

Bergische Universität Wuppertal
Fachbereich G – Human- und Sozialwissenschaften

Psychische Beanspruchung in der Arbeit

Theoretische Begründung, ökonomische Messung
und praxisnahe Anwendung – Bestimmung von Risikoprofilen
psychischer Belastung und Beanspruchung

– Teil 1: Haupttext –

Inaugural-Dissertation
zur Erlangung des Grades
Doktor der Philosophie (Dr. phil.)

vorgelegt von
Mike Hammes, Remscheid
September 2015

Gutachter:
Prof. Dr. phil. Rainer Wieland, Wuppertal
Prof. Dr. phil. Ralf Schulze, Wuppertal

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20160404-113610-5

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3A468-20160404-113610-5>]

INHALTSVERZEICHNIS

Inhaltsverzeichnis des Anhangs	VI
Tabellenverzeichnis.....	VIII
Abbildungsverzeichnis	XVI
Abkürzungsverzeichnis	XVIII
1 Einleitung.....	1
2 Psychische Belastung und Beanspruchung: Neurobiopsychologische Grundlagen	8
2.1 Beanspruchung als Inanspruchnahme und Anpassung von Organsystemen.....	9
2.1.1 Mechanische Belastung und Beanspruchung am Beispiel des Stützapparates ..	10
2.1.2 Inanspruchnahme und Anpassung von Überlebens-Schaltkreisen.....	12
2.2 Arousal: Die Dynamik von Valenz und Aktivierung	15
2.2.1 Inanspruchnahme und Anpassung des Zentralnervensystems unter Stress	15
2.2.2 Die Arousal-Uhr	19
2.2.3 Arousal und das Circumplex-Modell der Core Affects	27
2.2.4 Emotion, Motivation und Verhalten.....	35
2.3 Arousal und Gesundheit	38
2.3.1 ...bis sich das Selbst zersetzt: Zusammenhang zwischen Stress und Energie	39
2.3.2 Burnout: Wenn die Seele bricht?	42
2.3.3 Stress und Gesundheitsbeeinträchtigungen	43
2.3.4 Helfen Medikamente bei Depression? - Arousal als System	47
2.3.5 Verhaltensbezogene Förderung einer gesunden Arousal-Dynamik.....	51
2.4 Kognition, Emotion und die modulare Architektur des Gehirns.....	54
2.4.1 Drei weiträumig verteilte kortikale Netzwerke	55
2.4.2 Kortikale Funktionalität bei psychischen und Verhaltensstörungen.....	59
2.4.3 Doubleloop-Perturbation-Modell neuronaler Aktivität.....	61
3 Selbstorganisation in Arbeitssystemen	66
3.1 Sicherstellung der Homöostase	68

Inhaltsverzeichnis

3.1.1	Theoretische Grundlegung	68
3.1.2	Arbeitspsychologische Konsequenzen	70
3.2	Aufbau von Kompetenz	71
3.2.1	Theoretische Grundlegung	71
3.2.2	Arbeitspsychologische Konsequenzen	74
3.3	Soziale Teilhabe und Zugehörigkeit	77
3.3.1	Theoretische Grundlegung	78
3.3.2	Arbeitspsychologische Konsequenzen	81
3.4	Das Spannungsfeld von Autonomie und Selbstkontrolle	85
3.4.1	Theoretische Grundlegung	85
3.4.2	Arbeitspsychologische Konsequenzen	92
3.5	Unternehmenskultur als Kontext für Selbstorganisation	97
4	Begründung und Formulierung des empirisch-praktischen Untersuchungsanliegens	99
4.1	Das Wuppertaler Konzept	100
4.2	Hintergrund und Struktur der EEB	102
4.3	Formulierung des Forschungsanliegens	107
5	Klassifikator-Induktion: Methodisches Vorgehen und theoretische Begründung	108
5.1	Auswahl geeigneter Stichproben	110
5.1.1	Induktionsstichproben	110
5.1.2	Validierungsstichproben	111
5.1.3	Anwendungsstichprobe	112
5.2	Eingesetzte Analyse- und Entwicklungsumgebung	113
5.3	Aufbereitung der Daten	113
5.4	Einteilung in zwei Klassen: Dichotomisierung	114
5.4.1	Theoretische Begründung für die Dichotomisierung der Daten	115
5.4.2	Methodisches Vorgehen bei der Dichotomisierung	117
5.5	Deskriptiv-statistische Analyse	117
5.6	Theoretische Grundlagen der Klassifikator-Induktion	118

Inhaltsverzeichnis

5.6.1	Herausforderungen an die Klassifikator-Induktion.....	119
5.6.2	Formale Beschreibung eines Klassifikator-Induktors	123
5.6.3	Allgemeiner Aufbau eines Klassifikator-Induktors	124
5.7	Allgemeine Spezifikationen verwendeter Klassifikator-Induktoren	126
5.7.1	Beurteilung der Klassifikationsgüte	126
5.7.2	Kreuzvalidierung	129
5.7.3	Stabilität	131
5.8	Erste Experimentalreihe: Vollständige Suche und Bayes-Klassifikation.....	132
5.8.1	Vollständige Suche.....	132
5.8.2	Bayes-Klassifikation	133
5.9	Zweite Experimentalreihe: Heuristische Suche und binärlogistische Regression... 135	
5.9.1	Heuristische Suche	136
5.9.2	Binärlogistische Regression	138
5.9.3	Festlegung von Abbruchbedingungen.....	140
5.10	Finale Auswahl und Kalibrierung von Klassifikatoren: Bootstrapping	141
5.10.1	Vergleich von Klassifikator-Induktoren.....	142
5.10.2	Bootstrapping	142
5.10.3	Finale Auswahl von Klassifikatoren	143
5.10.4	Kalibrierung von Klassifikatoren	145
5.10.5	Vorhersagefehler von Erwartungswerten in Abhängigkeit des Stichprobenumfangs	146
5.11	Anwendung und Validierung der ermittelten Klassifikatoren.....	147
5.12	Automatisierte Darstellung der Ergebnisse	149
6	Ergebnisse	150
6.1	Deskriptive Statistik der EEB-Items.....	150
6.2	CV-T-Diagramme: Klassifikationsgüte in Abhängigkeit der Messkomplexität	154
6.3	Ergebnisse der Klassifikator-Induktion	167
6.3.1	Aufbau der Ergebnis-Berichte und Erläuterung verwendeter Kennwerte.....	168

Inhaltsverzeichnis

6.3.2	Aufgabenanforderungen.....	172
6.3.3	Kommunikation und Kooperation.....	184
6.3.4	Leistungs- und Zeitvorgaben.....	188
6.3.5	Regulationsbehinderungen.....	192
6.3.6	Tätigkeitsspielraum.....	196
6.3.7	Gesamtbedarf für Arbeitsgestaltung.....	200
6.3.8	Autoritäre Führung.....	204
6.3.9	Bedeutung von Anerkennung und Wertschätzung.....	208
6.3.10	Ganzheitliche Führung.....	212
6.3.11	Partizipative Unternehmenskultur.....	216
6.3.12	Zusammenarbeit.....	220
6.3.13	Gesundheitsbezogene Selbstwirksamkeit.....	224
6.3.14	Herz-Kreislauf-Beschwerden.....	228
6.3.15	Magen-Darm-Beschwerden.....	232
6.3.16	Muskel-Skelett-Beschwerden.....	236
6.3.17	Unspezifische Beschwerden.....	240
6.3.18	Absentismustage.....	244
6.3.19	Präsentismustage.....	248
6.4	Vorhersagefehler in Abhängigkeit des Stichprobenumfanges.....	252
6.5	Vergleich der Klassifikator-Induktoren.....	254
6.6	Validierung endgültig ausgewählter Klassifikatoren.....	259
6.7	Profil-Bestimmung: Demonstration der Anwendungsmöglichkeiten.....	265
7	Diskussion, Anwendungsbezug und Ausblick.....	274
7.1	Eingrenzung der Gültigkeitsbereiche vorliegender Untersuchungsbefunde.....	274
7.2	Methodische Diskussion vorliegender Untersuchungsbefunde.....	280
7.2.1	Diskussion und Interpretation der CV-T-Diagramme.....	280
7.2.2	Methodische Diskussion und Interpretation von Vorhersagegenauigkeiten....	282
7.2.3	Methodische Diskussion und Interpretation der ausgewählten Items.....	286

Inhaltsverzeichnis

7.2.4	Vergleich der Klassifikator-Induktoren.....	292
7.2.5	Diskussion von Kalibrierung und Validierung.....	293
7.3	Inhaltliche Interpretation der Klassifikatoren.....	296
7.3.1	Übersicht über ausgewählte Items.....	296
7.3.2	Inhaltliche Plausibilitätsbetrachtungen bezüglich der ausgewählten Items	301
7.3.3	Brauchen wir eine Theorie des „conceptual act“?.....	306
7.4	Anwendung der Klassifikatoren	307
7.4.1	WSIB Pro: Internetbasierte Anwendung zur Profil-Vorhersage.....	308
7.4.2	Eine realistische Zukunftsvision für WSIB Pro	311
7.4.3	...Beware of your Data!	313
	Literaturverzeichnis.....	316

