurteilung der Arbeitsleistung hingegen für das Kriterium *Arbeitsquantität* entschieden und von seinem Vorgesetzten die maximalen 8 Punkte erhalten, müssten auch diese zur Vergleichbarkeit in Prozent ausgedrückt werden. Bei der Wahl des Kriteriums Arbeitsquantität werden die erzielten Punkte mit dem betrieblichen Faktor 0,5625 multipliziert. Der betriebliche Faktor lässt sich wie folgt bestimmen:

$$\frac{2,25\% \text{ (Durchschnittlich zu erreichende außertarifliche Zulage in \%)}}{4 \text{ (Durchschnittlich zu erreichende Punktzahl im Hinblick auf die Beurteilung der Arbeitsquantität)}} = 0,5625\%.$$

Demnach würde im Beispiel die prozentuale Zulage ebenfalls 4,5% betragen (8*0,5625%). Im Gegensatz zum Zielvereinbarungsrating wird die Bewertung der Arbeitsquantität nicht auf der zweiten Seite des Beurteilungsbogens festgehalten, sondern gemeinsam mit dem Ergebnis der tariflichen Bewertung auf der ersten Seite. Zudem lässt sich der Bewertung der Arbeitsquantität aufgrund fehlender vorab festgelegter Quantitäten ein eher subjektiver Charakter zuschreiben.

Neben der tariflichen Beurteilung und der Bewertung des Arbeitsergebnisses werden auf der zweiten Seite des Beurteilungsbogens Ergänzungen zur Leistungsbeurteilung vermerkt. Diese richten sich an die Leistungsanforderungen des Beschäftigten und können

zum Beispiel Begründungen dafür sein, dass bestimmte Anforderungen nicht erfüllt wurden. Darüber hinaus werden – falls vereinbart – zwischen dem Beschäftigten und seinem Vorgesetzten Absprachen bis zur nächsten Beurteilung getroffen. Eine Absprache könnte sein, dass der Beschäftigte an einer ihm empfohlenen Weiterbildungsmaßnahme teilnimmt, um den zukünftigen Anforderungen an seine Tätigkeit gerecht zu werden.

Formalisierungsgrad

Ein wichtiger Faktor, den es vor dem Hintergrund des Forschungskontexts zu berücksichtigen gilt, ist der Formalisierungsgrad des Beurteilungssystems. In der vorliegenden Studie wird der Grad der Formalisierung durch die Unterscheidung zwischen Beschäftigten mit einer Bewertung der Arbeitsquantität und denen mit einer entgeltrelevanten Zielvereinbarung getroffen.

Die Verknüpfung der tariflichen Leistungsbeurteilung mit einer Bewertung des Arbeitsergebnisses stellt hierbei das weniger formalisierte Bewertungssystem dar, da die Führungskräfte die Arbeitsquantität subjektiv beurteilen. Die Bewertung einer entgeltrelevanten Zielvereinbarung als Ergänzung zur tariflichen Leistungsbeurteilung wird hingegen als vergleichsweise formalisiertes Beurteilungssystem betrachtet, da die Führungskräfte detaillierte Ziele und dazugehörige Kriterien mit den Beschäftigten vereinbaren. Wenn Beschäftigte ihre außertariflich vereinbarten Ziele erfüllen, ist entsprechend zu erwarten, dass sie auch bei den tariflichen Kriterien, wie zum Beispiel Anwendung der Kenntnisse und Fertigkeiten oder Arbeitseinsatz, gut abschneiden. Demnach haben die Vorgesetzten in Bezug auf die genannten Beurteilungsdimensionen nur einen geringen Evaluationsspielraum. Das formalisierte Beurteilungssystem erzeugt in der Wahrnehmung des Vorgesetzten keine mehrdeutigen Informationen, da die Führungskraft lediglich zu beurteilen hat, ob ein oder mehrere vereinbarte Ziele erfüllt oder nicht erfüllt wurden. Das weniger formalisierte Bewertungssystem mit einer subjektiven Beurteilung der Arbeitsquantität hingegen ist gekennzeichnet durch einen vergleichsweise größeren Evaluationsspielraum, der sich auch auf die tariflichen Beurteilungskriterien auswirken kann. Demzufolge bietet der vorliegende Fallbetrieb zwei unterschiedliche Bewertungssysteme, die sich anhand des Formalisierungsgrads voneinander unterscheiden.

5 Forschungssetting

5.1 Mixed-Methods-Ansatz

Die vorliegende Forschungsarbeit basiert auf einer einzelnen Fallstudie (n=1), die einen Mixed-Methods-Ansatz verfolgt und zwei Datentypen kombiniert. Nach Kuckartz (2014) wird unter Mixed-Methods „... die *Kombination und Integration von qualitativen und quantitativen Methoden im Rahmen des gleichen Forschungsprojekts verstanden*“ (Kuckartz, 2014: 33). Dieser Definition entsprechend wurden zum einen die gewonnenen Personaldaten (Daten zu Beschäftigten und deren Leistungsbeurteilungen) verwendet, um die in Kapitel 3.3 hergeleiteten Forschungshypothesen zu testen. Zum anderen wurden qualitative Experteninterviews durchgeführt, um das Verständnis für das im Fallstudienunternehmen praktizierte Beurteilungsverfahren zu vertiefen und um die Ergebnisse der quantitativen Analysen zu diskutieren.

Patton und Appelbaum (2003) weisen explizit darauf hin, dass Fallstudien nicht nur auf die Anwendung qualitativer Methoden zu beschränken sind. Sie ermöglichen eine ganzheitliche Sicht auf einen Untersuchungsgegenstand, da die gewonnene Evidenz – wie auch in der vorliegenden Fallstudie – beispielsweise auf der Auswertung von Dokumenten, Interviews und Beobachtungen basiert (Yin, 1984). Die Fallstudie berücksichtigt eine ganzheitliche Perspektive und wendet unter anderem quantitative Methoden (ökonometrische Verfahren und Hypothesentests) an, um Geschlechterunterschiede in tariflichen Leistungsbeurteilungen zu untersuchen. Die hieraus gewonnenen quantitativen Ergebnisse werden gemäß des Insider-Econometrics-Ansatzes um die Informationen betrieblicher Experten ergänzt: „... *insiders should be used to interpret the results* ...“ (Shaw, 2009: 614). Die betrieblichen Experten haben ein detailliertes Wissen über die Anwendung der tariflichen Leistungsbeurteilungen und sind daher als Insider zu sehen. Sie sollen insbesondere helfen, die kausalen Mechanismen zu verstehen (George & Bennett, 2005), die bei der Herleitung der Forschungshypothesen in Bezug auf Leistungsbeurteilungen beschrieben wurden. Die dargelegten Argumentationen hinsichtlich der hergeleiteten Hypothesen lassen darauf schließen, dass Geschlechtereffekte in Bewertungen heterogen sein können und durch organisationalen Mechanismen beeinflusst werden. Folglich empfiehlt es sich, das Wissen betrieblicher Insider zu analysieren, um die wirkenden Mechanismen näher zu beschreiben, die zu Geschlechtereffekten führen können.

Die inhaltliche Vorbereitung, die Leitfadenerstellung sowie die Durchführung der Experteninterviews hat sich an dem Leitfaden von Bähring, Hauff, Sossdorf und Thommes (2008) zur qualitativen Befragung von Experten in Unternehmen orientiert. Da die Experteninterviews nicht zur Hypothesengenerierung, sondern zur Diskussion der Ergebnisse und zu einer Erhöhung der internen Validität der quantitativen Befunde beitragen sollen, wurden die Inhalte der Interviews zwar transkribiert, jedoch nicht inhaltsanalytisch ausgewertet.

5.2 Experteninterviews

Es wurden insgesamt acht Interviews durchgeführt, wobei in zwei Interviewrunden jeweils drei und in einer Interviewrunde zwei Experten befragt wurden. Die Gespräche wurden im Rahmen von Einzelinterviews durchgeführt, um eine vertrauensvolle und anonyme Gesprächssituation herzustellen, da sich dies insbesondere bei heiklen Themen wie die tarifvertragliche Leistungsentlohnung von Beschäftigten anbietet. Die teilnehmenden Personen aus dem Fallunternehmen haben größtenteils Führungsverantwortung (nur eine Mitarbeiterin hat keine) und nehmen in der betrieblichen Praxis hinsichtlich der ERA-Leistungsbeurteilungen eine Doppelrolle ein. Da die Interviewteilnehmer tarifvertraglich angestellt sind, beurteilen sie ihre Beschäftigten, bekommen jedoch auch tarifvertragliche Leistungsbewertungen von ihren Vorgesetzten, die in der Regel hierarchisch höher positioniert sind. Durch die Einbeziehung von Führungskräften mit einer Doppelrolle werden ein Stück weit die Sichtweisen sowohl der Vorgesetzten als auch der Mitarbeiter beleuchtet.

Auswahl der Interviewteilnehmer

Neben den am Fallstudienprojekt beteiligten Experten wurden weitere Interviewpartner gewonnen, die nicht von Beginn an an dem Projekt beteiligt waren. Die Kriterien zur Auswahl der Interviewpartner wurden in Absprache mit den Projektpartnern des Unternehmens festgelegt. Diese waren zum einen der Bereich im Unternehmen und die Funktion des Experten im Betrieb. Zum anderen wurde darauf geachtet, dass beide Geschlechtergruppen einbezogen werden und dass das Alter sowie die Betriebszugehörigkeit unter den Experten variieren. Eine Vorselektion potentieller Interviewteilnehmer wurde durch die Projektpartner im Fallunternehmen vorgenommen. Über eine E-Mailanfrage wurde der ausgewählte Personenkreis an potentiellen Experten kontaktiert. Gegenstand dieser E-Mail waren allgemeine Inhalte zum Forschungsprojekt – eine Untersuchung zu den praktizier-

ten ERA-Leistungszulagen im Unternehmen – sowie die Intention der Interviews. Die Ergebnisse der quantitativen Studie hingegen waren nicht Bestandteil dieser E-Mail. Die Selektion der Interviewpartner erfolgte letztlich durch die im Projekt eingebundenen Beschäftigten des Fallstudienbetriebs. Tabelle 8 zeigt die Zusammensetzung der Interviewpartner, die aus den Bereichen Personal, Produktion, Arbeitsschutz, Fertigung, Compliance und Entwicklung stammen, deren demographische Merkmale, die im Hinblick auf das Geschlecht, die Betriebszugehörigkeit und das Alter variieren, sowie die Dauer der durchgeführten Interviews.

Tabelle 8: Informationen zu Experteninterviews und Teilnehmern

Funktion, Bereich	Geschlecht (m/w)	Betriebszugehörigkeit (Jahre)	Alter (Jahre)	Interviewdauer (Minuten)
Führungskraft, Personalwesen	m	7	32	28:59
Referentin, Personalwesen	w	10	40	14:51
Führungskraft, Personalwesen	w	8	37	29:15
Führungskraft, Produktion	m	23	54	23:08
Führungskraft, Arbeitsschutz	w	8	52	25:41
Führungskraft, Fertigung	w	29	46	13:05
Führungskraft, Compliance	w	8	55	24:38
Führungskraft, Entwicklung	m	18	49	30:37

Erläuterungen: Aus Gründen der Anonymisierung werden einzelne Abteilungen zu Unternehmensbereichen zusammengefasst und keine weiteren Informationen zu den Experten gegeben, die eindeutige Rückschlüsse auf deren Identität zulassen würden. Die Daten beziehen sich auf das Erhebungsjahr 2016.

Projektverlauf

Die Interviews mit den betrieblichen Experten fanden im Jahr 2016 an drei unterschiedlichen Terminen und Räumen des Fallstudienunternehmens statt. Das kürzeste Interview nahm ca. 13 Minuten in Anspruch, das längste dauerte etwas mehr als eine halbe Stunde. Alle Gespräche wurden mit einem digitalen Tonbandgerät aufgezeichnet und mithilfe des computergestützten Programms *f4transkript* transkribiert. Neben den erwähnten Interviews haben mehrere Vorklärungsgespräche mit den Projektpartnern zum Thema ERA-Leistungszulagen dazu beigetragen, dass wichtige Themengebiete identifiziert werden konnten, die die Grundlage des themenzentrierten Fragebogens bilden. In diesen Vorgesprächen wurden zudem Hintergrundinformationen zum Verfahren der tariflichen ERA-Leistungsbeurteilungen in der betrieblichen Praxis gegeben.

Durchführung der Interviews

Im Rahmen jedes Experteninterviews hat eine kurze Einführung stattgefunden, in der die Teilnehmer die Möglichkeit hatten, sich kurz vorzustellen. Es wurden der Forschungshintergrund und das Thema des Projekts angerissen und die Bedeutung der Einbindung betrieblicher Experten für den vorliegenden Forschungszweck erörtert. Im Anschluss daran wurde der Ablauf des Gesprächs erklärt und nach dem Einverständnis zur Aufzeichnung des Interviews gefragt. In diesem Zusammenhang ist auch darauf hingewiesen worden, dass die Inhalte der Gespräche anonym behandelt werden. Im letzten Schritt der Einführung wurde den Experten die Möglichkeit gegeben, Fragen zum Ablauf des Gesprächs zu stellen. Zum Abschluss der Interviews wurden Fragen zur Person geklärt. Diese bezogen sich auf die Funktion der Person im Unternehmen, die Übernahme von Führungsaufgaben, das Alter der Befragten und die Betriebszugehörigkeit – soweit dies noch nicht während der Vorstellungsrunde geklärt wurde. Der Ablauf der strukturierten Interviews hat sich an Themenblöcken orientiert.

Der themenzentrierte Leitfaden fokussiert sowohl generelle Wissens- und Einstellungsfragen zum Thema ERA-Leistungsbeurteilungen im Unternehmen als auch die empirischen Befunde aus der quantitativen Studie. Er setzt sich aus den folgenden Themengebieten zusammen:

- Themenblock 1: Allgemeine Fragen zu den ERA-Beurteilungen
- Themenblock 2: Fragen zur Interpretation der Fallstudienresultate
- Themenblock 3: Fragen zum Ausblick
- Themenblock 4: Fragen zur Person.

Die einzelnen Items des themenzentrierten Gesprächsleitfadens können Anhang 3 entnommen werden.

5.3 Quantitative Fallstudie

Neben den oben beschriebenen Experteninterviews wurde im Rahmen dieser Arbeit auch eine quantitative Studie durchgeführt, um die in Kapitel 3.3 hergeleiteten Forschungsfragen zu den Themen Formalisierung von Leistungsbeurteilungssystemen und Geschlechtereffekte in Bewertungen näher zu untersuchen. Die weiteren Erkenntnisse aus den Experteninterviews sollen im Rahmen der quantitativen Studie dazu beitragen, die empiri-

schen Ergebnisse besser zu interpretieren.

Die Daten der quantitativen Studie wurden auf Basis elektronisch erfasster Beurteilungsdaten des Fallunternehmens gewonnen. Die Analyse von Daten eines einzelnen Unternehmens bietet den Vorteil, dass aufgrund der Reichhaltigkeit des Datenmaterials die individuellen Effekte von Personalpraktiken im Detail untersucht werden können (siehe zum Beispiel für die Anwendung im Management Abraham, 2017; Fabel, Thomann & Zhang, 2014; Frick, Goetzen & Simmons, 2013 und im Bereich Soziologie Fernandez & Weinberg, 1997). Die Daten der vorliegenden Fallstudie sind zwar nicht repräsentativ für alle Unternehmen in jeglichen Branchen, jedoch sind sie typisch für Unternehmen in der M+E-Industrie in Deutschland, da die Beurteilungssystematik für die Unternehmen in dieser Branche, wie in Kapitel 4.1.2 dargelegt, tariflich vorgegeben wird.

Über den Fallstudiencharakter hinaus ist das vorliegende Setting der quantitativen Fallstudie vorteilhaft, da neben den Beurteilungsdaten der Beschäftigten auch deren Gehaltsdaten vorliegen. Dies bedeutet, dass potentielle Geschlechterunterschiede in den Leistungsbeurteilungen monetär beziffert werden können.

5.3.1 Operationalisierung der Variablen

Die Mitarbeiterbeurteilungen und -informationen stammen aus dem in Kapitel 4.4 näher beschriebenen Fallunternehmen, das in der M+E-Industrie in NRW ca. 3.500 Mitarbeiter beschäftigt. Der zugrunde liegende Paneldatensatz enthält mehr als 12.000 Leistungsbeurteilungen von Beschäftigten, die in den Jahren 2011 bis 2014 mindestens einmal von ihren Vorgesetzten nach den tarifrechtlichen ERA-Kriterien evaluiert wurden. Die tariflichen Beurteilungen, die einer prozentualen Leistungszulage entsprechen, bilden die abhängige Variable, auf die im Folgenden näher eingegangen wird. Anschließend werden die unabhängigen und die Kontrollvariablen der Fallstudie vorgestellt.

5.3.1.1 Abhängige Variable und unabhängige Variablen

Tarifliche Leistungszulage

In der vorliegenden Fallstudie bildet die Höhe der tariflichen Leistungszulage die abhängige Variable. Die Leistungszulage basiert auf den oben vorgestellten tariflichen Leistungskriterien und wird in Prozent des Bruttolohns der Beschäftigten bemessen. Im Fallunternehmen existieren 14 Einkommensgruppen und das Eckentgelt variierte in den Jahren 2011 bis 2014 je nach Lohngruppe zwischen durchschnittlich 2.067€ und 4.722€ brut-

to pro Monat (siehe letzte Spalte in Tabelle 9). Diese Durchschnittswerte lassen zu, dass die in der quantitativen Studie zu ermittelnden Geschlechtereffekte von prozentualen Angaben in absolute Geldbeträge umgerechnet werden können.

Tabelle 9: ERA-Monatseckentgelte NRW (2011 bis 2014)

Entgeltgruppe	Jahr				Ø
	2011	2012	2013	2014	
EG 1	1.958 €	2.042 €	2.112 €	2.158 €	2.067 €
EG 2	1.982 €	2.067 €	2.137 €	2.184 €	2.092 €
EG 3	2.005 €	2.091 €	2.162 €	2.210 €	2.117 €
EG 4	2.036 €	2.124 €	2.196 €	2.244 €	2.150 €
EG 5	2.078 €	2.168 €	2.241 €	2.291 €	2.194 €
EG 6	2.133 €	2.224 €	2.300 €	2.350 €	2.252 €
EG 7	2.200 €	2.294 €	2.372 €	2.424 €	2.322 €
EG 8	2.314 €	2.413 €	2.495 €	2.550 €	2.443 €
EG 9	2.501 €	2.608 €	2.697 €	2.756 €	2.640 €
EG 10	2.749 €	2.867 €	2.964 €	3.029 €	2.902 €
EG 11	3.081 €	3.214 €	3.323 €	3.396 €	3.253 €
EG 12*	3.356 €	3.495 €	3.614 €	3.694 €	3.540 €
EG 13*	3.827 €	3.991 €	4.127 €	4.217 €	4.040 €
EG 14*	4.472 €	4.665 €	4.823 €	4.929 €	4.722 €

Erläuterungen: Die abgebildeten Werte wurden auf Basis der ERA-Monatsentgelttabellen zusammengetragen, die der Vorstand IG Metall jährlich auf seiner Internetseite veröffentlicht (www.metall.nrw). *Für die Entgeltgruppen 12 bis 14 wurde auf Grundlage der unterschiedlichen Gehaltseingangsstufen, die für Mitarbeiter in NRW gelten, ein arithmetisches Mittel gebildet. Die Eingangsstufen ergeben sich aus unterschiedlichen Verweilzeiten in der jeweiligen ERA-Gruppe.

Geschlechtervariablen

Die unabhängigen Variablen der Fallstudie sind unter anderem mehrere Geschlechtervariablen, die einerseits den Haupteffekt des Geschlechts des Beschäftigten sowie des Vorgesetzten auf die Leistungsbeurteilung schätzen und andererseits die Untersuchung dyadischer Effekte ermöglichen sollen. In einem ersten Schätzmodell werden daher das Geschlecht des Beschäftigten und des Vorgesetzten berücksichtigt, um Hypothese 1 und Hypothese 2 zu testen. Die Kombination beider Geschlechter (Dyade aus Vorgesetzten und den Beschäftigten) wird in einem weiteren Schätzmodell eingeführt, um Hypothese 3 zu testen. Die Variablen *Frau beurteilt Frau*, *Frau beurteilt Mann* und *Mann beurteilt Frau* beschreiben die unterschiedlichen Kombinationen. Dementsprechend ist die Referenzkategorie: *Mann beurteilt Mann*.

Organisationale Hierarchiestufe

Die Forschung zur Wirkungsweise von Job-Komplexität auf Geschlechtereffekte zeigt, dass Vorgesetzte durch die Komplexität der Berufe in ihrem Evaluationsverhalten beeinflusst werden können (siehe Kapitel 2.3.2). Da der Datensatz aus Gründen der Anonymisierung keine expliziten Berufsbezeichnungen der Beschäftigten enthält, werden in der vorliegenden Arbeit die Komplexitätsniveaus aus den Entgeltstufen abgeleitet.¹⁸

Tabelle 9 zeigt, dass das monatlich gezahlte Entgelt mit einer steigenden Entgeltgruppenzugehörigkeit höher ausfällt. Es kann grundsätzlich angenommen werden, dass nicht nur das Gehalt mit steigender Entgeltstufe zunimmt, sondern auch die an die Mitarbeiter gestellten Anforderungen und damit auch die Komplexität des Jobs. Während beispielsweise in EG 1 bis EG 6 einfache Anlern Tätigkeiten ausgeführt werden, bilden für höhere EG (12 bis 14) ein Studium und eine gewisse Berufserfahrung Voraussetzungen für die Ausführung der Tätigkeit. In den EG 7 bis EG 11 wird eine abgeschlossene Berufsausbildung ohne bzw. mit Berufserfahrung vorausgesetzt.¹⁹ Es ist jedoch darauf hinzuweisen, dass Beschreibungen einzelner Entgeltgruppen nur in Tarifgebieten existieren, die im Rahmen der Arbeitsbewertung ein summarisches Verfahren anwenden. Tabelle 10 zeigt die Beschreibung der Anforderungen an die Beschäftigten und deren Qualifikationen für ausgewählte Entgeltgruppen im Tarifgebiet Thüringen. Der ERA-TV für Thüringen gibt eine summarische Arbeitsbewertung vor und hält fest, dass beispielweise für Ausübungen von Tätigkeiten in der Entgeltgruppe EG 1 keine Vorbildung notwendig ist. Die Einstufung in die EG 4 hingegen erfordert eine zweijährige fachspezifische Berufsausbildung, die Zuordnung zur höchsten Entgeltgruppe EG 12 sogar einen Abschluss einer vierjährigen Hochschulausbildung sowie Fachkenntnisse, die eine langjährige Berufserfahrung voraussetzen. Das Fallunternehmen hingegen wendet ein analytisches Verfahren an, indem einzelne Kriterien (zum Beispiel Fachkenntnisse, Berufserfahrung, Handlungs- und Entscheidungsspielraum) mit Punkten bewertet werden. Neben der Hierarchiestufe des Beschäftigten wird auch die Hierarchiestufe des Vorgesetzten in den nachfolgenden Panelschätzungen berücksichtigt.

¹⁸ Einen ähnlichen Ansatz verfolgen Joshi et al. (2015), die diversen Berufen unterschiedliche Komplexitätsniveaus anhand von Bewertungen der Problemlösungskompetenz zuordnen.

¹⁹ Die Informationen über die Komplexitäts- bzw. Qualifikationsniveaus der einzelnen Entgeltgruppen wurden von den Projektpartnern im Fallunternehmen mitgeteilt.

Tabelle 10: Beschreibung ausgewählter Entgeltgruppen am Beispiel ERA Thüringen

Entgelt-Gruppe	Anforderungen an Beschäftigte
EG 1	Einfache Tätigkeiten, deren Ausführung und Ablauf festgelegt sind und die nach einer zweckgerichteten Einarbeitung und Übung von bis zu vier Wochen verrichtet werden können. Es ist keine berufliche Vorbildung erforderlich.
EG 4	Tätigkeiten, deren Ausführung und Ablauf teilweise festgelegt sind. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine mindestens zweijährige fachspezifische Berufsausbildung erworben werden.
EG 8	Ein Aufgabengebiet, das im Rahmen von bestimmten Richtlinien erledigt wird, oder hochwertigste Facharbeiten, die hohes Dispositionsvermögen und umfassende Verantwortung erfordern. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine abgeschlossene mindestens dreijährige fachspezifische Berufsausbildung und eine mindestens zweijährige Fachausbildung oder durch eine abgeschlossene dreijährige Hochschulausbildung (zum Beispiel Bachelor) erworben werden.
EG 12	Ein erweiterter Aufgabenbereich, der teilweise im Rahmen von allgemeinen Richtlinien erledigt wird. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch den Abschluss einer mindestens vierjährigen Hochschulausbildung erworben werden sowie Fachkenntnisse, die langjährige spezifische Berufserfahrung voraussetzen.

Quelle: In Anlehnung an IG Metall (2014). Eine vollständige Übersicht der Anforderungen an Beschäftigte im Tarifgebiet Thüringen kann Anhang 4 entnommen werden.

Nachdem die abhängige und die unabhängigen Variablen der quantitativen Untersuchung vorgestellt wurden, sollen im nächsten Kapitel die Kontrollvariablen beschrieben werden.

5.3.1.2 Kontrollvariablen

Alter und Betriebszugehörigkeit

Da die Wirkungsweisen der Faktoren Alter und Betriebszugehörigkeit ähnlich sind, werden die Operationalisierungen beider Kontrollvariablen im Folgenden gemeinsam vorgestellt.

In Bezug auf das Alter der Beschäftigten kann beispielsweise vermuten werden, dass ältere Beschäftigte aufgrund ihres hohen Alters ungleich behandelt werden könnten. Ein Großteil der Forschung über Ungleichbehandlung im Alter basiert auf Basis von Kongruenzmodellen. Diese Modelle postulieren, dass der relative Fit zwischen dem Alter der Mitarbeiter und dem spezifischen Wissen, den Fähigkeiten und weiteren Merkmalen, die für die Ausführung einer Tätigkeit als erforderlich betrachtet werden, teilweise als Grundlage dafür dient, wie Beschäftigte beurteilt werden (Arvey, 1979; Heilman, 1983). Dabei wird die Einschätzung der Fähigkeiten und Mitarbeitercharakteristiken durch aktivierte Stereotype insofern beeinflusst, dass Beurteilungen besser ausfallen, wenn ein relativer Fit

zwischen den aktivierten Stereotypen hinsichtlich der personenbezogenen Eigenschaften und einer Tätigkeit gegeben ist (Gordon & Arvey, 2004).

Da nicht eindeutig geklärt ist, welche personenbezogenen Eigenschaften – negative oder positive – mit dem Alter oder der Betriebszugehörigkeiten von Beschäftigten assoziiert und welche Stereotype aktiviert werden, ist der Effekt des Alters und der Betriebszugehörigkeit nicht eindeutig zu bestimmen. Beispielsweise könnte ein hohes Alter oder eine lange Betriebszugehörigkeit ein Indikator dafür sein, dass Beschäftigte erfahren sind und bei der Ausführung ihrer Tätigkeiten von diesem Erfahrungswert profitieren, andererseits aufgrund ihres fortgeschrittenen Alters physische Einschränkungen haben und ihre Tätigkeit nicht mehr präzise ausführen. Diesen Überlegungen zufolge ist anzunehmen, dass zwischen dem Alter bzw. der Betriebszugehörigkeit und der Leistungsbeurteilung der Mitarbeiter ein nicht-linearer Zusammenhang besteht. Um dieser quadratischen Beziehung Rechnung zu tragen, werden in der vorliegenden Studie neben dem Alter und der Betriebszugehörigkeit auch die quadrierten Terme der Variablen berücksichtigt.

Bei Leistungsbeurteilungen kann auch das Alter der Vorgesetzten eine Rolle spielen (Landy, 2010). So könnten beispielsweise jüngere Vorgesetzte zu Beginn ihrer Karriere strenger beurteilen, um nicht in den Augen anderer Führungskräfte im Unternehmen als zu milde zu gelten. Alternativ ist auch denkbar, dass Vorgesetzte am Anfang ihrer beruflichen Laufbahn tendenziell besser bewerten, um nicht das Verhältnis mit der Belegschaft zu schädigen. Beide diskutierten Wirkungszusammenhänge sind also möglich.²⁰

Abteilungsmerkmale

Weitere Faktoren, die einen Einfluss auf die Leistungsbeurteilungen von Beschäftigten haben können, bilden die diversen Bewertungskulturen und die Geschlechteranteile in den Funktionsbereichen des Fallunternehmens. Die Funktionsbereiche des Unternehmens gliedern sich in Abteilungen, zu denen beispielsweise klassische Produktionsbereiche als auch Support Units gehören, wie die Personalabteilung, die arbeitstechnisch mehrere Unternehmensbereiche unterstützt. Die einzelnen Unternehmensabteilungen bilden trotz ihrer

²⁰ Siehe Landy (2010) für eine systematische Übersicht von Metaanalysen und individuellen Studien, die den Einfluss des Faktors Alter auf die Leistungsbeurteilung von Individuen untersuchen. Die Überblickstudie kommt zu dem Ergebnis, dass ältere Beschäftigte im Vergleich zu jüngeren nicht systematisch schlechter beurteilt werden, sondern bessere Bewertungen erhalten, wenn sie von jüngeren Vorgesetzten beurteilt werden. Allerdings werden die Ergebnisse der berücksichtigten Untersuchungen hinterfragt, da sie auf der Grundlage von Querschnittstudien basieren und keine Längsschnittdaten berücksichtigen.

Eingliederung in die Gesamtorganisation selbstständige Einheiten mit eigenen Bewertungskulturen. Dies liegt darin begründet, dass Vorgesetzte beispielsweise in kleineren Abteilungen ihre Beschäftigten vergleichsweise besser beurteilen, da sie häufiger mit ihnen in Kontakt treten und soziale Bindungen aufbauen (Breuer et al., 2013). Dieser Argumentation zufolge können in größeren Abteilungen vergleichsweise schwache soziale Bindungen zwischen Führungskräften und Beschäftigten aufgebaut werden, da Führungskräfte beispielweise eine hohe Anzahl an Leistungsbeurteilungen durchführen und kaum die Möglichkeit haben, sich intensiv mit den Bewertungen der einzelnen Beschäftigten auseinanderzusetzen.

Die Gespräche mit den Experten aus dem Fallunternehmen haben bestätigt, dass in der Tat unterschiedliche Bewertungskulturen im Unternehmen existieren. Demnach beurteilen Führungskräfte in einigen Bereichen zu streng bzw. zu milde. Dies liegt unter anderem daran, dass einige Führungskräfte in den Produktionsbereichen des Betriebs mehr als 30 Personen bewerten, deren Leistungen allerdings nicht ausreichend observiert werden können, um differenzierte Bewertungen durchzuführen. Eine mögliche Folge hiervon sind tendenziell zu gute oder schlechte Beurteilungen, die im Rahmen weiterer Beurteilungsrunden nachkalibriert werden müssen, um die tariflichen Bestimmungen in Bezug auf die ERA-Leistungsbeurteilungen einzuhalten. Auf den Einfluss der Bewertungskulturen in Abteilungen wird kontrolliert, in dem in den durchgeführten Schätzmodellen für jeden Unternehmensbereich Dummyvariablen als Kontrollen eingeführt werden.

Neben den unterschiedlichen Bewertungskulturen können auch die Geschlechteranteile in den Abteilungen den Leistungsbeurteilungsprozess beeinflussen. Diverse Forschungsarbeiten zeigen auf, dass der prozentuale Anteil von Frauen in spezifischen Berufsgruppen die Wirkung von Geschlechterstereotypen im Rahmen von Leistungsbeurteilungen und anderer Personalpraktiken verstärken bzw. schwächen kann (Joshi et al., 2015; Koch et al., 2015). Da dem vorliegenden Fallstudien Datensatz aufgrund betrieblicher Regelungen zum Datenschutz keine Informationen über die spezifischen Berufsbezeichnungen der Beschäftigten vorliegen, werden in der quantitativen Untersuchung nicht wie bei Joshi et al. (2015) und Koch et al. (2015) die Geschlechteranteile in Berufen bzw. Berufsfeldern berücksichtigt, sondern die prozentualen Anteile der Geschlechter in den 27 Abteilungen des

Unternehmens.²¹ Im Fallbetrieb liegt der Frauenanteil bei durchschnittlich ca. 31%. Dies bedeutet, dass die Abteilungen im Mittel männerdominiert sind. Die Berücksichtigung der Geschlechteranteile in den Abteilungen ist relevant, da beispielsweise in einer Untersuchung der Stadt München herausgefunden wurde, dass Leistungsbeurteilungen in Berufen mit hohen Frauenanteilen vergleichsweise besser ausfallen als in von Männern dominierten Berufen (Schreyögg, 1998). Da der Frauenanteil in den Abteilungen zu ähnlichen Effekten führen kann, wird in der quantitativen Studie die Wirkung von Geschlechteranteilen auf die Leistungsbeurteilung durch die Aufnahme einer zusätzlichen Kontrollvariable berücksichtigt.