INHALTSVERZEICHNIS DES ANHANGS

A1. Mathematischer Anhang	1
A2. Dichotomisierung der EEB-Items	8
A3. Vollständige Ergebnisberichte	26
A3.1 Aufgabenanforderungen	26
A3.2 Kommunikation und Kooperation	38
A3.3 Leistungs- und Zeitvorgaben	50
A3.4 Regulationsbehinderungen	62
A3.5 Tätigkeitsspielraum	74
A3.6 Gesamtbedarf für Arbeitsgestaltung	86
A3.7 Autoritäre Führung	98
A3.8 Bedeutung von Anerkennung und Wertschätzung	110
A3.9 Ganzheitliche Führung	122
A3.10 Partizipative Unternehmenskultur	134
A3.11 Zusammenarbeit	146
A3.12 Gesundheitsbezogene Selbstwirksamkeit	158
A3.13 Herz-Kreislauf-Beschwerden	170
A3.14 Magen-Darm-Beschwerden	182
A3.15 Muskel-Skelett-Beschwerden	194
A3.16 Unspezifische Beschwerden	206
A3.17 Absentismustage	218
A3.18 Präsentismustage	230
A4. Definition der Klassifikatoren	242
A4.1 Arbeitsanforderungen	242
A4.2 Kommunikation und Kooperation	244
A4.3 Leistungs- und Zeitvorgaben	246
A4.4 Regulationsbehinderungen	247
A4.5 Tätigkeitsspielraum	249

Inhaltsverzeichnis des Anhangs

A4.6	Gesamtbedarf für Arbeitsgestaltung	251
A4.7	Autoritäre Führung	253
A4.8	Ganzheitliche Führung	255
A4.9	Partizipative Unternehmenskultur	256
A4.10	Bedeutung von Anerkennung und Wertschätzung	258
A4.11	Zusammenarbeit	260
A4.12	Gesundheitsbezogene Selbstwirksamkeit.....	262
A4.13	Herz-Kreislauf-Beschwerden	264
A4.14	Magen-Darm-Beschwerden.....	266
A4.15	Muskel-Skelett-Beschwerden	268
A4.16	Unspezifische Beschwerden	270
A4.17	Absentismustage	271
A4.18	Präsentismustage	272
A5.	Vergleich von Messung und Vorhersage für 12 Stichproben	273

TABELLENVERZEICHNIS

Tabelle 1. Substanzen der Arousal-Systeme und assoziierte Situationen bzw. Aktivitäten, Schlaf-Wach-Zustände und sensumotorische (SM) bzw. kortikale (CTX) Aktivität.	21
Tabelle 2. Struktur der Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben während der Arbeit (EEB; Scherrer, 2002; Wieland, 2014b; Wieland & Hammes, 2014; Wieland, Metz & Richter, 2001).	106
Tabelle 3. Vorherzusagende Kriterien, unterteilt nach Arbeitsaufgabe und -situation, Führung und Zusammenarbeit sowie Gesundheitsindikatoren.	168
Tabelle 4. Deskriptiv-statistische Kennwerte der Variablen Aufgabenanforderungen.	172
Tabelle 5. Korrelationen (Spearman ρ) der Variablen Aufgabenanforderungen mit den einzelnen Items der EEB.	172
Tabelle 6. Selegierte Eigenschaftswörter der vollständigen EEB bei drei Bayes-Selektions-Experimenten (bestes Item, bestes Paar und bestes Tripel) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.	173
Tabelle 7. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Bayes-Selektions-Experimente (bestes Item, bestes Paar, bestes Tripel).	174
Tabelle 8. Selegierte Eigenschaftswörter des WSIB bei drei Bayes-Selektions-Experimenten (bestes Item, bestes Paar und bestes Tripel) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.	175
Tabelle 9. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Bayes-Selektions-Experimente (bestes Item, bestes Paar, bestes Tripel).	176
Tabelle 10. Selegierte Eigenschaftswörter der vollständigen EEB bei drei Selektions-Experimenten (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.	177
Tabelle 11. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Selektions-Experimente (Binärlogistische Regression, Sequential	

Tabellenverzeichnis

Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]).	178
Tabelle 12. Selegierte Eigenschaftswörter des WSIB bei drei Selektions-Experimenten (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.	179
Tabelle 13. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Selektions-Experimente (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]).	180
Tabelle 14. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	181
Tabelle 15. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	182
Tabelle 16. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	183
Tabelle 17. Deskriptiv-statistische Kennwerte der Variablen Kommunikation und Kooperation.	184
Tabelle 18. Korrelationen (Spearman ρ) der Variablen Kommunikation und Kooperation mit den einzelnen Items der EEB.	184
Tabelle 19. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt.	185
Tabelle 20. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Kommunikation und Kooperation (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	186
Tabelle 21. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Kommunikation und Kooperation (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	187
Tabelle 22. Deskriptiv-statistische Kennwerte der Variablen Leistungs- und Zeitvorgaben.	188
Tabelle 23. Korrelationen (Spearman ρ) der Variablen Leistungs- und Zeitvorgaben mit den einzelnen Items der EEB.	188
Tabelle 24. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	189

Tabellenverzeichnis

Tabelle 25. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Leistungs- und Zeitvorgaben (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	190
Tabelle 26. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Leistungs- und Zeitvorgaben (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	191
Tabelle 27. Deskriptiv-statistische Kennwerte der Variablen Regulationsbehinderungen. ...	192
Tabelle 28. Korrelationen (Spearman ρ) der Variablen Regulationsbehinderungen mit den einzelnen Items der EEB.	192
Tabelle 29. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	193
Tabelle 30. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Regulationsbehinderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	194
Tabelle 31. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Regulationsbehinderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	195
Tabelle 32. Deskriptiv-statistische Kennwerte der Variablen Tätigkeitsspielraum.	196
Tabelle 33. Korrelationen (Spearman ρ) der Variablen Tätigkeitsspielraum mit den einzelnen Items der EEB.	196
Tabelle 34. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	197
Tabelle 35. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Tätigkeitsspielraum (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	198
Tabelle 36. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Tätigkeitsspielraum (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	199
Tabelle 37. Deskriptiv-statistische Kennwerte der Variablen Gesamtbedarf für Arbeitsgestaltung.	200
Tabelle 38. Korrelationen (Spearman ρ) der Variablen Gesamtbedarf für Arbeitsgestaltung mit den einzelnen Items der EEB.	200
Tabelle 39. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	201
Tabelle 40. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Arbeitsgestaltung gesamt (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.	202

Tabellenverzeichnis

Tabelle 41. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Arbeitsgestaltung gesamt (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.....	203
Tabelle 42. Deskriptiv-statistische Kennwerte der Variablen Autoritäre Führung.....	204
Tabelle 43. Korrelationen (Spearman ρ) der Variablen Autoritäre Führung mit den einzelnen Items der EEB.	204
Tabelle 44. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	205
Tabelle 45. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von autoritärer Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	206
Tabelle 46. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von autoritärer Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	207
Tabelle 47. Deskriptiv-statistische Kennwerte der Variablen Bedeutung von Anerkennung und Wertschätzung.....	208
Tabelle 48. Korrelationen (Spearman ρ) der Variablen Bedeutung von Anerkennung und Wertschätzung mit den einzelnen Items der EEB.....	208
Tabelle 49. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	209
Tabelle 50. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Bedeutung von Anerkennung und Wertschätzung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	210
Tabelle 51. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Bedeutung von Anerkennung und Wertschätzung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	211
Tabelle 52. Deskriptiv-statistische Kennwerte der Variablen Ganzheitliche Führung.....	212
Tabelle 53. Korrelationen (Spearman ρ) der Variablen Ganzheitliche Führung mit den einzelnen Items der EEB.....	212
Tabelle 54. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.....	213
Tabelle 55. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von ganzheitlicher Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	214
Tabelle 56. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von ganzheitlicher Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.....	215

Tabellenverzeichnis

Tabelle 57. Deskriptiv-statistische Kennwerte der Variablen Partizipative Unternehmenskultur.	216
Tabelle 58. Korrelationen (Spearman ρ) der Variablen Partizipative Unternehmenskultur mit den einzelnen Items der EEB.	216
Tabelle 59. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	217
Tabelle 60. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von partizipativer Unternehmenskultur (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	218
Tabelle 61. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von partizipativer Unternehmenskultur (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	219
Tabelle 62. Deskriptiv-statistische Kennwerte der Variablen Zusammenarbeit.	220
Tabelle 63. Korrelationen (Spearman ρ) der Variablen Zusammenarbeit mit den einzelnen Items der EEB.	220
Tabelle 64. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	221
Tabelle 65. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Zusammenarbeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	222
Tabelle 66. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Zusammenarbeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	223
Tabelle 67. Deskriptiv-statistische Kennwerte der Variablen Gesundheitsbezogene Selbstwirksamkeit.	224
Tabelle 68. Korrelationen (Spearman ρ) der Variablen Gesundheitsbezogene Selbstwirksamkeit mit den einzelnen Items der EEB.	224
Tabelle 69. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	225
Tabelle 70. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von gesundheitsbezogener Selbstwirksamkeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	226
Tabelle 71. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von gesundheitsbezogener Selbstwirksamkeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.	227
Tabelle 72. Deskriptiv-statistische Kennwerte der Variablen Herz-Kreislauf-Beschwerden.	228

Tabellenverzeichnis

Tabelle 73. Korrelationen (Spearman ρ) der Variablen Herz-Kreislauf-Beschwerden mit den einzelnen Items der EEB.	228
Tabelle 74. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	229
Tabelle 75. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Herz-Kreislauf-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	230
Tabelle 76. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Herz-Kreislauf-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	231
Tabelle 77. Deskriptiv-statistische Kennwerte der Variablen Magen-Darm-Beschwerden. .	232
Tabelle 78. Korrelationen (Spearman ρ) der Variablen Magen-Darm-Beschwerden mit den einzelnen Items der EEB.	232
Tabelle 79. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	233
Tabelle 80. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Magen-Darm-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	234
Tabelle 81. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Magen-Darm-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	235
Tabelle 82. Deskriptiv-statistische Kennwerte der Variablen Muskel-Skelett-Beschwerden.	236
Tabelle 83. Korrelationen (Spearman ρ) der Variablen Muskel-Skelett-Beschwerden mit den einzelnen Items der EEB.	236
Tabelle 84. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	237
Tabelle 85. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	238
Tabelle 86. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.	239
Tabelle 87. Deskriptiv-statistische Kennwerte der Variablen Unspezifische Beschwerden. .	240
Tabelle 88. Korrelationen (Spearman ρ) der Variablen Unspezifische Beschwerden mit den einzelnen Items der EEB.	240
Tabelle 89. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	241

Tabellenverzeichnis

Tabelle 90. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	242
Tabelle 91. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	243
Tabelle 92. Deskriptiv-statistische Kennwerte der Variablen Absentismustage.	244
Tabelle 93. Korrelationen (Spearman ρ) der Variablen Absentismustage mit den einzelnen Items der EEB.	244
Tabelle 94. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	245
Tabelle 95. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Absentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	246
Tabelle 96. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Absentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	247
Tabelle 97. Deskriptiv-statistische Kennwerte der Variablen Präsentismustage.	248
Tabelle 98. Korrelationen (Spearman ρ) der Variablen Präsentismustage mit den einzelnen Items der EEB.	248
Tabelle 99. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren.	249
Tabelle 100. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Präsentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	250
Tabelle 101. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Präsentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.....	251
Tabelle 102. Parameter und Fit-Indizes der logarithmischen Regressionen der Standardabweichungen von Vorhersagefehlern in Abhängigkeit des Stichprobenumfangs $n \in \{10, \dots, 100\}$ für alle 18 Kriterien.....	253
Tabelle 103. Vergleich der Cross-Validation-Korrektklassifikationen bei Bayes-Klassifikation in Kombination mit vollständiger Suche und mit Sequential Forward Selection (SFS).....	255
Tabelle 104. Vergleich der CV-Korrektklassifikationen bei binärlogistischer Regression in Kombination mit Sequential Forward Selection (BLR/SFS) und Bayes-Klassifikation mit vollständiger Suche.	256

Tabellenverzeichnis

Tabelle 105. Vergleich der Cross-Validation-Korrektklassifikationen bei binärlogistischer Regression (BLR) und Bayes-Klassifikation in Verbindung mit Sequential Forward Selection.	257
Tabelle 106. Vergleich der Cross-Validation-Korrektklassifikationen bei Bayes-Klassifikation in Verbindung mit Sequential Forward Selection (SFS) und Sequential Forward Floating Selection (SFFS).	258
Tabelle 107. Deskriptiv-statistische Kennwerte der Validierungsstudie und Vergleich mit den Induktionsstichproben (t-Test für zwei unabhängige Stichproben).	260
Tabelle 108. Anzahl korrekt vorhergesagter Erwartungswerte für 16 Kriterien unter Anwendung des funktionalen Zusammenhanges zwischen dem entsprechenden Vorhersagefehler und dem Stichprobenumfang.	264
Tabelle 109. Nummerierungen der Items der vollständigen EEB. Angegeben sind auch die theoriegeleiteten Dimensionen, denen diese Items zugeordnet sind.	297
Tabelle 110. Ausgewählte Items aus dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB) zur Vorhersage der 18 Kriterien.	298
Tabelle 111. Ausgewählte Items aus Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben (EEB) zur Vorhersage der 18 Kriterien.	299

ABBILDUNGSVERZEICHNIS

Abbildung 1. Arousal-Systeme (eigene Darstellung, vereinfacht nach den Angaben in Jones [2003] und Brown et al. [2012]).....	22
Abbildung 2. Die Arousal-Uhr.....	24
Abbildung 3. Circumplex-Modell nach Posner, Russell und Peterson (2005, S. 716, Figure 1)	29
Abbildung 4. Synthese verschiedener dimensionaler Modelle emotionalen Erlebens im Circumplex-Modell (nach Yik, Russell, & Feldman Barrett, 1999, S. 601, Figure 1)	33
Abbildung 5. Doubleloop-Perturbation-Modell neuronaler Aktivität. Eigene Darstellung in Anlehnung an Menon (2011).	62
Abbildung 6. Fünf x Fünf-Wirkungsmodell und eingesetzte Instrumente.....	102
Abbildung 7. Grundlegender Aufbau eines Klassifikator-Induktors (Flussdiagramm) (eigene Darstellung).....	125
Abbildung 8. Schematische Darstellung einer 2 x 2-Kontingenztafel.	128
Abbildung 9. Grundlegender Aufbau eines Klassifikator-Induktors bei 10-fach Kreuzvalidierung (Flussdiagramm).....	131
Abbildung 10. Schematische Darstellung des Vorgehens beim Bootstrapping zur Schätzung statistischer Kennwerte unter Verwendung eines Klassifikator-Induktors (eigene Darstellung).	144
Abbildung 11. Deskriptive Statistik der einzelnen Items der Eigenschaftswörter zu Emotionen und Beanspruchungserleben während der Arbeit (EEB).	151
Abbildung 12. P-P-Plot für das Item „zufrieden“. Die gepunktete Linie stellt die Regressions- Gerade dar.	153
Abbildung 13. CV-T-Diagramme für Bayes-Klassifikatoren auf Basis sämtlicher 25 EEB- Items für jedes der 18 Merkmale.....	155
Abbildung 14. CV-T-Diagramme für Bayes-Klassifikatoren auf Basis der neun WSIB-Items für jedes der 18 Merkmale.....	158
Abbildung 15. CV-T-Diagramme für binärlogistische Regression auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale.....	161
Abbildung 16. CV-T-Diagramme für binärlogistische Regression auf Basis der neun WSIB- Items für jedes der 18 Merkmale.....	164
Abbildung 17. Vorhersagefehler und die Standardabweichung dieses Fehlers in Abhängigkeit des Stichprobenumfanges am Beispiel von Aufgabenanforderungen. Dargestellt sind auch die	

Abbildungsverzeichnis

Anpassung an die Anpassung an die empirischen Daten durch logarithmische Regression sowie die Anpassung, die sich aus der Betrachtung aller 18 Kriterien ergibt.....	252
Abbildung 18. Validierung der Klassifikatoren bezüglich des WSIB zur Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium.	261
Abbildung 19. Validierung der Klassifikatoren bezüglich des WSIB zur Vorhersage der Erwartungswerte je Kriterium.	262
Abbildung 20. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die gesamte Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	266
Abbildung 21. Vorhersage der Erwartungswerte je Kriterium für die gesamte Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	267
Abbildung 22. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die Teilstichprobe „Führungskräfte“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	268
Abbildung 23. Vorhersage der Erwartungswerte je Kriterium für die Teilstichprobe „Führungskräfte“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	269
Abbildung 24. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die Teilstichprobe „Produktion“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	270
Abbildung 25. Vorhersage der Erwartungswerte je Kriterium für die Teilstichprobe „Produktion“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB.	271

ABKÜRZUNGSVERZEICHNIS

- a. a. O. *am angegebenen Ort*
- ACTH *Adrenocorticotropes Hormon*
- AD-ACL *Activation-Deactivation
Adjective Check List*
- ATP *Adenosintriphosphat*
- BLR *Binärlogistische Regression*
- CEN *Central Executive Network*
- CMQ *Current Mood Questionnaire*
- CRH *corticotropin releasing hormone*
- CSV *Comma Separated Values*
- d. h. *das heißt*
- DMN *Default Mode Network*
- dMPFC *Dorsaler Anteil des medialen
Präfrontalkortex*
- EEB *Eigenschaftswörterliste zur
Erfassung von Emotionen und Befinden
während der Arbeit*
- etc. *et cetera (und so fort)*
- FFZ *Fragebogen zu Führung und
Zusammenarbeit*
- GABA *Gamma-Aminobuttersäure*
- ggf. *gegebenenfalls*
- GKF *Gesundheitskompetenz-Fragebogen*
- HkB *Fragebogen zu Häufigkeiten
körperlicher Beschwerden*
- HPA-System *Hypothalamic Pituitary
Adrenal System*
- ICD *International Classification of
Diseases*
- MACL *Mood Adjective Checklist*
- MTL *medial temporal lobe*
- NA *Negative Affect*
- NADH *Nikotinamid-Adenin-Dinucleotid-
Hydrid*
- PA *Positive Affect*
- PANAS *Positive and Negative Affect
Schedule*
- PTBS *Posttraumatische
Belastungsstörung*
- REM-Schlaf *Repid-Eye-Movement-Schlaf*
- s. *siehe*
- S. *Seite*
- s. o. *siehe oben*
- s. u. *siehe unten*
- SAM-System *Sympathisch-Adreno-
Medulläres System*
- SFFS *Sequential Forward Floating
Selection*
- SFS *Sequential Forward Selection*
- SN *Salient Network*
- SRP *Stressreaktionsprozess*
- SW-Schlaf *Slow-Wave-Schlaf*
- SynBA *Synthetische Belastungs- und
Arbeitsanalyse*
- u. a. *unter anderem*
- ugs. *umgangssprachlich*
- WITAL *Wuppertaler Inventar zur
ressourcenorientierten Arbeits- und
Organisationsgestaltung*
- WSIB *Wuppertaler Screening Instrument
psychische Beanspruchung*
- WSIB Pro *Wuppertaler Screening
Instrument psychische Beanspruchung -
Profilanalyse*

z. T. *zum Teil*

1 EINLEITUNG

Mit der vorliegenden Arbeit werden folgende Ziele verfolgt: Erstens, soll dem zunehmenden Bedarf von Unternehmen an Unterstützung bezüglich des Umgangs mit psychischen Belastungen und Beanspruchungen entsprochen werden.¹ Zweitens werden die theoretischen, d. h. neurobiopsychologischen Grundlagen des dafür in dieser Arbeit entwickelten Instruments, dem „Wuppertaler Screening Instrument psychische Beanspruchung – Profilanalyse (WSIB Pro)“ ausführlich dargestellt. Drittens, soll mit dem hier beschriebenen Ansatz u. a. gezeigt werden, in welcher Weise die Arbeits- und Organisationspsychologie von der theoretischen Begründung und einer gründlichen empirischen Fundierung im Kontext psychischer Belastung und Beanspruchung profitieren kann. Zur Einführung der Leserinnen und Leser in die Thematik werden in diesem einleitenden Kapitel Aufbau und Herangehensweise der vorliegenden Arbeit etwas ausführlicher dargestellt.