Eine weitere Kontrollvariable, die im Zusammenhang mit den Abteilungsmerkmalen steht, ist der prozentuale Anteil von Frauen auf den Hierarchiestufen. Diesbezüglich zeigen Untersuchungen, dass ein steigender Anteil von Frauen in höheren Entgeltgruppen einer Geschlechterungleichbehandlung von Führungskräften wesentlich entgegenwirkt (Cohen & Broschak, 2013; Cohen, Broschak & Haveman, 1998; Huffman, Cohen & Pearlman, 2010). Der Grund hierfür ist, dass die Präsenz von Frauen in Top-Positionen einen symbolischen Wert für andere Frauen im Unternehmen haben kann, da dies als ein Signal für Statuserrungenschaften verstanden wird (Joshi et al., 2015). Zudem impliziert ein steigender Anteil von Frauen in höheren Entgeltgruppen möglicherweise, dass eine größere Kontrolle über die Allokation von Gehaltszulagen und Leistungsbeurteilungspraktiken vorliegt (Huffman et al., 2010; Joshi et al., 2006).

Da das Fallunternehmen aus einer männerdominierten Branche stammt und der Frauenanteil als auch der Anteil weiblicher Führungskräfte gering ist, liegt die Vermutung nahe, dass Frauen vergleichsweise schlechtere Leistungsbeurteilungen erhalten, da sie insbesondere in den niedrigeren Entgeltgruppen überrepräsentiert sind. Dieser vermutete Effekt wird beispielsweise in einer Studie aus der Schweiz bestätigt, in der Personen mit einer niedrigen Eingruppierung – Frauen waren in diesen Gruppen häufiger vertreten – schlechtere Bewertungen von ihren Vorgesetzten erhalten haben (Arni, Bigler-Wälti, Hösli, Kessler & Morandi, 2002).

Der Einfluss des Frauenanteils in den Entgeltgruppen auf die Leistungsbeurteilung wird in der quantitativen Studie folgendermaßen operationalisiert: Mitarbeiterinnen in EG 1 bis

²¹ Die genauen Bezeichnungen der Abteilungen werden in der vorliegenden Arbeit aufgrund datenschutzrechtlicher Bestimmungen nicht aufgeführt.

EG 6 werden als geringqualifiziert eingestuft, da sie eher Anlern Tätigkeiten ausführen. Durchschnittlich qualifizierte Mitarbeiterinnen hingegen haben mindestens eine zwei- bis dreijährige Berufsausbildung und können gegebenenfalls Berufserfahrung vorweisen (EG 7 bis EG 11). Hochqualifizierte weibliche Beschäftigte sind vorrangig in EG 12 bis EG 14 vertreten, da sie in der Regel neben Berufserfahrung auch eine universitäre Ausbildung vorweisen sollten. Das mittlere Qualifikationsniveau dient in den Schätzungen der quantitativen Studie als Referenzkategorie.

Proxy für die Arbeitsleistung der Beschäftigten

Um von vornherein auszuschließen, dass Geschlechtereffekte durch tatsächliche Leistungsunterschiede zwischen Mitarbeiterinnen und Mitarbeitern beeinflusst werden, wird in den quantitativen Schätzungen der vorliegenden Fallstudie auch das Arbeitsergebnis der Beschäftigten kontrolliert (Arbeitsquantität versus Zielvereinbarungsrating). Es ist anzunehmen, dass das Arbeitsergebnis der Beschäftigten auch die tarifliche Beurteilung beeinflusst, da das Arbeitsergebnis beispielsweise mit den tariflichen Kriterien Qualitätsorientierung und Effizienz der auszuführenden Tätigkeit in einer Beziehung steht. Dies lässt sich an zwei Beispielen in Bezug auf die erwähnten Zusammenhänge wie folgt erklären: Wenn ein Beschäftigter ein hohes mengenmäßiges Arbeitsergebnis abgeliefert hat, ist anzunehmen, dass Einbußen bei der Sorgfalt, Genauigkeit und Zuverlässigkeit vorliegen (Unterkriterien des Beurteilungsmerkmals Anwendung der Kenntnisse und Fertigkeiten). Ähnliche Auswirkungen einer hohen mengenmäßigen Arbeitsleistung sind auch hinsichtlich des tariflichen Leistungskriteriums Beweglichkeit denkbar (Überblick, Setzen von Prioritäten, Verhalten bei verschiedenen Arbeitssituationen). Um diesen begründeten Zusammenhängen Rechnung zu tragen, werden in den quantitativen Schätzungen die Faktoren Arbeitsquantität bzw. der Erreichungsgrad einer Zielvereinbarung als Proxy für die Mitarbeiterleistung operationalisiert. Diesbezüglich ist anzumerken, dass diese Proxyvariablen für das Arbeitsergebnis letztlich auch subjektive Indikatoren darstellen, da sowohl die zu evaluierende Arbeitsquantität als auch das Zielvereinbarungsrating gegebenenfalls von subjektiven Einschätzungen der Vorgesetzten abhängen.

Ein weiteres Argument für die Berücksichtigung der oben beschriebenen Faktoren Arbeitsquantität und der Erreichungsgrad von Zielen ist, dass das Arbeitsergebnis eines Beschäftigten auch einen Ankereffekt verursachen kann. Zu Beginn eines Beurteilungszeitraums wird zwischen den Beschäftigten und Vorgesetzten festgelegt, ob das außertariflich

bewertete Arbeitsergebnis in Form einer subjektiven Einschätzung der Arbeitsquantität oder einer Zielvereinbarung erfolgt. Obwohl die Bewertung des Arbeitsergebnisses nicht zur Bestimmung der tariflichen Leistungszulage beiträgt, kann angenommen werden, dass durch diese zusätzliche Bewertung weitere Informationen zur Verfügung gestellt werden, auf die Vorgesetzte bei der Bewertung der tariflichen Kriterien zurückgreifen. Diese zusätzlichen Informationen hinsichtlich des Arbeitsergebnisses können Ankereffekte hervorrufen. Demnach wirkt das Abschneiden in der Beurteilung des quantitativen Arbeitsergebnisses oder in der Zielvereinbarung als Anker, der einen Einfluss auf die tarifliche Leistungsbeurteilung ausübt – zumal zwischen den tariflichen und außertariflichen Bewertungskriterien Beziehungen vermutet werden. Diverse Forschungsarbeiten deuten darauf hin, dass Anker die Wahrnehmung von beurteilenden Individuen beeinflussen, unter anderem auch im Kontext von Leistungsbeurteilungen (Epley & Gilovich, 2001; Furnham & Boo, 2011; Thorsteinson et al., 2008).

Die Bewertung des Arbeitsergebnisses kann demnach aufgrund von zwei Wirkmechanismen Anker bzw. zusätzliche Informationen erzeugen, die die Wahrnehmungen der Vorgesetzten beeinflussen und die Leistungsbeurteilungen von Frauen und Männern verzerren. Zum einen signalisiert die Wahl der Bewertung des Arbeitsergebnisses, ob ein Mitarbeiter die Bewertung auf Grundlage der geleisteten Arbeitsquantität oder anhand einer Zielvereinbarung präferiert. Zum anderen erhält die Führungskraft Informationen über die Arbeitsleistung des Beschäftigten im vergangenen Beurteilungszeitraum. Informationen über die geleistete Arbeit wiederum können die tarifliche Beurteilung beeinflussen, da bei einer hohen Arbeitsleistung zum Beispiel der Arbeitseinsatz (Kriterium 2) oder die Beweglichkeit (Kriterium 3) eingeschränkt sein kann. Es lässt sich somit festhalten, dass das geleistete Arbeitsergebnis einen Einfluss auf die Bewertung der tariflichen Leistungskriterien haben kann.

5.3.2 Methodische Vorgehensweise

Um geschlechterspezifische Bewertungsverzerrungen in Leistungsbeurteilungen empirisch zu untersuchen, wurde eine analytische Strategie entwickelt, die in diesem Kapitel näher erläutert wird. Nach der Beschreibung der zugrunde liegenden Schätzstrategie wird das Regressionsmodell vorgestellt, das die Unterschiede von Frauen und Männern in einem formalisierten als auch in einem weniger formalisierten Beurteilungssystem untersucht und die in Kapitel 5.3.1 vorgestellten Variablen enthält.

Analytische Strategie

Da mehr als 12.000 Leistungsbeurteilungen von Beschäftigten über mehrere aufeinanderfolgende Jahre vorliegen, bietet sich die Durchführung von Paneldatenanalysen an. Der zugrundeliegende Paneldatensatz ist nicht balanciert, da Ein- und Austritte von Beschäftigten während des Beobachtungszeitraums möglich sind. Es werden vier separate Panelschätzungen durchgeführt. Zunächst berücksichtigen die ersten beiden Schätzmodelle lediglich die Geschlechtervariablen der Beschäftigten und der Vorgesetzten. Die weiteren Schätzmodelle beinhalten im Gegensatz dazu die kombinierten Geschlechtervariablen und berücksichtigen somit den Einfluss spezifischer Dyaden aus Vorgesetzten und Beschäftigten. Von besonderem Interesse ist hierbei die Kombination, bei der eine männliche Führungskraft eine Mitarbeiterin beurteilt. Als Referenzkategorie dient die Dyade, die sich aus einer männlicher Führungskraft und einem männlichen Beschäftigten zusammensetzt. Um zu überprüfen, wie sich der Formalisierungsgrad des Beurteilungssystems auf die Leistungsbeurteilungen auswirkt, wurden für die beiden oben beschriebenen Beurteilungssysteme separate Schätzungen durchgeführt. Zur Untersuchung der Einflüsse von Gehaltsgruppen auf geschlechterspezifische Effekte wurden auf Basis der Panelschätzungen weitere Analysen mit Interaktionen zwischen der Entgeltgruppe und dem Geschlecht der Beschäftigten durchgeführt.

Regressionsbasierte Schätzmodelle

Um die Auswirkungen von Formalisierung und Geschlechterdyaden auf die Leistungsbeurteilungen von Mitarbeitern zu untersuchen und den Panelcharakter der Fallstudien Daten auszunutzen, bietet sich die Schätzung mehrerer Panelregressionen an, die auf der Logik der oben genannten Modelle aufbauen. Da die zu schätzenden Regressionsmodelle Variablen enthalten, die nicht über die Beurteilungszeiträume variieren, wie beispielsweise die haupterklärenden Geschlechtervariablen (Geschlecht der Mitarbeiter und Vorgesetzten sowie Dyaden), wird in der vorliegenden Untersuchung nach Greene (2012), Verbeek (2012) und Wooldridge (2013) die Random-Effects-Methode angewendet. Dabei bilden die beobachteten Dyaden aus Vorgesetzten und Mitarbeitern die Panelvariable. Diese Vorgehensweise hat den Vorteil, dass soziale Beziehungen zwischen den Dyaden berücksichtigt werden, die sich über die Zeit entwickeln können. Im zugrunde liegenden Regressionsmodell sind die Standardfehler auf der Ebene der Dyaden geclustert, um robuste Schätzer zu erhalten.

Die vier oben beschriebenen Modelle basieren auf folgender Schätzgleichung:

$$Y_{it} = \beta_1 * \text{Geschlecht}_{it} + \beta_2 * \text{Beschäftigter}_{it} + \beta_3 * \text{Vorgesetzter}_{it} + \beta_4 * \text{Abteilung}_{it} + \beta_5 * \text{Abteilungsdummy}_{it} + \alpha + u_{it} + \varepsilon_{it}$$

Die abhängige Variable ist die prozentuale Leistungszulage, die sich aus der tariflichen Leistungsbeurteilung ergibt. Diese wird im obigen Basismodell durch Y_{it} abgebildet, wobei i eine Beobachtung einer individuellen Dyade i zum Zeitpunkt t darstellt. Der Vektor Geschlecht_{it} wiederum stellt abhängig von den oben beschriebenen Schätzmodellen entweder das Geschlecht des Beschäftigten sowie Vorgesetzten dar oder alternativ einen Indikator für die unterschiedlichen Geschlechterdyaden. Mögliche Ausprägungen sind: Frau beurteilt Frau, Frau beurteilt Mann, Mann beurteilt Frau und Mann beurteilt Mann, wobei die letztere Geschlechterkombination die Referenzkategorie darstellt. $\text{Beschäftigter}_{it}$ und Vorgesetzter_{it} repräsentieren Vektoren, die die oben beschriebenen demographischen sowie weitere personenbezogene Variablen beinhalten. Ferner enthält das obige Modell Kontrollvariablen, die auf die Eigenschaften der *Abteilungen* kontrollieren, wie zum Beispiel Frauenanteile in den Abteilungen und der Anteil an Frauen, die geringqualifiziert (EG 1 bis EG 6), durchschnittlich qualifiziert (EG 7 bis EG 11) und hochqualifiziert (EG 12 bis EG 14) sind. Da der Fallstudien datensatz Informationen von Beschäftigten aus allen 27 Abteilungen des Unternehmens enthält und jede Abteilung eine eigene Bewertungskultur hat, sollen durch die Variable $\text{Abteilungsdummy}_{it}$ neben den Geschlechteranteilen auf weitere Abteilungsunterschiede kontrolliert werden. α repräsentiert den Achsenabschnitt, u_{it} einen Fehlerterm, der für die Variation innerhalb einer Dyade i über die Zeit t und ε_{it} einen Fehlerterm, der für die Variation zwischen den unterschiedlichen Dyaden über die Zeit kontrolliert.

Im folgenden Kapitel 6 werden zunächst wichtige Insiderinformationen zur ERA-Leistungsbeurteilung und anschließend die Ergebnisse der Fallstudie präsentiert. In diesem Zusammenhang werden in einem ersten Schritt Einstellungs- und Wissensfragen über die Bewertungen geklärt. Im Anschluss daran wird auf deskriptive Statistiken der berücksichtigten Variablen eingegangen. Es werden die quantitativen Ergebnisse präsentiert und die in Kapitel 3.3 hergeleiteten Hypothesen zu Formalisierung und organisationale Mechanismen mithilfe des oben beschriebenen Basismodells getestet. Die Aussagen der Experten aus den Interviews, die im Rahmen der qualitativen Studie durchgeführt wurden, werden hierbei verwendet, um die quantitativen Ergebnisse zu interpretieren.

6 Ergebnisse und Interpretationen

6.1 Insiderinformationen zur ERA-Leistungsbeurteilung

Dieses Kapitel soll die in Kapitel 4.4 erörterten Merkmale des im Fallstudienunternehmen praktizierten Beurteilungsverfahrens aufgreifen, indem allgemeine Fragen zu dem angewendeten Verfahren anhand von Aussagen der betrieblichen Insider geklärt werden. Die weiteren Ergebnisse der Experteninterviews werden mit der Diskussion der Ergebnisse aus dem quantitativen Teil der Arbeit vorgestellt, da die Antworten der Experten unter anderem mögliche Interpretationen für die aufgedeckten statistischen Zusammenhänge bieten.

In Themenblock 1 (siehe Gesprächsleitfaden in Anhang 3) wurden zunächst Einstellungs- und Wissensfragen zum Thema ERA-Leistungsbeurteilung behandelt. Die erste Frage an die Experten soll Aufschluss darüber liefern, wie eine übliche Leistungsbeurteilung im Fallunternehmen abläuft.

Die Mitarbeitergespräche finden einmal im Jahr statt und werden in dem Zeitraum 1. Januar bis 31. März durchgeführt. Beurteilt wird jeweils die relative Leistung sowie das Leistungsverhalten der zurückliegenden 12 Monate. Die Durchführung der Beurteilung läuft laut der Antworten der Experten von Führungskraft zu Führungskraft unterschiedlich ab. Die interviewte männliche Führungskraft im Personalwesen meint beispielsweise: *„Ich glaube, es gibt kein Standardgespräch ... und jede Führungskraft führt dieses Gespräch anders“*. Unterschiede sind zum Beispiel: der Beurteilungsbogen wird im Vorfeld des Gesprächs oder während des Gesprächs ausgefüllt; Mitarbeiter führen eine Selbsteinschätzung durch und/oder erhalten eine Fremdeinschätzung; zwei der befragten Führungskräfte (Produktion und Compliance) erwarten ferner ein Feedback zu ihrem Führungsstil. Gemeinsamkeiten im Verfahren hingegen sind, dass die Punktevergabe und die Beurteilungsentscheidungen mit Argumenten begründet und die einzelnen Beurteilungskategorien und -ergebnisse tendenziell systematisch aufbereitet werden. Darüber hinaus bekommen die Mitarbeiter in der Regel die Möglichkeit, im Dialog mit ihren Vorgesetzten die Punktevergabe zu besprechen und ihre Sichtweise über ihre geleistete Performance zu erläutern. Sind die Mitarbeiter mit ihrer Beurteilung nicht einverstanden, können sie ihre Leistungsbeurteilung reklamieren. Im Durchschnitt werden im Fallbetrieb nur ca. 10 Mitarbeiterbewertungen pro Jahr reklamiert.

Die tarifliche Vorgabe besagt, dass die variablen Leistungszulagen im betrieblichen

Durchschnitt 10% der Monatsgrundentgeltsummen aller Mitarbeiter entsprechen müssen. Auch im Hinblick auf diese Herausforderungen gehen die Meinungen der Experten auseinander. Die Hälfte der befragten Führungskräfte sieht keine Problematiken, die sich aus der tariflich vorgegebenen Quote ergeben können. Zum Beispiel meint die männliche Führungskraft im Personalwesen, dass die Vorgesetzten im Unternehmen seit der Einführung des ERA in 2011 Erfahrungen mit dem Prozess der Leistungsbeurteilungen sammeln konnten und mittlerweile vertraut sind mit dem Verfahren: *„Also, es weiß jeder, dass das Unternehmen diesen Mittelwert erreichen muss und wenn der nicht erreicht wird, gibt es irgendwelche Anpassungsmechanismen und die möchte keiner haben.“* Auch die weibliche Führungskraft im Personalwesen bestätigt diese Ansichten und resümiert: *„Alle wissen um den Umstand, dass man im Schnitt auf die 10% kommen muss.“* Laut den Aussagen der Referentin im Personalwesen verursacht die Einhaltung der tariflichen Vorgaben auch Schwierigkeiten in der betrieblichen Praxis. Sie sieht es problematisch, *„... dass der Tarifvertrag ... diese engen Leitplanken vorgibt. Das heißt, man hat eigentlich kaum Flexibilität ...“*. Die Führungskraft im Funktionsbereich Arbeitsschutz thematisiert ebenfalls das Problem der tariflichen Quote und sieht die Beurteilungspraxis kritisch, da die Vorgesetzten nicht vorerst ihre Bewertungen vergeben und anschließend eventuelle Anpassungsmechanismen nutzen, sondern mit der Vorgabe im Hinterkopf die Beurteilungen ihrer Mitarbeiter gestalten. Jedoch führt auch die übliche Vorgehensweise zu Problemen. Die Führungskraft im Compliance meint beispielsweise, dass die tarifliche Vorgabe nachträglich zu Veränderungen der Leistungsbeurteilungen geführt hat. In ihrem Fall wurde das Thema aktiv aufgenommen und es sind letztlich im Nachhinein Korrekturen vorgenommen worden. Auch die Führungskraft in der Entwicklung kritisiert das Verfahren der tariflichen Beurteilung. Die ERA-Bewertungen und die durchschnittliche Vorgabe in Höhe von 10% sind seiner Meinung nach kritisch zu sehen, da eine Leistungssteigerung im Vergleich zum Vorjahreszeitraum nicht automatisch mit einer besseren Leistungsbeurteilung seiner Mitarbeiter einhergehe. Darüber hinaus sei es ein relatives Verfahren und führe auch unter den Mitarbeitern zu einem Wettbewerb um die entgeltlichen Zulagen – in anderen Abteilungen sei in diesem Zusammenhang sogar von „Grabenkämpfen“ unter Kollegen die Rede.

Bis auf einen Insider stimmen alle Interviewteilnehmer zu, dass die ERA-Evaluationen keine bzw. nur eine geringe Anreizfunktion haben. Diesbezüglich führen die Insider zwei Gründe an: Zum einen spiegeln sich Leistungssteigerungen kaum in der Vergütung der

Mitarbeiter wider, zum anderen ermöglicht die oben erwähnte Vorgabe nur eine geringe Differenzierung zwischen Hoch- und Geringleistern. Die weibliche Führungskraft im Personalwesen thematisiert die enge Bandbreite der Notenverteilung und ist der Meinung, dass ein Großteil der Mitarbeiter eine Leistungszulage in Höhe von ca. 9 bis 11% erhält. *„Und dann habe ich natürlich eine nur noch sehr schwache Differenzierung zwischen einem Geringleister, sage ich mal, bei 9% und einem guten Performer bei 11%.“* Auch die Führungskraft im Compliance ist sich sicher, dass der Bewertungsspielraum für Vorgesetzte eingeschränkt ist, um nach den Leistungen ihrer Mitarbeiter zu differenzieren und dementsprechend bessere oder schlechtere Leistungsbeurteilungen zu vergeben. Die befragten Experten sind mehrheitlich der Auffassung, dass durch die eingeschränkte Differenzierung bei der Notenvergabe keine Leistungsanreize geschaffen werden. Lediglich zwei Führungskräfte nennen Faktoren, wie die allgemeine Jobzufriedenheit oder die intrinsische Motivation, die im Gegensatz zu den ERA-Leistungsbeurteilungen mehr Anreize für Leistungssteigerungen bieten.

In Hinblick auf die Objektivität der tariflichen Evaluationen sind sich alle Experten einig, dass die ERA-Evaluationen tendenziell subjektiv sind. Hierzu merkt die männliche Führungskraft im Personalwesen Folgendes an: *„Es gibt keine Parameter außer Stückzahl und irgendwie eine messbare Qualität oder sowas, die objektiv sind. Und die Kriterien, die ... in der täglichen Leistungsbeurteilung sind, sind nicht objektiv“.* Zudem argumentiert er, dass Beurteilungsfehler trotz der angebotenen Schulungen und Schulungsunterlagen für Vorgesetzte auftreten. Auch die Führungskraft in der Produktion sieht die Problematik in der Messbarkeit der tariflichen Leistungs- und Verhaltenskriterien: *„Viele Kriterien sind schlecht messbar.“* Die Experten sind sich mehrheitlich einig, dass die Leistungsbeurteilungen durch die subjektiven Wahrnehmungen der Vorgesetzten beeinflusst werden. Zum Beispiel sind Faktoren, die ungewollt eine Bedeutung haben: Sympathie oder Antipathie der Mitarbeiter. Auch der Einfluss einer subjektiven Wertung einzelner Kriterien ist im Rahmen des Leistungsbeurteilungsprozesses nicht zu vernachlässigen. Kriterien, wie beispielsweise der Arbeitseinsatz oder der Umgang mit Sorgfalt am Arbeitsplatz werden von Führungskraft zu Führungskraft anders interpretiert und dementsprechend individuell gewichtet.

Die Insider sind überwiegend der Meinung, dass die Objektivität der tarifvertraglichen Leistungsbeurteilung nicht durch zusätzliche Zielvereinbarungen zwischen Vorgesetzten

und deren Mitarbeitern beeinflusst wird. Zum Beispiel meinen die beiden Führungskräfte in den Bereichen Arbeitsschutz und Fertigung, dass die tarifliche Leistungsbeurteilung und die Zielvereinbarungsprozesse nicht in einem Zusammenhang stehen. Die Führungskraft in der Entwicklung hingegen kann eine mögliche Beziehung nicht ausschließen.

6.2 Deskriptive Statistiken

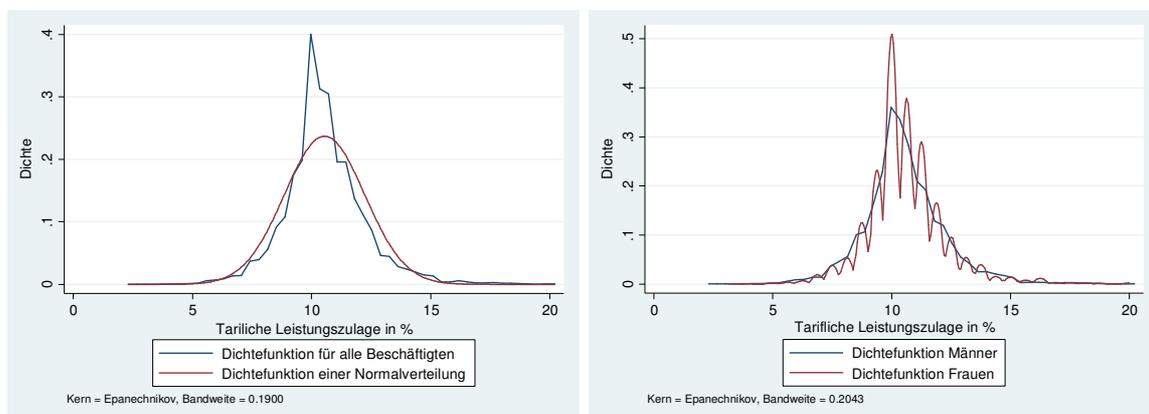
In den nachfolgenden Unterkapiteln werden zunächst die deskriptiven Statistiken zu der abhängigen, den unabhängigen und Kontrollvariablen beschrieben. Im Anschluss daran werden in Kapitel 6.3 die Ergebnisse der multivariaten Analysen beschrieben und interpretiert.

6.2.1 Abhängige Variable und unabhängige Variablen

Tarifliche Leistungszulage

Abbildung 7 zeigt die Wahrscheinlichkeitsverteilung der tariflichen Leistungszulagen in Prozent als abhängige Variable (linke Graphik) und entsprechende Verteilungen der Zulagen getrennt nach Geschlechtern (rechte Graphik).

Abbildung 7: Verteilungen der prozentualen tariflichen Leistungszulage



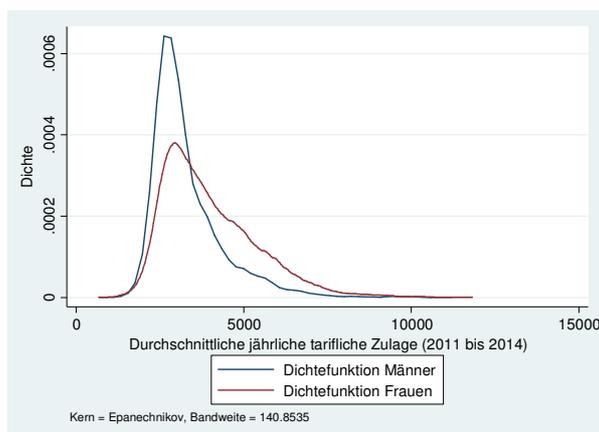
Erläuterungen: Kerndichteschätzungen basieren auf den tariflichen Zulagen in Prozent (links: gesamt; rechts: getrennt nach Geschlecht). Darstellungen wurden mit STATA 13 erstellt.

Im Datensatz liegen die Leistungszulagen zwischen 0% und 20% des tariflichen Bruttoeinkommens. Die tarifliche Verteilungsempfehlung, die in § 10 ERA Abs. 4 geregelt ist, gibt vor, dass die gezahlten Leistungszulagen im Unternehmen ca. 10% der tariflichen Monatsgrundentgeltsummen entspricht. Da es sich um ein relatives Beurteilungsverfahren handelt (Landesarbeitsgericht Baden-Württemberg, 2012), streut die abhängige Variable

um den Mittelwertwert 10%. Die Verläufe beider Dichtefunktionen lassen darauf schließen, dass die Ergebnisse der Leistungsbeurteilungen annähernd normalverteilt sind, jedoch eine starke Tendenz zur Mitte aufweisen. Dies könnte ein Anzeichen dafür sein, dass das Unternehmen zwar die tariflich vorgegebene Bandbreite an durchschnittlichen Beurteilungen anwendet, aber die durchschnittliche Vorgabe in Höhe von 10% zu einer deutlichen Mittelwerttendenz bei den Vorgesetztenbewertungen führt. Die stärkere Streuung der Dichtefunktion für Frauen (rechte Graphik) ist durch eine vergleichsweise geringe Anzahl an Beobachtungen mit Frauen zu erklären (Fälle in denen Frauen beurteilt werden $n=3.721$ und mit Männern $n=8.488$). Davon abgesehen zeigt sich, dass bei Betrachtung nach Geschlecht im Mittel für Frauen und Männer keine statistisch signifikanten Unterschiede bestehen. Frauen erhalten eine durchschnittliche Zulage in Höhe von 10,59% und Männer in Höhe von 10,52% ($t=-1,84$; $p>0,10$).

Abbildung 8 zeigt die Dichtefunktionen der tatsächlich gezahlten Leistungsentgelte an beide Geschlechtergruppen in Eurobeträgen. Der Abbildung ist zu entnehmen, dass die Leistungsentgelte – wie auch die prozentualen Zulagen – sich an einem Mittelwert orientieren, wobei Frauen vergleichsweise höhere Zulagen erhalten als ihre Kollegen. Im Durchschnitt betrachtet zeigen sich jedoch keine statistisch signifikanten Unterschiede in den entgeltlichen Leistungszulagen. Dies schließt jedoch nicht aus, dass Geschlechterunterschiede dennoch unter spezifischen Bedingungen auftreten können.

Abbildung 8: Tatsächliche tarifliche Leistungszulage nach Geschlecht



Erläuterungen: Kerndichteschätzungen basieren auf den jährlichen tariflichen Zulagen in Eurobeträgen (getrennt nach Geschlecht). Darstellung wurde mit STATA 13 erstellt.

Die oben dargelegten prozentualen und entgeltlichen Leistungszulagen basieren auf der Bewertung der vier tariflichen Kriterien. Als abhängige Variable wurde die prozentuale

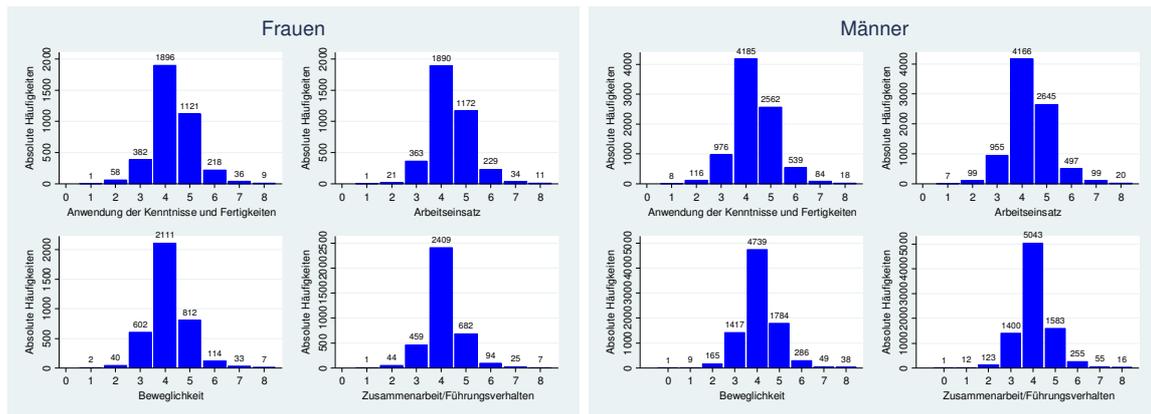
Leistungszulage, die sich aus der Bewertung der Kriterien ergibt, gewählt, da zum einen die theoretisch zu erwartenden Geschlechterunterschiede entgeltlich quantifiziert werden sollen, zum anderen haben Analysen bestätigt, dass die tariflichen Bewertungskriterien untereinander positiv korrelieren. Tabelle 11 zeigt die Korrelationen zwischen den tariflichen Kriterien. Die mittel- bis hohen Korrelationswerte ($r=0,49$ bis $r=0,60$) bestätigen, dass die Merkmalskriterien nicht einzeln als abhängige Variable berücksichtigt werden sollten. Die Auswahl der prozentualen Zulage als abhängige Variable, die sich aus den Bewertungskriterien ergibt, erscheint daher als sinnvoll.

Tabelle 11: Korrelationen zwischen tariflichen Beurteilungsmerkmalen

	Beurteilungsmerkmale	Mittelwert	σ	Min	Max	1	2	3	4
1	Anwendung der Kenntnisse und Fertigkeiten	4,32	0,86	1	8	1,00			
2	Arbeitseinsatz	4,34	0,85	1	8	0,57	1,00		
3	Beweglichkeit	4,11	0,83	0	8	0,50	0,60	1,00	
4	Zusammenarbeit/Führungsverhalten	4,09	0,77	0	8	0,49	0,54	0,55	1,00

Erläuterungen: Korrelation der tariflichen Beurteilungsmerkmale, die addiert die Grundlage für die Berechnungen der tariflichen Leistungszulagen bilden.

Auch hinsichtlich der Verteilungen der Merkmalskriterien zeigen sich keine wesentlichen Geschlechterunterschiede (siehe Abbildung 9). Die linke Graphik zeigt die Häufigkeitsverteilungen der tariflichen Kriterien für Frauen, die rechte Graphik die für Männer. Die Untersuchung der Geschlechterverteilungen für die einzelnen Merkmale bestätigt, dass die einzelnen tariflichen Beurteilungskriterien nicht geschlechterverzerrt sind. Von diesem Aspekt abgesehen ist auffällig, dass die volle Bandbreite an Bewertungspunkten nicht ausgeschöpft wird. Extremwerte, wie 0 Punkte oder 8 Punkte, werden von den Vorgesetzten kaum vergeben.

Abbildung 9: Punkteverteilungen der Beurteilungsmerkmale nach Geschlecht

Erläuterungen: Häufigkeitsverteilungen der tariflichen Beurteilungsmerkmale für weibliche Beschäftigte (links) und für männliche Beschäftigte (rechts). Darstellungen wurden mit STATA 13 erstellt.

Frauen erreichen bezüglich der Merkmale Arbeitseinsatz ($t=-1,96$; $p=0,05$) und Zusammenarbeit/Führungsverhalten ($t=-2,55$; $p=0,01$) im Mittel leicht bessere Noten als männliche Beschäftigte (siehe Tabelle 12). Allerdings sind diese Geschlechterdifferenzen aufgrund des geringen Ausmaßes zu vernachlässigen.