Im Zentrum der vorliegenden Arbeit steht die psychische Beanspruchung während der Arbeit, die in Form von verschiedenen Beanspruchungszuständen erhoben wird. Messinstrument ist die „Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchung in der Arbeit (EEB)“ von Wieland (Wieland, 2004, 2006a; Wieland, Krajewski & Memmou, 2006; Wieland, Metz & Richter, 2001), die inzwischen auch als Kurzform, dem „Wuppertaler Screening Instrument psychische Beanspruchung (WSIB)“ vorliegt (Hammes & Wieland, 2012a, b; Wieland, 2013; Wieland & Hammes, 2014b). Diese Instrumente gehen auf die Forschung von Rainer Wieland zurück.² Ausgangspunkt seiner Forschung waren die Erkenntnisse der „mental-workload Forschung“ sowie dimensionaler Modelle von Emotionen (Thayer, 1967; Watson, Clark & Tellegen, 1988; Wieland-Eckelmann, 1992; Wieland-Eckelmann & Bösel, 1987). Die EEB wird in der Arbeitsgruppe von Rainer Wieland als Bestandteil eines Fragebogeninventars, dem „Wuppertaler Inventar zur ressourcenorientierten Arbeits- und Organisationsgestaltung (WITAL)“ (Wieland, 2014b), eingesetzt. Konzeptueller

¹ Dieser Bedarf ergibt sich aus zum einen aus zunehmenden Fehlzeiten auf Grund psychischer und Verhaltensstörungen (Bundesministerium für Arbeit und Soziales [BMAS], 2013; Lohmann-Haislah, 2012; Seiler & Jansing, 2014), zum anderen aus der inzwischen auch durch das Arbeitsschutzgesetz vorgeschriebenen Forderung, psychische Belastungen in Gefährdungsbeurteilungen zu berücksichtigen sowie einer damit einhergehenden Neuausrichtung des deutschen Systems des Arbeits- und Gesundheitsschutzes (BMAS, 2014; Bechmann, Jäckle, Lück & Herdegen, 2011; Beck, Richter, Ertel & Morschhäuser, 2012; Berufsverband Deutscher Psychologinnen und Psychologen, 2008; Cosmar, Eichendorf & Portuné, 2014; Eigenstetter, Kunz, Portuné & Trimpop, 2014; Joiko, Schmauder & Wolff, 2010; Länderausschuss für Arbeitsschutz und Sicherheitstechnik, 2002, 2003, 2009; Lohmann-Haislah, 2012; Nationale Arbeitsschutzkonferenz, 2012; Satzer & Langhoff, 2010; Scholz, 2013; Schütte & Köper, 2013; Trimpop et al., 2012; Wieland, 2013).

² Zahlreiche Einzelbeiträge von und mit Rainer Wieland zur EEB werden im Methodenteil zur Darstellung der Entwicklung und Validierung dieses Instrumentes verarbeitet.

Rahmen dieses Inventars ist das „Fünf x Fünf-Wirkungsmodell“ (Wieland, 2008, 2009, 2013, 2014a), in dem die Wechselwirkungen von fünf Merkmalsbereichen eines Arbeitssystems abgebildet werden: Arbeitsaufgabe und Arbeitsplatz, Zusammenarbeit und Führung, Merkmale der Person, Erleben und Befinden (Beanspruchung) sowie Leistung und Gesundheit. Psychische Beanspruchung steht als Prozessvariable im Zentrum dieses Wirkungsgefüges. Neben anwendungsbezogenen Zielstellungen, d. h. der praktischen Anwendung der EEB, besteht die wissenschaftliche Zielsetzung dieser Arbeit in folgenden Punkten: (a) der Ausarbeitung eines neurobiopsychologischen Verständnisses von psychischer Beanspruchung, (b) einer darauf aufbauenden Begründung für die Eignung der EEB als Instrument zur Erfassung psychischer Beanspruchung, (c) einer theoretischen Begründung für die zentrale Bedeutung von psychischer Beanspruchung in dem Wirkungsgefüge des Fünf x Fünf Wirkungsmodells, sowie schließlich (d) der empirischen Überprüfung der theoretisch zentralen Bedeutung, die der psychischen Beanspruchung im Arbeitsprozess zukommt.

Im Theorieteil der vorliegenden Arbeit erfolgt eine vertiefte Auseinandersetzung mit den neurobiopsychologischen Grundlagen von psychischer Belastung und psychischer Beanspruchung. Hiermit wird eine Anknüpfung an die psychisch relevanten physiologischen Grundlagen der „mental-workload Forschung“ sowie dimensionaler Modelle der Emotion hergestellt (Hockey, 1997; LeDoux, 2012; Nowlis & Nowlis, 1956; Posner, Russell & Peterson, 2005; Pribram & McGuinness, 1975; Rohmert, 1984; Sanders, 1983; Schönplflug, 1991; Schulz, 2012; Thayer, 1967). Damit erfolgt zugleich eine theoretische Begründung aus neurobiopsychologischer Perspektive, warum ein Instrument wie die EEB geeignet ist, psychische Beanspruchung zu erfassen und warum der psychischen Beanspruchung in der Erwerbsarbeit eine zentrale Bedeutung zugeschrieben wird (Wieland, 2008, 2009, 2013, 2014a). Diese neurobiopsychologische Ausrichtung ist neben den Arbeiten von und dem persönlichen Austausch mit Rainer Wieland von vier miteinander gut vereinbaren Grundanschauungen geleitet: (a) Einer persönlichen naturalistischen Grundüberzeugung. Meiner Überzeugung nach manifestiert sich psychische Beanspruchung in Form körperlicher Prozesse, die eine Widerspiegelung der Interaktion des Menschen mit seiner Umwelt in seinen körperlichen Strukturen beinhalten (vgl. Birbaumer & Schmidt, 2010; Leontjew, 1982; Rohmert, 1984; Schulz, 2012).³ (b) Einer systemischen Betrachtung der relevanten Prozesse, wobei vom menschlichen Organismus als offenem System ausgegangen wird. Seine Strukturelemente stehen in verstärkender und hemmender wechselseitiger Interaktion. Seine

³ Auch, wenn es weitere Autoren gibt, die eine vergleichbare Anschauung vertreten, waren die beiden zitierten Texte für mich in dieser Beziehung besonders inspirierend.

übergeordneten Strukturen (Organe, Organsysteme) stabilisieren sich selbstorganisiert bei beständigem Austausch von Energie und Stoff mit der Umwelt (Birbaumer & Schmidt, 2010; Penzlin, 2014; Strunk & Schiepek, 2006). (c) Der Annahme einer psychischen Tiefenstruktur bei jedweder menschlicher Tätigkeit. Während das äußerlich beobachtbare Verhalten die Oberflächenstruktur einer Tätigkeit beschreibt, umfasst die nicht beobachtbare psychische Tiefenstruktur „die *regulativ* wirksamen psychischen Vorgänge [Kognition, Motivation und Emotion] und Repräsentationen in der durch die Regulationsfunktion bedingten Ordnung“ (Hacker & Sachse, 2014, S. 120, 569; vgl. auch Miller, Galanter & Pribram, 1960). (d) Der Normreihe DIN EN ISO 10075, in der psychische Beanspruchung wie folgt definiert wird: „Psychische Beanspruchung ist die unmittelbare (nicht langfristige) Auswirkung der psychischen Belastung im Individuum in Abhängigkeit von seinen jeweiligen überdauernden und augenblicklichen Voraussetzungen, einschließlich der individuellen Bewältigungsstrategien“ (Joiko, Schmauder & Wolff, 2010, S. 10). Diese Grundanschauungen werden – mit Ausnahme von Punkt (c) – zu Beginn des Theorieteils am Beispiel der mechanischen Belastung und Beanspruchung des Muskel-Skelett-Systems veranschaulicht. Zugleich wird das Beispiel der mechanischen Belastung und Beanspruchung sowohl als (praxistaugliche) Metapher als auch als adaptiertes theoretisches Vorbild für die Konzeption von psychischer Belastung und Beanspruchung eingeführt (vgl. Joiko et al., 2010; Rohmert, 1984).