Tabelle 12: Mittelwerte der tariflichen Beurteilungsmerkmale nach Geschlecht

Merkmal	Mittelwert	σ	95%-Konfidenzintervall	t-Wert	Signifikanz
Anwendung der Kenntnisse und Fertigkeiten	Männer= 4,32	Männer=0,87	Männer=[4,03 – 4,34]	-0,04	0,97
	Frauen= 4,32	Frauen=0,85	Frauen= [4,26 – 4,35]		
Arbeitseinsatz	Männer=4,33	Männer=0,86	Männer=[4,32 – 4,35]	-1,96	0,05
	Frauen=4,37	Frauen=0,82	Frauen= [4,34 – 4,39]		
Beweglichkeit	Männer=4,10	Männer=0,85	Männer=[4,06 – 4,12]	-1,55	0,12
	Frauen=4,13	Frauen=0,80	Frauen= [4,10 – 4,15]		
Zusammenarbeit/ Führungsverhalten	Männer=4,07	Männer=0,79	Männer=[4,06 – 4,09]	-2,55	0,01
	Frauen=4,11	Frauen=0,73	Frauen= [4,09 – 4,14]		

Erläuterungen: Schätzungen wurden durchgeführt für $n=8.488$ Mitarbeiter-Jahres-Beobachtungen mit männlichen und $n=3.721$ weiblichen Beschäftigten. Zahl der Freiheitsgrade beträgt $n=12.207$.

Geschlechtervariablen

Dass das Fallunternehmen in einer männerdominierten Branche agiert, bestätigt sich in den Statistiken zu den Geschlechteranteilen des Fallunternehmens. Der Anteil an Mitarbeiterinnen liegt bei ca. einem Drittel. Unter den Führungskräften ist der Frauenanteil noch geringer. Der Datensatz enthält Mitarbeiterbeurteilungen von 40 weiblichen und 399 männlichen Vorgesetzten, wobei die weiblichen Führungskräfte 608 und die männlichen 11.601 Leistungsevaluierungen vergeben haben. Diese ungleichen Geschlechterverteilun-

gen haben Auswirkungen auf die Zusammensetzungen der Geschlechterdyaden, die neben dem Formalisierungsgrad der Leistungsbeurteilung eine weitere haupterklärende Variable der Fallstudie bilden.

Tabelle 13: Aufbau des Datensatzes nach Geschlecht

	Weiblich	Männlich	Σ
Anzahl Beschäftigte	1.086	2.406	3.492
Anzahl Beobachtungen mit Beschäftigten	3.721	8.488	12.209
Anzahl Vorgesetzte	40	399	439
Anzahl Beobachtungen mit Vorgesetzten	608	11.601	12.209

Erläuterungen: Aufgrund der nicht-balancierten Panelstruktur werden die Anzahl der Individuen sowie der Leistungsbewertungen über die Jahre nicht separat dargestellt.

Der vorliegende Datensatz enthält in Summe 12.209 individuelle Leistungsbewertungen, wobei in 67% der beobachteten Fälle männliche Vorgesetzte einen Mitarbeiter und in 28% der Fälle eine Mitarbeiterin evaluiert haben. Im Gegensatz dazu stehen die Beobachtungen mit weiblichen Führungskräften: diesbezüglich zeigt sich, dass in nur 3% der Fälle Männer und in 2% der Fälle Frauen beurteilt wurden. Der Einfluss der Variable Geschlechterdyaden sollte daher in den anschließenden quantitativen Analysen mit Vorsicht interpretiert werden. Tabelle 14 zeigt die Mittelwerte der prozentualen Leistungszulagen für die jeweiligen Geschlechterdyaden aus Vorgesetzten und Beschäftigten, deren Standardabweichungen sowie Häufigkeiten.

Tabelle 14: Statistiken zu Geschlechterdyaden

	Mitarbeiter	Mitarbeiterin
Männliche Führungskraft	Leistungszulage in %=10,54 (1,69) n=8.174 (66,95%)	Leistungszulage in %=10,59 (1,64) n=3.427 (28,07%)
Weibliche Führungskraft	Leistungszulage in %=10,24 (1,95) n=314 (2,57%)	Leistungszulage in %=10,49 (1,70) n=294 (2,41%)

Erläuterungen: Zellen enthalten die Mittelwerte der Leistungszulagen in Prozent (Standardabweichungen in Klammern) sowie die Beobachtungszahl n und relative Anteile der Dyaden (in Klammern).

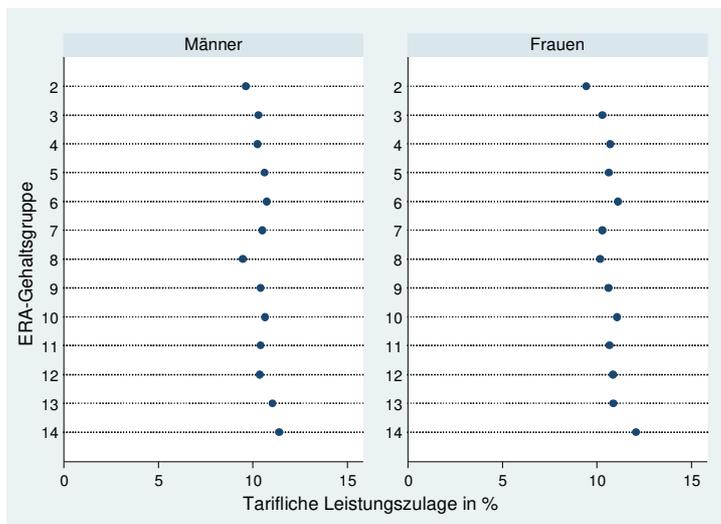
Die Mittelwerte der prozentualen Leistungszulagen für die jeweiligen Geschlechterdyaden aus Tabelle 14 bestätigen, dass diese nicht systematisch voneinander abweichen. Die aufgezeigten Unterschiede in den Mittelwerten sind statistisch nicht signifikant. Den deskriptiven Statistiken zufolge erhalten jedoch Frauen im Durchschnitt leicht höhere Leistungs-

zulagen als ihre Kollegen und weibliche Führungskräfte sind bei ihren Bewertungen augenscheinlich etwas strenger als männliche Vorgesetzte.

Organisationale Hierarchiestufe

In Bezug auf die organisationale Hierarchiestufe wurde argumentiert, dass mit steigender Job-Komplexität die Geschlechterdifferenzen in den Leistungszulagen zunehmen, da komplexere Tätigkeiten schwieriger zu beurteilen sind und daher Geschlechterstereotype eher eine Bedeutung haben als bei Bewertungen einfacherer Tätigkeiten. Abbildung 10 zeigt auf, dass deskriptiv keine systematischen Geschlechterunterschiede in den Gehaltsgruppen festzustellen sind. Auffällig ist jedoch, dass Frauen und Männer in der niedrigsten tariflichen Entgeltstufe (EG 2) die schlechtesten und in der höchsten Entgeltgruppe (EG 14) im Mittel die besten durchschnittlichen Leistungszulagen erhalten. Dies ist ein Indiz dafür, dass die Entgeltgruppe die Leistungsbeurteilung beeinflussen könnte und daher in den quantitativen Analysen als Kontrollvariable zu berücksichtigen ist.

Abbildung 10: ERA-Gehaltsgruppen und Leistungszulagen nach Geschlecht



Erläuterungen: Durchschnittliche Leistungszulagen in Prozent für Männer (links) und für Frauen (rechts) in den jeweiligen Gehaltsgruppen. Entgeltgruppen liegen im Fallunternehmen zwischen EG 2 und EG 14. Darstellungen wurden mit STATA 13 erstellt.

Im vorliegenden Datensatz sind Beschäftigte im Mittel in EG 9 und Führungskräfte in EG 13 eingestuft.

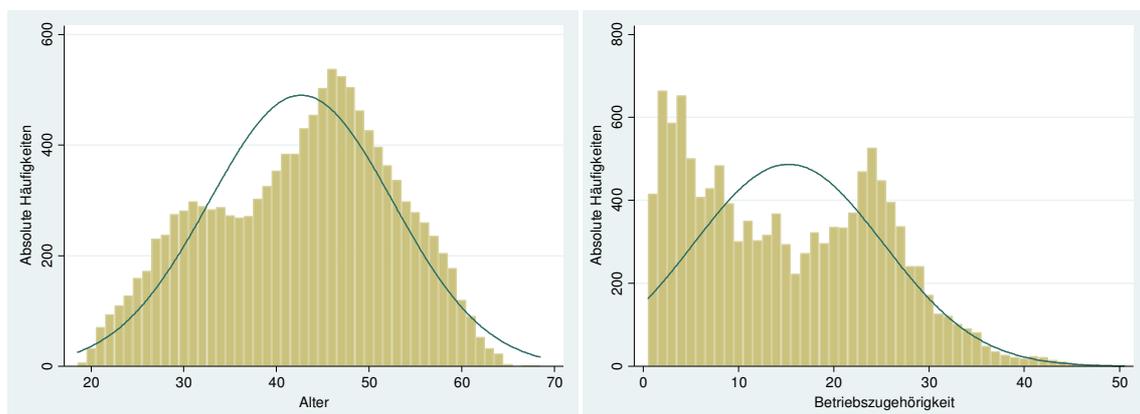
6.2.2 Kontrollvariablen

Alter und Betriebszugehörigkeit

In der Fallstudie beträgt das Durchschnittsalter der Beschäftigten ca. 43 Jahre. Die Führungskräfte sind im Mittel ca. ein Jahr älter. Bei der Betriebszugehörigkeit zeigt sich ein ähnliches Muster: Die Beschäftigten im Fallstudienunternehmen weisen im Vergleich zu den Vorgesetzten eine um ca. drei Jahre geringere Betriebszugehörigkeit auf. Vorgesetzte sind im Mittel ca. 18 Jahre im Unternehmen und Beschäftigte ca. 15 Jahre. Somit lässt sich festhalten, dass die Führungskräfte im Unternehmen im Vergleich zu den Beschäftigten die erfahrenere Personengruppe bilden.

An der Häufigkeitsverteilung in Abbildung 11 wird deutlich, dass sich die Verteilungen beider Variablen für Beschäftigte unterscheiden. Auch die Korrelation zwischen dem Alter und der Betriebszugehörigkeit ($r=0,56$) erlaubt die Berücksichtigung beider Variablen in den multivariaten Analysen. Nach Wirtz und Nachtigall (2012) ist die „... Güte einer Korrelation ... daran zu messen, welche ... Zusammenhänge in dem betreffenden Forschungsbereich gemessen wurden“ (Wirtz & Nachtigall, 2012: 107). Da die Studien von Abraham (2017) und Elvira und Graham (2002) einen ähnlichen Forschungskontext wie die vorliegende Arbeit haben und auch ähnliche Korrelationen zwischen den beiden genannten Faktoren messen ($r=0,55$ bzw. $r=0,61$), stellt sich die Berücksichtigung der beiden Variablen in den multivariaten Analysen der vorliegenden Studie als unproblematisch dar.

Abbildung 11: Verteilungen des Alters und der Betriebszugehörigkeit



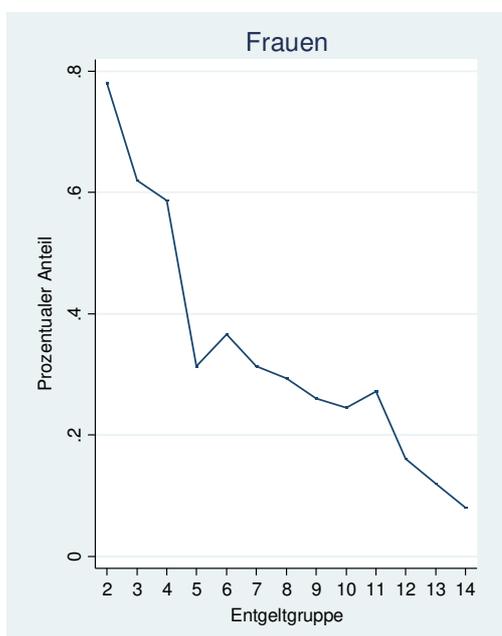
Erläuterungen: Darstellungen beziehen sich auf die Beschäftigten. Blaue Linien zeigen Normalverteilungen der Variablen. Darstellungen wurden mit STATA 13 erstellt.

Abteilungsmerkmale

Neben den Bewertungskulturen und den Geschlechteranteilen in den einzelnen Abteilungen, bildet der Geschlechteranteil in den jeweiligen Gehaltsgruppen eine weitere Kontrollvariable der vorliegenden Fallstudie. Da die spezifischen Abteilungsbezeichnungen aufgrund des betrieblichen Datenschutzes nicht genannt werden dürfen, wird im Folgenden lediglich auf die Geschlechteranteile in den Entgeltgruppen als Kontrollvariable näher eingegangen. Dummyvariablen für die einzelnen Abteilungen als auch die Geschlechteranteile in den einzelnen Abteilungen bleiben jedoch in der quantitativen Studie als Kontrollvariablen weiterhin berücksichtigt.

Die deskriptiven Statistiken zu den Anteilen von Frauen in den einzelnen Entgeltgruppen zeigt die Tendenz auf, dass Frauen insbesondere in den niedrigen Entgeltgruppen deutlich überrepräsentiert und in den höheren unterrepräsentiert sind. Beispielsweise beträgt der Anteil an Mitarbeiterinnen in EG 2 ca. 78% und in EG 14 nur ca. 8%. Diese Geschlechteranteile in den unterschiedlichen Gehaltsstufen werden in Abbildung 12 veranschaulicht.²²

Abbildung 12: Geschlechteranteile in den ERA-Gehaltsgruppen

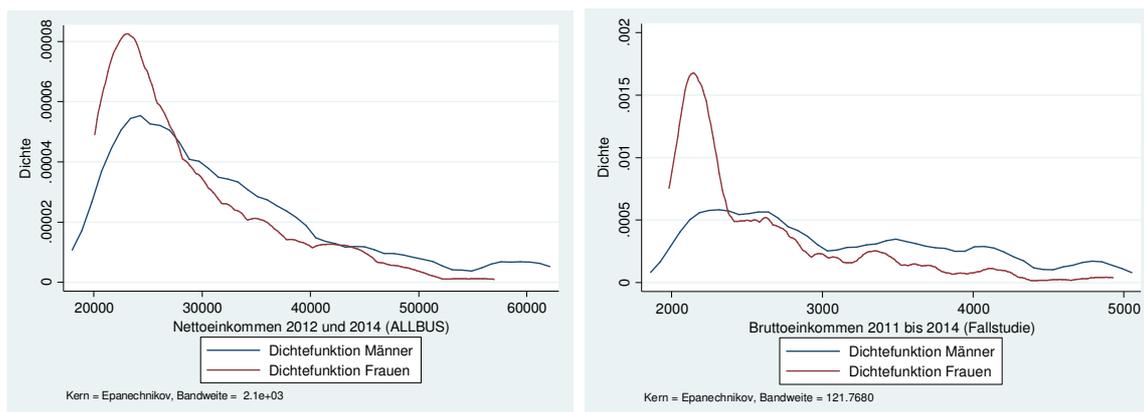


Erläuterungen: Abbildung zeigt den relativen Frauenanteil in den einzelnen Entgeltgruppen. Darstellung wurde mit STATA 13 erstellt.

²² Anhang 5 zeigt die absoluten Häufigkeiten von Frauen und Männern in den entsprechenden Entgeltgruppen.

Die oben angesprochene Ungleichverteilung der Geschlechter in den ERA-Entgeltgruppen wirkt sich auf die Einkommen der weiblichen und männlichen Beschäftigten aus. Analog zu der Analyse der Geschlechteranteile in den jeweiligen Hierarchiegruppen zeigt sich bei näherer Betrachtung der Einkommensverteilung im Unternehmen, dass Frauen häufiger ein geringeres Einkommen als Männer beziehen. Abbildung 13 stellt vergleichend die Einkommensverteilungen zwischen den Geschlechtern in der Gesamtbevölkerung in Deutschland mit der im Fallunternehmen gegenüber. Die linke Graphik stellt die deutsche Einkommensverteilung in Form von Kerndichteschätzern für beide Geschlechtergruppen dar und wurde auf Basis der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) für die Jahre 2012 und 2014 erstellt.²³ Die rechte Graphik hingegen zeigt die Einkommensverteilungen in den Jahren 2011 bis 2014 für Männer und Frauen im Fallbetrieb.

Abbildung 13: Einkommensverteilungen (ALLBUS und Fallstudie)



Erläuterungen: Kerndichteschätzungen basieren auf der repräsentativen Stichprobe der Nettojahreseinkommen aus dem ALLBUS für die Jahre 2012 und 2014 (linke Graphik) und auf der Stichprobe der Bruttojahreseinkommen aus der Fallstudie für die Jahre 2011 bis 2014 (rechte Graphik). Darstellungen wurden mit STATA 13 erstellt.

Obwohl die linke Graphik auf der Grundlage von Jahreseinkommen und die Graphik mit den Fallstudien Daten auf Basis von Monateinkommensdaten erstellt wurden, ist dies für den Vergleich von Verteilungen unproblematisch. Der Vergleich zeigt, dass die Einkommensdaten der Fallstudie vergleichbar sind mit dem deutschen Durchschnitt. Darüber hinaus verdeutlichen die Abbildungen, dass analog zu den Geschlechterverteilungen in den ERA-Gehaltsgruppen die Einkommensverteilungen ebenfalls Geschlechterunterschiede aufweisen: Männer beziehen im Fallbetrieb höhere Durchschnittseinkommen als Frauen.

²³ Die ALLBUS-Datensätze sind über die Internetseite www.gesis.org zugänglich.

Proxy für die Arbeitsleistung der Beschäftigten

Eine weitere Kontrollvariable in der Fallstudie bilden die Bewertungen des Arbeitsergebnisses im Rahmen einer formalisierten Zielvereinbarung oder einer vergleichsweise weniger formalisierten Beurteilung des mengenmäßigen Arbeitsergebnisses. In Bezug auf diese Faktoren wurde in Kapitel 5.3.1.2 argumentiert, dass die Beurteilungen des Arbeitsergebnisses als Anker wirken und die tarifliche Leistungsbeurteilung beeinflussen können. Im vorliegenden Fallstudien Datensatz zeigt sich, dass Beschäftigte in Bezug auf das Arbeitsergebnis vergleichsweise besser beurteilt werden, wenn sie mit ihren Vorgesetzten Ziele vereinbaren. Beschäftigte im Betrieb erhalten im Fall von Zielvereinbarungen im Mittel eine außertarifliche Leistungszulage in Höhe von 4,29% und im Fall des quantitativen Arbeitsergebnisses lediglich 3,46%. In Bezug auf diese Werte zeigen sich in den deskriptiven Analysen keine statistisch signifikanten geschlechterspezifischen Unterschiede.

Formalisierungsgrad

Laut Auskunft der Experten im Unternehmen können die Beschäftigten auf Basis einer geltenden Betriebsvereinbarung frei darüber entscheiden, ob das Arbeitsergebnis im Rahmen einer formalisierten Zielvereinbarung oder einer weniger formalisierten quantitativen Bewertung evaluiert wird. Deskriptive Statistiken zur Wahl dieser beiden Bewertungssysteme deuten darauf hin, dass von den 1.243 Frauen im Fallunternehmen ca. 37% eine Zielvereinbarung bekommen und 63% keine Ziele mit ihren Vorgesetzten vereinbaren. Bei den Männern (n=2.934) hingegen erhält ca. die Hälfte eine Zielvereinbarung (49%) und die übrige Hälfte keine (51%). Tabelle 15 stellt diese Verteilung gegenüber.

Tabelle 15: Statistiken zur Wahl des Beurteilungssystems

	Männliche Beschäftigte	Weibliche Beschäftigte	Σ
Arbeitsquantität (weniger formalisiert)	n=1.508	n=783	n=2.291
	51,40%	62,99%	54,85%
Zielvereinbarung (formalisiert)	n=1.426	n=460	n=1.886
	48,60%	37,01%	45,15%
Σ	n=2.934	n=1.243	n=4.177
	100%	100%	100%

Erläuterungen: Der Formalisierungsgrad ist in Verbindung mit der tariflichen Leistungszulage zu verstehen. Die Beobachtungszahlen beziehen sich auf sämtliche Fälle, in denen zwischen den beiden Beurteilungssystemen gewählt wurde.

Die Ungleichverteilung von Frauen in den beiden Beurteilungssystemen ist laut den Experten im Betrieb dadurch bedingt, dass Frauen eher in den unteren Entgeltstufen beschäftigt sind, Ziele jedoch tendenziell in den oberen Entgeltgruppen vereinbart werden. Die zuvor beschriebenen Verteilungen der Geschlechter über die Entgeltgruppen (siehe Abbildung 12) und die ungleichen Einkommensverteilung von Frauen und Männern im Betrieb (siehe Abbildung 13 rechte Graphik) bestätigen diese Vermutung und somit auch die Relevanz des Insiderwissens der Experten im Betrieb. In der vorliegenden Untersuchung wird der Faktor Formalisierungsgrad des Leistungsbeurteilungssystems dadurch kontrolliert, dass für beide Formalisierungsarten getrennte Schätzungen durchgeführt werden.

6.3 Multivariate Analysen

Nachdem in den vorangegangenen Kapiteln 6.2.1 und 6.2.2 auf die deskriptiven Statistiken der abhängigen, unabhängigen und Kontrollvariablen eingegangen wurde, sollen in diesem Kapitel die multivariaten Analysen der quantitativen Studie vorgestellt und interpretiert werden. Hierfür werden zunächst die in Kapitel 3.3 aufgestellten Hypothesen über Geschlechtereffekte hinsichtlich der Formalisierung von Leistungsbeurteilungssystemen und den organisationalen Mechanismen getestet. Im Anschluss daran werden weitere relevante Ergebnisse der vorliegenden Fallstudie beschrieben und interpretiert. Für die Interpretation der Ergebnisse werden auch die Aussagen aus den Experteninterviews herangezogen.²⁴

6.3.1 Hypothesentests

Ergebnisse zu Hypothese 1

Die Ergebnisse der Panelregressionen, die als Grundlage zur Überprüfung der ersten Hypothesen dienen, sind in Tabelle 16 gegenübergestellt. Die Schätzungen deuten darauf hin, dass in dem vergleichsweise formalisierten Beurteilungssystem mit einer Zielvereinbarung (Modell (1)) der Geschlechternachteil für Frauen kleiner ist als in dem weniger formalisierten System mit einer Bewertung der Arbeitsquantität (Modell (2)). Im formalisierten Leistungsbeurteilungssystem erhalten Frauen sogar höhere tarifliche Leistungszulagen als vergleichbare Männer ($\beta=0,18$; $p<0,05$). Im weniger formalisierten Bewertungssystem

²⁴ Die Interviewpartner wurden im Rahmen der Gespräche darüber informiert, dass die Effekte unter statistisch kontrollierten Bedingungen geschätzt wurden. Da es während der Gespräche mehrere Rückfragen zu den geschätzten Effekten gab, kann nicht ganz ausgeschlossen werden, dass die Ergebnisse der multivariaten Analyse teilweise missverstanden wurden.

hingegen erhalten Frauen *ceteris paribus* schlechtere Leistungsbeurteilungen ($\beta=-0,19$; $p<0,01$). Beide angesprochenen Effekte sind statistisch signifikant und haben eine vergleichbare Effektgröße in Höhe von ca. 0,2%-Punkten der gezahlten Leistungszulagen. Da das Grundentgelt von Frauen im Fallunternehmen bei durchschnittlich 31.032€ (brutto) liegt, entspricht dieser geschlechterspezifische Effekt umgerechnet ca. 62€ pro Jahr. Diese Ergebnisse aus beiden Regressionsmodellen zeigen, dass der Formalisierungsgrad die Wahrnehmungen der Führungskräfte beeinflusst und die Leistungszulagen von Frauen und Männern verzerrt.

Tabelle 16: Panelschätzungen zu Geschlechtereffekten in Leistungszulagen (H1)

Erklärende Variablen	Abhängige Variable: Tarifliche Leistungszulage in %	
	(1) Formalisiert	(2) Weniger formalisiert
<i>Geschlecht</i>		
Mitarbeiterin	0,184** (0,079)	-0,191*** (0,062)
Weibliche Führungskraft	-0,088 (0,128)	-0,200 (0,165)
<i>Organisationale Hierarchiestufe</i>		
Entgeltgruppe Mitarbeiter/in	0,080*** (0,018)	0,037*** (0,012)
Entgeltgruppe Führungskraft	0,100*** (0,026)	-0,006 (0,030)
<i>Merkmale Mitarbeiter/in</i>		
Alter	1,338*** (0,329)	2,000*** (0,182)
Alter*Alter	-0,146*** (0,041)	-0,212*** (0,021)
Betriebszugehörigkeit	1,039*** (0,128)	1,021*** (0,088)
Betriebszugehörigkeit*Betriebszugehörigkeit	-0,240*** (0,038)	-0,249*** (0,026)
Zulage Beurteilung Zielvereinbarung	0,072 (0,044)	./. ./.
Zulage Beurteilung Quantität	./. ./.	0,486*** (0,046)
<i>Merkmale Führungskraft</i>		
Alter	-0,412 (0,459)	0,814** (0,337)
Alter*Alter	0,051 (0,052)	-0,085** (0,040)
Betriebszugehörigkeit	-0,266** (0,130)	-0,178* (0,092)
Betriebszugehörigkeit*Betriebszugehörigkeit	0,083** (0,037)	0,077*** (0,026)
<i>Merkmale Abteilung</i>		
Frauenanteil	-0,308 (0,550)	0,260 (1,194)
Frauenanteil in EG 1 bis EG 6	-0,787** (0,337)	-2,816*** (0,613)
Frauenanteil in EG 12 bis EG 14	1,404** (0,690)	5,611*** (1,327)
Abteilungsdummies	Berücksichtigt	Berücksichtigt
Konstante	5,045*** (1,178)	0,952 (0,852)
N	5.429	6.646
Anzahl an Beschäftigter-Vorgesetzter-Dyaden	2.572	3.259
R ² (Individuen)	0,034	0,120
R ² (Dyaden)	0,177	0,248
R ² (Insgesamt)	0,172	0,244

Erläuterungen: Modell (1) stellt das formalisierte und Modell (2) das weniger formalisierte Beurteilungssystem dar. Ergebnisse basieren auf Random-Effects-Schätzungen. Robuste Standardfehler in Klammern. Signifikanzniveaus: *** p<0,01, ** p<0,05, * p<0,1. Die Alters- und Betriebszugehörigkeitsvariablen (Beschäftigte sowie Führungskräfte) wurden aus Interpretationsgründen mit dem Faktor 10 multipliziert. Männer und der Frauenanteil in EG 7 bis EG 11 bilden die entsprechenden Referenzkategorien. Die deskriptiven Statistiken und Korrelationen der im Modell berücksichtigten Variablen sind in Anhang 6 und Anhang 7 enthalten.

Aufgrund der dargelegten Ergebnisse kann Hypothese 1 nicht abgelehnt werden („Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem formalisierten Bewertungssystem kleiner als in einem weniger formalisierten Bewertungssystem“). Die

Ergebnisse bestätigen vielmehr, dass der Unterschied in der Leistungszulage für Frauen in formalisierten Leistungsbeurteilungssystemen nicht nur geringer ist, sondern Frauen im System mit einer Zielvereinbarung tendenziell sogar höhere tarifliche Leistungszulagen erhalten als ihre Kollegen.

Die Experten im Unternehmen erörtern diese Ergebnisse folgendermaßen. Drei der befragten Experten bringen das Ergebnis mit der Branche, den Gegebenheiten des Unternehmens und der beruflichen Position in Verbindung. Beispielsweise ist die Führungskraft in der Produktion der Ansicht, dass „... *Frauen in Führungspositionen oder in einer männerdominierten Welt, wie es in manchen Bereichen ist, ... einen anderen Ehrgeiz entwickelt haben. (...) in Ingenieurberufen, wo ja der Frauenanteil gefühlt kleiner ist, glaube ich ... schon, ... dass dort Frauen andere Leistung bringen, die häufig deutlich besser ist*“. Die männliche Führungskraft aus dem Personalwesen findet eine ähnliche Erklärung und sieht ebenfalls einen potentiellen Grund in der Struktur der Belegschaft. Im Unternehmen seien deutlich mehr Männer in Führungspositionen und dementsprechend werden in den meisten Fällen weibliche Beschäftigte von männlichen Vorgesetzten bewertet. Für ihn sei vorstellbar, dass „... *die ein oder andere männliche Führungskraft einen anderen Maßstab ansetzt bei einer beurteilten Frau als bei einem beurteilten Mann*“. Die Personalreferentin spricht die technische Orientierung des Unternehmens an und kann sich vorstellen, „... *dass man vielleicht engagierte Frauen in einem technischen Unternehmen mehr wertschätzt als Männer*“. Die weibliche Führungskraft im Personalwesen hingegen hat für den vorliegenden Befund einen weiteren Erklärungsansatz. Für sie gestalten sich die Beurteilungsgespräche mit Mitarbeiterinnen schwieriger, da Frauen emotionaler reagieren als Männer. Die Folge hiervon sei eine mögliche geschlechterspezifische Tendenz zur Milde.

Die Antworten der Experten bestätigen, dass hinsichtlich des Verfahrens der tariflichen Leistungsbeurteilung geschlechterspezifische Differenzen existieren. Allerdings geht aus den Antworten nicht hervor, in welcher Weise der Formalisierungsgrad des Beurteilungssystems Unterschiede in den Leistungsbeurteilungen von Frauen und Männern hervorruft.

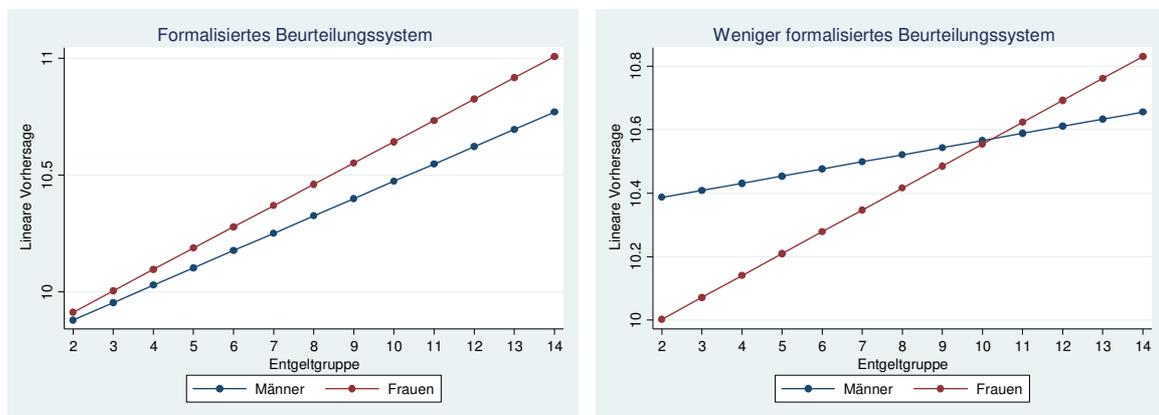
Ergebnisse zu Hypothese 2

Hinsichtlich Hypothese 2 wurde argumentiert, dass der Geschlechterunterschied für Frauen in weniger formalisierten Leistungsbeurteilungssystemen durch die organisationale Hierarchiestufe beeinflusst wird (*„Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem weniger formalisierten Bewertungssystem auf höheren organisationa-*

len Hierarchiestufen kleiner als auf niedrigeren“). Um diese Annahme zu überprüfen, wurden auf Basis der obigen Panelschätzungen (siehe Tabelle 16) weitere Analysen mit einer zusätzlichen Interaktionsvariable durchgeführt, die das Beschäftigtengeschlecht mit der organisationalen Hierarchiestufe interagiert.

Die marginalen Effekte der Entgeltgruppe für Frauen und Männer sind in Abbildung 14 illustriert. Zunächst ist auffällig, dass generell in beiden Beurteilungssystemen Beschäftigte auf niedrigeren Hierarchiestufen schlechtere Bewertungen erhalten als Beschäftigte in höheren Entgeltgruppen. Die linke Graphik verdeutlicht, dass in dem formalisierten Beurteilungssystem Frauen tendenziell auf allen Hierarchiestufen bessere Leistungsbeurteilungen erhalten. Der Geschlechterunterschied ist jedoch nur in den höheren Entgeltgruppen (EG 10 bis EG 14) statistisch signifikant. Darüber hinaus verdeutlicht die rechte Graphik, dass die geschätzten Werte der Leistungsbeurteilungen für Frauen in dem weniger formalisierten System auf den niedrigeren organisationalen Hierarchiestufen unter denen der Männer liegen, und dass sich dieser Effekt in den oberen Stufen umkehrt.

Abbildung 14: Interaktion zwischen Geschlecht und organisationaler Hierarchiestufe



Erläuterungen: Darstellungen enthalten die durchschnittlich geschätzten Werte der tariflichen Leistungsbeurteilung in Prozent für Frauen und Männer. Darstellungen wurden mit STATA 13 erstellt.

Die Ergebnisse des Interaktionseffekts in Tabelle 17 unterstützen die Hypothese, dass Frauen in dem weniger formalisierten Beurteilungssystem insbesondere in niedrigen Entgeltgruppen schlechtere Bewertungen erhalten. Dieser Geschlechterunterschied ist für die EG 2 bis EG 7 negativ und statistisch signifikant. Beispielsweise erhalten Frauen in EG 2 im weniger formalisierten Beurteilungssystem im Vergleich zu Männern eine um ca. 0,4%-Punkte geringere tarifliche Leistungszulage. Die entsprechenden Konfidenzintervalle (EG 2 bis EG 7) enthalten ausschließlich negative Werte. Demzufolge tragen eine nied-

rige organisationale Stellung und ein weniger formalisiertes Beurteilungssystem zu einer größeren Geschlechterungleichheit bei.