Der erste Teil der theoretischen Auseinandersetzung handelt von Arousal. Arousal kann vereinfacht als Ausmaß der Handlungs- und Reaktionsbereitschaft verstanden werden und hängt eng mit dem Grad der Wachheit zusammen. Auf der Grundlage aktueller neurobiologischer Kenntnisse (z. B. Brown, Basheer, McKenna, Strecker & McCarley, 2012; Hüther, 2014; Jones, 2003; Rasch & Born, 2013) wird mit der „Arousal-Uhr“ ein erstes heuristisches, theoretisches Modell abgeleitet. Dieses unterscheidet sechs grundlegende Verhaltenskategorien: Wachen, Bewältigen, Konsumieren, mentales Simulieren, Schlafen und Aufgeben. Über die Dimensionen Valenz und Aktivierung lässt sich ein direkter Bezug zu dimensional Modellen der Emotion (Circumplex-Modell der Core Affects, siehe z. B. Yik, Russell & Feldman Barrett, 1999) herstellen. Darüber hinaus lassen sich über neurobiologische Stresskonzepte (z. B. Hüther, 2014) Bezüge zu psychosozial (mit-) verursachten körperlichen und psychischen Beschwerde- und Krankheits- bzw. Störungsbildern herstellen. Ein weiteres heuristisches Modell wird aus aktuellen Erkenntnissen über großräumig verteilte Netzwerke der Großhirnrinde abgeleitet (z. B. Andrews-Hanna, 2012; Fox et al., 2005; Menon, 2011; Miller & Cohen, 2001). Grundlage des „Doubleloop-Perturbation-Modells“ (kurz „ ∞ -Modell“) ist die Annahme dreier großräumig verteilter kortikaler Netzwerke: Dem Central Executive

Network, (CEN; zuständig für Problemlösung), dem Default Mode Network (DMN; zuständig für mentale Simulationen der sozialen Rolle des Selbst in Vergangenheit, Gegenwart und Zukunft) sowie dem Salient Network (SN; zuständig für die Entdeckung von Fehlern in der Handlungsausführung, Konflikten und Bedrohungen) (vgl. Menon, 2011). Dieses Tripel-Netzwerk-Modell wird erweitert um die Arousal-Systeme (s. o.). Das ∞ -Modell postuliert, dass sich das menschliche Gehirn – sofern es nicht durch externe (sensumotorische oder viszerale) Eingaben gestört wird – in einer Doppelschleife befindet: Mental simulierte Situationen im Default Mode Network (DMN) werden bei Feststellung eines Konfliktes (Arousal-Systeme und Salient Network; SN) zur Problemlösung an das Central Executive Network (CEN) weitergeleitet. Sobald der Konflikt unter Beanspruchung des CEN gelöst werden konnte, stellen sich neue mentale Simulationen (DMN) ein, bis wieder ein Konflikt entdeckt wird etc. (vgl. Fox et al., 2005; Menon, 2011). Sensumotorische oder viszerale (körperliche) Signale, die von den Arousal-Systemen und dem SN als Fehler, Konflikt oder Gefahr bewertet werden, führen zu Störungen dieser Doppelschleife. In diesem Falle wird das CEN für die Lösung des aktuellen Problems in Dienst genommen, während Simulationen im DMN gehemmt werden. Das ∞ -Modell konkretisiert mit den beschriebenen Prozessen, was aus neurobiopsychologischer Perspektive unter psychischer Belastung und psychischer Beanspruchung zu verstehen ist. Zugleich wird über die Integration der Arousal-Uhr und ihrem engen Bezug zu dimensional Modellen der Emotion deutlich, warum ein Instrument wie die EEB geeignet ist, psychische Beanspruchung zu erfassen (vgl. Feldman Barrett, 2006b; LeDoux, 2012; Nowlis & Nowlis, 1956; Rohmert, 1984; Thayer, 1978).

Im verbleibenden Teil der theoretischen Auseinandersetzung werden grundsätzliche menschliche Bedürfnisse besprochen. Den wichtigsten Ausgangspunkt stellt die Self-Determination Theory von Deci und Ryan (2000) dar. Ihr werden die Grundbedürfnisse nach Kompetenz sowie sozialer Teilhabe und Zugehörigkeit entnommen und vor dem Hintergrund des ∞ -Modells diskutiert. Es ergibt sich, dass beide Grundbedürfnisse (a) im Sinne des ∞ -Modells interpretiert werden können (Heatherton, 2011; Lord, Diefendorff, Schmidt & Hall, 2010) und (b) als (intrinsische) „Wachstumsmotive“ aufgefasst werden können, die auf mentalen Simulationen des DMN beruhen (a. a. O.; Baumeister, Masicampo & Vohs, 2011). Die beiden Grundbedürfnisse werden um das Grundbedürfnis nach Sicherstellung der Homöostase (Integration und Funktionsfähigkeit des Organismus) ergänzt (vgl. Maslow, 1943). Dieses wird bei homöostatischen Krisen aktiviert, die direkt in den Arousal-Systemen als Störungen abgebildet werden. Die theoretische Betrachtung schließt ab mit einer Diskussion des Spannungsfeldes zwischen Autonomie und Selbstkontrolle. Autonomie wird hier, anders

als bei Deci und Ryan (2000), definiert als die *umfassende Erfüllung persönlicher Bedürfnisse in der persönlich bevorzugten Art und Weise*. Selbstkontrolle wird notwendig, wenn die persönlichen Bedürfnisse oder die persönlich bevorzugte Art und Weise ihrer Erfüllung mit den äußeren physischen oder sozialen Bedingungen in Konflikt geraten. Da Selbstkontrolle die Aktivität des CEN voraussetzt, und diese energetisch sehr aufwändig ist, bevorzugt der menschliche Organismus im Allgemeinen autonomes Handeln. Im Rahmen der Betrachtung der Grundbedürfnisse werden Bezüge zu etablierten Konzepten aus der Arbeits- und Organisationspsychologie hergestellt, die auch aus der hier eingenommenen Perspektive als „menschengerecht“ bzw. „human“ bewertet werden können.

Die in der vorliegenden Arbeit verwendeten Methoden sind der statistischen Mustererkennung bzw. Klassifikation sowie der Merkmalsselektion zuzuordnen (z. B. Clarke, Fokoué & Zhang, 2009; Hastie, Tibshirani & Friedman, 2009; Maimon & Rokach, 2005). Sie werden unter dem Stichwort der Klassifikatorinduktion gemeinsam behandelt. Ziel dieser Methoden ist es, ausgehend von den 25 (EEB) bzw. 9 (WSIB) Eigenschaftswörtern auf die Ausprägung von Arbeitsplatzmerkmalen zu schließen, wie sie im Fünf x Fünf-Wirkungsmodell expliziert sind. Insgesamt sollen 18 Arbeitsplatzmerkmale aus den Bereichen *Arbeitsgestaltung, Führung und Zusammenarbeit* sowie *Gesundheitsindikatoren* vorhergesagt werden. Im Methodenteil dieser Arbeit werden die notwendigen Schritte für die Klassifikatorinduktion erläutert sowie die jeweils vorgenommene Auswahl an Algorithmen und Verfahren begründet. Notwendige Schritte sind (vgl. Jain, Duin & Mao, 2000; Kotsiantis, Kanellopoulos & Pintelas, 2006; Maimon & Rokach, 2005): (1) Auswahl geeigneter Stichproben, (2) Aufbereitung der Daten, (3) Induktion und (Kreuz-) Validierung von Klassifikatoren, (4) endgültige Auswahl geeigneter Klassifikatoren, (5) Kalibrierung und (6) externe Validierung der endgültig ausgewählten Klassifikatoren. Die zur Klassifikatorinduktion verwendeten Daten entstammen vier Umsetzungsprojekten zur betrieblichen Gesundheitsförderung in unterschiedlichen Branchen und für unterschiedliche Berufsgruppen (Hammes, Wieland & Winizuk, 2009; Wieland & Hammes, 2014; Wieland, Winizuk & Hammes, 2009). Für die 18 Arbeitsplatzmerkmale stehen zwischen 523 und 1549 Datensätze zur Verfügung, die zur Klassifikatorinduktion verwendet werden. In Schritt (3) werden Algorithmen und Verfahren in verschiedenen Experimenten zur Klassifikatorinduktion variiert. Verglichen werden zum einen verschiedene Verfahren zur Merkmalsselektion (vollständige Suche, schrittweise Vorwärtssuche, schrittweise „fließende“ Vorwärtssuche; Somol, Novovicova & Pudil, 2010). Diese Verfahren dienen dazu, aus den 25 bzw. 9 Eigenschaftswörtern jene auszuwählen, die zu einer besonders guten Vorhersage der 18 Arbeitsplatzmerkmale führen. Es wird sich zeigen,