Tabelle 17: Marginale Effekte im weniger formalisierten Beurteilungssystem (H2)

EG	Geschlechterunterschied	Standardfehler	z	P> z	95%-Konfidenzintervall
2	-0,385	0,089	-4,33	0,00	-0,56 bis -0,21
3	-0,338	0,075	-4,53	0,00	-0,48 bis -0,19
4	-0,291	0,064	-4,53	0,00	-0,42 bis -0,17
5	-0,245	0,060	-4,11	0,00	-0,36 bis -0,13
6	-0,198	0,062	-3,20	0,00	-0,32 bis -0,08
7	-0,152	0,071	-2,15	0,03	-0,29 bis -0,01
8	-0,105	0,084	-1,26	0,21	-0,27 bis 0,06
9	-0,058	0,099	-0,59	0,56	-0,25 bis 0,14
10	-0,012	0,117	-0,10	0,92	-0,24 bis 0,22
11	0,035	0,135	0,26	0,80	-0,23 bis 0,30
12	0,081	0,154	0,53	0,60	-0,22 bis 0,38
13	0,128	0,173	0,74	0,46	-0,21 bis 0,47
14	0,175	0,193	0,91	0,37	-0,20 bis 0,55

Erläuterungen: Schätzungen basieren auf 6.646 Beobachtungen. Der Geschlechterunterschied spiegelt die durchschnittlichen marginalen Effekte wider. Es werden robuste Standardfehler ausgegeben (Spalte 3).

Ergebnisse zu Hypothese 3

In Bezug auf Hypothese 3 wurde hergeleitet, dass bei einem niedrigen Formalisierungsgrad Geschlechterkombinationen zwischen Beschäftigten und Vorgesetzten unterschiedliche Auswirkungen auf die Leistungsbeurteilungen haben können („*Der Geschlechternachteil in Leistungsbeurteilungen ist für Frauen in einem weniger formalisierten Bewertungssystem vergleichsweise kleiner, wenn sie von weiblichen und nicht von männlichen Vorgesetzten beurteilt werden*“). Die Panelmodelle (1) und (2) in Tabelle 18 ermöglichen die Überprüfung der Hypothese 3. Die geschätzten Koeffizienten der Geschlechterdyaden deuten darauf hin, dass der Geschlechterunterschied für Frauen in einem weniger formalisierten System (Modell (2)) nicht automatisch geringer ist, wenn sie von weiblichen Führungskräften evaluiert werden. Der Koeffizient der Variable *Mann beurteilt Frau* ($\beta=-0,19$ $p<0,01$) in dem Panelmodell (2) ist zwar in der Tat negativ, jedoch hat auch der Koeffizient der Variable *Frau beurteilt Frau* ($\beta=-0,42$; $p>0,10$) ein negatives Vorzeichen. Auffällig ist, dass der geschätzte Effekt der letzteren Kombination deutlich größer ist, jedoch keine statistische Signifikanz aufweist. In beiden Modellen sind die geschätzten Koeffizienten der Geschlechterdyaden mit männlichen Führungskräften, die Frauen beurteilen,

statistisch signifikant. In diesen Fällen zeigt sich, dass Frauen vergleichsweise besser als ihre Kollegen beurteilt werden, wenn sie im Rahmen des formalisierten Bewertungssystems evaluiert werden, und schlechter, falls ihre Bewertungen in einem weniger formalisierten Beurteilungssystem durchgeführt werden. Die Ergebnisse machen zudem deutlich, dass Frauen die tendenziell strengeren Führungskräfte sind und ihre Beschäftigten kritischer beurteilen. Dies drückt sich durch die negativen Vorzeichen der Geschlechterdyaden mit weiblichen Vorgesetzten aus.

Tabelle 18: Panelschätzungen zu Geschlechtereffekten in Leistungszulagen (H3)

Erklärende Variablen	Abhängige Variable:	
	(1) Formalisiert	(2) Weniger formalisiert
<i>Geschlecht</i>		
Frau beurteilt Frau	0,130 (0,169)	-0,423 (0,316)
Frau beurteilt Mann	-0,123 (0,166)	-0,180 (0,161)
Mann beurteilt Frau	0,176** (0,084)	-0,190*** (0,063)
<i>Organisationale Hierarchiestufe</i>		
Entgeltgruppe Mitarbeiter/in	0,080*** (0,018)	0,037*** (0,012)
Entgeltgruppe Führungskraft	0,101*** (0,026)	-0,007 (0,030)
<i>Merkmale Mitarbeiter/in</i>		
Alter	1,340*** (0,329)	2,001*** (0,183)
Alter*Alter	-0,146*** (0,041)	-0,212*** (0,021)
Betriebszugehörigkeit	1,039*** (0,128)	1,021*** (0,088)
Betriebszugehörigkeit*Betriebszugehörigkeit	-0,240*** (0,038)	-0,249*** (0,026)
Zulage Beurteilung Zielvereinbarung	0,072 (0,044)	./. ./.
Zulage Beurteilung Quantität	./. ./.	0,486*** (0,046)
<i>Merkmale Führungskraft</i>		
Alter	-0,418 (0,460)	0,813** (0,337)
Alter*Alter	0,052 (0,052)	-0,085** (0,040)
Betriebszugehörigkeit	-0,267** (0,130)	-0,177* (0,092)
Betriebszugehörigkeit*Betriebszugehörigkeit	0,083** (0,037)	0,077*** (0,026)
<i>Merkmale Abteilung</i>		
Frauenanteil	-0,311 (0,550)	0,261 (1,194)
Frauenanteil in EG 1 bis EG 6	-0,788** (0,337)	-2,814*** (0,612)
Frauenanteil in EG 12 bis EG 14	1,407** (0,690)	5,606*** (1,326)
Abteilungsdummies	Berücksichtigt	Berücksichtigt
Konstante	5,052*** (1,178)	0,949 (0,852)
N	5.429	6.646
Anzahl an Beschäftigter-Vorgesetzter-Dyaden	2.572	3.259
R ² (Individuen)	0,034	0,121
R ² (Dyaden)	0,177	0,248
R ² (Insgesamt)	0,172	0,244

Erläuterungen: Modell (1) stellt das formalisierte und Modell (2) das weniger formalisierte Beurteilungssystem dar. Ergebnisse basieren auf Random-Effects-Schätzungen. Robuste Standardfehler in Klammern. Signifikanzniveaus: *** p<0,01, ** p<0,05, * p<0,1. Die Alters- und Betriebszugehörigkeitsvariablen (Beschäftigte sowie Führungskräfte) wurden aus Interpretationsgründen mit dem Faktor 10 multipliziert. Die Geschlechterkombination Mann beurteilt Mann sowie der Frauenanteil in EG 7 bis EG 11 bilden die entsprechenden Referenzkategorien.

Die Experteninterviews greifen auch das Ergebnis auf, dass weibliche Beschäftigte vergleichsweise bessere tarifliche Leistungsbeurteilungen erhalten, wenn sie mit ihren Vorge-

setzten Ziele vereinbaren und die beurteilenden Vorgesetzten männlich sind. Die Führungskraft in der Entwicklung erklärt diesen Befund mit den folgenden Worten: *„Also denkbar wäre an der Stelle für mich, dass wenn die Zielvereinbarung erfüllt wird, dann dem Mann klarer und präsenter ist, ... dass positive Arbeitsergebnisse vorhanden sind, die sonst vielleicht eher geringschätziger gewertet werden“*. Die Führungskraft in der Fertigung hingegen weist im Rahmen der ERA-Beurteilungen auf eine mögliche Bevorzugung der männlichen Belegschaft hin und meint, dass Mitarbeiterinnen tendenziell eher Ziele vereinbaren, *„ ... weil sie vielleicht in der Vergangenheit das Gefühl hatten, dass die männlichen Kollegen bevorzugt werden. Und das ist dann natürlich auch ein Hilfsmittel ... für die männlichen Vorgesetzten, da ganz klar zu sagen: ‚Du hast da die und die Leistung erbracht, also sehe ich dich auch da und da‘“*. Interessant ist auch der Aspekt, dass für zwei der befragten Experten die angesprochenen Geschlechterunterschiede keine Überraschung darstellen.

Außerdem wurden die Experten mit dem Ergebnis konfrontiert, dass die weiblichen Führungskräfte im Unternehmen vergleichsweise strenger beurteilen als die männlichen Vorgesetzten. Die weibliche Führungskraft im Personalwesen ist der Meinung, dass Frauen es insbesondere in technischen Berufen schwer haben, eine höhere Position in der Hierarchie des Unternehmens zu erreichen. Die Frauen, die es letztlich auf höhere Positionen schaffen, hätten vergleichsweise eine höhere Erwartungshaltung gegenüber ihren Beschäftigten. Die Führungskraft im Arbeitsschutz hat eine ähnliche Erklärung und meint, dass nicht viele Frauen im Unternehmen Leistungsbeurteilungen durchführen. Zudem sei der Karriereweg für weibliche Führungskräfte ein anderer als für männliche und dies führe dazu, dass sich Frauen in Führungspositionen an männliche Strukturen anpassen: *„Ja, dieses Typische, was man ja Frauen in Führungspositionen nachsagt, diese Anpassung an männliche Strukturen: Ich bin hart. Ich weiß nicht, ob das eine Rolle spielt, aber dass solche Dinge vielleicht da im Unterbewusstsein eine Rolle spielen.“* Weitere Erklärungsansätze für den angesprochenen Effekt sind: Frauen könnten einen höheren Anspruch haben; sie beurteilen möglicherweise sachlicher oder differenzierter; sie sind eher dazu bereit, individuelle Schwächen der Beschäftigten anzusprechen.

6.3.2 Weitere Einflussgrößen

Neben den Resultaten der Hypothesentests, die im vorangegangenen Kapitel 6.3.1 erörtert wurden, enthalten die Panelschätzungen aus Tabelle 16 und Tabelle 18 weitere Ergebnis-

se, auf die im Folgenden näher eingegangen wird. Diese Ergebnisse beziehen sich auf die drei Merkmalskategorien: organisationale Hierarchiestufe, Eigenschaften der Beschäftigten und Vorgesetzten sowie die Gegebenheiten in den Abteilungen des Fallstudienunternehmens.

Organisationale Hierarchiestufe

Die geschätzten Koeffizienten der Variable *Entgeltgruppe Mitarbeiter/in* bestätigen einen statistisch signifikanten Hierarchieeffekt, der sich in den Leistungsbeurteilungen der Beschäftigten widerspiegelt. Die positiven Vorzeichen der Koeffizienten der Entgeltgruppenvariable bestätigen, dass sich eine höhere organisationale Hierarchiestufe positiv auf die Beurteilungen von Beschäftigten auswirkt. Der geschätzte Koeffizient der Variable *Entgeltgruppe Mitarbeiter/in* in Modell (1) der Tabelle 16 in Höhe von 0,08 sagt beispielweise aus, dass ein Anstieg der Entgeltgruppe um eine Stufe die tarifliche Leistungszulage im Durchschnitt um ca. 0,1%-Punkte erhöht. Die relativ geringe Effektgröße ist jedoch nicht zu vernachlässigen, da Beschäftigte in ihrer beruflichen Laufbahn die Karriereleiter theoretisch um mehr als eine Hierarchiestufe hinaufsteigen können. Im Gegensatz zu dem Effekt der organisationalen Hierarchiestufe der Beschäftigten lässt sich der Einfluss der organisationalen Hierarchiestufe der Führungskraft auf die tarifliche Leistungszulage der Beschäftigten nicht eindeutig interpretieren, da die geschätzten Koeffizienten der Variable *Entgeltgruppe Führungskraft* nicht in allen Schätzmodellen positiv und statistisch signifikant sind. Jedoch bestätigen die statistisch signifikanten Effekte im formalisierten System, dass es notwendig ist, die erwähnte Variable als Kontrolle in den Schätzungen zu berücksichtigen.

Das im vorangegangenen Absatz beschriebene Phänomen, dass mit steigenden Gehaltsstufen der Beschäftigten die ERA-Bewertungen positiver ausfallen, war ebenfalls Bestandteil der Experteninterviews. Dieser Effekt wird in der Literatur über Leistungsbeurteilungen auch als Hierarchieeffekt bezeichnet. Der Führungskraft im Personalwesen ist der Effekt bekannt, da dieser auch in den Schulungen für Vorgesetzte thematisiert wird. Ihm zufolge zeigt sich unter den Vorgesetzten die Tendenz „... *die vermeintlich einfacheren Tätigkeiten unterdurchschnittlich zu bewerten und die höherwertigen ... überdurchschnittlich*“. Die Personalreferentin führt ein ähnliches Argument an, das den Hierarchieeffekt in den ERA-Beurteilungen erklären könnte: „... *wenn einer auf einem gewissen Entgeltniveau ist, geht man einfach davon aus, dass der auch in der Leistung gut sein muss, was ja de*

facto Blödsinn ist“. Die Führungskraft in der Entwicklung weist auf den Aspekt hin, dass vergleichsweise wenige Beschäftigte in den höheren Entgeltgruppen existieren. Demzufolge könne er es sich nicht leisten, „ ... *Konflikte aufgrund dieser Leistungsbeurteilung zu erzeugen ...* “. Er kann sich jedoch vorstellen, „ ... *dass man ... wesentlich vorsichtiger wird und das lieber positiv hält, als wirklich mal negative Rückkopplung zu geben*“. Die weibliche Führungskraft im Personalwesen hingegen spricht auch die Konsequenzen an, die sich insbesondere für die Beschäftigten auf den unteren Hierarchiestufen ergeben: „*Ich glaube, da kommt schon ein subjektiver Einfluss rein, dass man quasi mehr von jemandem erwartet, weil die Aufgaben eben ... teilweise sehr einfach sind und das [dies] dann zu einer Abwertung führt. (...) Es ist halt insofern schwierig, weil sich dieser Unterschied eben in der Entgeltgruppe ... und sich nicht in der Leistungsbeurteilung manifestieren sollte*“. Auch die Führungskraft im Compliance fände es schlecht, wenn die Tarifstufe der Beschäftigten an das Ergebnis der Leistungsbeurteilung gekoppelt wäre, „ ... *weil es natürlich auch niedrigtarifizierte Mitarbeiter gibt, die eine sehr gute Leistung abliefern. (...) Es sollte auch in allen Tarifgruppen möglich sein, eine entsprechende Leistungsbewertung zu bekommen, wenn man die [entsprechende] Leistung abliefern*“.

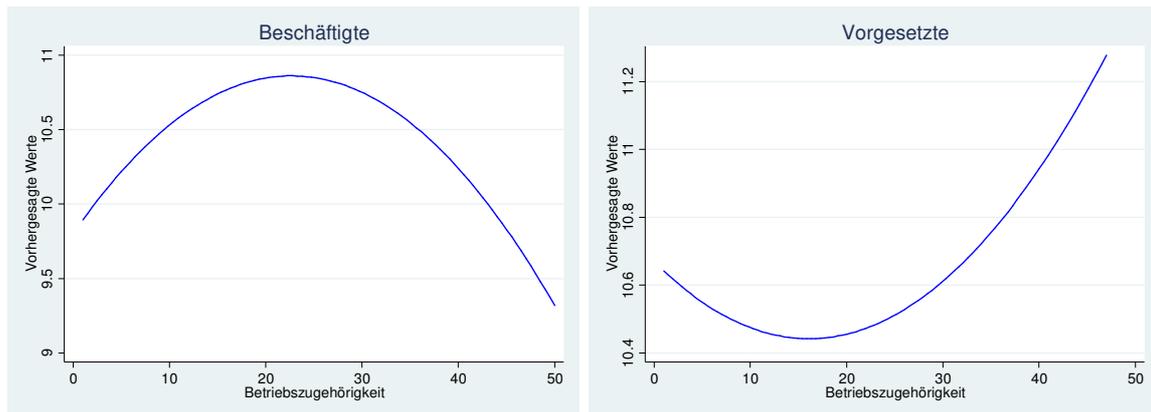
Merkmale der Beschäftigten und Führungskräfte

Die geschätzten Koeffizienten der Alters- und Betriebszugehörigkeitsvariablen der Beschäftigten stellen ebenfalls interessante Ergebnisse der quantitativen Fallstudie dar. Da sowohl die Richtungen, die Effektstärken als auch die vermuteten Wirkungszusammenhänge der genannten Variablen (siehe Kapitel 5.3.1.2) auf die Leistungsbeurteilungen der Beschäftigten ähnlich sind, sollen die Ergebnisse der geschätzten Koeffizienten gemeinsam beschrieben und interpretiert werden.

Die relativ hohen Koeffizienten der Alters- und Betriebszugehörigkeitsvariablen sowie deren quadrierte Terme zeigen auf, dass diese Variablen einen wichtigen Beitrag zur Erklärung der tariflichen Leistungszulagen beitragen. Die geschätzten Koeffizienten der Altersvariablen in Tabelle 16 und Tabelle 18 sagen beispielsweise voraus, dass Beschäftigte *ceteris paribus* ca. 1%- bis 2%-Punkte höhere Leistungszulagen erhalten, wenn sie im Durchschnitt zehn Jahre älter sind. Der Einfluss der Betriebszugehörigkeitsdauer auf die Leistungsbeurteilungen der Beschäftigten ist ähnlich zu interpretieren, wobei die Effektgröße der Variable bei ca. einem Prozentpunkt liegt. Im Gegensatz zu den Koeffizienten der Alters- und Betriebszugehörigkeit haben die geschätzten Koeffizienten der quadrierten

Terme ein negatives Vorzeichen. Dies bedeutet, dass der Zusammenhang zwischen dem Alter bzw. der Betriebszugehörigkeit und der tariflichen Leistungszulage nicht linear, sondern wie vermutet umgekehrt U-förmig ist. Beispielsweise bedeutet dies, dass Beschäftigte mit steigender Betriebszugehörigkeit zunächst höhere Leistungsbewertungen erhalten, sich der Effekt aber nach einer gewissen Betriebszugehörigkeitsdauer umkehrt.

Abbildung 15: Betriebszugehörigkeit und tarifliche Leistungszulage



Erläuterungen: Darstellungen zeigen den nicht-linearen Zusammenhang zwischen der Betriebszugehörigkeit und der tariflichen Leistungszulage für Beschäftigte (links) sowie Vorgesetzte (rechts) in beiden Beurteilungssystemen. Darstellungen wurden mit STATA 13 erstellt.

Abbildung 15 verdeutlicht graphisch die geschätzten Werte für den Zusammenhang zwischen der Betriebszugehörigkeitsdauer (Beschäftigte und Vorgesetzte) und der tariflichen Leistungszulage. Der Verlauf der Kurve in der rechten Graphik verdeutlicht, dass Führungskräfte insbesondere in späteren Jahren ihrer Betriebszugehörigkeit milder beurteilen als zu Beginn ihrer Tätigkeit im Unternehmen. Mit anderen Worten: Die Führungskräfte tendieren dazu, mit steigender Betriebszugehörigkeit²⁵ die Leistungen ihrer Beschäftigten besser zu beurteilen. Diese Ergebnisse lassen darauf schließen, dass trotz Berücksichtigung einer Vielzahl von Faktoren, insbesondere die Betriebszugehörigkeit der Beurteilten als auch der Vorgesetzten nicht zu vernachlässigende Einflussgrößen auf die tarifliche Leistungsbeurteilung darstellen.

In den Experteninterviews wurden auch die Einflüsse der Betriebszugehörigkeit auf die Leistungsbeurteilung der Beschäftigten thematisiert. In diesem Zusammenhang war von Interesse, welche Gründe die Experten dafür anführen, dass sich die Leistungsbeurteilungen im Karriereverlauf ändern, d.h., neue Beschäftigte mit steigender Betriebszugehörig-

²⁵ Anhang 8 illustriert den Zusammenhang zwischen Alter und tariflicher Leistungszulage.

keit tendenziell positiver bewertet werden, langjährige negativer. Sieben der acht befragten Experten sind der Meinung, dass dieser Effekt durch die individuellen Entwicklungen ihrer Beschäftigten bedingt wird, die sich im Verlauf der Betriebszugehörigkeit ändern. In diesem Kontext haben auch die jeweiligen Stellenbilder eine nicht unwesentliche Bedeutung, wie die Antwort der weiblichen Führungskraft im Personalwesen bestätigt: *„Die Bewertung ist ja so aufgebaut, dass ich quasi so bewerten muss, als wenn jemand sein Stellenbild schon jahrelang ausführt und gerade ... neue Mitarbeiter können das ja am Anfang noch gar nicht, also gerade, weil sie die Betriebsabläufe, die Prozesse und so weiter nicht kennen“*. Auch die Führungskraft im Compliance verdeutlicht die Beziehung zwischen der persönlichen Entwicklung der Beschäftigten und der ERA-Bewertungen: *„Der Mitarbeiter muss Erfahrungen mit seiner Aufgabe sammeln und deshalb wird da sicherlich die Anfangsbeurteilung grundsätzlich geringer ausfallen. Die muss sicherlich nicht absolut unterdurchschnittlich ausfallen, aber da bin auch ich eher zurückhaltender.“* Aus den Antworten der Experten wird deutlich, dass die angesprochenen Lerneffekte der Beschäftigten die Leistungsbeurteilungen beeinflussen können, jedoch sind diese über die Karriere hinweg unterschiedlich ausgeprägt. In Bezug auf Beschäftigte mit längeren Betriebszugehörigkeiten, die relativ gesehen schlechter bei den Beurteilungen abschneiden, wird von zwei Befragten als Argument angeführt, dass diese sich nicht entsprechend weiterentwickeln, um den steigenden Anforderungen gerecht zu werden. Sie hätten zudem eine Routine erreicht, die dazu führt, dass Anreize zur Weiterentwicklung fehlen. Neben diesen Lerneffekten spielt auch das Bewertungsverhalten der Vorgesetzten eine Rolle. Beispielsweise meint die Führungskraft in der Entwicklung, dass neue Beschäftigte eher niedrig beurteilt werden, *„ ... um ... erfahrenen Mitarbeitern auch noch einen Punkt mehr geben zu können“*.

Gegenstand der Experteninterviews war auch das weiter oben beschriebene Ergebnis, dass Vorgesetzte zum Ende ihrer beruflichen Laufbahn tendenziell positivere Leistungsbeurteilungen vergeben. Für die befragten Experten liegt der Grund in der Erfahrung der Vorgesetzten mit dem tariflichen Leistungsbeurteilungsverfahren. Laut der männlichen Führungskraft im Personalwesen führe ein geringer Erfahrungsstatus im Umgang mit dem Verfahren dazu, *„ ... dass eine junge Führungskraft ... tendenziell noch ein bisschen strenger ist und mit einem anderen Maßstab rangeht ...“* als Vorgesetzte, die hinsichtlich des Bewertungsverfahrens erfahrener sind. Die Personalreferentin kann sich auch vorstellen, dass ältere Führungskräfte eine gewisse Systemmüdigkeit erfahren. Diese Ansicht

wird von der weiblichen Führungskraft im Personalwesen geteilt. Sie ist der Meinung, dass ältere Vorgesetzte „... *sich nicht mehr dem Stress aussetzen wollen* ...“ Diskussionen über Leistungsbeurteilungen zu führen. Die Folge hiervon sei eine Tendenz zur Milde. Die Führungskraft in der Entwicklung führt eine weitere Erklärung an: „*Man wird ruhiger. Der eigene Karrierezeit ist erreicht. Ich muss mich nicht mehr gegenüber weiteren Vorgesetzten ... beweisen, indem ich den harten Mann spiele*“.

Neben dem Alter und der Betriebszugehörigkeit wurde in den Panelschätzungen das Arbeitsergebnis der Beschäftigten berücksichtigt. Aus diesem Grund enthalten die Modelle (1) der geschätzten Panelregressionen eine Variable, die die außertarifliche Zulage aus der Zielvereinbarung und die Modelle (2) eine Variable, die die außertarifliche Zulage aus der Bewertung der Arbeitsquantität berücksichtigt. Bezüglich dieser Faktoren zeigen die Fallstudien, dass nur in dem weniger formalisierten Beurteilungssystem die subjektive Bewertung des Arbeitsergebnisses einen positiven und statistisch signifikanten Einfluss auf die tarifliche Leistungsbeurteilung hat. Eine Begründung hierfür könnte sein, dass das subjektiv bewertete Arbeitsergebnis einen Ankereffekt auf die tarifliche Leistungsbeurteilung der Beschäftigten ausübt. Ein statistisch signifikanter Zusammenhang zwischen dem Ergebnis von Zielvereinbarungen und der tariflichen Leistungszulage ist hingegen nicht gegeben. Indirekt hat das Ergebnis einer Zielvereinbarung jedoch einen Einfluss auf die Leistungsbeurteilung der Beschäftigten, da andernfalls die Koeffizienten der geschätzten Modelle sich nicht voneinander unterscheiden würden. Dies lässt vermuten, dass Führungskräfte je nach Formalisierungsgrad der Leistungsbeurteilungen die Leistungen der Beschäftigten unterschiedlich wahrnehmen, d.h. ihre Bewertungen auf unterschiedlichen Ankern basieren, und Männer und Frauen anders beurteilen.

Abteilungsmerkmale

Um den Einfluss abteilungsspezifischer Effekte auf die Leistungsbeurteilung der Belegschaft zu kontrollieren, wurden neben den 27 Dummyvariablen für die einzelnen Abteilungen des Unternehmens der Frauenanteil in den Abteilungen und der prozentuale Anteil geringqualifizierter (EG 1 bis EG 6), durchschnittlicher (EG 7 bis EG 11) und hochqualifizierter (EG 12 bis EG 14) Frauen in den Abteilungen als Einflussfaktoren berücksichtigt. Auf die Ergebnisse zu diesen Einflussfaktoren soll im Folgenden eingegangen werden.

Die Ergebnisse der Panelschätzungen aus Tabelle 16 und Tabelle 18 zeigen, dass die Variable *Frauenanteil* keinen systematischen und statistisch signifikanten Einfluss auf die

Leistungsbeurteilungen der Beschäftigten hat. Die Koeffizienten der Variablen, die den Anteil an Frauen in den unterschiedlichen Entgeltgruppen kontrollieren, sind hingegen in allen Panelmodellen statistisch signifikant (zur Erinnerung: die Referenzkategorie ist der Frauenanteil in EG 7 bis EG 11). Die geschätzten Koeffizienten deuten darauf hin, dass die Leistungsbeurteilungen der Beschäftigten vergleichsweise negativer ausfallen, wenn der Anteil an geringqualifizierten Frauen in den Abteilungen steigt bzw. positiver ausfallen, wenn der Anteil an hochqualifizierten Frauen zunimmt. Allerdings sind die Effektgrößen mit Vorsicht zu interpretieren, da der Frauenanteil in den höheren Entgeltgruppen tendenziell gering ist und vergleichsweise wenige Beobachtungen hochqualifizierter Frauen vorliegen (siehe Abbildung 12 in Kapitel 6.2.2).

Die Experten wurden in den Interviews gefragt, wie sie das Ergebnis interpretieren, dass die Leistungsbeurteilungen positiver ausfallen, wenn der Anteil hochqualifizierter Frauen in einer Abteilung steigt. Die Führungskräfte im Personalwesen sind der Meinung, dass heterogene Teams vergleichsweise bessere Leistungen erbringen als homogenen Arbeitsgruppen. Diesbezüglich sagt die männliche Führungskraft: *„Je bunter das Team, desto besser [ist dies] für das Team. Ich bin davon überzeugt, dass das tatsächlich der Fall ist. (...) Ob das dann dazu führt, dass bessere Leistungsbeurteilungsergebnisse da sind, ist mir nicht bekannt, aber wenn Sie sagen, dass das so ist, dann glaube ich das mal.“* Die weibliche Führungskraft ist ebenfalls der Ansicht, dass heterogene Teams leistungsfähiger sind als andere. Hinsichtlich des qualifizierten Fachpersonals meint sie: *„... gerade wenn man dann über hochqualifizierte Personen spricht, ist es natürlich so, dass ja die Zusammenarbeit von solchen Berufsgruppen ... eine ganz andere ist, als wenn ich jetzt in der Montage in einer Abteilung bin oder so“*. Darüber hinaus weist die Führungskraft in der Produktion auf den Aspekt hin, dass Frauen eine andere Herangehensweise an Arbeitsaufgaben hätten. Ihm zufolge könne dies *„... für eine Gruppe sehr befruchtend sein ...“*. Die Führungskraft aus dem Compliance argumentiert, dass der Frauenanteil insbesondere auf den höheren Hierarchiestufen relativ gering sei und fügt hinzu: *„... wenn Frauen sich hier entsprechend etablieren, dann sind sie in der Regel auch ganz gut“*. Ihrer Meinung nach erfahren Frauen im Berufsleben Nachteile, wie zum Beispiel Ausfälle durch in Anspruch genommene Elternzeit und müssen sich *„... trotzdem gegen einen gleich guten Bewerber ...“* durchsetzen. Sie glaubt, *„... dass die, die es dann schaffen anzukommen, auch recht gut sind“*.

Im Rahmen weiterer Analysen wurden einzelne Panelregressionen für alle 27 Abteilungen des Fallunternehmens geschätzt. Da die Fallstudien­daten spezifische Bezeichnungen der Abteilungen enthalten, die Rückschlüsse auf den Fallbetrieb ermöglichen, werden die Ergebnisse der Abteilungsregressionen aus Datenschutzgründen nicht abgebildet. Auch eine Zuordnung der Abteilungen zu klassischen Unternehmensbereichen, wie zum Beispiel Produktion, Marketing, Vertrieb oder Logistik, ist nach Rücksprache mit den betrieblichen Experten des Unternehmens problematisch, da die 27 Funktionsbereiche unterschiedliche Aufgaben erfüllen und keinen klassischen Unternehmensbereichen zugeordnet werden können. Die Ergebnisse der Panelregressionen für die 27 Abteilungen des Unternehmens weisen darauf hin, dass die angesprochenen geschlechterspezifischen und weiteren Effekte nicht für jede der 27 Abteilungen beobachtbar sind. Die Ergebnisse lassen vielmehr darauf schließen, dass – wie vermutet – in den Funktionsbereichen des Betriebs unterschiedliche Bewertungskulturen existieren. Dementsprechend macht es Sinn, in den betrieblichen Gesamtschätzungen Abteilungs­dummies als Kontrollvariablen zu berücksichtigen.

Die Experten wurden auch darauf angesprochen, ob in den Abteilungen des Unternehmens unterschiedliche Bewertungskulturen existieren. Sie wurden darüber informiert, dass die oben angesprochenen Effekte abteilungsspezifisch sind, d.h., nur in bestimmten Unternehmensbereichen auftreten. Die männliche Führungskraft im Personalwesen berichtet, dass die Leistungsbeurteilungen im Durchschnitt je nach Unternehmensbereich unterschiedlich ausfallen. Jedoch würde er nicht sagen, „ ... *dass das an der Leistungsfähigkeit der Mitarbeiter liegt, sondern ... eher an dem Führungsstil der Führungskraft ...*“. Seiner Meinung nach bewerten die Vorgesetzten aus den Produktionsbereichen tendenziell unterdurchschnittlich. Auch die weibliche Führungskraft im Personalwesen weist darauf hin, dass das Bewertungsverhalten „ ... *stark von der Persönlichkeit der Führungskraft ...*“ abhängt. Sie kann ebenfalls Unternehmensbereiche benennen, die durch ihre Bewertungskulturen auffallen: „ ... *mir fallen spontan zwei Beispiele ein, wo die Bewertungen sehr, sehr positiv sind. Das ist zum einen der Betriebsrat und zum anderen sind es in der Regel die Geschäftsleitungssekretärinnen.*“ Auch die Führungskräfte in der Produktion und in der Fertigung sowie die Personalreferentin betonen, dass das Bewertungsverhalten von Bereich zu Bereich unterschiedlich ist und vom Führungsstil der Vorgesetzten abhängt. Zudem erklärt die Führungskraft in der Entwicklung, dass sich das Bewertungsverhalten von Bereichsleitern auch auf die nachgeschalteten Führungskräfte auswirkt: „*Ich glaube schon, dass sich da wesentliche Unterschiede entwickeln können, wenn ich einen Be-*

reichsleiter habe, der sehr stringent beurteilt, da würde ich mal sagen: ‚Wie der Herr so ist’s Gescherr‘ ... Und wenn ich daneben einen anderen Bereichsleiter habe, der das relativ lasch handhabt, dann wird das für die nachgeschalteten Führungskräfte in Anführungszeichen Vorbild sein, damit lascher umzugehen.“

6.3.3 Kausalität der Geschlechtereffekte

Die empirische Evidenz zeigt, dass die tariflichen Leistungsbeurteilungen durch den Formalisierungsgrad des Beurteilungssystems beeinflusst werden. Demnach haben Frauen in dem vergleichsweise weniger formalisierten System einen Nachteil und in dem formalisierten System einen Vorteil gegenüber ihren Kollegen (Hypothese 1). Es stellt sich die Frage, ob dieses Ergebnis kausal interpretiert werden darf oder es auf unbeobachteter Heterogenität beruht. Mit anderen Worten: Lassen sich die geschlechterspezifischen Effekte auf den Formalisierungsgrad des Beurteilungssystems zurückführen oder hängen sie vielmehr damit zusammen, dass sich beispielsweise die leistungsfähigeren Frauen in das formalisierte System mit Zielvereinbarungen hinein selektieren? Es ist prinzipiell denkbar, dass ein Selektionseffekt vorliegt und die Geschlechterunterschiede in der Fallstudie nicht kausal sind.