dass zur Vorhersage eines einzigen Arbeitsplatzmerkmals zwischen zwei und vier Eigenschaftswörter genügen. Zum anderen kommen binärlogistische Regression (Agresti, 2002, 2007) und Bayes-Klassifikation (Duda, 1970; Wald, 1949) als zwei Verfahren zur Klassifikation zum Einsatz. Im Vergleich, so die Ergebnisse, haben Bayes-Klassifikatoren vor allem einen Effizienz- und Stabilitäts-Vorteil gegenüber binärlogistischer Regression. Hinsichtlich der Vorhersagegenauigkeit (Korrektklassifikation) sind beide Verfahren jedoch gleichwertig. Diese Schlussfolgerungen gründen auf 10-fach Kreuzvalidierung (s. Efron, 1983; Stone, 1974) der jeweiligen Verfahren. Für die Schritte (4) und (5) kommt Bootstrapping (Efron, 1979, 2003) zum Einsatz. Die endgültige Auswahl von Klassifikatoren ist notwendig, da in den verschiedenen Experimentalreihen sowie bei der 10-fach Kreuzvalidierung je Arbeitsplatzmerkmal eine Reihe von Klassifikatoren induziert werden, die als Kandidaten in Frage kommen. Bootstrapping führt zu einer validen Einschätzung ihrer Vorhersagegenauigkeiten. Diese können als Kriterium für die endgültige Auswahl herangezogen werden. Unter Berücksichtigung möglicher Überanpassung an die zu Grunde gelegten Daten erreichen die ausgewählten Klassifikatoren eine Vorhersagegenauigkeit zwischen 70 und 75 %, was dem theoretischen Optimum entspricht (Hughes, 1968). Kalibrierung (Schritt 5) meint die Erfassung systematischer Klassifikationsfehler sowie deren anschließende Berücksichtigung bei der Anwendung. Die Vorhersagegenauigkeit von Ausprägungen der Arbeitsplatzmerkmale (Erwartungswerte) erfolgt in 99,8 %-Vertrauensintervallen, die bei einem Stichprobenumfang von $n = 100$ maximal 18 % der jeweiligen Gesamtskala umfassen. Darüber hinaus besteht ein funktionaler (logarithmischer) Zusammenhang zwischen den Vorhersagefehlern für Erwartungswerte und dem Stichprobenumfang. Je geringer der Stichprobenumfang ist, desto größer ist dieser Vorhersagefehler. Für die externe Validierung (Schritt 6) liegen zwei weitere Datensätze aus Umsetzungsprojekten zur betrieblichen Gesundheitsförderung mit 159 bzw. 160 Teilnehmerinnen und Teilnehmern vor. Diese Datensätze werden nicht in die Entwicklung und Auswahl der Klassifikatoren einbezogen, sie sind diesen Klassifikatoren also vollkommen „unbekannt“. Eine der beiden Stichproben ist in 11 Bereiche mit 7 bis 22 Teilnehmern unterteilt. Werden die Klassifikatoren auf diese Datensätze angewandt, so liegen die tatsächlich gemessenen Merkmalsausprägungen in 201 von 210 Fällen innerhalb der vorhergesagten 99,8 %-Vertrauensintervalle. Dies entspricht einer Quote von 95,7 % korrekter Vorhersagen.

Produkt der oben beschriebenen Verfahrensweisen ist das „Wuppertaler Screening Instrument psychische Beanspruchung – Profilanalyse (WSIB Pro)“. Als Erhebungsinstrument kommt das Wuppertaler Screening Instrument psychische Beanspruchung (WSIB; Wieland & Hammes, 2014) zum Einsatz. WSIB Pro ist eine internetbasierte Anwendung, in der die oben

ermittelten Klassifikatoren angewendet werden. Mit WSIB Pro können auf Grundlage von Befragungen mittels des WSIB (bestehend aus nur 9 Eigenschaftswörtern) Profile bezüglich der 18 Arbeitsplatzmerkmale (Arbeitsgestaltung, Führung und Zusammenarbeit, Gesundheitsindikatoren) bestimmt und mit Vergleichsmittelwerten („Benchmarks“) verglichen werden. Auf Grundlage dieser Profile kann im Rahmen betrieblichen Gesundheitsmanagements entschieden werden, ob in einer Abteilung Interventionsbedarf besteht oder nicht. Außerdem gibt WSIB Pro erste Hinweise darüber, bei welchen der 18 Arbeitsplatzmerkmale besonderer Bedarf zur (Um-) Gestaltung zu erwarten ist. Somit wurde ein niederschwelliges, ökonomisches und valides Screening-Instrument entwickelt, das betrieblichen Akteuren erste Hinweise zum Interventionsbedarf und Orientierung bietet. WSIB Pro wurde während der Entstehung dieser Arbeit in ein bereits bestehendes Projekt zum betrieblichen Gesundheitsmanagement anwendungsreif integriert. Im Diskussionsteil der vorliegenden Arbeit wird diskutiert, inwieweit WSIB Pro um sinnvolle Funktionen für Praxis und Forschung erweitert werden kann. Außerdem wird diskutiert, inwieweit die hier verwendeten Verfahren auch in anderen Forschungs- und Anwendungsfeldern sinnvoll eingesetzt werden können.

Auf Grund der zahlreichen Einzelbefunde, die bei der Klassifikatorinduktion angefallen sind, ist diese Arbeit in zwei separate Teile gegliedert: Diesen vorliegenden Haupttext sowie einen separaten, recht umfangreichen Anhang. Der Anhang enthält zum einen Details zu mathematischen und algorithmischen Notationen. Zum anderen finden sich dort sämtliche Parameter der Klassifikatoren, vollständige Ergebnisberichte der Klassifikatorinduktion sowie Detailergebnisse der Validierung von Klassifikatoren. Im vorliegenden Haupttext wurden die Ergebnisberichte mit Ausnahme eines Beispiels auf die wesentlichen Einzelbefunde beschränkt. Der Anhang kann als elektronische Ressource beim Autor angefragt werden.

2 PSYCHISCHE BELASTUNG UND BEANSPRUCHUNG: NEUROBIOPSYCHOLOGISCHE GRUNDLAGEN

Das Begriffspaar *psychische Belastung* und *psychische Beanspruchung* hat sich nicht nur in den Arbeits- und Sozialwissenschaften (Rohmert, 1984; Rohmert & Rutenfranz, 1975), sondern auch bei den Akteuren der Arbeitssicherheit und des Gesundheitsschutzes etabliert. Die Norm DIN EN ISO 10075 (Demerouti et al., 2012; Joiko, Schmauder & Wolff, 2010; Nachreiner, 1999; Richter, 2000) stellt eine europaweite Übereinkunft über dessen Gebrauch dar. Psychische Belastung wird dort definiert als „die Gesamtheit aller erfassbaren Einflüsse, die von außen auf den Menschen zukommen und psychisch auf ihn einwirken“ (Joiko et al., 2010, S. 9). Psychische Belastung ist zunächst weder gut noch schlecht. Sie bewirkt jedoch im betroffenen Menschen eine psychische Gegenreaktion, die von dessen überdauernden und augenblicklichen Voraussetzungen abhängt. In der DIN EN ISO 10075 wird daher definiert: „Psychische Beanspruchung ist die unmittelbare (nicht langfristige) Auswirkung der psychischen Belastung im Individuum in Abhängigkeit von seinen jeweiligen überdauernden und augenblicklichen Voraussetzungen, einschließlich der individuellen Bewältigungsstrategien“ (Joiko et al., 2010, S. 10).

Peter Schulz (2012) fasst Beanspruchung in Anlehnung an Rohmert (1984) und Schönpflug (1991) als „Inanspruchnahme der Funktionen von Organen bzw. Organsystemen zur Bewältigung der unterschiedlichen Arten von Anforderungen“ (Schulz, 2012, S. 34) auf. Beispielsweise meint er mit körperlicher Beanspruchung in der Hauptsache die Inanspruchnahme des Stütz- und Bewegungsapparates (a. a. O., S. 27). Für die weiteren Beanspruchungsarten, mentale, sensorische, sensumotorische, emotionale, motivationale und sozial-kommunikative Beanspruchung, stellt er die Inanspruchnahme neurobiologischer Strukturen bzw. des Gehirns und des Zentralnervensystems heraus (a. a. O., S. 27-29). Über die genannten Beanspruchungsarten hinaus führt Schulz außerdem die Inanspruchnahme der biochemischen bzw. zellulären Regenerationsfunktionen auf (a. a. O., S. 29).

Schulz ist der Ansicht, „dass sich Beanspruchung über die Anregung und Hemmung von Organfunktionen realisiert“ (a. a. O., S. 34). Gesundheitliche Beeinträchtigungen bzw. „funktionelle Einbußen und [langfristig] strukturelle Schäden an Organen“ (a. a. O., S. 40) führt er auf *Fehlbeanspruchung* zurück. Damit meint er eine übermäßige oder unvollständige Anregung bzw. Hemmung von Organfunktionen. Positive Gesundheitseffekte setzen demgegenüber *optimale Beanspruchung* voraus. Unter qualitativen Gesichtspunkten sollte die Inanspruchnahme von Organfunktionen zur erfolgreichen Bewältigung von Alltags-

anforderungen gelingen. Unter quantitativen Gesichtspunkten sollten „alle verfügbaren Organfunktionen [...] ausreichend, aber maßvoll dosiert (nicht übermäßig)“ (a. a. O., S. 34) angeregt werden.

Im vorliegenden Kapitel werden die Inanspruchnahme von Organfunktionen sowie die sie auslösenden Belastungsfaktoren spezifiziert. Als erstes Beispiel wird die Belastung und Beanspruchung des Stütz- und Bewegungsapparates beschrieben. Im Zuge dessen wird das mechanische Belastungs-Beanspruchungs-Paradigma dargestellt, das als theoretisches Vorbild für das psychische Belastungs-Beanspruchungs-Paradigma diene (vgl. Rohmert, 1984). Der Schwerpunkt dieses Kapitels liegt jedoch - wie bei Schulz (2012) - bei der Inanspruchnahme des Zentralnervensystems. Es wird sich zeigen, dass das biopsychologische Konzept des Arousal bzw. der generellen Aktiviertheit eine zentrale Stellung bei der Begründung psychischer Kategorien wie Emotion, Motivation, Verhalten und Lernen einnimmt. Zugleich kann mit diesem Konzept theoretisch begründet werden, warum arbeitspsychologische Prävention wesentlich ist, um arbeitsbedingte körperliche und psychische Beeinträchtigungen und Störungen zu vermeiden.

2.1 BEANSPRUCHUNG ALS INANSPRUCHNAHME UND ANPASSUNG VON ORGANSYSTEMEN

Das psychologische Paradigma der Belastung und Beanspruchung lehnt sich an dem Paradigma der mechanischen Belastung und Beanspruchung an (vgl. Rohmert, 1984). Dass bereits dieses mechanische Paradigma für die körperliche und psychische Gesundheit eines Menschen relevant ist, wird im ersten Abschnitt dargestellt. Dem Kern des vorliegenden Kapitels - der Inanspruchnahme des Zentralnervensystems - werde ich mich anschließend mit LeDoux (2012) über die Inanspruchnahme und Anpassung sogenannter Überlebens-Schaltkreise annähern.