Da die Beschäftigten im Fallunternehmen tendenziell freiwillig zwischen den beiden oben beschriebenen Bewertungssystemen (Arbeitsquantität versus Zielvereinbarung) wählen können, stellt sich die Frage, ob ein Selektionseffekt vorliegt. Dieser Argumentation zufolge ist nicht auszuschließen, dass Frauen möglicherweise das eher formalisierte System mit einer Zielvereinbarung wählen, da sie antizipieren, dass sie in diesem Leistungsbeurteilungssystem vergleichsweise besser abschneiden als in einem weniger formalisierten System. Um zu überprüfen, ob ein Selektionseffekt vorliegt, wurden über die oben beschriebenen Panelregressionen hinaus weitere Schätzungen mit Logit-Modellen durchgeführt. Die Ergebnisse dieser Schätzung sind in Tabelle 19 abgebildet. Zudem wurden unterschiedliche Testverfahren angewendet, um ein Vorliegen von Endogenität auszuschließen.

Tabelle 19: Logit-Schätzung zur Wahl des Formalisierungsgrads

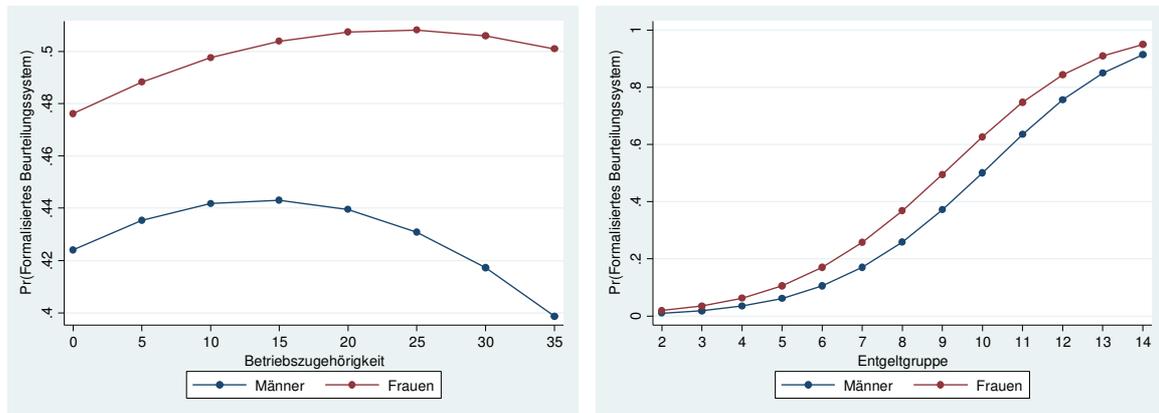
Erklärende Variablen	Abhängige Variable: Dummy für Wahl einer Zielvereinbarung
<i>Geschlecht</i>	
Mitarbeiterin	1,910*** (0,207)
Weibliche Führungskraft	1,796*** (0,300)
<i>Organisationale Hierarchiestufe</i>	
Entgeltgruppe Mitarbeiter/in	1,964*** (0,041)
Entgeltgruppe Führungskraft	1,105*** (0,031)
<i>Merkmale Mitarbeiter/in</i>	
Alter	0,701 (0,216)
Alter*Alter	1,018 (0,038)
Betriebszugehörigkeit	1,476*** (0,209)
Betriebszugehörigkeit*Betriebszugehörigkeit	0,918** (0,036)
<i>Merkmale Führungskraft</i>	
Alter	0,362* (0,188)
Alter*Alter	1,112* (0,065)
Betriebszugehörigkeit	1,056 (0,156)
Betriebszugehörigkeit*Betriebszugehörigkeit	0,915** (0,037)
<i>Merkmale Abteilung</i>	
Frauenanteil	0,078* (0,118)
Frauenanteil in EG 1 bis EG 6	0,243 (0,257)
Frauenanteil in EG 12 bis EG 14	17,720 (39,270)
Abteilungsdummies	Berücksichtigt
<i>Weitere Kontrolle</i>	
Beurteilungsjahr	Berücksichtigt
Konstante	0,044** (0,060)
Beobachtungen	12.085
Pseudo-R ²	0,491

Erläuterungen: Standardfehler sind auf Ebene von Dyaden aus Beschäftigten-Vorgesetzten geclustert. Ergebnisse basieren auf Logit-Schätzungen. Signifikanzniveaus: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$. Die abhängige Variable ist wie folgt kodiert: 1=Beschäftigte erhalten formalisierte Beurteilungen, sonst 0. Die geschätzten Effekte sind als Odds-Ratios ausgegeben. Als zusätzliche Kontrolle wird das Beurteilungsjahr berücksichtigt.

Die Ergebnisse der Schätzung des Logit-Modells aus Tabelle 19 bestätigen, dass die Zuordnung der Beschäftigten zu einem weniger formalisierten oder formalisierten Beurteilungssystem unter anderem durch demographische Merkmale der Beschäftigten beeinflusst wird, wie zum Beispiel die Merkmale Geschlecht, Entgeltgruppe und Betriebszuge-

hörigkeit.²⁶ Die berechnete Odds-Ratio der Variable *Mitarbeiterin* bedeutet beispielsweise, dass Frauen im Unternehmen gegenüber Männern eine ca. 2-mal höhere Chance haben, sich einem formalisiertem Leistungsbeurteilungssystem zuzuordnen.

Abbildung 16: Marginale Effekte der Zuordnung zu dem formalisierten System



Erläuterungen: Die Darstellungen enthalten die durchschnittlichen Wahrscheinlichkeiten für männliche und weibliche Beschäftigte, abhängig von der Betriebszugehörigkeit (links) und der Entgeltgruppe (rechts). Darstellungen wurden mit STATA 13 erstellt.

Die beiden Graphiken in Abbildung 16 zeigen die durchschnittlichen marginalen Effekte bzw. geschätzten Wahrscheinlichkeiten für die Zuordnung von Frauen und Männern zu einem formalisierten Leistungsbeurteilungssystem über unterschiedliche Betriebszugehörigkeiten (linke Graphik) und Karrierestufen (rechte Graphik). Die Verläufe der unterschiedlichen Kurven für Frauen und Männer zeigen die Tendenz auf, dass Frauen sich generell eher einem formalisierten Leistungsbeurteilungssystem zuordnen, unabhängig von der Betriebszugehörigkeit und der hierarchischen Position im Unternehmen. Diese Geschlechterunterschiede sind durchweg statistisch signifikant (siehe Anhang 9 und Anhang 10).

Eine Frage des Experteninterviews zielt auf Faktoren ab, die einen Einfluss auf die Zuordnung der Beschäftigten zu einem der beiden Beurteilungssysteme ausüben. Die Aussagen aus den Experteninterviews verdeutlichen, dass die Entscheidung für oder gegen eine formalisierte Bewertung wesentlich durch die unterschiedlichen Unternehmensbereiche und die auszuführenden Tätigkeiten der Beschäftigten bestimmt wird. Einen klassischen Unternehmensbereich, in dem tendenziell formalisierte Leistungsbeurteilungen durchgeführt

²⁶ Die deskriptiven Statistiken und Korrelationen der im Modell berücksichtigten Variablen sind in Anhang 11 enthalten.

werden, bildet demnach der Beschäftigtenbereich. Im Gegensatz zum Produktionsbereich, in dem Facharbeiter tätig sind und häufig Anlern Tätigkeiten ausgeübt werden, können für die Beschäftigten in diesem Bereich einfacher Ziele ausfindig gemacht und vereinbart werden. Laut Aussage der Führungskraft im Arbeitsschutz umfassen Jobs, die sich für weniger formalisierte Beurteilungen des Arbeitsergebnisses eignen, Tätigkeiten, die „ ... *in der Regel stark abarbeitungsgeprägt...*“ sind. Die Führungskraft in der Fertigung präferiert zum Beispiel „ ... *immer die Quantitätsbewertung..., weil ... [Ziele] ... in der Produktion schlecht umzusetzen sind*“. Neben diesen Aspekten hängt die Zuordnung zu einem Beurteilungssystem auch von den tariflichen Eingruppierungen der Beschäftigten und somit auch von deren Aufgaben ab. Dass üblicherweise für höher eingruppierte Beschäftigte der Prozess der Leistungsbeurteilung stärker formalisiert ist, bestätigt die Führungskraft im Compliance: „*Ich sage mal, ab E-11 ... [wird eine] Zielvereinbarung [gewählt] ... und sonst ein Arbeitskriterium ...*“.

Des Weiteren wurden die Experten mit der Frage konfrontiert, ob es den Beschäftigten frei überlassen ist, sich einem formalisierten oder weniger formalisierten Leistungsbeurteilungssystem zuzuordnen. Diesbezüglich gehen die Meinungen der Experten auseinander. Die Mehrheit der Führungskräfte bestätigt, dass ihre Beschäftigten prinzipiell frei darüber entscheiden, ob sie sich einem formalisierten Beurteilungssystem zuordnen oder nicht. Zum Beispiel meint die männliche Führungskraft im Personalwesen: „*Ich mache das auch da mit offenen Karten. (...) Auch wer sich gegen entgeltrelevante Ziele entscheidet, wird trotzdem nach Zielen geführt.*“ Darüber hinaus bestätigt die Referentin im Personal die freie Wahl über den Formalisierungsgrad der Leistungsbeurteilungen: „*Also ein Mitarbeiter kann erstmal frei wählen und es ist eine Vereinbarung zwischen Führungskraft und Mitarbeiter, wobei man abwägen muss, inwieweit ... Ziele ... sinnvoll sind*“. Auch die weibliche Führungskraft im Personalwesen bekräftigt: „*Schlussendlich fällt man immer darauf zurück, was der Mitarbeiter entscheidet.*“ Bei den Führungskräften, die nicht dem Bereich Personalwesen zuzuordnen sind, ist die Sichtweise auf die Wahlentscheidung eine andere. Aus den Antworten wird nochmals deutlich, dass die Entscheidung durch weitere Faktoren bedingt wird. Beispielsweise bestätigt die Führungskraft in der Produktion, dass die Wahl einer Zielvereinbarung prinzipiell frei ist, „ ... *wobei Führungskräfte schon die Zielvereinbarung bekommen*“. Weitere Faktoren, die diese Wahl beeinflussen, sind: übergeordnete Abteilungsziele, die zur Vorgabe individueller Ziele für die Beschäftigten führen sowie die eigenen Erfahrungen und Einstellungen in Bezug auf Zielvereinbarungspro-

zesse. Zum Beispiel meint die Führungskraft im Compliance: „*Mir ist diese Entscheidung nie gestellt worden, ob ich eine [Zielvereinbarung] gerne hätte oder nicht. Es war selbstverständlich, dass ich eine bekomme. Und so vermittele ich das entsprechend auch. Also das ist gelerntes Verhalten.*“ Die Führungskraft in der Entwicklung erwartet ebenfalls, dass grundsätzlich Ziele vereinbart werden, jedoch würde er „... *nicht auf Ziele drängen*“.

Aus den Aussagen der Experten wird deutlich, dass die Wahl des Formalisierungsgrads der Beurteilungen von mehreren Faktoren abhängt und dass die Beschäftigten mehrheitlich frei darüber entscheiden können, welches Beurteilungssystem sie wählen. Jedoch wird auch offensichtlich, dass einige Führungskräfte eine Vereinbarung formalisierter Ziele mit ihren Beschäftigten erwarten. Die deskriptiven Statistiken zur Wahl zwischen den beiden Leistungsbeurteilungssystemen zeigen, dass ein höherer Prozentsatz an Frauen ihre Leistungsbeurteilungen in dem vergleichsweise weniger formalisierten Setting ohne Zielvereinbarung erhält (siehe Tabelle 15 in Kapitel 6.2.2). Demnach erhalten lediglich ca. 37% der Frauen ihre Bewertungen im Rahmen einer formalisierten Beurteilung mit einer Zielvereinbarung, bei den Männern sind es hingegen ca. 49%. Im Gegensatz dazu stehen die statistischen Wahrscheinlichkeiten, die unter kontrollierten Bedingungen geschätzt wurden (Abbildung 16).

Um der oben angesprochenen Frage nach der Kausalität der aufgezeigten Geschlechtereffekte auf den Grund zu gehen, wurden die weiblichen Beschäftigten in die Gruppen der Leistungsträgerinnen und Nicht-Leistungsträgerinnen eingeteilt. Dies ist sinnvoll, da leistungsstarke Frauen die zugrunde liegenden Bewertungsmechanismen antizipieren könnten und sich bewusst für das formalisierte Beurteilungssystem entscheiden, um bessere Leistungsbeurteilungen zu erhalten. Als Leistungsträgerinnen und Nicht-Leistungsträgerinnen wurden diejenigen Mitarbeiterinnen identifiziert, die – unabhängig von der Art des Beurteilungssystems – mehr als 10% bzw. bis zu 10% tarifliche Leistungszulage erhalten. Der Schwellenwert für die Einteilung der Beschäftigten wurde auf 10% festgelegt, da die Regelungen im ERA-TV vorgeben, dass durchschnittlich leistungsfähige Beschäftigte im Mittel 10% Leistungszulage erhalten.

Nach der obigen Einteilung der weiblichen Beschäftigten in die zwei Gruppen wurde in einem zweiten Schritt überprüft, ob diese sich bezüglich der Wahrscheinlichkeiten unterscheiden, sich entweder in das formalisierte oder weniger formalisierte Leistungsbeurtei-

lungssystem hinein zu selektieren. Um dies zu überprüfen, wurde ein Chi-Quadrat-Test durchgeführt. Die deskriptiven Statistiken zeigen, dass ca. 52% der Frauen im Unternehmen zu den Leistungsträgerinnen zählen, ca. 48% hingegen sind nach der oben erwähnten Einteilung keine Leistungsträgerinnen und erhalten demnach weniger als 10% Leistungszulage. Ein Chi-Quadrat-Test, der die Unabhängigkeit zwischen der Art des Beurteilungssystems und den zwei bewerteten Personengruppen testet, bestätigt, dass zwischen den Merkmalen keine statistisch signifikante Beziehung besteht ($\chi^2=0,125$; $p=0,72$). Folglich existieren keine statistisch signifikanten Gruppenunterschiede.

In einem weiteren Schritt wurde geprüft, ob aufgrund der gegebenen potentiellen Wahlfreiheit zwischen beiden Beurteilungssystemen Beschäftigte von einem Bewertungssystem in das andere wechseln. Analog zu der Begründung für die Unterscheidung nach Leistungsträgerinnen und Nicht-Leistungsträgerinnen kann angenommen werden, dass Leistungsträgerinnen eher in das formalisierte Beurteilungssetting wechseln, da sie bessere Leistungsbeurteilungen erwarten. Die deskriptiven Statistiken belegen, dass in Bezug auf Frauen in lediglich 217 Fällen (dies entspricht einem Anteil von 8,22% aller Beobachtungen) eine andere Beurteilungsform vorliegt als im Vorjahr. In 96 Fällen (8,19%) sind Leistungsträgerinnen in das weniger formalisierte Beurteilungssystem gewechselt, wohingegen in 121 Fällen (8,24%) vergleichsweise weniger produktive Mitarbeiterinnen in das formalisierte Beurteilungssystem mit einer Zielvereinbarung gewechselt sind. In diesem Zusammenhang bestätigt ein durchgeführter Chi-Quadrat-Test, dass keine systematische Beziehung zwischen den Leistungsträgerinnen sowie Nicht-Leistungsträgerinnen und einem Wechsel zwischen den Systemen vorliegt ($\chi^2=0,002$; $p=0,96$).

Da die Ergebnisse nicht auf einen möglichen Selektionseffekt hindeuten, kann tendenziell davon ausgegangen werden, dass die aufgedeckten geschlechterspezifischen Ergebnisse der Fallstudie auf kausalen Zusammenhängen beruhen. Aus diesem Grund ist es sehr unwahrscheinlich, dass Leistungsträgerinnen sich in das formalisierte Beurteilungssystem hinein selektieren, um im Rahmen dieses Systems vergleichsweise bessere Leistungsbeurteilungen zu erhalten. Für diese Implikationen sprechen auch die durchgeführten Panelanalysen (siehe Tabelle 16 und Tabelle 18), die ein Stück weit das Problem der Selektion einzudämmen versuchen und die Schätzung kausaler Geschlechtereffekte ermöglichen.

Die dargelegten Argumente belegen jedoch nicht, dass die Zuordnung zu einer der beiden Beurteilungsformen gänzlich exogen ist. Es könnte auch vermutet werden, dass einige

weibliche Beschäftigte sich bewusst für das formalisierte Beurteilungssystem mit einer Zielvereinbarung entscheiden, da sie in diesem System fairere Leistungsbeurteilungen erwarten. D.h., dass die Frauen im Fallstudienunternehmen die in dieser Arbeit adressierten Geschlechtereffekte antizipieren und die Mechanismen verstehen, die für das Auftreten geschlechterspezifischer Einflüsse auf die Leistungsbeurteilung verantwortlich sind.

7 Diskussion und Schlussbetrachtung

7.1 Hauptergebnisse der quantitativen Fallstudie

In der vorliegenden Arbeit wurde untersucht, wie spezifische organisationale Mechanismen zum Auftreten von Geschlechterunterschieden in tariflichen Leistungsbeurteilungen und im Gehalt beitragen. Die Auswertungen der Fallstudien­daten zeigen, dass Frauen im Fallunter­nehmen im Mittel nicht systematisch schlechter beurteilt werden als vergleichbare Männer. Jedoch führen indirekte Effekte, die durch organisationale Mechanismen bedingt werden, zu Differenzen in tariflichen Beurteilungen zwischen den Geschlechtern. Die vorliegende Dissertation hat aus den im Folgenden erörterten Gründen zu einem besseren Verständnis von Geschlechterunterschieden in Leistungsbeurteilungen beigetragen.

Im Theorieteil dieser Arbeit ist herausgestellt worden, dass zur Aufklärung von Ungleichbehandlung im Kontext von Leistungsbeurteilungen unterschiedliche Forschungsdisziplinen herangezogen werden können, unter anderem ökonomische (präferenzbasierte und statistische Diskriminierung), sozialpsychologische (Stereotype) und soziologische Ansätze (zum Beispiel die Soziale Identität und die LMX-Theorie). Auf Grundlage dieser Ansätze wurde argumentiert, dass Frauen schlechtere Bewertungen erhalten, wenn sie von männlichen Führungskräften evaluiert werden. Im Gegensatz zu dieser Annahme stehen zum Beispiel die Ergebnisse von Metaanalysen, die auf eine positive Verzerrungen für Frauen hindeuten (Bowen et al., 2000; Roth et al., 2012). Auch die Analyse von Moderatoren, wie beispielsweise Geschlechterverteilung in Berufsgruppen, Job-Komplexität als auch Geschlechterdyaden, zeigt keine konsistenten Ergebnisse auf, die auf eine systematische Benachteiligung von Frauen in Leistungsbewertungen schließen lassen. Die Erkenntnisse deuten eher darauf hin, dass geschlechterspezifische Effekte indirekt wirken und nur unter bestimmten Bedingungen auftreten. In Bezug hierauf argumentieren Abraham (2017) und Reskin (2000), dass zum Beispiel organisationale Mechanismen in Unternehmen im Hinblick auf eine Ungleichbehandlung von Frauen und Männern näher untersucht werden sollten.

Ein organisationaler Mechanismus, der im Kontext von Leistungszulagen näher beleuchtet wurde (Abraham, 2017; Elvira & Graham, 2002) und den Fokus dieser Arbeit bildet, ist der Formalisierungsgrad von Leistungsbeurteilungssystemen. Eine Gegenüberstellung

diverser Untersuchungen zu den Wirkungsweisen von Formalisierungsgraden in Gehaltssystemen deutet auf eine Verringerung von Geschlechtereffekten durch Formalisierung hin (Elvira & Graham, 2002). Daher wurde im Rahmen der Hypothese 1 der vorliegenden Dissertation argumentiert, dass der Geschlechternachteil für Frauen in einem formalisierten Beurteilungssystem vergleichsweise kleiner ist als in einem weniger formalisierten. Die Ergebnisse der Fallstudien bestätigen diese Vermutung und zeigen, dass Frauen *ceteris paribus* schlechter beurteilt werden, wenn sie im Rahmen eines weniger formalisierten Leistungsbewertungssystems evaluiert werden. Diese geschlechterspezifischen Nachteile in den Beurteilungen heben sich allerdings auf bzw. wandeln sich in Vorteile um, wenn Frauen ihre Leistungsbewertungen in einem vergleichsweise formalisierten Setting mit einer objektiven Evaluation ihrer Arbeitsergebnisse erhalten. Die Erkenntnisse aus dieser Studie lassen daher darauf schließen, dass in der Tat organisationale Mechanismen in Zusammenhang stehen mit dem Auftreten von Geschlechterungleichbehandlung. Der Formalisierungsgrad als potentielle Einflussgröße sollte den Fallstudienresultaten zufolge als ein Moderator in Studien über Geschlechterdifferenzen berücksichtigt werden.

Neben dem Formalisierungsgrad können Geschlechtereffekte durch die organisationale Hierarchiestufe der Beschäftigten beeinflusst werden. Da beispielweise Frauen, die bereits erfolgreich die Karriereleiter aufgestiegen sind, zu einer Abschwächung von Geschlechterstereotypen in Unternehmen beitragen können, wurde im Rahmen der Hypothese 2 argumentiert, dass Frauen in einem weniger formalisierten Beurteilungssystem insbesondere auf niedrigen organisationalen Hierarchiestufen Nachteile erfahren. Den Auswertungen der Fallstudien zufolge lässt sich auch diese Vermutung bestätigen. Im Gegensatz zu Abraham (2017) lassen die Resultate dieser Fallstudie jedoch darauf schließen, dass Frauen in weniger formalisierten Gehaltssystemen insbesondere auf niedrigen organisationalen Hierarchiestufen (und nicht höheren) benachteiligt werden. Die Ergebnisse der vorliegenden Fallstudie verdeutlichen, dass auch die Interaktion zwischen dem Geschlecht und der Hierarchiestufe berücksichtigt werden sollte, um Geschlechterunterschiede in Abhängigkeit des Formalisierungsgrads von Beurteilungssystemen zu untersuchen. In der Studie von Elvira und Graham (2002) wird die organisationale Hierarchiestufe beispielweise nur als Kontrollvariable berücksichtigt und nicht mit dem Geschlecht der Beschäftigten interagiert.

Neben den zuvor genannten Einflussfaktoren sind auch spezifische Geschlechterdyaden

zwischen Beschäftigten und Vorgesetzten relevant. Beispielweise sagen die LMX- als auch die Theorie der Sozialen Identität, dass weibliche Führungskräfte Frauen tendenziell besser beurteilen, da sie zu ihrer Ingroup gehören. Dementsprechend wurde im Rahmen von Hypothese 3 postuliert, dass der Nachteil für Frauen in einem weniger formalisierten Beurteilungssystem kleiner ist, wenn sie von weiblichen und nicht von männlichen Führungskräften Leistungsbeurteilungen erhalten. Die Ergebnisse der Fallstudie deuten darauf hin, dass auch spezifische Geschlechterdyaden zwischen Führungskräften und Beschäftigten Auswirkungen auf die Bewertungen haben. Anders als in den Untersuchungen von Furnham und Stringfield (2001) sowie Varma und Stroh (2001) zeigen sich allerdings keine konsistenten Ergebnisse hinsichtlich gleichgeschlechtlicher Dyaden. Die Analyse der Fallstudien Daten zeigt auf, dass in dem weniger formalisierten Beurteilungssystem Frauen sowohl von männlichen als auch von weiblichen Führungskräften eher schlechter beurteilt werden als vergleichbare Männer.²⁷ Der statistisch nicht signifikante und negative Koeffizient der Dyade *Frau beurteilt Frau* in der Schätzung des weniger formalisierten Leistungsbeurteilungssystems lässt – entgegen der theoretischen Argumente – darauf schließen, dass die Benachteiligung für Frauen tendenziell am größten ist, wenn sie von weiblichen Führungskräften beurteilt werden. Dieses Ergebnis ist jedoch mit Vorsicht zu interpretieren, da nur relativ wenige Beobachtungen mit weiblichen Führungskräften vorliegen.

Es lässt sich zusammenfassend festhalten, dass Geschlechtereffekte in Leistungsbeurteilungen von diversen unternehmens- und kontextspezifischen Bedingungen abhängen. Zwei Wirkmechanismen, die in dieser Arbeit näher fokussiert wurden, sind der Formalisierungsgrad von Leistungsbeurteilungen und die organisationale Hierarchiestufe der Beschäftigten. Die übergeordnete Forschungsfrage der vorliegenden Dissertation (*Wie wirken sich die Formalisierung eines Beurteilungssystems sowie Geschlechterdyaden auf die Geschlechterunterschiede in Leistungsbeurteilungen und auf die Vergütung von Beschäftigten aus?*) lässt sich wie folgt beantworten. Bedingungen für das Vorhandensein von Geschlechtereffekten in Leistungsbeurteilungen sind in erster Linie: ein geringer oder hoher Formalisierungsgrad eines Beurteilungssystems und eine vergleichsweise niedrige oder hohe organisationale Hierarchiestufe der beurteilten Beschäftigten. Die Andersbeurteilung von Frauen geht überwiegend von männlichen Vorgesetzten aus. Die Ergebnisse der Fall-

²⁷ Anzumerken ist, dass die angesprochenen Geschlechterkombinationen einen geringen Erklärungsbeitrag hinsichtlich der Variation in den tariflichen Leistungsbeurteilungen haben – ein Vergleich der berechneten Gütemaße (R^2) für die geschätzten Regressionsmodelle mit und ohne Geschlechterdyaden bestätigt dies.

studie liefern auch einen Beitrag zu den Auswirkungen von Geschlechtereffekten in Leistungsbeurteilungen auf die Vergütung der Beschäftigten. Die geschätzten Koeffizienten der Variable *Mitarbeiterin* lassen darauf schließen, dass Frauen je nach Formalisierungsgrad im Vergleich zu Männern im Durchschnitt eine um ca. 0,2%-Punkte höhere oder geringere Leistungszulage erhalten. Die sich ergebenden individuellen Lohnunterschiede für Männer und Frauen sind allerdings moderat. Obwohl die oben erwähnten geschlechterspezifischen Effekte in der Fallstudie relativ gering sind, bleibt zu beachten, dass diese sich über weitere Beurteilungsjahre kumulieren und letztlich zu beachtlichen Gehaltsunterschieden zwischen Frauen und Männern beitragen können.

Einen weiteren Diskussionspunkt in Bezug auf die Ergebnisse der Fallstudie bieten die geschlechterspezifischen Zuordnungen zu Jobs, in denen Beschäftigte auf Basis subjektiver oder objektiver Lohnkomponenten vergütet werden. Ein bekanntes Phänomen ist, dass Frauen in Jobs, in denen ein fixer Stücklohn gezahlt wird, überrepräsentiert sind (Heywood & Jirjahn, 2002). Dies lässt sich damit begründen, dass Frauen sich eher in Berufe hinein selektieren, in denen objektivere Vergütungsinstrumente verwendet werden und somit den Führungskräften ein vergleichsweise kleiner Spielraum für Geschlechterungleichbehandlung geboten wird (Jirjahn & Stephan, 2004). Darüber hinaus bestätigen zum Beispiel Maas und Torres-González (2011), dass Mitarbeiterinnen eher positiv verzerrte subjektive Beurteilungen erwarten, wenn die Wahrscheinlichkeit steigt, dass sie von weiblichen Vorgesetzten evaluiert werden. Im Rahmen dieser Fallstudie wurde aufgezeigt, dass Frauen sich tendenziell in das formalisierte Beurteilungssystem hinein selektieren und in diesem System von männlichen Führungskräften bessere Leistungsbeurteilungen erhalten als ihre Kollegen. Dies bedeutet, dass weibliche Beschäftigte, unabhängig davon, ob sie zu der Gruppe der Leistungsträgerinnen oder Nicht-Leistungsträgerinnen gehören, die zugrunde liegenden Bewertungsmechanismen verstehen und sich für das Beurteilungssystem entscheiden, in dem sie vergleichsweise positivere Beurteilungen erhalten.

7.2 Ergebnisse der Experteninterviews

Neben den Ergebnissen der quantitativen Studie bieten die durchgeführten Interviews mit den betrieblichen Experten weitere Diskussionspunkte, auf die nachfolgend näher eingegangen werden soll.

In Bezug auf die generellen Einstellungs- und Wissensfragen zum Thema ERA-Leistungsbeurteilungen zeigt sich, dass die Durchführung der tariflichen Leistungsbeurtei-

lungen Herausforderungen mit sich bringt. In einer Frage wurde beispielsweise thematisiert, ob die Einhaltung der tariflichen Mittelwertvorgabe zu praktischen Schwierigkeiten führt. Hierzu gehen die Meinungen der Experten auseinander: Die Hälfte der Befragten sieht keine Problematiken, da die Führungskräfte sowohl mit der Vorgabe als auch den Anpassungsmechanismen, die bei Nichteinhaltung der Vorgabe Anwendung finden, vertraut sind. Andererseits wird kritisiert, dass die Vorgabe zu einer geringen Leistungsdifferenzierung zwischen den Beschäftigten führt. Erstaunlich ist auch der Aspekt, dass sieben der acht befragten Experten die ERA-Leistungsbeurteilungen nicht als Anreizinstrument sehen, um die Leistungen ihrer Beschäftigten zu erhöhen. Darüber hinaus sind sie auch mehrheitlich der Auffassung, dass die tariflichen Evaluationen tendenziell subjektiv sind.

Ein weiteres Hauptaugenmerk der Experteninterviews lag auf der Interpretation der quantitativen Fallstudienresultate. Die Antworten der Experten deuten darauf hin, dass hinsichtlich des Beurteilungsverfahrens geschlechterspezifische Differenzen existieren. Diese werden offenbar durch mehrere Faktoren beeinflusst. Die Erklärungen der Experten für die festgestellten Geschlechtereffekte in der vorliegenden Arbeit sind, dass Frauen in einem männerdominierten Umfeld einen anderen Ehrgeiz entwickelt haben können und vergleichsweise bessere Leistungen erbringen oder sie werden in einem technisch orientierten Unternehmen mehr wertgeschätzt als Männer. Als weitere Gründe werden angeführt: für Frauen wird ein anderer Bewertungsmaßstab herangezogen; Vorgesetzte neigen womöglich zu einer Milde-Tendenz bei Frauen, da sie emotionaler auf schlechtere Beurteilungen reagieren. Im Übrigen bestätigen die Experteninterviews die Vermutung, dass die Bewertungen durch diverse Faktoren beeinflusst werden, wie beispielsweise Geschlecht, Betriebszugehörigkeit, Entgeltgruppen- und Abteilungszugehörigkeit sowie spezifische Gegebenheiten in den Funktionsbereichen (zum Beispiel die Frauenanteile in den Entgeltgruppen). Zudem wird aus den Antworten der Experten ein Aspekt sehr deutlich: Jede Führungskraft hat einen eigenen Führungsstil, der sich auf das Bewertungsverhalten und letztlich auch auf die Bewertungen der Beschäftigten auswirkt. Demnach hat die Persönlichkeit des Vorgesetzten einen entscheidenden Einfluss auf die Bewertungen.

Die Ergebnisse aus dem quantitativen Teil als auch den Experteninterviews lassen schlussfolgern, dass neben der Leistung weitere Faktoren die Bewertungen der Beschäftigten beeinflussen. Die anhaftende Eigenschaft der Subjektivität der tariflichen Leistungsbeurtei-

lungen²⁸ wird auch durch die eingangs zitierte Umfrage zum Leistungsentgelt in der Tarifregion Baden-Württemberg bestätigt (Bahnmüller & Hoppe, 2016b). Die Umfrage zeigt, dass neben der Leistung die Budgetierung des Leistungszulagenvolumens, Sympathien und Antipathien sowie die Dauer der Zugehörigkeit zu einer Abteilung die Leistungsbeurteilungen beeinflussen, jedoch erstaunlicherweise nicht das Geschlecht. Nur eine Minderheit der befragten Betriebsräte und ca. ein Viertel der befragten Personalmanager sind anderer Auffassung, da sie einen gewissen Einfluss von geschlechterspezifischen Effekten vermuten (Bahnmüller & Hoppe, 2016b). Diese Vermutung wird durch die quantitativen Ergebnisse der vorliegenden Untersuchung unterstützt. Sie bietet empirische Evidenz dafür, dass Leistungsbeurteilungen durch geschlechterspezifische Effekte beeinflusst werden.

Die Befunde der Fallstudie bieten möglicherweise auch Gründe dafür, weshalb die praktizierten Leistungsbeurteilungen bei Beschäftigten als auch deren Führungskräften zu Unzufriedenheit führen. Einige Aussagen der Experten deuten darauf hin, dass Beschäftigte und Vorgesetzte gegenüber den ERA-Beurteilungen überwiegend eine kritische Haltung einnehmen, da sie eher als subjektiv wahrgenommen werden und keine Anreize zur Leistungssteigerung bieten. Auch die öffentliche Debatte um Beurteilungen und daran gekoppelte Leistungszulagen bestätigt, dass Leistungsbewertungen als unfair und verzerrt wahrgenommen werden (Blume, Seyerlein & Fröndhoff, 2016).