Obwohl die vorliegende Arbeit von psychischer Belastung und Beanspruchung handelt, wird im Folgenden zu Beginn in gewisser Ausführlichkeit auf mechanische Belastung und Beanspruchung des Stützapparates eingegangen. Wie bereits in der Einleitung formuliert, erfolgt dies aus zwei Gründen. Zum einen ist mechanische Belastung und Beanspruchung das adaptierte theoretische Vorbild für psychische Belastung und Beanspruchung und kann zugleich als praxistaugliche Metapher herangezogen werden (Joiko et al., 2010; Rohmert, 1984). Zum anderen sollen drei der vier wesentlichen Grundannahmen, die dieser Arbeit zu Grunde liegen, an diesem Konzept verdeutlicht werden: Einerseits die Widerspiegelung der Interaktion von Individuum und Umwelt in organismischen Strukturen sowie die systemische Interaktion der beteiligten Systembestandteile. Andererseits die Unterscheidung von Belastung

als Summe aller äußerer Einflussgrößen und Beanspruchung als innerorganismische Reaktion und Anpassung auf diese Einflussgrößen.

2.1.1 Mechanische Belastung und Beanspruchung am Beispiel des Stützapparates

Das Begriffspaar Belastung und Beanspruchung entstammt dem Teilgebiet Festigkeitslehre der technischen Mechanik (Holzmann, Meier & Schumpich, 2012). Eine Definition für mechanische Belastung könnte folgendermaßen lauten: Mechanische Belastung ist die Gesamtheit aller Kräfte, die von außen auf einen Körper zukommen und mechanisch auf ihn einwirken (vgl. a. a. O., S. 11). Mechanische Belastung ist zunächst weder gut noch schlecht. Sie bewirkt jedoch im Inneren des betroffenen Körpers eine Gegenkraft. Diese Kraftwirkung der mechanischen Belastung im Inneren des Körpers wird mechanische Beanspruchung genannt (a. a. O., S. 1-2). Im Falle eines starren Körpers herrscht zwischen den eingepprägten Kräften (Belastung) und den Kräften in dessen Innerem (Beanspruchung) mechanisches Gleichgewicht. Der starre Körper ist formstabil. Verformbare feste Körper verfügen über individuelle Elastizität, die auf den Wechselwirkungen atomarer Bestandteile beruht. Wird einem verformbaren festen Körper eine Kraft eingepprägt (Belastung), so werden dessen atomare Bestandteile soweit gegeneinander verschoben, bis die resultierende innere Kraft (Beanspruchung) der Belastung in Größe und Richtung genau entgegengerichtet ist (a. a. O.).

Für die Elastizität macht es einen Unterschied, ob ein Körper aus kristallinem Kohlenstoff (Diamant) oder aus Aluminium besteht und ob die Atome, z. B. in Abhängigkeit der Temperatur, augenblicklich mehr oder weniger stark aneinander gebunden sind. Mechanische Beanspruchung ist somit abhängig von den überdauernden und augenblicklichen Eigenschaften des betroffenen Körpers. Ist die mechanische Belastung so stark, dass Atome dauerhaft gegeneinander verschoben oder voneinander getrennt werden, liegt Überbeanspruchung vor. Der Körper verformt sich dauerhaft, bekommt Risse oder bricht (a. a. O.).

Wie die beschriebene innere Dynamik und mechanische Verformung eines Körpers zu bewerten ist, hängt selbstverständlich von den Perspektiven und Interessen verschiedener Beobachter ab. Im Falle eines Beinbruchs ist die Bewertungstendenz vermutlich einhellig: Einem solchen Ereignis beizuwohnen, ist weder für den betroffenen Menschen, noch für außenstehende Beobachter angenehm. Offenbar verdeutlicht ein Knochenbruch jedoch die physische und psychische Relevanz des mechanischen Belastungs-Beanspruchungs-Modells für die Gesundheit des Menschen. Die physische Relevanz dieses Modells untersuchte Friedrich Pauwels eingehend in einer Reihe von Schriften zur "funktionellen Anatomie und kausalen Morphologie des Stützapparates". In seinem ersten Beitrag legt er dar, dass der

gesamte Stützapparat in seinem Aufbau ingenieurstechnischen Grundsätzen folgt: "Kurz zusammengefaßt, lehrt die quantitative mechanische Analyse, daß im Stütz- und Bewegungsapparat Konstruktionsprinzipien verankert sind, durch welche die gefährliche Komponente der Beanspruchung des Röhrenknochens - die Biegebeanspruchung - möglichst klein gehalten ist" (Pauwels, 1948, S. 166). In seinem zweiten Beitrag leitet er unter Verwendung dieser Konstruktionsprinzipien ein Modell der Anatomie und Morphologie der unteren Extremitäten ab (Pauwels, 1950). Seine Ableitungen resultieren in einigen schematischen Darstellungen, in denen der Länge des Beines nach frontal und sagittal (seitlich) mechanische Belastung (eingeprägte Kräfte) und mechanische Beanspruchung (Gesamt-, Druck-, Scher- und Biegespannung) gegenüber gestellt sind. Diese Darstellungen veranschaulichen „unmittelbar die Größe der Bruchgefährdung, die durch die einzelnen Beanspruchungsarten hervorgerufen wird“ (a. a. O., S. 537). Die höchste Bruchgefährdung liegt an Stellen mit Beanspruchungsspitzen vor. Auch mit der Anwendung des Konzepts der mechanischen Belastung und Beanspruchung auf die Muskulatur konnten kritische Belastungskonstellationen ermittelt werden. Denoth und Stacoff (1991) stellen fest, dass die Wahrscheinlichkeit für Muskelverletzungen mit der Höhe der Dehnungsgeschwindigkeit wächst. Sie verweisen als Beleg unter anderem auf die Beobachtung, dass Muskelkater „(fast) ausschließlich“ (a. a. O., S. 17) bei mechanischer Belastung auftritt, die zu einer passiven Dehnung der Muskulatur führt (etwa beim Fallenlassen einer Hantel mit dem Unterarm).

Die psychische Komponente eines Beinbruchs äußert sich beim betroffenen Menschen sehr wahrscheinlich akut in „einer Art von ‚Betäubung‘, einer gewissen Bewusstseinsengung und eingeschränkter Aufmerksamkeit, einer Unfähigkeit, Reize zu verarbeiten und Desorientiertheit“ (Dilling, Mombour & Schmidt, 2011, S. 206). Zusätzlich können Symptome panischer Angst auftreten sowie Unruhe, Überaktivität oder Rückzug bzw. Flucht folgen. Dieses Syndrom wird als akute Belastungsreaktion bezeichnet und in Abschnitt F43.0 der International Classification of Diseases (ICD) 10 beschrieben (Dilling et al., 2011, S. 205-207).

Pauwels hat sich nicht auf die Beschreibung eines Modells beschränkt, bei dem der Stützapparat aus mechanisch verformbaren Gliedern zusammengesetzt ist. In seinem zehnten Beitrag ist er auf die Plastizität bzw. Anpassungsfähigkeit des Stützapparates eingegangen. Er führt an zahlreichen Beispielen auf, dass Zug-, Druck- und Schubspannungen (Beanspruchungsarten) die Neubildung von Knochen- und Knorpelgewebe anregen können (Pauwels, 1960). Innerhalb eines Knochens bilden sich feine Strukturen, die sogenannten Spongiosabögen, aus. Diese zeichnen Zug- und Drucklinien nach, die typischen Belastungen eines Knochens entsprechen. Die Plastizität des Stützapparates zeigt sich u. a. nach operativen

Veränderungen der Knochen-Geometrie. Diese führen zu neuen typischen Belastungen, die sich nach und nach in einer angepassten Struktur der Spongiosa widerspiegeln (z. B. a. a. O., S. 484, Abb. 5). Pauwels führt die Plastizität des Stützapparates auf zwei mechanische Reize (Beanspruchungsarten) zurück, die auf die Zellen entsprechender Gewebe (Knochen, Knorpel, Bänder) wirken: Dehnung in eine bestimmte Richtung und hydrostatischer Druck (a. a. O., S. 494). Im weiteren Verlauf seines Beitrages stellt er u. a. dar, dass die vermuteten physiologischen Prozesse die Knochenheilung nach einem Bruch erklären können.

Folgt man den Betrachtungen Pauwels, so wohnt der mechanischen Belastung und Beanspruchung des Stützapparates ein Doppelcharakter inne. Eine mechanische Überbeanspruchung (als beliebige Kombination von Druck, Zug, Schub, Biegung oder Scherung) kann zu einer Verletzung bis hin zu einem Bruch eines Knochens führen und stellt somit eine Beeinträchtigung der Gesundheit dar. Andererseits führt eine maßvolle neue Belastungskonstellation zur Anpassung eines Knochens an die neuen ökologischen Bedingungen. Die mechanische Beanspruchung der Gewebezellen im Stützapparat trägt zu dieser Anpassung ebenso bei, wie zur Heilung eines Knochenbruchs. Heinonen, Sievänen, Kannus, Oja und Vuori (2002) stellen beispielsweise bei einer Gegenüberstellung von Gewichtheberinnen und einer Vergleichsgruppe fest, dass Gewichtheben mit einer etwa 10 % dichteren Spongiosa in Speiche (*radius*) und Oberschenkelknochen (*femur*) einhergeht. Außerdem fanden sie dickere Knochenwände (*substantia compacta*) im Querschnitt der Speiche (38 % am körperfernen Ende [*distal*], 26 % am Knochenschaft) und des Schienbeins (*tibia*, 9 % am Knochenschaft).