7.3 Praktische Implikationen

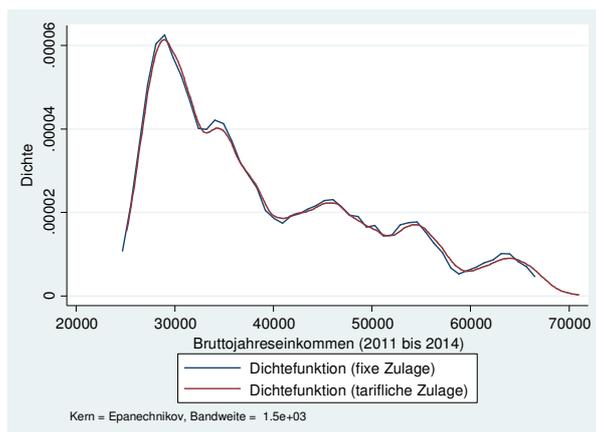
Die diskutierte Subjektivität des tariflichen Leistungsbewertungsverfahrens lässt den Gedanken aufkommen, ob das Beurteilungssystem in dieser Form weiterhin praktiziert werden sollte. Im Gespräch mit den Experten im Fallunternehmen wurde auch die Entwicklung eines alternativen Beurteilungssystems besprochen, bei dem der Ausgang der Beurteilung nicht an die Vergütung der Beschäftigten gekoppelt ist, sondern an eine pauschale Leistungszulage, die der tariflichen Vorgabe in Höhe von 10% des Bruttoeinkommens (siehe Grundsätze zu § 10 ERA-TV) entspricht und an alle Beschäftigten in gleicher Höhe ausgezahlt wird. Die Folge dieses alternativen Beurteilungssystems wäre, dass die Bewertungen nicht mehr leistungsrelevant wären und eher die Funktion eines reinen Feedbacks

²⁸ Auch in der zu Beginn dieser Arbeit zitierten Studie zum Thema ERA der Hans-Böckler-Stiftung liegt der Hauptkritikpunkt des tariflichen Leistungsbeurteilungsverfahrens auf der Subjektivität (Kratzer & Nies, 2009).

der Führungskräfte an die Beschäftigten wahrnehmen.

Abbildung 17 zeigt die Verteilung des tatsächlichen Jahreseinkommens (brutto) der Beschäftigten im Unternehmen, inklusive der tatsächlich gezahlten tariflichen Leistungszulagen (siehe rote Funktion) und die alternative Verteilung (blaue Funktion) mit dem üblichen Einkommen und einer fixen Leistungszulage in Höhe von 10%. Die beiden Beurteilungsalternativen zeigen ganz offensichtlich, dass bei einer Einführung einer Leistungszulage in Form einer Pauschale sich zumindest an der Einkommensverteilung der Beschäftigten im Fallbetrieb kaum etwas ändern würde.²⁹ Die Verläufe der Einkommensverteilungen (rote und blaue Funktion) sind über das Jahreseinkommen der Beschäftigten betrachtet nahezu identisch. Dies trifft auch auf die separaten Einkommensverteilungen von Frauen (siehe Anhang 12) und Männern (siehe Anhang 13) zu.

Abbildung 17: Pauschale und ERA-Leistungszulage für Frauen und Männer



Erläuterungen: Kerndichteschätzungen basieren auf den Bruttojahreseinkommen der Beschäftigten inklusive einer fixen Leistungszulage (blaue Funktion) und inklusive der tariflichen Zulage (rote Funktion). Darstellung wurde mit STATA 13 erstellt.

Die Führungskräfte wurden im Rahmen der Interviews gefragt, was sie von der oben diskutierten Idee halten würden, die Beschäftigten im Unternehmen nicht variabel in Abhängigkeit ihrer Leistungsbeurteilungen zu entlohnen, sondern den Beschäftigten eine Art pauschale Zulage zu zahlen. Sechs der acht befragten Experten würden grundsätzlich die Ausschüttung einer Pauschalzulage an die Beschäftigten befürworten. Zum Beispiel würde die männliche Führungskraft im Personalwesen die tarifliche Leistungsbeurteilung ab-

²⁹ Im alternativen Szenario mit einer fixen Leistungszulage würden die Kosten, die sich durch das jährliche Verfahren der Leistungsbewertung ergeben, vermutlich erheblich reduzieren, da die Beurteilungen der Beschäftigten nicht nachkalibriert werden müssten, um die tariflich vorgegebene Quote hinsichtlich des Zulagenvolumens einzuhalten.

schaffen und sich ein Instrument wünschen, das seiner Meinung nach tatsächlich die Leistung der Beschäftigten misst und Potenzialträger im Unternehmen identifiziert. In diesem Zusammenhang argumentiert er: *„Und wenn man das über eine pauschale Zahlung und ein vernünftiges Potenzialidentifikationsgespräch oder ... Entwicklungsgespräch ersetzen kann, dann bin ich davon überzeugt, dass es die bessere Variante ist als das heutige System“*. Auch die Personalreferentin sieht die Ausschüttung einer pauschalen Zulage als eine mögliche Alternative zu den praktizierten ERA-Bewertungen. Sie und drei weitere Experten betonen jedoch die Notwendigkeit von institutionalisierten Mitarbeitergesprächen zwischen den Führungskräften und Beschäftigten, die ähnlich wie die ERA-Bewertungen jährlich durchgeführt werden. Die Personalreferentin findet es wichtig, *„ ... dass es dieses Gespräch zwischen Mitarbeiter und Führungskraft gibt. Gerade in den ... eher gewerblich-technischen Bereichen kommt das [Bewertungsgespräch] oft zu kurz“*. Die Führungskraft im Compliance sieht das tarifliche Bewertungsverfahren zwar kritisch, würde aber auch in einem System mit Pauschalzulage die Mitarbeitergespräche beibehalten: *„ ... es wäre ... schade um die institutionalisierten Mitarbeitergespräche, die ich für sehr wertvoll halte“*. Zwei der befragten Experten würden für eine Beibehaltung des ERA-Evaluationssystem plädieren. Beispielsweise hält die Führungskraft aus der Produktion nicht viel von einer einmaligen Ausschüttung einer Leistungszulage, da der Motivationseffekt einer Einmalzahlung nicht zu vergleichen sei mit einer monatlich gezahlten Leistungszulage, so wie sie im ERA-System praktiziert wird. Bezüglich der Pauschalzulage argumentiert er: *„Das zählt nur einmal und der Effekt ist relativ schnell verflogen ... Und deswegen ist die Leistungsbeurteilung schon ein sehr gutes Instrument, [um] die Motivation ... oben zu halten“*. Die Führungskraft aus der Entwicklung würde ebenfalls die ERA-Bewertung aufrechterhalten und die Implementierung einer fixen jährlichen Leistungszulage ablehnen. Seiner Meinung nach würde die Abschaffung der tariflichen Bewertung diejenigen Beschäftigten demotivieren, die gute Arbeit leisten.

Die Interviews zeigen, dass das Leistungsbeurteilungssystem trotz der genannten Kritikpunkte nicht von allen Experten grundsätzlich in Frage gestellt wird, jedoch sind die Einstellungen gegenüber dem System sehr unterschiedlich. Den Führungskräften ist tendenziell von Bedeutung, dass zumindest einmal im Jahr ein Austausch mit den Beschäftigten über die erbrachten Leistungen sichergestellt wird, ungeachtet dessen, ob die Beschäftigten hinsichtlich der Leistungszulagen letztendlich pauschal oder variabel vergütet werden. Interessant ist auch der Aspekt, dass nur zwei Experten das derzeit praktizierte ERA-

Beurteilungsverfahren aufgrund von Motivationseffekten beibehalten würden.

Bereits in Vorgesprächen mit den Projektpartnern des Fallstudienbetriebs wurde deutlich, dass die Grundhaltung der Führungskräfte im Unternehmen zum Thema ERA-Leistungsbeurteilung tendenziell kritisch ist. Die Intention der letzten Frage des Experteninterviews war es daher zu erfahren, ob Änderungen am derzeit durchgeführten tariflichen Bewertungssystem gewünscht werden. Dass das angewendete System kritisch hinterfragt wird, wird durch die Äußerungen der Experten bestätigt. Lediglich einer der acht befragten Experten (Führungskraft aus der Produktion) hat an dem derzeitigen ERA-System nichts auszusetzen und sieht auch kein Verbesserungspotenzial: *„Also ich kann nicht sagen, dass das System schlecht ist. Ich denke, man kann gut damit arbeiten.“* Die übrigen Experten sind hingegen anderer Meinung, wie die Auswertung der Interviews zeigt. Die Hauptkritiken sowie die Änderungswünsche am System sind in Tabelle 20 gegenübergestellt. Diese beziehen sich hauptsächlich auf die im Beurteilungsformular aufgelisteten Kriterien, die die Vorgesetzten bewerten, und den Prozess der Leistungsbeurteilung.

Tabelle 20: Kritiken und Änderungswünsche am ERA-Beurteilungssystem

Kritik	Änderungswunsch
<p>Kriterien:</p> <ul style="list-style-type: none"> ▪ wie Arbeitsquantität, Beweglichkeit und Führungsverhalten sind schlecht bewertbar, da sie subjektiv sind und Interpretationsspielraum bieten ▪ wie Zusammenarbeit bzw. Führungsverhalten sind für einzelne Beschäftigte irrelevant ▪ haben nicht viel mit dem Unternehmen gemeinsam und sind abstrakt ▪ Unterkriterien bzw. Beschreibungen der Hauptkriterien führen zu weiterer Komplexität (Unterkriterien müssen auch evaluiert werden) <p>Beurteilungsprozess:</p> <ul style="list-style-type: none"> ▪ Beurteilungen basieren auf subjektiven Beobachtungen der Vorgesetzten ▪ Tarifliche Vorgabe eines Mittelwerts führt zu geringer Leistungsdifferenzierung ▪ Abstufungen bzw. Notenschritte sind nicht optimal 	<p>Kriterien:</p> <ul style="list-style-type: none"> ▪ Anpassung der allgemeinen tariflichen Kriterien an das Unternehmen ▪ Ausrichtung der Beurteilungen an die Profile der Beschäftigten ▪ Implementierung objektiver Kriterien ▪ Einführung einer pauschalen Zulage in Form einer jährlichen fixen Zahlung an die Beschäftigten <p>Beurteilungsprozess:</p> <ul style="list-style-type: none"> ▪ Objektivierung der Leistungsbeurteilung ▪ Abschaffung der tariflichen Mittelwertvorgabe, die zu einer geringen Streuung der Leistungsbeurteilungen führt ▪ Etablierung des Beurteilungsgesprächs als Dialog zwischen Führungskräften und Beschäftigten

Erläuterungen: Tabelle enthält zusammenfassend die Kritikpunkte und Änderungswünsche der befragten Experten hinsichtlich des praktizierten ERA-Leistungsbeurteilungsverfahrens.

Die Antworten der Führungskräfte verdeutlichen den Wunsch nach Verbesserungen am ERA-Leistungsbeurteilungsverfahren. Bemerkenswert ist, dass die Experten mehrheitlich das System kritisieren, jedoch nicht tendenziell dazu raten, das ERA-Beurteilungssystem gänzlich abzuschaffen. Beispielweise meint die Führungskraft aus der Entwicklung diesbezüglich: *„Ich würde das ERA auf keinen Fall abschaffen. Es ist ein schlechtes System, aber das Beste, glaube ich, was man wählen kann. (...) Ich kann mir ... vorstellen, dass es da beliebig Probleme geben kann“*.

Zusammenfassend lässt sich sagen, dass nicht nur das Vorgehen der Führungskräfte im Unternehmen zur tariflichen Leistungsbeurteilung unterschiedlich ist, sondern auch die Einstellungen gegenüber dem Verfahren. Die in dieser Arbeit aufgezeigten Kritikpunkte am ERA-Verfahren und insbesondere die geäußerten Änderungswünsche der beteiligten Experten sollen einen Denkanstoß über Neuerungen hinsichtlich der ERA-Leistungszulagen geben. Neuerungen sind erforderlich, da die Grundlage des heute praktizierten tariflichen Leistungsbeurteilungsverfahrens auf Konzepten beruht, die in den 1970er Jahren von den beteiligten Tarifparteien entwickelt wurden. Es lässt sich darüber diskutieren, ob die Bewertungskriterien der damaligen Zeit noch zeitgemäß sind und den heutigen Leistungsanforderungen an die Beschäftigten der M+E-Industrie gerecht werden.³⁰ Der § 10 Nr. 12 ERA-TV nennt zum Beispiel explizit die Möglichkeit, auf Basis einer freiwilligen Betriebsvereinbarung zwischen dem Arbeitgeber und dem Betriebsrat, ein anderes Beurteilungsverfahren einzuführen, das andere Kriterien als die tariflich empfohlenen enthält. Die Gewichtungen der einzelnen Merkmale können ebenfalls angepasst werden (siehe § 10 Nr. 13 ERA-TV).

Abgesehen von den Bewertungskriterien sollten auch Führungskräfte besser im Hinblick auf die vorhandenen Verzerrungseffekte geschult werden. Im Fallbetrieb ist zwar eine Schulung für neue Führungskräfte im Hinblick auf klassische Verzerrungseffekte, wie zum Beispiel Stereotype, obligatorisch, jedoch werden im Rahmen dieser Schulungen zum Beispiel nicht die geschlechterspezifischen Effekte angesprochen, die aufgrund unterschiedlicher Formalisierungsarten von Leistungsbeurteilungssystemen entstehen können. Die Ergebnisse dieser Arbeit bieten Evidenz dafür, dass auch der Formalisierungsgrad

³⁰ Beispielweise erwähnt Breisig (1989), dass in den 1950er und 1960er Jahren im Rahmen von Beurteilungsverfahren persönlichkeitsbezogene Merkmale im Vordergrund standen. Im Laufe der 1970er Jahre folgten eine Aufwertung von leistungs- und verhaltensbezogenen Merkmalen und eine Abwertung von Persönlichkeitsmerkmalen.

eines Bewertungssystems die Entstehung von Geschlechtereffekten bedingt und daher Gegenstand der Schulungen sein sollte. Auch der aufgedeckte Hierarchieeffekt, der sich in den Beurteilungen der Beschäftigten widerspiegelt, sollte als Bestandteil der Schulungen mit Vorgesetzten etabliert werden, sodass Beschäftigte nicht in Abhängigkeit ihrer organisationalen Position, sondern bezüglich ihrer tatsächlichen Leistung evaluiert werden.

Im Hinblick auf den Formalisierungsgrad und die geschlechterspezifischen Effekte ist auch empfehlenswert, die tarifliche Leistungsbeurteilung klarer von der außertariflichen zu trennen. Beispielsweise zeigen die Daten der Studie, dass die außertarifliche Zulage in Beziehung steht mit der tariflichen Leistungsbeurteilung. Damit die Führungskräfte zukünftig die beiden Beurteilungsdimensionen eindeutig voneinander trennen und nicht von Ankereffekten beeinflusst werden, ist beispielsweise notwendig, dass die tarifliche Leistungsbeurteilung nicht auf demselben Formular wie das Zielvereinbarungsrating bzw. die subjektiven Bewertungen des Arbeitsergebnisses abgebildet werden.

7.4 Limitationen und Anknüpfungspunkte für weitere Forschung

Die vorliegende Fallstudie hat einige Limitationen, auf die abschließend eingegangen werden soll, um zugleich Vorschläge für die Herangehensweise an zukünftige Forschungsarbeiten zu bieten.

An erster Stelle ist zu erwähnen, dass nicht eindeutig aufgeklärt werden konnte, was die Ursache für die in der Arbeit aufgedeckten Ankereffekte bildet, die durch den Formalisierungsgrad des Beurteilungssystems angestoßen werden und sich auf die Bewertungsunterschiede von Frauen und Männern auswirken. Den Fallstudienresultaten zufolge wirken zwei Ankereffekte, die in Zusammenhang stehen mit der tariflichen Leistungsbeurteilung der Beschäftigten. Zum einen bildet das Ergebnis der außertariflichen Zulage (im Fall der subjektiven Beurteilung der Quantität), zum anderen die Zuordnung der Beschäftigten zu einem formalisierten oder weniger formalisierten Bewertungssystem einen Anker, der einen Einfluss auf das Bewertungsverhalten der Führungskräfte hat. In Bezug auf die Wirkungsweise von Ankereffekten resümieren Belle, Cantarelli und Belardinelli (2017) in ihrem Übersichtsartikel über kognitive Verzerrungen in Leistungsbeurteilungen: „*The presence of anchoring effects in performance appraisals seems to hold across [different] types of ... anchors*“ (Belle, Cantarelli & Belardinelli, 2017: 281). Zukünftige Forschungsarbeiten sollten sich den Resultaten der vorliegenden Fallstudie zufolge den Wirkungsmechanismen von Ankern widmen, die im Rahmen von Leistungsbeurteilungen

wirken (Furnham & Boo, 2011) und zur Aufklärung von indirekten Geschlechtereffekten in Bewertungen beitragen, wie beispielweise der Formalisierungsgrad von Beurteilungssystemen.

Eine weitere Limitation der Fallstudie ist eine mögliche Endogenität des Faktors Formalisierungsgrad. Obwohl die Experteninterviews zeigen, dass der Zuordnung der Beschäftigten zu einem der beschriebenen Leistungsbeurteilungssysteme keine systematischen Mechanismen zugrunde liegen, ist nicht auszuschließen, dass die Beschäftigten hinsichtlich ihrer Wahl eines Beurteilungssystems eingeschränkt sind. Die Interviews mit den Führungskräften zeigen auf, dass die Zuordnung zu einem Beurteilungssystem durch mehrere Faktoren bestimmt wird und nicht stets frei ist. Obwohl die Ergebnisse einen möglichen Selektionseffekt nicht unterstützen, könnte es dennoch sein, dass sich leistungsfähigere Frauen in das formalisierte Beurteilungssystem hinein selektieren und vergleichsweise bessere Beurteilungen erhalten als Männer. Zukünftige Forschungsarbeiten sollten daher Geschlechtereffekte in Settings untersuchen, in der die Zuordnung der Beschäftigten zu einem formalisierten oder weniger formalisierten Bewertungssystem eindeutig zufällig ist.

Aufgrund der Tatsache, dass das Fallunternehmen in der männerdominierten M+E-Industrie angesiedelt ist, enthält der Datensatz relativ wenige Beobachtungen weiblicher Führungskräfte, die ihre Beschäftigten bewertet haben (Anzahl an Beobachtungen mit weiblichen Führungskräften $n=608$ im Vergleich zu $n=11.601$ Beobachtungen mit männlichen Vorgesetzten). Um den Effekt von Geschlechterdyaden auf die Leistungsbeurteilungen von Beschäftigten tiefergehend zu erforschen, wäre es vorteilhaft, weitere Fallstudien mit Unternehmen durchzuführen, in denen eine vergleichsweise hohe Anzahl an weiblichen Führungskräften die Leistungen ihrer weiblichen und männlichen Beschäftigten beurteilen.

Eine weitere Einschränkung dieser Studie stellt auch die Übertragbarkeit der Ergebnisse auf andere Unternehmen dar. Da das Fallunternehmen aus der männerdominierten M+E-Industrie stammt, ist fraglich, ob die aufgezeigten geschlechterspezifischen Einflüsse auch in frauendominierten Unternehmen bzw. Industrien auftreten. Um diese Frage zu beantworten, sind weitere Studien mit Paneldaten aus unterschiedlichen Unternehmen notwendig. Denkbar ist auch, dass die in dieser Arbeit festgestellten Geschlechtereffekte in anderen Unternehmen aus der M+E-Industrie abwesend sein können, da beispielsweise bestimmte organisationale Mechanismen, die eine Ungleichbehandlung zwischen Geschlech-

tern verstärken, nicht vorhanden sind. Da allerdings das Fallunternehmen typisch für seine Industrie ist, die praktizierten Leistungsbeurteilungen weit verbreitet sind und darüber hinaus für tarifvertragliche Unternehmen vorgeschrieben werden, ist das Vorhandensein von geschlechterspezifischen Effekten auch in weiteren Unternehmen zumindest nicht auszuschließen. Abgesehen vom Nachteil des Fallstudiencharakters der vorliegenden Untersuchung bietet sie den Vorteil, dass organisationale Mechanismen und deren Wirkungsweisen auf Geschlechtereffekte im Detail analysiert werden konnten und zum besseren Verständnis dieser Effekte beigetragen haben. Zukünftige Studien sollten Geschlechtereffekte in Unternehmen untersuchen, die zum Beispiel aus unterschiedlichen Industrien stammen und Unterschiede in Formalisierungsgraden ihrer Beurteilungssysteme oder weitere – als die in dieser Arbeit fokussierten – organisationale Mechanismen aufweisen.

Darüber hinaus ist nicht auszuschließen, dass die in der vorliegenden Fallstudie aufgedeckten Geschlechterunterschiede in Bewertungen auf tatsächliche Leistungsunterschiede zwischen Frauen und Männern zurückzuführen sind. Im Rahmen der Panelschätzungen wurden Leistungsunterschiede durch die Aufnahme der außertariflichen Zulage berücksichtigt. Diese außertarifliche Leistungszulage beinhaltet entweder das Ergebnis einer subjektiven Beurteilung des Arbeitsergebnisses oder einer vergleichsweise objektiven Bewertung von Zielen, die zwischen Beschäftigten und Führungskräften vereinbart wurden. Die subjektive Beurteilung des Arbeitsergebnisses als auch die Bewertung der Zielvereinbarung könnten verzerrt sein. Allerdings wurde im Rahmen dieser Dissertation aufgezeigt, dass ein möglicher Selektionseffekt, der zu den angesprochenen Verzerrungen führen könnte, tendenziell nicht zu bestätigen ist. Nichtsdestotrotz sollten zukünftige Untersuchungen alternative Leistungsmaße berücksichtigen, um den Einfluss von tatsächlichen Leistungsunterschieden oder Selektionseffekten auszuschließen.

8 Anhang

Anhangsverzeichnis

Anhang 1: Beschäftigte nach Tarifgebieten der M+E-Industrie	143
Anhang 2: ERA-Leistungsentgelte in der M+E-Industrie	144
Anhang 3: Gesprächsleitfaden der Experteninterviews	145
Anhang 4: Beschreibung der Entgeltgruppen am Beispiel ERA Thüringen.....	147
Anhang 5: Verteilung von Frauen und Männern in den Entgeltgruppen.....	148
Anhang 6: Deskriptive Statistiken (weniger formalisiertes Beurteilungssystem)	149
Anhang 7: Deskriptive Statistiken (formalisiertes Beurteilungssystem)	150
Anhang 8: Zusammenhang zwischen Alter und tariflicher Leistungszulage.....	151
Anhang 9: Effekte der Betriebszugehörigkeit auf die Wahl einer Zielvereinbarung.....	151
Anhang 10: Effekte der Entgeltgruppe auf die Wahl einer Zielvereinbarung	152
Anhang 11: Deskriptive Statistiken und Korrelationen (Logit-Schätzung).....	153
Anhang 12: Pauschale und ERA-Leistungszulage (Frauen).....	154
Anhang 13: Pauschale und ERA-Leistungszulage (Männer).....	154

Anhang 1: Beschäftigte nach Tarifgebieten der M+E-Industrie

Tarifgebiet	Anzahl der Beschäftigten
Baden-Württemberg	500.000
Bayern	475.000
Berlin	41.000
Berlin/Brandenburg	19.000
Hessen	213.000
Niedersachsen	66.500
NRW	693.000
Nordverbund	115.000
Osnabrück/Emsland	13.500
Pfalz/Rheinland-Rheinessen	116.000
Saarland	57.000
Sachsen	54.500
Sachsen-Anhalt	12.500
Thüringen	17.000
Gesamt:	2.392.500

Erläuterungen und Quelle: Das Tarifgebiet Nordverbund setzt sich zusammen aus den Gebieten: Hamburg, Mecklenburg-Vorpommern, Schleswig-Holstein, Nordwestliches Niedersachsen und Unterweser. In Anlehnung an IG Metall (2015).

Anhang 2: ERA-Leistungsentgelte in der M+E-Industrie

Tarifgebiet	Entgeltmethoden	Durchschnittliche(s) Leistungsentgelt/Leistungszulage
Baden-Württemberg	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	mindestens 15%, individuell 0 bis 30%
Bayern	Beurteilen/Leistungs-, Akkord-, Prämien-, Zielentgelt	Zeitentgelt 0 bis 28%
Berlin/Brandenburg	Beurteilen/Prämien-, Akkord-, Zielentgelt, Provision	Zeitentgelt 10%
Hessen	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10%
Niedersachsen	Beurteilen/Prämien-, Akkord-, Zielentgelt	mindestens 10% in drei Entgeltbereichen/mindestens 10%
Nordverbund	Beurteilen/Prämien-, Akkord-, Zielentgelt, Provision	Zeitentgelt mindestens 6%
NRW	Beurteilen/Prämie, Akkord, Zielvereinbarung I und II	Zeitentgelt 9 bis 11%/Leistungsentgelt mindestens 10%
Osnabrück-Emsland	Beurteilen/Prämien-, Akkord-, Zielentgelt	mindestens 10%
Pfalz	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10%
Rheinland- Rheinhessen	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10%
Saarland	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10%
Sachsen	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10 %
Sachsen-Anhalt	Beurteilen/Prämien-, Zielentgelt	Zeitentgelt mindestens 8%
Thüringen	Beurteilen/Kennzahlenvergleich, Zielvereinbarung	Zeitentgelt und Zielvereinbarung mindestens 10%

Quelle: In Anlehnung an IG Metall (2015). Tabelle bezieht sich auf Beschäftigte im Zeit-/Leistungsentgelt.

Anhang 3: Gesprächsleitfaden der Experteninterviews**Themenblock 1: ERA-Beurteilungen**

- 1) Wie läuft eine übliche ERA-Leistungsbeurteilung ab?
- 2) Die Mitarbeiter können ja zwischen einem weiteren Kriterium und einer Zielvereinbarung wählen.
 - a. Wovon hängt diese Wahl Ihrer Meinung nach ab?
 - b. Ist es dem/der Mitarbeiter/in frei überlassen, ob er/sie sich für oder gegen eine Zielvereinbarung entscheidet?
- 3) Der Tarifvertrag besagt ja, dass die Gesamtsumme der Leistungszulagen ca. 10% der tariflichen Monatsgrundentgeltsummen ausmacht. Gibt es jährlich Herausforderungen diese tarifliche Vorgabe für die variable Vergütung der Mitarbeiter/innen einzuhalten?
- 4) Wirken die Leistungsbeurteilungen Ihrer Meinung nach als Anreizinstrument, um die Leistung der Mitarbeiter/innen zu steigern?
- 5) Glauben Sie, dass die Leistungsbeurteilungen objektiv sind?
 - a. Wenn nicht, warum nicht?
 - b. Hängt die Objektivität davon ab, ob eine Zielvereinbarung gewählt wird oder nicht?

Themenblock 2: Interpretation der Fallstudienresultate

- 6) Die Ergebnisse zeigen Unterschiede zwischen Frauen und Männern. Mich würde interessieren, wie Sie diese Unterschiede erklären würden.
 - a. Frauen wählen eher eine Zielvereinbarung.
 - b. Frauen erhalten vergleichsweise bessere Leistungsbeurteilungen, wenn sie eine Zielvereinbarung wählen und schlechtere, wenn sie sich gegen eine Zielvereinbarung entscheiden.
 - c. Diese Unterschiede zeigen sich eher, wenn Männer beurteilen.
 - d. Frauen scheinen tendenziell strenger zu beurteilen.
- 7) Die Leistungsbeurteilung wird den Ergebnissen zufolge von weiteren Faktoren beeinflusst. Wie würden Sie diese Effekte erklären?

- a. Leistungsbeurteilungen ändern sich im Karriereverlauf: Neue Mitarbeiter werden tendenziell positiver beurteilt, langjährige negativer.
- b. Vorgesetzte sind am Anfang ihrer Laufbahn strenger und zum Ende ihrer Karriere eher milder in ihrem Bewertungsverhalten.
- c. Je höher die Hierarchiestufe des Mitarbeiters, desto besser ist die Leistungsbeurteilung.
- d. Je höher der Anteil hochqualifizierter Frauen in der Abteilung, desto besser fallen die Beurteilungen aus.
- e. Die Effekte scheinen abteilungsspezifisch: Existieren unterschiedliche Bewertungskulturen in den Abteilungen?

Themenblock 3: Ausblick

- 8) Was halten Sie von der Idee, die Mitarbeiter nicht variabel in Abhängigkeit ihrer Leistungsbeurteilung zu entlohnen, sondern eine Pauschalzulage an alle Arbeitnehmer zu verteilen?
- 9) Welche Änderungen würden Sie sich hinsichtlich der ERA-Beurteilungen wünschen?

Themenblock 4: Zur Person

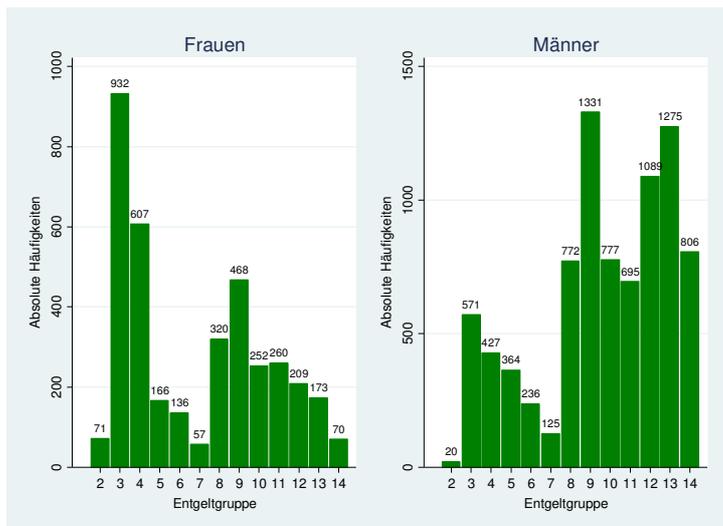
- 10) Was ist Ihre Funktion im Unternehmen?
- 11) Übernehmen Sie Führungsaufgaben?
- 12) Wie alt sind Sie und seit wann arbeiten Sie für den Betrieb?

Erläuterungen: Transkription der Experteninterviews wurde mittels des computergestützten Programms *f4transkript* durchgeführt.

Anhang 4: Beschreibung der Entgeltgruppen am Beispiel ERA Thüringen

EG	Beschreibung der Anforderungen
EG 1	Einfache Tätigkeiten, deren Ausführung und Ablauf festgelegt sind und die nach einer zweckgerichteten Einarbeitung und Übung von bis zu 4 Wochen verrichtet werden können. Es ist keine berufliche Vorbildung erforderlich.
EG 2	Tätigkeiten, deren Ausführung und Ablauf weitgehend festgelegt sind. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch ein systematisches Anlernen von bis zu 6 Monaten erworben werden.
EG 3	Tätigkeiten, deren Ausführung und Ablauf überwiegend festgelegt sind. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch ein systematisches Anlernen von mehr als 6 Monaten erworben werden.
EG 4	Tätigkeiten, deren Ausführung und Ablauf teilweise festgelegt sind. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine mindestens 2-jährige fachspezifische Berufsausbildung erworben werden.
EG 5	Fachspezifische Aufgaben oder Facharbeiten, deren Erledigung weitgehend festgelegt ist. Erforderlich sind Kenntnisse und Fertigkeiten wie sie in der Regel durch eine mindestens 3-jährige abgeschlossene fachspezifische Berufsausbildung erworben werden.
EG 6	Erweiterte fachspezifische Aufgaben oder schwierige Facharbeiten, deren Erledigung überwiegend festgelegt ist. Erforderlich sind Kenntnisse und Fertigkeiten wie sie in der Regel durch eine mindestens 3-jährige abgeschlossene fachspezifische Berufsausbildung und mehrjährige Berufserfahrung erworben werden.
EG 7	Umfassende fachspezifische Aufgaben im Rahmen eines Sachgebiets oder hochwertige Facharbeiten, deren Erledigung teilweise festgelegt ist. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine mindestens 3-jährige abgeschlossene fachspezifische Berufsausbildung und langjährige Berufserfahrung oder durch eine mindestens 3-jährige abgeschlossene fachspezifische Berufsausbildung und mehrjährige Berufserfahrung und eine mindestens 1-jährige spezifische berufliche Weiterbildung erworben werden.
EG 8	Ein Aufgabengebiet, das im Rahmen von bestimmten Richtlinien erledigt wird, oder hochwertigste Facharbeiten, die hohes Dispositionsvermögen und umfassende Verantwortung erfordern. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine abgeschlossene mindestens 3-jährige fachspezifische Berufsausbildung und eine mindestens 2-jährige Fachausbildung oder durch eine abgeschlossene 3-jährige Hochschulausbildung (zum Beispiel Bachelor) erworben werden.
EG 9	Ein erweitertes Aufgabengebiet, das überwiegend im Rahmen von bestimmten Richtlinien erledigt wird. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch eine mindestens 3-jährige Berufsausbildung, eine mindestens zweijährige Fachausbildung und langjährige Berufserfahrungen oder durch eine abgeschlossene 3-jährige Hochschulausbildung (zum Beispiel Bachelor) und langjährige Berufserfahrungen erworben werden.
EG 10	Ein umfassendes Aufgabengebiet, das teilweise im Rahmen von bestimmten Richtlinien erledigt wird. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch den Abschluss einer mindestens 4-jährigen Hochschulausbildung erworben werden.
EG 11	Ein Aufgabenbereich, der im Rahmen von allgemeinen Richtlinien erledigt wird. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch den Abschluss einer mindestens 4-jährigen Hochschulausbildung erworben werden, so wie Fachkenntnisse, die mehrjährige spezifische Berufserfahrung voraussetzen.
EG 12	Ein erweiterter Aufgabenbereich, der teilweise im Rahmen von allgemeinen Richtlinien erledigt wird. Erforderlich sind Kenntnisse und Fertigkeiten, wie sie in der Regel durch den Abschluss einer mindestens 4-jährigen Hochschulausbildung erworben werden sowie Fachkenntnisse, die langjährige spezifische Berufserfahrung voraussetzen.

Quelle: In Anlehnung an IG Metall (2014).

Anhang 5: Verteilung von Frauen und Männern in den Entgeltgruppen

Erläuterungen: Darstellungen beziehen sich auf die Beschäftigten. Darstellungen wurden mit STATA 13 erstellt.

Anhang 6: Deskriptive Statistiken (weniger formalisiertes Beurteilungssystem)

Variablen	Mittelwert	σ	Min.	Max.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 Mitarbeiterin	0,36	0,48	0,00	1,00	1,00																		
2 Frau beurteilt Frau	0,01	0,10	0,00	1,00	0,13	1,00																	
3 Frau beurteilt Mann	0,01	0,12	0,00	1,00	-0,09	-0,01	1,00																
4 Mann beurteilt Frau	0,35	0,48	0,00	1,00	0,98	-0,07	-0,09	1,00															
5 Alter Beschäftigter	4,37	1,02	1,90	6,70	0,25	0,02	-0,12	0,25	1,00														
6 Alter Beschäftigter ²	20,18	8,62	3,61	44,89	0,25	0,01	-0,11	0,24	0,99	1,00													
7 Betriebsz. Beschäftigter	1,58	1,00	0,10	5,00	0,20	0,00	-0,07	0,21	0,55	0,54	1,00												
8 Betriebsz. Beschäftigter ²	3,51	3,56	0,01	25,00	0,18	0,01	-0,06	0,18	0,53	0,53	0,96	1,00											
9 Entgeltgruppe Mitarbeiter/in	6,67	3,17	2,00	14,00	-0,40	0,01	0,08	-0,40	-0,30	-0,30	-0,03	-0,01	1,00										
10 Zulage Beurteilung Quantität	3,46	0,68	0,56	4,50	-0,04	0,05	0,03	-0,05	0,10	0,08	0,11	0,08	0,13	1,00									
11 Weibliche Führungskraft	0,02	0,15	0,00	1,00	0,01	0,62	0,78	-0,11	-0,08	-0,08	-0,05	-0,04	0,07	0,06	1,00								
12 Alter Führungskraft	4,40	0,68	2,50	6,20	0,00	-0,06	-0,04	0,01	0,06	0,06	0,01	0,02	0,06	-0,03	-0,07	1,00							
13 Alter Führungskraft ²	19,83	5,97	6,25	38,44	0,00	-0,05	-0,04	0,01	0,07	0,06	0,02	0,03	0,05	-0,02	-0,07	0,99	1,00						
14 Betriebsz. Führungskraft	2,00	1,09	0,10	4,70	-0,11	-0,07	0,01	-0,10	-0,07	-0,07	0,05	0,03	0,09	-0,01	-0,04	0,44	0,44	1,00					
15 Betriebsz. Führungskraft ²	5,17	4,20	0,01	22,09	-0,12	-0,08	0,00	-0,10	-0,05	-0,05	0,08	0,07	0,08	-0,01	-0,05	0,54	0,55	0,96	1,00				
16 Entgeltgruppe Führungskraft	12,75	1,13	3,00	16,00	-0,02	0,02	-0,02	-0,03	-0,05	-0,06	-0,08	-0,05	0,39	0,01	-0,01	0,26	0,23	0,00	0,00	1,00			
17 Frauenanteil	0,30	0,11	0,00	1,00	0,38	0,16	0,01	0,35	0,09	0,08	-0,07	-0,08	-0,30	-0,11	0,11	0,00	0,00	-0,15	-0,16	0,01	1,00		
18 Frauenanteil in EG 1 bis 6	0,18	0,16	0,00	9,20	0,23	-0,02	0,07	0,24	0,05	0,04	-0,02	-0,04	-0,28	-0,08	0,04	-0,06	-0,05	-0,01	-0,03	-0,13	0,51	1,00	
19 Frauenanteil in EG 12 bis 14	0,04	0,07	0,00	4,60	0,01	0,16	0,13	-0,03	-0,01	-0,02	-0,06	-0,04	0,19	0,01	0,20	0,06	0,06	-0,07	-0,07	0,22	0,06	0,34	1,00

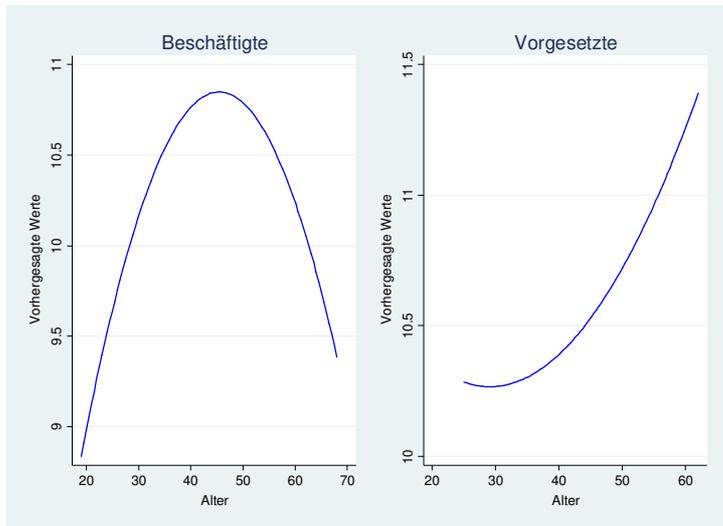
Erläuterungen: Die abgebildeten Statistiken beziehen sich auf n=6.646 Beobachtungen. Die Alters- und Betriebszugehörigkeitsvariablen (Mitarbeiter sowie Führungskräfte) wurden aus Interpretationsgründen mit dem Faktor 10 multipliziert.

Anhang 7: Deskriptive Statistiken (formalisiertes Beurteilungssystem)

Variablen	Mittelwert	σ	Min.	Max.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 Mitarbeiterin	0,24	0,43	0,00	1,00	1,00																		
2 Frau beurteilt Frau	0,04	0,20	0,00	1,00	0,38	1,00																	
3 Frau beurteilt Mann	0,04	0,19	0,00	1,00	-0,11	-0,04	1,00																
4 Mann beurteilt Frau	0,20	0,40	0,00	1,00	0,88	-0,10	-0,10	1,00															
5 Alter Beschäftigter	4,13	0,94	2,00	6,80	-0,03	-0,03	-0,02	-0,02	1,00														
6 Alter Beschäftigter ²	17,97	7,79	4,00	46,24	-0,04	-0,03	-0,02	-0,02	0,99	1,00													
7 Betriebsz. Beschäftigter	1,46	0,99	0,10	4,50	-0,04	-0,08	-0,06	0,00	0,57	0,57	1,00												
8 Betriebsz. Beschäftigter ²	3,09	3,43	0,01	20,25	-0,05	-0,08	-0,06	-0,01	0,57	0,57	0,96	1,00											
9 Entgeltgruppe Mitarbeiter/in	11,42	2,06	3,00	14,00	-0,33	-0,11	0,06	-0,30	0,14	0,12	0,00	0,00	1,00										
10 Zulage Beurteilung Zielvereinbarung	4,29	0,50	0,75	4,50	0,09	0,03	-0,04	0,09	0,01	0,01	0,03	0,01	-0,06	1,00									
11 Weibliche Führungskraft	0,08	0,27	0,00	1,00	0,20	0,70	0,68	-0,15	-0,04	-0,04	-0,10	-0,10	-0,04	0,00	1,00								
12 Alter Führungskraft	4,35	0,73	2,50	6,20	0,05	-0,05	-0,06	0,08	0,12	0,12	0,08	0,07	0,07	0,09	-0,08	1,00							
13 Alter Führungskraft ²	19,50	6,49	6,25	38,44	0,05	-0,05	-0,06	0,08	0,12	0,11	0,08	0,06	0,08	0,09	-0,08	1,00	1,00						
14 Betriebsz. Führungskraft	1,65	0,95	0,10	4,70	0,03	-0,03	-0,10	0,05	0,06	0,05	0,15	0,12	-0,10	0,14	-0,10	0,50	0,50	1,00					
15 Betriebsz. Führungskraft ²	3,64	3,53	0,01	22,09	0,02	-0,05	-0,10	0,05	0,07	0,07	0,16	0,13	-0,09	0,12	-0,11	0,54	0,55	0,96	1,00				
16 Entgeltgruppe Führungskraft	13,77	1,58	3,00	16,00	-0,05	-0,07	0,01	-0,01	0,02	0,02	-0,06	-0,04	0,25	-0,05	-0,04	0,20	0,20	0,12	0,10	1,00			
17 Frauenanteil	0,31	0,15	0,00	1,00	0,22	0,24	0,06	0,11	-0,04	-0,04	-0,17	-0,17	-0,13	-0,03	0,22	-0,01	-0,01	-0,09	-0,13	0,01	1,00		
18 Frauenanteil in EG 1 bis 6	0,13	0,28	0,00	9,20	-0,05	-0,03	-0,04	-0,03	-0,04	-0,04	-0,05	-0,05	-0,03	-0,08	-0,05	-0,02	-0,02	-0,04	-0,05	0,04	0,20	1,00	
19 Frauenanteil in EG 12 bis 14	0,09	0,15	0,00	4,60	0,10	0,11	0,08	0,06	0,00	0,00	-0,06	-0,06	0,07	-0,04	0,14	0,04	0,04	-0,03	-0,04	0,00	0,13	0,62	1,00

Erläuterungen: Die abgebildeten Statistiken beziehen sich auf n=5.429 Beobachtungen. Die Alters- und Betriebszugehörigkeitsvariablen (Mitarbeiter sowie Führungskräfte) wurden aus Interpretationsgründen mit dem Faktor 10 multipliziert.

Anhang 8: Zusammenhang zwischen Alter und tariflicher Leistungszulage



Erläuterungen: Darstellungen zeigen den nicht-linearen Zusammenhang zwischen dem Alter und der tariflichen Leistungszulage für Beschäftigte (links) sowie Vorgesetzte (rechts) in beiden Beurteilungssystemen. Darstellungen wurden mit STATA 13 erstellt.

Anhang 9: Effekte der Betriebszugehörigkeit auf die Wahl einer Zielvereinbarung

Betriebszugehörigkeit (Jahre)	Geschlechterunterschied	Standardfehler	z	P> z	95%-Konfidenzintervall
0	0,052	0,029	1,83	0,07	0,00 bis 0,11
5	0,053	0,016	3,23	0,00	0,02 bis 0,09
10	0,056	0,014	4,06	0,00	0,03 bis 0,08
15	0,061	0,015	4,06	0,00	0,03 bis 0,09
20	0,068	0,015	4,59	0,00	0,04 bis 0,10
25	0,077	0,014	5,34	0,00	0,05 bis 0,11
30	0,089	0,021	4,30	0,00	0,05 bis 0,13
35	0,102	0,036	2,84	0,00	0,03 bis 0,17

Erläuterungen: Schätzungen basieren auf 12.085 Beobachtungen. Der Geschlechterunterschied spiegelt die durchschnittlichen marginalen Effekte wider. Es werden robuste Standardfehler ausgegeben (Spalte 3).

Anhang 10: Effekte der Entgeltgruppe auf die Wahl einer Zielvereinbarung

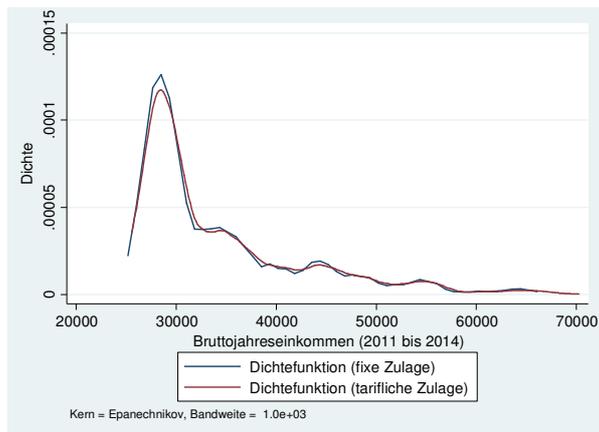
EG	Geschlechterunterschied	Standardfehler	Z	P> z	95 %-Konfidenzintervall
2	0,010	0,004	2,24	0,03	0,00 bis 0,02
3	0,017	0,006	2,62	0,01	0,00 bis 0,03
4	0,028	0,009	3,11	0,00	0,01 bis 0,05
5	0,044	0,012	3,71	0,00	0,02 bis 0,07
6	0,065	0,015	4,39	0,00	0,04 bis 0,09
7	0,087	0,017	5,10	0,00	0,05 bis 0,12
8	0,108	0,019	5,68	0,00	0,07 bis 0,15
9	0,123	0,021	5,86	0,00	0,08 bis 0,16
10	0,126	0,023	5,53	0,00	0,08 bis 0,17
11	0,113	0,023	4,91	0,00	0,07 bis 0,16
12	0,087	0,021	4,21	0,00	0,05 bis 0,13
13	0,059	0,017	3,55	0,00	0,03 bis 0,09
14	0,036	0,012	2,97	0,00	0,01 bis 0,06

Erläuterungen: Schätzungen basieren auf 12.085 Beobachtungen. Der Geschlechterunterschied spiegelt die durchschnittlichen marginalen Effekte wider. Es werden robuste Standardfehler ausgegeben (Spalte 3)

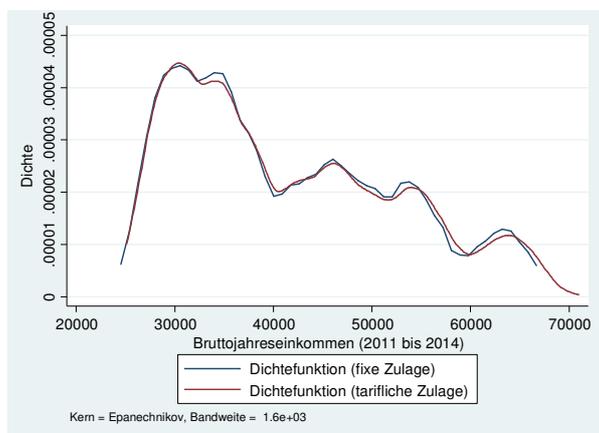
Anhang 11: Deskriptive Statistiken und Korrelationen (Logit-Schätzung)

Variablen	Mittelwert	σ	Min.	Max.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Mitarbeiterin	0,30	0,46	0,00	1,00	1,00														
2 Alter Beschäftigter	4,27	0,99	1,90	6,80	0,15	1,00													
3 Alter Beschäftigter ²	19,19	8,33	3,61	46,24	0,15	0,99	1,00												
4 Betriebsz. Beschäftigter	1,53	1,00	0,10	5,00	0,11	0,56	0,55	1,00											
5 Betriebsz. Beschäftigter ²	3,32	3,51	0,01	25,00	0,09	0,55	0,55	0,96	1,00										
6 Entgeltgruppe Mitarbeiter/in	8,79	3,61	2,00	14,00	-0,37	-0,20	-0,21	-0,06	-0,04	1,00									
7 Weibliche Führungskraft	0,05	0,22	0,00	1,00	0,09	-0,07	-0,07	-0,09	-0,08	0,10	1,00								
8 Alter Führungskraft	4,38	0,70	2,50	6,20	0,02	0,09	0,09	0,05	0,04	0,03	-0,08	1,00							
9 Alter Führungskraft ²	19,68	6,21	6,25	38,44	0,02	0,09	0,09	0,05	0,05	0,03	-0,07	0,99	1,00						
10 Betriebsz. Führungskraft	1,84	1,04	0,10	4,70	-0,03	0,00	0,00	0,10	0,08	-0,09	-0,09	0,46	0,46	1,00					
11 Betriebsz. Führungskraft ²	4,48	3,99	0,01	22,09	-0,04	0,02	0,02	0,12	0,11	-0,10	-0,10	0,53	0,54	0,96	1,00				
12 Entgeltgruppe Führungskraft	13,20	1,44	3,00	16,00	-0,08	-0,05	-0,06	-0,08	-0,06	0,45	0,02	0,20	0,19	0,00	-0,02	1,00			
13 Frauenanteil	0,30	0,13	0,00	1,00	0,29	0,02	0,02	-0,12	-0,12	-0,16	0,18	-0,01	0,00	-0,12	-0,14	0,01	1,00		
14 Frauenanteil in EG 1 bis 6	0,16	0,22	0,00	9,20	0,09	0,01	0,01	-0,03	-0,04	-0,19	-0,04	-0,03	-0,03	-0,01	-0,02	-0,06	0,30	1,00	
15 Frauenanteil in EG 12 bis 14	0,06	0,11	0,00	4,60	0,03	-0,03	-0,03	-0,07	-0,06	0,22	0,17	0,04	0,04	-0,07	-0,08	0,13	0,11	0,50	1,00

Erläuterungen: Die abgebildeten Statistiken beziehen sich auf n=12.085 Beobachtungen. Die Alters- und Betriebszugehörigkeitsvariablen (Mitarbeiter sowie Führungskräfte) wurden aus Interpretationsgründen mit dem Faktor 10 multipliziert.

Anhang 12: Pauschale und ERA-Leistungszulage (Frauen)

Erläuterungen: Kerndichteschätzungen basieren auf den Bruttojahreseinkommen der Frauen inklusive fixer Zulage (blaue Funktion) und inklusive tariflicher Zulage (rote Funktion). Darstellung wurde mit STATA 13 erstellt.

Anhang 13: Pauschale und ERA-Leistungszulage (Männer)

Erläuterungen: Kerndichteschätzungen basieren auf den Bruttojahreseinkommen der Männer inklusive fixer Zulage (blaue Funktion) und inklusive tariflicher Zulage (rote Funktion). Darstellung wurde mit STATA 13 erstellt.

Literaturverzeichnis

- Abraham, M. 2017. Pay formalization revisited: considering the effects of manager gender and discretion on closing the gender wage gap. *Academy of Management Journal*, 60(1): 29–54.
- Acker, J. 1990. Hierarchies, jobs, bodies: a theory of gendered organizations. *Gender and Society*, 4(2): 139–158.
- Addison, J. T. & Belfield, C. R. 2008. The determinants of performance appraisal systems: a note (Do Brown and Heywood's results for Australia hold up for Britain?). *British Journal of Industrial Relations*, 46(3): 521–531.
- Agars, M. D. 2004. Reconsidering the impact of gender stereotypes on the advancement of women in organizations. *Psychology of Women Quarterly*, 28(2): 103–111.
- Ahn, T. S., Hwang, I. & Kim, M.-I. 2010. The impact of performance measure discriminability on ratee incentives. *The Accounting Review*, 85(2): 389–417.
- Anderson, C. D. & Tomaskovic-Devey, D. 1995. Patriarchal pressures: an exploration of organizational processes that exacerbate and erode gender earnings inequality. *Work and Occupations*, 22(3): 328–356.
- Arni, S., Bigler-Wälti, M., Hösli, M., Kessler, M. & Morandi, C. 2002. *Geschlechts- und kulturspezifische Unterschiede in der Mitarbeitendenbeurteilung?*
http://www.femtech.at/sites/default/files/Geschlechtsspezifische_Unterschiede_in_der_Mitarbeitendenbeurteilung.pdf. Zuletzt geprüft am 28. November 2017.
- Arrow, K. J. 1971. *The theory of discrimination*. Working Paper No. 30A. Princeton University: Princeton.
- Arvey, R. D. 1979. Unfair discrimination in the employment interview: legal and psychological aspects. *Psychological Bulletin*, 86(4): 736–765.
- Arvey, R. D. & Murphy, K. R. 1998. Performance evaluation in work settings. *Annual Review of Psychology*, 49(1): 141–168.
- Ashforth, B. E. & Mael, F. 1989. Social identity theory and the organization. *The Academy of Management Review*, 14(1): 20–39.
- Auster, E. R. 1989. Task characteristics as a bridge between macro- and microlevel research on salary inequality between men and women. *The Academy of Management*

- Review*, 14(2): 173–193.
- Ayers, R. S. 2015. Aligning individual and organizational performance: goal alignment in federal government agency performance appraisal programs. *Public Personnel Management*, 44(2): 169–191.
- Bahn Müller, R. & Hoppe, M. 2016a. Ten years after – Langzeiteffekte von ERA in der baden-württembergischen Metall- und Elektroindustrie. *WSI Mitteilungen*, 2016(8): 631–640.
- Bahn Müller, R. & Hoppe, M. 2016b. „Wenn der Staub sich gelegt hat“. *Die Langzeitwirkung von ERA in der Metall- und Elektroindustrie Baden-Württemberg: Ergebnisse der Befragung von Betriebsräten und Personalverantwortlichen im Herbst 2015*. http://www.fatk.uni-tuebingen.de/files/era_ii-befragung_das_wichtigste_in_k__r__ze.pdf. Zuletzt geprüft am 28. November 2017.
- Bahn Müller, R. & Schmidt, W. 2009. *Riskante Modernisierung des Tarifsystems: Die Reform der Entgelttarifverträge am Beispiel der Metall- und Elektroindustrie Baden-Württembergs*. Berlin: edition sigma.
- Bähring, K., Hauff, S., Sossdorf, M. & Thommes, K. 2008. Methodologische Grundlagen und Besonderheiten der qualitativen Befragung von Experten in Unternehmen: Ein Leitfaden. *Die Unternehmung*, 92(1): 89–111.
- Banaji, M. R., Hardin, C. & Rothman, A. J. 1993. Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, 65(2): 272–281.
- Banaji, M. R. & Hardin, C. D. 1996. Automatic stereotyping. *Psychological Science*, 7(3): 136–141.
- Barnett, W. P., Baron, J. N. & Stuart, T. E. 2000. Avenues of attainment: occupational demography and organizational careers in the California Civil Service. *American Journal of Sociology*, 106(1): 88–144.
- Baron, J. N., Davis-Blake, A. & Bielby, W. T. 1986. The structure of opportunity: how promotion ladders vary within and among organizations. *Administrative Science Quarterly*, 31(2): 248–273.
- Baron, J. N., Hannan, M. T., Hsu, G. & Koçak, Ö. 2007. In the company of women: gender inequality and the logic of bureaucracy in start-up firms. *Work and Occupations*, 34(1): 35–66.

- Baron, J. N. & Newman, A. E. 1990. For what it's worth: organizations, occupations, and the value of work done by women and nonwhites. *American Sociological Review*, 55(2): 155–175.
- Bayo-Moriones, A., Galdon-Sanchez, J. E. & Martinez-de-Morentin, S. 2016. *Competitive strategy, performance appraisal and firm results*. Discussion Paper No. 10041. Forschungsinstitut zur Zukunft der Arbeit (IZA): Bonn.
- Becker, G. S. 1971. *The economics of discrimination* (2. Aufl.). Chicago: University of Chicago Press.
- Belle, N., Cantarelli, P. & Belardinelli, P. 2017. Cognitive biases in performance appraisal: experimental evidence on anchoring and halo effects with public sector managers and employees. *Review of Public Personnel Administration*, 37(3): 275–294.
- Bento, R. F., White, L. F. & Zacur, S. R. 2012. The stigma of obesity and discrimination in performance appraisal: a theoretical model. *The International Journal of Human Resource Management*, 23(15): 3196–3224.
- Berger, J., Harbring, C. & Sliwka, D. 2013. Performance appraisals and the impact of forced distribution: an experimental investigation. *Management Science*, 59(1): 54–68.
- Bielby, W. T. 2008. Applying social research on stereotyping and cognitive bias to employment discrimination litigation: the case of allegations of systematic gender bias at Wal-Mart stores. In L. B. Nielsen & R. L. Nelson (Hg.), *Handbook of employment discrimination research: rights and realities*: 395–408. New York: Springer-Verlag.
- Biernat, M. & Ma, J. E. 2005. Stereotypes and the confirmability of trait concepts. *Personality and Social Psychology Bulletin*, 31(4): 483–495.
- Bigoness, W. J. 1976. Effect of applicant's sex, race, and performance on employers' performance ratings: some additional findings. *Journal of Applied Psychology*, 61(1): 80–84.
- Blair, I. V. & Banaji, M. R. 1996. Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, 70(6): 1142–1163.
- Blau, F. D. & DeVaro, J. 2007. New evidence on gender differences in promotion rates: an empirical analysis of a sample of new hires. *Industrial Relations*, 46(3): 511–550.

- Blume, B. D., Baldwin, T. T. & Rubin, R. S. 2009. Reactions to different types of forced distribution performance evaluation systems. *Journal of Business and Psychology*, 24(1): 77–91.
- Blume, B. D., Rubin, R. S. & Baldwin, T. T. 2013. Who is attracted to an organisation using a forced distribution performance management system? *Human Resource Management Journal*, 23(4): 360–378.
- Blume, J., Seyerlein, C. & Fröndhoff, B. 2016. Sechs, setzen! *Handelsblatt*, 2016(187): 14–15.
- BMAS (Bundesministerium für Arbeit und Soziales) 2013. *Erster Zwischenbericht im Projekt: Arbeitsqualität und wirtschaftlicher Erfolg – Längsschnittstudie in deutschen Betrieben*: http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/Forschungsberichte/forschungsbericht-fb-442-arbeitsqualitaet.pdf?__blob=publicationFile. Zuletzt geprüft am 28. November 2017.
- BMBF (Bundesministerium für Bildung und Forschung) 2014. *Bildung in Deutschland 2014*. <https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2014/pdf-bildungsbericht-2014/bb-2014.pdf>. Zuletzt geprüft am 28. November 2017.
- Böhner, H. & Ziepke, J. 1972. *Handkommentar zum Lohnrahmenabkommen vom 26.9.1967/bzw. 15.4.1970 (Teile I und II) sowie zum Tarifvertrag zur Leistungsbeurteilung von Zeitlohnarbeitern vom 7.9.1970* (2. Aufl.). Düsseldorf: Rheindruck.
- Böhner, H. & Ziepke, J. 1975. *Handkommentar zum Tarifvertrag zur Leistungsbeurteilung von Angestellten vom 19. 2. 1975*. Wuppertal: W. Girardet.
- Bol, J. C. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5): 1549–1575.
- Bol, J. C. & Smith, S. D. 2011. Spillover effects in subjective performance evaluation: bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4): 1213–1230.
- Boll, C., Leppin, J., Rossen, A. & Wolf, A. 2016. *Magnitude and impact factors of the gender pay gap in EU countries*. Luxemburg: Publication Office of the European Union.
- Booth, A. L., Francesconi, M. & Frank, J. 2003. A sticky floors model of promotion, pay, and gender. *European Economic Review*, 47(2): 295–322.

- Bowen, C.-C., Swim, J. K. & Jacobs, R. R. 2000. Evaluating gender biases on actual job performance of real people: a meta-analysis. *Journal of Applied Social Psychology*, 30(10): 2194–2215.
- Breisig, T. 1989. *Personalbeurteilung als Führungsinstrument. Eine betriebliche Sozialtechnik – wie sie funktioniert und wie man sie pariert*. Berlin: edition sigma.
- Brenner, O. C., Tomkiewicz, J. & Schein, V. E. 1989. The relationship between sex role stereotypes and requisite management characteristics revisited. *Academy of Management Journal*, 32(3): 662–669.
- Bretz, R. D., Milkovich, G. T. & Read, W. 1992. The current state of performance appraisal research and practice: concerns, directions, and implications. *Journal of Management*, 18(2): 321–352.
- Breuer, K., Nieken, P. & Sliwka, D. 2013. Social ties and subjective performance evaluations: an empirical investigation. *Review of Managerial Science*, 7(2): 141–157.
- Breuer, K. & Zimmermann, J.-H. 2010. *Determinants and effects of performance appraisal and target agreement systems: an empirical investigation of German firms*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.8444&rep=rep1&type=pdf>. Zuletzt geprüft am 28. November 2017.
- Breuer, K. A. 2011. *The evaluation of human resource management practices – empirical studies on subjective performance evaluation, target agreements, and intra-firm trainings*. Dissertation. Universität zu Köln: Köln.
- Brewer, M. B. & Miller, N. 1984. Beyond the contact hypothesis: theoretical perspectives on desegregation. In N. Miller & M. B. Brewer (Hg.), *Groups in contact: the psychology of desegregation*: 281–302. Orlando: Academic Press.
- Brief, A. P. & Wallace, M. J. 1976. The impact of employee sex and performance on the allocation of organizational rewards. *The Journal of Psychology*, 92(1): 25–34.
- Brown, M. & Heywood, J. S. 2005. Performance appraisal systems: determinants and change. *British Journal of Industrial Relations*, 43(4): 659–679.
- Brown, M., Hyatt, D. & Benson, J. 2010. Consequences of the performance appraisal experience. *Personnel Review*, 39(3): 375–396.
- Brutus, S., Fletcher, C. & Baldry, C. 2009. The influence of independent self-construal on

- rater self-efficacy in performance appraisal. *The International Journal of Human Resource Management*, 20(9): 1999–2011.
- Butler, A. B. & Skattebo, A. 2004. What is acceptable for women may not be for men: the effect of family conflicts with work on job-performance ratings. *Journal of Occupational and Organizational Psychology*, 77(4): 553–564.
- Byrne, D. & Neuman, J. H. 1992. The implications of attraction research for organizational issues. In K. Kelley (Hg.), *Issues, theory, and research in industrial/organizational psychology*: 29–70 (82. Aufl.). North-Holland Publishing: Amsterdam.
- Byrne, D. 1971. *The attraction paradigm*. New York: Academic Press.
- Capozza, D. & Brown, R. 2000. *Social identity processes: trends in theory and research*. Thousand Oaks: Sage Publications.
- Cappelli, P. & Conyon, M. 2016. *What do performance appraisals do?* NBER Working Paper No. 22400. National Bureau of Economic Research: Cambridge.
- Carli, L. L. & Eagly, A. H. 1999. Gender effects on social influence and emergent leadership. In G. N. Powell (Hg.), *Handbook of gender and work*: 203–221. Thousand Oaks: Sage Publications.
- Castilla, E. J. 2008. Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 113(6): 1479–1526.
- Charles, M. T. 1981. The performance and socialization of female recruits in the Michigan State Police Training Academy. *Journal of Police Science and Administration*, 9(2): 209–223.
- Chicha, M.-T. 2006. *A comparative analysis of promoting pay equity: models and impacts*. Working Paper No. 49. International Labour Office (ILO): Genf.
- Cobb-Clark, D. A. 2001. Getting ahead: the determinants of and payoffs to internal promotion for young U.S. men and women. In S. Polachek (Hg.), *Worker wellbeing in a changing labor market*: 339–372 (20. Aufl.). Bingley: Emerald Group Publishing Limited.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Erlbaum.

- Cohen, L. E. & Broschak, J. P. 2013. Whose jobs are these? The impact of the proportion of female managers on the number of new management jobs filled by women versus men. *Administrative Science Quarterly*, 58(4): 509–541.
- Cohen, L. E., Broschak, J. P. & Haveman, H. A. 1998. And then there were more? The effect of organizational sex composition on the hiring and promotion of managers. *American Sociological Review*, 63(5): 711–727.
- Dansereau, F., Graen, G. & Haga, W. J. 1975. A vertical dyad linkage approach to leadership within formal organizations. *Organizational Behavior and Human Performance*, 13(1): 46–78.
- Davison, H. K. & Burke, M. J. 2000. Sex discrimination in simulated employment contexts: a meta-analytic investigation. *Journal of Vocational Behavior*, 56(2): 225–248.
- Deaux, K. 1995. How basic can you be? The evolution of research on gender stereotypes. *Journal of Social Issues*, 51(1): 11–20.
- Deaux, K. & Kite, M. 1993. Gender stereotypes. In F. L. Denmark & M. A. Paludi (Hg.), *Psychology of women: a handbook of issues and theories*: 107–139. Westport: Greenwood.
- Dencker, J. C. 2008. Corporate restructuring and sex differences in managerial promotion. *American Sociological Review*, 73(3): 455–476.
- Destatis (Statistisches Bundesamt) 2013. *Hochschulen auf einen Blick*.
https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/BroschuereHochschulenBlick0110010137004.pdf?__blob=publicationFile. Zuletzt geprüft am 28. November 2017.
- Destatis (Statistisches Bundesamt) 2017. *Drei Viertel des Gender Pay Gap lassen sich mit Strukturunterschieden erklären*.
https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2017/03/PD17_094_621.html. Zuletzt geprüft am 28. November 2017.
- Devine, P. G. 1989. Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1): 5–18.
- Diaye, M.-A. & Greenan, N. 2012. *The economics of performance appraisals*. Working Paper No. 2012-15. University of Evry Val d'Essonne: Evry.

- Dienesch, R. M. & Liden, R. C. 1986. Leader-member exchange model of leadership: a critique and further development. *Academy of Management Review*, 11(3): 618–634.
- Drogosz, L. M. & Levy, P. E. 1996. Another look at the effects of appearance, gender, and job type on performance-based decisions. *Psychology of Women Quarterly*, 20(3): 437–445.
- Dusterhoff, C., Cunningham, J. B. & MacGregor, J. N. 2014. The effects of performance rating, leader-member exchange, perceived utility, and organizational justice on performance appraisal satisfaction: applying a moral judgment perspective. *Journal of Business Ethics*, 119(2): 265–273.
- Eagly, A. H. 1987. *Sex differences in social behavior: a social-role interpretation*. Hillsdale: Erlbaum.
- Eagly, A. H. & Karau, S. J. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3): 573–598.
- Eagly, A. H., Karau, S. J. & Makhijani, M. G. 1995. Gender and the effectiveness of leaders: a meta-analysis. *Psychological Bulletin*, 117(1): 125–145.
- Eagly, A. H., Makhijani, M. G. & Klonsky, B. G. 1992. Gender and the evaluation of leaders: a meta-analysis. *Psychological Bulletin*, 111(1): 3–22.
- Eagly, A. H. & Mladinic, A. 1989. Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15(4): 543–558.
- Ederer, F. 2010. Feedback and motivation in dynamic tournaments. *Journal of Economics and Management Strategy*, 19(3): 733–769.
- Elicker, J. D., Levy, P. E. & Hall, R. J. 2006. The role of leader-member exchange in the performance appraisal process. *Journal of Management*, 32(4): 531–551.
- Elliott, J. R. & Smith, R. A. 2004. Race, gender, and workplace power. *American Sociological Review*, 69(3): 365–386.
- Elvira, M. M. & Graham, M. E. 2002. Not just a formality: pay system formalization and sex-related earnings effects. *Organization Science*, 13(6): 601–617.
- Ely, R. J. 1994. The effects of organizational demographics and social identity on relationships among professional women. *Administrative Science Quarterly*, 39(2): 203–238.

- Epley, N. & Gilovich, T. 2001. Putting adjustment back in the anchoring and adjustment heuristic: differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5): 391–396.
- Eremin, D. V., Wolf, J. F. & Woodard, C. A. 2010. Systemic bias in federal performance evaluations: does hierarchy trump a performance management process? *Public Performance and Management Review*, 34(1): 7–21.
- Etaugh, C. & Kasley, H. C. 1981. Evaluating Competence: effects of sex, marital status, and parental status. *Psychology of Women Quarterly*, 6(2): 196–203.
- Fabel, O., Thomann, C. J. & Zhang, Y. 2014. *Pay inequity and job performance: an insider-econometrics study*.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350770. Zuletzt geprüft am 28. November 2017.
- Farndale, E. & Kelliher, C. 2013. Implementing performance appraisal: exploring the employee experience. *Human Resource Management*, 52(6): 879–897.
- Feldman, J. M. 1981. Beyond attribution theory: cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66(2): 127–148.
- Fernandez, R. M. & Weinberg, N. 1997. Sifting and sorting: personal contacts and hiring in a retail bank. *American Sociological Review*, 62(6): 883–902.
- Ferris, G. R., Munyon, T. P., Basik, K. & Buckley, M. R. 2008. The performance evaluation context: social, emotional, cognitive, political, and relationship components. *Human Resource Management Review*, 18(3): 146–163.
- Fiske, S. T. 1998. Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Hg.), *The handbook of social psychology*: 357–411 (4. Aufl.). New York: McGraw-Hill.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K. & Heilman, M. E. 1991. Social science research on trial: use of sex stereotyping research in Price Waterhouse v. Hopkins. *American Psychologist*, 46(10): 1049–1060.
- Fiske, S. T., Cuddy, A. J. C., Glick, P. & Xu, J. 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6): 878.

- Fiske, S. T. & Neuberg, S. L. 1994. A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. In M. Zanna (Hg.), *Advances in experimental social psychology*: 1–74 (23. Aufl.). San Diego: Academic Press.
- Frank, F. D. & Drucker, J. 1977. The influence of evaluatee's sex on evaluations of a response on a managerial selection instrument. *Sex Roles*, 3(1): 59–64.
- Frederiksen, A., Lange, F. & Kriechel, B. 2017. Subjective performance evaluations and employee careers. *Journal of Economic Behavior and Organization*, 134(1): 408–429.
- Frick, B. J., Goetzen, U. & Simmons, R. 2013. The hidden costs of high-performance work practices: evidence from a large german steel company. *Industrial and Labor Relations Review*, 66(1): 198–224.
- Furnham, A. & Boo, H. C. 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1): 35–42.
- Furnham, A. & Stringfield, P. 2001. Gender differences in rating reports: female managers are harsher raters, particularly of males. *Journal of Managerial Psychology*, 16(4): 281–288.
- George, A. L. & Bennett, A. 2005. *Case studies and theory development in the social sciences*. Cambridge etc.: MIT Press.
- Gerdes, E. P. & Garber, D. M. 1983. Sex bias in hiring: effects of job demands and applicant competence. *Sex Roles*, 9(3): 307–319.
- Gesamtmetall 2011. *Frauen in Führungspositionen in der M+E-Industrie: Ergebnisse einer Unternehmensbefragung im April/Mai 2011*.
<https://www.gesamtmetall.de/sites/default/files/downloads/frauen-in-fuehrungspositionen.pdf>. Zuletzt geprüft am 28. November 2017.
- Gesamtmetall 2016. *Zahlen 2016: Die Metall- und Elektro-Industrie in der Bundesrepublik Deutschland*. https://www.gesamtmetall.de/sites/default/files/downloads/me-zahlenheft_2016.pdf. Zuletzt geprüft am 28. November 2017.
- Gibbs, M., Merchant, K. A., Van der Stede, W. A. & Vargus, M. E. 2004. Determinants and effects of subjectivity in incentives. *The Accounting Review*, 79(2): 409–436.

- Glick, P. 1991. Trait-based and sex-based discrimination in occupational prestige, occupational salary, and hiring. *Sex Roles*, 25(5): 351–378.
- Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., Adetoun, B., Osagie, J. E., Akande, A., Alao, A., Annetje, B., Willemsen, T. M., Chipeta, K., Dardenne, B., Dijksterhuis, A., Wigboldus, D., Eckes, T., Six-Materna, I., Expósito, F., Moya, M., Foddy, M., Kim, H.-J., Lameiras, M., Sotelo, M. J., Mucchi-Faina, A., Romani, M., Sakalli, N., Udegbe, B., Yamamoto, M., Ui, M., Ferreira, M. C. & López, W. L. 2000. Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, 79(5): 763–775.
- Golman, R. & Bhatia, S. 2012. Performance evaluation inflation and compression. *Accounting, Organizations and Society*, 37(8): 534–543.
- Goodwin, S. A., Gubin, A., Fiske, S. T. & Yzerbyt, V. Y. 2000. Power can bias impression processes: stereotyping subordinates by default and by design. *Group Processes and Intergroup Relations*, 3(3): 227–256.
- Goodwin, S. A., Operario, D. & Fiske, S. T. 1998. Situational power and interpersonal dominance facilitate bias and inequality. *Journal of Social Issues*, 54(4): 677–698.
- Gordon, R. A. & Arvey, R. D. 2004. Age bias in laboratory and field settings: a meta-analytic investigation. *Journal of Applied Social Psychology*, 34(3): 468–492.
- Gorman, E. H. 2005. Gender stereotypes, same-gender preferences, and organizational variation in the hiring of women: evidence from law firms. *American Sociological Review*, 70(4): 702–728.
- Graen, G. B., Liden, R. C. & Hoel, W. 1982. Role of leadership in the employee withdrawal process. *Journal of Applied Psychology*, 67(6): 868–872.
- Graen, G. B. & Scandura, T. A. 1987. Toward a psychology of dyadic organizing. *Research in Organizational Behavior*, 9: 175–208.
- Graves, L. M. & Powell, G. N. 1995. The effect of sex similarity on recruiters' evaluations of actual applicants: a test of the similarity-attraction paradigm. *Personnel Psychology*, 48(1): 85–98.
- Graves, L. M. & Powell, G. N. 1996. Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational and Organizational Psychology*, 69(3): 243–261.

- Grean, G. 1976. Role making processes within complex organizations. In M. D. Dunnette (Hg.), *Handbook of industrial and organizational psychology*: 1201–1245. Chicago: Rand McNally.
- Grean, G. & Cashman, J. 1975. A role-making model of leadership in formal organizations: a developmental approach. In J. G. Hunt & L. L. Larson (Hg.), *Leadership frontiers*: 143–165. Kent: Kent State University Press.
- Greene, W. H. 2012. *Econometric analysis* (7. Aufl.). Boston: Prentice Hall.
- Grund, C. & Przemeczek, J. 2012. Subjective performance appraisal and inequality aversion. *Applied Economics*, 44(17): 2149–2155.
- Grund, C. & Sliwka, D. 2009. The anatomy of performance appraisals in Germany. *The International Journal of Human Resource Management*, 20(10): 2049–2065.
- Hamner, W. C., Kim, J. S., Baird, L. & Bigoness, W. J. 1974. Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. *Journal of Applied Psychology*, 59(6): 705–711.
- Hausmann, A.-C. & Kleinert, C. 2014. *Berufliche Segregation auf dem Arbeitsmarkt: Männer- und Frauendomänen kaum verändert (IAB-Kurzbericht 9/2014)*. <http://doku.iab.de/kurzber/2014/kb0914.pdf>. Zuletzt geprüft am 28. November 2017.
- Heilman, M. E. 1983. Sex bias in work settings: the lack of fit model. *Research in Organizational Behavior*, 5: 269–298.
- Heilman, M. E. 1994. Affirmative action: some unintended consequences for working women. In B. Straw & L. Cummings (Hg.), *Research in organizational behavior*: 125–169. Greenwich: JAI Press.
- Heilman, M. E., Block, C. J. & Martell, R. F. 1995. Sex stereotypes: do they influence perceptions of managers? *Journal of Social Behavior and Personality*, 10(4): 237.
- Heilman, M. E. & Eagly, A. H. 2008. Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology*, 1(4): 393–398.
- Heilman, M. E. & Haynes, M. C. 2005. No credit where credit is due: attributional rationalization of women's success in male-female teams. *Journal of Applied Psychology*, 90(5): 905–916.

- Heywood, J. S. & Jirjahn, U. 2002. Payment schemes and gender in Germany. *Industrial and Labor Relations Review*, 56(1): 44–64.
- Hill, A. D., Upadhyay, A. D. & Beekun, R. I. 2015. Do female and ethnically diverse executives endure inequity in the CEO position or do they benefit from their minority status? An empirical examination. *Strategic Management Journal*, 36(8): 1115–1134.
- Hilton, J. L. & Hippel, W. Von 1996. Stereotypes. *Annual Review of Psychology*, 47(1): 237–271.
- Hmurovic, J. L. 2012. *Faulted when females? Gender-biased evaluative judgements and attributional rationalization in response to leader failure*. Masterarbeit. Michigan State University: East Lansing.
- Huffman, M. L., Cohen, P. N. & Pearlman, J. 2010. Engendering change: organizational dynamics and workplace gender desegregation, 1975-2005. *Administrative Science Quarterly*, 55(2): 255–277.
- Huffman, M. L. & Velasco, S. C. 1997. When more is less: sex composition, organizations, and earnings in U.S. firms. *Work and Occupations*, 24(2): 214–244.
- IG Metall (Industriegewerkschaft Metall) 2014. *Daten, Fakten, Informationen* (43. Aufl.). Dresden: Neue Druckhaus Dresden.
- IG Metall (Industriegewerkschaft Metall) 2015. *Daten, Fakten, Informationen* (44. Aufl.). Dresden: Neue Druckhaus Dresden.
- IG Metall (Industriegewerkschaft Metall) 2017. *Daten, Fakten, Informationen* (46. Aufl.). Dresden: Neue Druckhaus Dresden.
- Ilgen, D. R. & Youtz, M. A. 1986. Factors affecting the evaluation and development of minorities in organizations. In K. M. Rowland & G. R. Ferris (Hg.), *Research in personnel and human resources management*: 307–337 (4. Aufl.). Greenwich: JAI Press.
- Ito, T. A. & Urland, G. R. 2003. Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85(4): 616–626.
- Jennings, T., Palmer, J. K. & Thomas, A. 2004. Effects of performance context on processing speed and performance ratings. *Journal of Business and Psychology*, 18(4): 453–

463.

Jirjahn, U. & Poutsma, E. 2013. The use of performance appraisal systems: evidence from Dutch establishment data. *Industrial Relations*, 52(4): 801–828.

Jirjahn, U. & Stephan, G. 2004. Gender, piece rates and wages: evidence from matched employer-employee data. *Cambridge Journal of Economics*, 28(5): 683–704.

Jochmann-Döll, A. & Ranftl, E. 2010. *Impulse für die Entgeltgleichheit: Die ERA und ihre betriebliche Umsetzung auf dem gleichstellungspolitischen Prüfstand*. Berlin: edition sigma.

Johnson, J. W. & Cochran, C. C. 2008. Studying the influence of stereotypes on personnel decisions in the real world. *Industrial and Organizational Psychology*, 1(4): 423–425.

Joshi, A., Jackson, S. E. & Liao, H. 2006. Cross-level effects of workplace diversity on sales performance and pay. *Academy of Management Journal*, 49(3): 459–481.

Joshi, A., Son, J. & Roh, H. 2015. When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, 58(5): 1516–1545.

Jost, J. T. & Banaji, M. R. 1994. The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, 33(1): 1–27.

Kalev, A., Dobbin, F. & Kelly, E. 2006. Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, 71(4): 589–617.

Kampkötter, P. 2017. Performance appraisals and job satisfaction. *The International Journal of Human Resource Management*, 28(5): 750–774.

Kampkötter, P. & Sliwka, D. 2016. The complementary use of experiments and field data to evaluate management practices: the case of subjective performance evaluations. *Journal of Institutional and Theoretical Economics*, 172(2): 364–389.

Kanter, R. M. 1993. *Men and women of the corporation* (2. Aufl.). New York: Basic Books.

Kirnan, J. P., Farley, J. A. & Geisinger, K. F. 1989. The relationship between recruiting source, applicant quality, and hire performance: an analysis by sex, ethnicity, and age. *Personnel Psychology*, 42(2): 293–308.

- Knez, I. & Enmarker, I. 1998. Effects of office lighting on mood and cognitive performance and a gender effect in work-related judgment. *Environment and Behavior*, 30(4): 553–567.
- Koch, A. J., D’Mello, S. D. & Sackett, P. R. 2015. A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1): 128–161.
- Koenig, A. M., Eagly, A. H., Mitchell, A. A. & Ristikari, T. 2011. Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4): 616–642.
- Konrad, A. M. & Linnehan, F. 1995. Formalized HRM structures: coordinating equal employment opportunity or concealing organizational practices? *Academy of Management Journal*, 38(3): 787–820.
- Kratzer, N. & Nies, S. 2009. *Neue Leistungs politik bei Angestellten: ERA, Leistungssteuerung, Leistungsentgelt*. Berlin: edition sigma.
- Kreimer, M. 2009. *Ökonomie der Geschlechterdifferenz: Zur Persistenz von Gender Gaps*. Wiesbaden: Springer VS.
- Krell, G. 2011. Diskriminierungs- und Gleichstellungspotenzial von Leistungsbeurteilungen. In G. Krell, R. Ortlieb & B. Sieben (Hg.), *Chancengleichheit durch Personalpolitik: Gleichstellung von Frauen und Männern in Unternehmen und Verwaltungen*: 261–274 (6. Aufl.). Wiesbaden: Gabler.
- Krell, G. & Winter, R. 2011. Anforderungsabhängige Entgeltdifferenzierung: Orientierungshilfen auf dem Weg zu einer diskriminierungsfreieren Arbeitsbewertung. In G. Krell, R. Ortlieb & B. Sieben (Hg.), *Chancengleichheit durch Personalpolitik: Gleichstellung von Frauen und Männern in Unternehmen und Verwaltungen*: 343–360 (6. Aufl.). Wiesbaden: Gabler.
- Kuckartz, U. 2014. *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer VS.
- Kuvaas, B. 2006. Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation. *The International Journal of Human Resource Management*, 17(3): 504–522.
- Landesarbeitsgericht Baden-Württemberg 2012. Urteil vom 19. Juni 2013, Aktenzeichen 2

Sa 2/13.

- Landy, F. J. 2008. Stereotypes, bias, and personnel decisions: strange and stranger. *Industrial and Organizational Psychology*, 1(4): 379–392.
- Landy, F. J. 2010. Performance ratings: then and now. In J. L. Outtz (Hg.), *Adverse impact: implications for organizational staffing and high stakes selection*: 227–248. New York: Taylor and Francis.
- Landy, F. J. & Farr, J. L. 1980. Performance rating. *Psychological Bulletin*, 87(1): 72–107.
- Landy, F. J. & Farr, J. L. 1983. *The measurement of work performance: methods, theory and applications*. New York: Academic Press.
- Lau, C. M., Wong, K. M. & Eggleton, I. R. 2008. Fairness of performance evaluation procedures and job satisfaction: the role of outcome-based and non-outcome-based effects. *Accounting and Business Research*, 38(2): 121–135.
- Lee, H. & Dalal, R. S. 2011. The effects of performance extremities on ratings of dynamic performance. *Human Performance*, 24(2): 99–118.
- Levy, P. E. & Williams, J. R. 2004. The social context of performance appraisal: a review and framework for the future. *Journal of Management*, 30(6): 881–905.
- Liden, R. C. & Mitchell, T. R. 1988. Ingratiation behaviors in organizational settings. *Academy of Management Review*, 13(4): 572–587.
- Lyness, K. S. & Heilman, M. E. 2006. When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4): 777–785.
- Maas, V. S. & Torres-González, R. 2011. Subjective performance evaluation and gender discrimination. *Journal of Business Ethics*, 101(4): 667–681.
- Macrae, C. N., Milne, A. B. & Bodenhausen, G. V. 1994. Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66(1): 37–47.
- Martell, R. F. 1991. Sex bias at work: the effects of attentional and memory demands on performance ratings of men and women. *Journal of Applied Social Psychology*, 21(23): 1939–1960.

- Massengill, D. & Di Marco, N. 1979. Sex-role stereotypes and requisite management characteristics: a current replication. *Sex Roles*, 5(5): 561–570.
- Metall NRW 2004. *ERA-Infobrief Nr. 12: ERA – Verhandlungsergebnis liegt vor*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Metall NRW 2006. *ERA-Infobrief Nr. 14: Die betriebliche ERA-Einführung – Anregungen für eine Mitarbeiter-Information*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Metall NRW 2007a. *Handreichung Entgeltabrechnung ERA NRW*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Metall NRW 2007b. *Leitfaden II: Auf dem Weg zum Entgeltraahmenabkommen*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Metall NRW 2008. *ERA – Tarifliche Leistungsbeurteilung: Leitfaden für Führungskräfte*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Metall NRW 2009. *ERA – Tarifliche Leistungszulage: Orientierungshilfe für das betriebliche Personalwesen*. Düsseldorf: Verband der Metall- und Elektro-Industrie Nordrhein-Westfalen e.V.
- Mettrick, J. & Cowan, G. 1996. Individuating information, gender, and the 1995 America's Cup. *Journal of Social Behavior and Personality*, 11(2): 399–410.
- Mishra, V. & Roch, S. G. 2013. Cultural values and performance appraisal: assessing the effects of rater self-construal on performance ratings. *The Journal of Psychology*, 147(4): 325–344.
- Mobley, W. H. 1982. Supervisor and employee race and sex effects on performance appraisals: a field study of adverse impact and generalizability. *Academy of Management Journal*, 25(3): 598–606.
- Moers, F. 2005. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1): 67–80.
- Moore, D. P. 1984. Evaluating in-role and out-of-role performers. *Academy of Management Journal*, 27(3): 603–618.
- Narcisse, S. & Harcourt, M. 2008. Employee fairness perceptions of performance ap-

- praisal: a Saint Lucian case study. *The International Journal of Human Resource Management*, 19(6): 1152–1169.
- Niedersächsisches Ministerium für Soziales, Frauen, Familie und Gesundheit 2005. *Grundlagen und Beispiele: Gender Mainstreaming in Niedersachsen – Gleiche Chancen, gleiche Rechte, Politik mit Konsequenz*.
<http://www.niedersachsen.de/download/9585>. Zuletzt geprüft am 28. November 2017.
- Nieva, V. F. & Gutek, B. A. 1980. Sex effects on evaluation. *Academy of Management Review*, 5(2): 267–276.
- Nisar, T. M. 2007. Evaluation of subjectivity in incentive pay. *Journal of Financial Services Research*, 31(1): 53–73.
- Norton, M. I., Vandello, J. A. & Darley, J. M. 2004. Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87(6): 817–831.
- Norton, S. D., Gustafson, D. P. & Foster, C. E. 1977. Assessment for management potential: scale design and development, training effects and rater/ratee sex effects. *Academy of Management Journal*, 20(1): 117–131.
- Nurse, L. 2005. Performance appraisal, employee development and organizational justice: exploring the linkages. *The International Journal of Human Resource Management*, 16(7): 1176–1194.
- O'Reilly, J., Smith, M., Deakin, S. & Burchell, B. 2015. Equal pay as a moving target: international perspectives on forty-years of addressing the gender pay gap. *Cambridge Journal of Economics*, 39(2): 299–317.
- Ogunfowora, B., Bourdage, J. & Lee, K. 2010. Rater personality and performance dimension weighting in making overall performance judgments. *Journal of Business and Psychology*, 25(3): 465–476.
- Ohl, K. 2011. Was hat die IG Metall mit ERA gewollt und wie wurden diese Ziele im Verlauf der Einführung umgesetzt? In S. Brandl & H. Wagner (Hg.), *Ein „Meilenstein der Tarifpolitik“ wird besichtigt. Die Entgelttarifverträge in der Metall- und Elektroindustrie: Erfahrungen Resultate Auseinandersetzungen*: 23–26. Berlin: edition sigma.
- Olian, J. D., Schwab, D. P. & Haberfeld, Y. 1988. The impact of applicant gender compared to qualifications on hiring recommendations: a meta-analysis of experimental

- studies. *Organizational Behavior and Human Decision Processes*, 41(2): 180–195.
- Patton, E. & Appelbaum, S. H. 2003. The case for case studies in management research. *Management Research News*, 26(5): 60–71.
- Payne, M.-L. 2011. The impact of rater's gender on performance evaluations of HR managers. *International Management Review*, 7(2): 5–12.
- Pazy, A. 1986. The persistence of pro-male bias despite identical information regarding causes of success. *Organizational Behavior and Human Decision Processes*, 38(3): 366–377.
- Peters, L. H., O'Connor, E. J., Weekley, J., Pooyan, A., Frank, B. & Erenkrantz, B. 1984. Sex bias and managerial evaluations: a replication and extension. *Journal of Applied Psychology*, 69(2): 349–352.
- Pfeffer, J. & Cohen, Y. 1984. Determinants of internal labor markets in organizations. *Administrative Science Quarterly*, 29(4): 550–572.
- Phelps, E. S. 1972. The statistical theory of racism and sexism. *The American Economic Review*, 62(4): 659–661.
- Pichler, S. 2012. The social context of performance appraisal and appraisal reactions: a meta-analysis. *Human Resource Management*, 51(5): 709–732.
- Plake, B. S., Murphy-Berman, V., Derscheid, L. E., Gerber, R. W., Miller, S. K., Speth, C. A. & Tomes, R. E. 1987. Access decisions by personnel directors: subtle forms of sex bias in hiring. *Psychology of Women Quarterly*, 11(2): 255–264.
- Poon, J. M. L. 2004. Effects of performance appraisal politics on job satisfaction and turnover intention. *Personnel Review*, 33(3): 322–334.
- Powell, G. N. & Butterfield, D. A. 1994. Investigating the “glass ceiling“ phenomenon: an empirical study of actual promotions to top management. *Academy of Management Journal*, 37(1): 68–86.
- Powell, G. N. & Graves, L. M. 2003. *Women and men in management* (3. Aufl.). Thousand Oaks: Sage Publications.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37(1): 7–63.
- Pulakos, E. D. & Wexley, K. N. 1983. The relationship among perceptual similarity, sex,

- and performance ratings in manager-subordinate dyads. *Academy of Management Journal*, 26(1): 129–139.
- Pulakos, E. D., White, L. A., Oppler, S. H. & Borman, W. C. 1989. Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74(5): 770–780.
- Ranftl, E., Buchinger, B., Gschwandtner, U. & Meggeneder, O. 2002. *Gleicher Lohn für gleichwertige Arbeit: Praktische Beispiele diskriminierungsfreier analytischer Arbeitsbewertung*. München etc.: Rainer Hampp.
- Ransom, M. & Oaxaca, R. L. 2005. Intrafirm mobility and sex differences in pay. *Industrial and Labor Relations Review*, 58(2): 219–237.
- Ray, K. 2007. Performance evaluations and efficient sorting. *Journal of Accounting Research*, 45(4): 839–882.
- Reskin, B. F. 2000. The proximate causes of employment discrimination. *Contemporary Sociology*, 29(2): 319–328.
- Reskin, B. F. 2008. Including mechanisms in our models of ascriptive inequality. In L. B. Nielsen & R. L. Nelson (Hg.), *Handbook of employment discrimination research: rights and realities*: 75–97. New York: Springer.
- Reskin, B. F. & McBrier, D. B. 2000. Why not ascription? Organizations' employment of male and female managers. *American Sociological Review*, 65(2): 210–233.
- Ridgeway, C. 1991. The social construction of status value: gender and other nominal characteristics. *Social Forces*, 70(2): 367–386.
- Ridgeway, C. L. 1997. Interaction and the conservation of gender inequality: considering employment. *American Sociological Review*, 62(2): 218–235.
- Riegelhaupt, B. J. 1984. *Discrimination in personnel decisions: the effects of applicant sex and physical attractiveness*. Dissertation. Virginia State University: Petersburg.
- Robbins, T. L. 1993. Moderators of sex bias in the performance appraisal process: a cognitive analysis. *Journal of Management*, 19(1): 113–126.
- Roth, P. L., Purvis, K. L. & Bobko, P. 2012. A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management*, 38(2): 719–739.

- Rudman, L. A. & Glick, P. 1999. Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, 77(5): 1004–1010.
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E. & Nauts, S. 2012. Status incongruity and backlash effects: defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, 48(1): 165–179.
- Saffie-Robertson, C. M. & Brutus, S. 2014. The impact of interdependence on performance evaluations: the mediating role of discomfort with performance appraisal. *The International Journal of Human Resource Management*, 25(3): 459–473.
- Schein, V. E. 2001. A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, 57(4): 675–688.
- Schmitt, N. & Lippin, M. 1980. Race and sex as determinants of the mean and variance of performance ratings. *Journal of Applied Psychology*, 65(4): 428–435.
- Schreyögg, F. 1998. Der Einfluss von Geschlechterstereotypen auf die Beurteilung von Arbeitsleistung. *Organisationsberatung, Supervision, Clinical Management*, 5(1): 27–46.
- Schreyögg, F. 2011. Praxisbeispiel Stadt München: Beurteilungsverfahren sind nicht geschlechtsneutral. In G. Krell, R. Ortlieb & B. Sieben (Hg.), *Chancengleichheit durch Personalpolitik: Gleichstellung von Frauen und Männern in Unternehmen und Verwaltungen*: 275–282. Wiesbaden: Gabler.
- Selvarajan, T. T. & Cloninger, P. A. 2012. Can performance appraisals motivate employees to improve performance? A Mexican study. *The International Journal of Human Resource Management*, 23(15): 3063–3084.
- Shaw, K. 2009. Insider econometrics: a roadmap with stops along the way. *Labour Economics*, 16(6): 607–617.
- Sholihin, M. & Pike, R. 2009. Fairness in performance evaluation and its behavioural consequences. *Accounting and Business Research*, 39(4): 397–413.
- Smeets, V., Waldman, M. & Warzynski, F. 2013. *Performance, career dynamics, and span of control*. Discussion Paper 2013-001. Cornell University: Ithaca.
- Spence, J. T. & Hahn, E. D. 1997. The attitudes toward women scale and attitude change

- in college students. *Psychology of Women Quarterly*, 21(1): 17–34.
- Stangor, C. & Lange, J. 1994. Mental representations of social groups: advances in understanding stereotypes and stereotyping. In M. Zanna (Hg.), *Advances in experimental social psychology*: 357–416 (26. Aufl.). San Diego: Academic Press.
- Stangor, C., Lynch, L., Duan, C. & Glas, B. 1992. Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, 62(2): 207–218.
- Steensma, H. & Visser, E. 2007. Procedural justice and supervisors' personal power bases: effects on employees' perceptions of performance appraisal sessions, commitment, and motivation. *Journal of Collective Negotiations*, 31(2): 101–118.
- Sudin, S. 2011. Fairness of and satisfaction with performance appraisal process. *Journal of Global Management*, 2(1): 66–83.
- Sutton, A. W., Baldwin, S. P., Wood, L. & Hoffman, B. J. 2013. A meta-analysis of the relationship between rater liking and performance ratings. *Human Performance*, 26(5): 409–429.
- Sutton, J. R., Dobbin, F., Meyer, J. W. & Scott, W. R. 1994. The legalization of the workplace. *American Journal of Sociology*, 99(4): 944–971.
- Swiercz, P. M., Bryan, N. B., Eagle, B. W., Bizzotto, V. & Renn, R. W. 2012. Predicting employee attitudes and performance from perceptions of performance appraisal fairness. *Business Renaissance Quarterly*, 7(1): 25–46.
- Swim, J., Borgida, E., Maruyama, G. & Myers, D. G. 1989. Joan McKay versus John McKay: do gender stereotypes bias evaluations? *Psychological Bulletin*, 105(3): 409–429.
- Tajfel, H. 1981. *Human groups and social categories: studies in social psychology*. Cambridge: Cambridge University Press.
- Tajfel, H. & Turner, J. C. 1986. The social identity theory of intergroup behavior. In W. G. Austin & S. Worchel (Hg.), *Psychology of intergroup relations*: 7–24 (2. Aufl.). Chicago: Nelson-Hall.
- Taylor, S. E., Fiske, S. T., Etoff, N. L. & Ruderman, A. J. 1978. Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psy-*

- chology*, 36(7): 778–793.
- Terborg, J. R. & Ilgen, D. R. 1975. A theoretical approach to sex discrimination in traditionally masculine occupations. *Organizational Behavior and Human Performance*, 13(3): 352–376.
- Terjesen, S., Aguilera, R. V. & Lorenz, R. 2015. Legislating a woman's seat on the board: institutional factors driving gender quotas for boards of directors. *Journal of Business Ethics*, 128(2): 233–251.
- Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C. & Privette, M. 2008. Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes*, 107(1): 29–40.
- Thurston, P. W. & McNall, L. 2010. Justice perceptions of performance appraisal practices. *Journal of Managerial Psychology*, 25(3): 201–228.
- Tondorf, K. & Jochmann-Döll, A. 2005. *(Geschlechter-)Gerechte Leistungsvergütung? Vom (Durch-)Bruch des Leistungsprinzips in der Entlohnung*. Hamburg: VSA-Verlag.
- Tsui, A. S., Egan, T. D. & O'Reilly, C. A. 1992. Being different: relational demography and organizational attachment. *Administrative Science Quarterly*, 37(4): 549–579.
- Turban, D. B. & Jones, A. P. 1988. Supervisor-subordinate similarity: types, effects, and mechanisms. *Journal of Applied Psychology*, 73(2): 228–234.
- Turner, J. C. 1985. Social categorization and the self-concept: a social theory of group behavior. In E. J. Lawler (Hg.), *Advances in group processes: a research annual*: 77–122. Greenwich: JAI Press.
- Umair, T., Javaid, M. F., Amir, H. & Luqman, M. K. 2016. Effect of perceived appraisal fairness on job satisfaction. *Journal of Applied Environmental and Biological Sciences*, 6(2): 12–20.
- Varma, A., Pichler, S. & Srinivas, E. S. 2005. The role of interpersonal affect in performance appraisal: evidence from two samples – the US and India. *The International Journal of Human Resource Management*, 16(11): 2029–2044.
- Varma, A. & Stroh, L. K. 2001. The impact of same-sex LMX dyads on performance evaluations. *Human Resource Management*, 40(4): 309–320.

- Vecchio, R. P. & Gobdel, B. C. 1984. The vertical dyad linkage model of leadership: problems and prospects. *Organizational Behavior and Human Performance*, 34(1): 5–20.
- Verbeek, M. 2012. *A guide to modern econometrics* (4. Aufl.). Hoboken: Wiley.
- Wayne, S. J., Shore, L. M. & Liden, R. C. 1997. Perceived organizational support and leader-member exchange: a social exchange perspective. *The Academy of Management Journal*, 40(1): 82–111.
- Welle, B. & Heilman, M. E. 2007. Formal and informal discrimination against women at work: the role of gender stereotypes. In D. D. Steiner, D. P. Skarlicki & S. W. Gilliland (Hg.), *Managing social and ethical issues in organizations*: 229–252. Charlotte: Information Age Publishers.
- Wharton, A. S. 1992. The social construction of gender and race in organizations: a social identity and group mobilization perspective. *Research in the Sociology of Organizations*, 10(1): 55–84.
- Williams, J. E. & Best, D. L. 1990. *Measuring sex stereotypes: a multinational study*. Newbury Park: Sage Publications.
- Williams, K. Y. & O'Reilly, C. A. 1997. Demography and diversity in organizations: a review of 40 years of research. *Research in Organizational Behavior*, 20(1): 77–140.
- Wirtz, M. A. & Nachtigall, C. 2012. *Deskriptive Statistik: Statistische Methoden für Psychologen Teil 1* (6. Aufl.). Weinheim: Beltz Juventa.
- Wood, M. E. 1999. *Examination of the role of sexism in the overvaluation of female leaders*. Dissertation Abstracts International 60(8-B), 4285. Section B: Sciences and Engineering.
- Woods, A. 2012. Subjective adjustments to objective performance measures: the influence of prior performance. *Accounting, Organizations and Society*, 37(6): 403–425.
- Wooldridge, J. M. 2013. *Introductory econometrics: a modern approach* (5. Aufl.). Boston etc.: Cengage Learning.
- Yammarino, F. J. & Dubinsky, A. J. 1988. Employee responses: gender-or job-related differences? *Journal of Vocational Behavior*, 32(3): 366–383.
- Yin, R. K. 1984. *Case study research: design and methods*. Beverly Hills: Sage Publica-

tions.

Ziepke, J. 1975. *Handkommentar zum Gehaltsrahmenabkommen in der Eisen-, Metall-, Elektro- und Zentralheizungsindustrie Nordrhein-Westfalens vom 19.02.1975 und Gehaltsabkommen vom 19.02.1975* (8. Aufl.). Wuppertal: W. Girardet.