Der Stütz- und Bewegungsapparat stellt das mechanische Gedächtnis eines Organismus dar. In diesem Gedächtnis werden Reaktionspotenziale in Bezug auf typische mechanische Belastungskonstellationen gespeichert. Diese können in Bezug auf die akute mechanische Belastungskonstellation angepasst und abgerufen werden. Im Falle der Fehlbeanspruchung reicht das Reaktionspotenzial nicht aus und der Stütz- und Bewegungsapparat nimmt Schaden.

2.1.2 Inanspruchnahme und Anpassung von Überlebens-Schaltkreisen

Die materielle Grundlage für das Erleben und Verhalten des Menschen bildet in erster Linie das Zentralnervensystem und das Hormonsystem. Für Joseph LeDoux (2012) ist die Aktivität neuroendokriner Überlebens-Schaltkreise und ihrer Funktionen (*survival circuits and functions*) wesentlich für Emotion, Motivation, Verstärkung und Aktiviertheit (*arousal*)

während der Interaktion zwischen Individuum⁴ und ökologischer Umwelt. “Focusing on survival functions and circuits allows phenomena related to emotion, motivation, reinforcement, and arousal to be treated as components of a unified process that unfolds when an organism faces a challenge or opportunity” (a. a. O., S. 654). Ihm ist es, wie Lisa Feldman Barrett (2009), ein Anliegen, Neurologie und Psychologie einander anzunähern. Feldman Barrett (2009) formuliert die Herausforderung an die Psychologie dabei wie folgt: „Psychology, in walking a tightrope between the social world and the natural world, tries to map observer-dependent categories [z. B. psychologische Konzepte] to observer-independent categories [z. B. neurologische Prozesse]“ (a. a. O., S. 328). Psychologische Konzepte sind ihrer Ansicht nach mental und sozial konstruiert und akzeptiert. Sie haben zwar eine neurologische Ursache, jedoch nicht in einem reduktionistischen Eins-zu-Eins-Verhältnis. Sie sind nicht einfach „neurons firing“ (a. a. O., S. 330). Feldman Barrett (2009) schlägt vor, psychologische Konzepte und neurologische Prozesse über „psychological primitives“ (a. a. O., S. 331; vgl. Lindquist, Wager, Kober, Bliss-Moreau, & Feldman Barrett, 2012) miteinander in Beziehung zu setzen. Hiermit meint sie psychologisch irreduzible Prozesse und nennt Bewertung, Kategorisierung und Aufmerksamkeitssteuerung. Diese haben eine neurologische Grundlage und führen in ihrem Zusammenspiel zu einem „mental state“ (Feldman Barrett, 2009, S. 331, Fig. 1), der schließlich in psychologischen Begriffen (z. B. Emotion, Kognition) beschrieben und erklärt werden kann.

LeDoux (2012) unterstützt diese Sichtweise weitgehend, hebt aber die biologische Bedeutung identifizierbarer Überlebens-Schaltkreise hervor. Er unterscheidet Überlebens-Schaltkreise für Verteidigung, Aufnahme von Energie und Nährstoffen, Flüssigkeitsausgleich, Temperaturregulation und Fortpflanzung. Jeder dieser Schaltkreise ist darauf ausgelegt, funktionsspezifische Stimuli zu entdecken und entsprechendes Antwortverhalten zu steuern. Überlebens-Schaltkreise aktivierende Stimuli können in Auslösereize, Anreize und Verstärker unterteilt werden (vgl. a. a. O., S. 661, Table 1). Auslösereize (unkonditioniert oder konditioniert) lösen mit hoher Wahrscheinlichkeit angeborenes Antwortverhalten aus. Anreize steuern instrumentelles, zielgerichtetes Verhalten. Solches Verhalten hilft, Herausforderungen oder Möglichkeiten zu meistern. Ein unkonditionierter appetitiver (ugs. “angenehmer”) Anreiz löst Annäherung, ein unkonditionierter aversiver (ugs. “unangenehmer”) Anreiz löst Vermeidung aus. Ein konditionierter Anreiz löst ein Verhalten aus, durch das eine Situation

⁴ LeDoux (2012) beschränkt sich bei seinen Betrachtungen nicht auf Menschen. Vielmehr ist es ihm ein Anliegen, die funktionalen Gemeinsamkeiten zwischen Tieren im Allgemeinen und Säugetieren im Speziellen hervorzuheben.

entsteht, in der eine gegebene Herausforderung oder Möglichkeit gemeistert werden kann. Verstärker unterstützen das Lernen von Assoziationen. Primäre Verstärker unterstützen Assoziationen zwischen Auslösereizen und neutralen Reizen (Reizkonditionierung) sowie zwischen Stimuli und Verhaltensweisen, die zur Präsenz appetitiver oder Entfernung aversiver Auslösereize führen (instrumentelle Konditionierung). Sekundäre Verstärker unterstützen die weitere assoziative Verkettung von Stimuli oder instrumentellen Verhaltensweisen.

Überlebens-Schaltkreise werden bei Herausforderungen und Möglichkeiten aktiviert. Für LeDoux (2012) sind dies “emotional or motivated situations” (a. a. O., S. 662). Daher nennt er Stimuli, die Überlebens-Schaltkreise aktivieren, “‘Emotional’ Stimuli” (a. a. O., S. 661). Über die Einteilung der Stimuli verdeutlicht er, dass Emotion, Motivation (Anreize) und Lernen (Verstärkung) innerhalb von Überlebens-Schaltkreisen untrennbar miteinander verknüpft sind. Darüber hinaus erläutert er, dass die verschiedenen Überlebens-Schaltkreise über (zeitversetzte) gegenseitige Aktivierung oder Hemmung in komplexer Weise miteinander wechselwirken. Beispielsweise hemmt der Verteidigungs-Schaltkreis (Sympathikus und Hypothalamus-Hypophysen-Nebennierenrinden-System, s. u.) bei akuter Präsenz einer Gefahr zunächst sämtliche anderen Überlebens-Schaltkreise. Energieverbrauch und Temperaturerhöhung beim Verteidigungsverhalten führen jedoch zu einer zunehmenden Aktivierung der entsprechenden Überlebens-Schaltkreise. Bald konkurrieren Verteidigungsverhalten, Nahrungsaufnahme und Temperaturregulation um benötigte Ressourcen. Generelle Aktiviertheit (*generalized arousal*) übernimmt in der Koordination der Überlebens-Schaltkreise eine modulierende Funktion. LeDoux (2012) beschreibt am Beispiel des Verteidigungs-Schaltkreises, dass die Aktivierung eines Überlebens-Schaltkreises zunächst die Freisetzung und weiträumige Verteilung bestimmter Neurotransmitter und Hormone (Neuromodulatoren) veranlasst. Diese bewirken eine verstärkte Aktivierung des auslösenden Überlebens-Schaltkreises sowie der sensorischen Stimulus-Verarbeitung, der Aufmerksamkeit und der Gedächtnis-Funktionen. Dies wiederum verstärkt die Freisetzung von Neuro-modulatoren. Erst das erfolgreiche instrumentelle Verhalten beendet diese positive Rückkopplung. Die Gesamtheit aller neuroendokrinen Prozesse, die bei der Aktivierung eines Überlebens-Schaltkreises zur Bewältigung einer emotionalen und motivierenden Herausforderung oder Möglichkeit ablaufen, nennt LeDoux (2012) einen globalen organismischen Zustand (“*global organismic state*”; a. a. O., S. 663). Die bewusste Repräsentation und Benennung eines solchen Zustandes setzt er mit einem menschlichen, durch die situationsbezogene Aktivierung eines Überlebens-Schaltkreises ausgelösten Gefühl gleich.

Das Hauptanliegen seines Beitrages besteht in der Stimulierung der Emotionsforschung. Hierauf wird später zurückzukommen sein.

2.2 AROUSAL: DIE DYNAMIK VON VALENZ UND AKTIVIERUNG

LeDoux (2012) hebt in seinem Beitrag die modulierende Bedeutung des Arousal für die Koordination konkurrierender Überlebens-Schaltkreise hervor. Als einen bedeutenden Überlebens-Schaltkreis führt er den Verteidigungs-Schaltkreis an. Dieser mobilisiert im Individuum Prozesse für die Bewältigung aversiver Herausforderungen. Situationen, die den Verteidigungs-Schaltkreis aktivieren, werden häufig unter dem Begriff *Stressor* eingeordnet. Die im Inneren des Individuums ablaufenden neuroendokrinen Prozesse in Gegenwart eines Stressors heißen *Stressreaktionsprozesse*. Die dauerhafte Präsenz eines unbewältigten Stressors oder die dauerhafte Aktivierung von Stressreaktionsprozessen sowie das Aufgeben von Bewältigungsversuchen sind wesentliche Mitursachen für die Entstehung körperlicher und psychischer Beeinträchtigungen und Störungen. Das *Bewältigen* von Stressoren oder das *Aufgeben* von Bewältigungsversuchen sind jedoch nur zwei verhaltensbezogene Komponenten der Arousal-Dynamik. Weitere Komponenten bestehen im Wachen, Konsumieren, mentalen Simulieren und im Schlafen. Im Folgenden wird dargestellt, dass diese Verhaltenskomponenten stark aufeinander bezogen und eng mit emotionsbezogenem Erleben⁵ verknüpft sind.

2.2.1 Inanspruchnahme und Anpassung des Zentralnervensystems unter Stress

Gerald Hüther stellt gemeinsam mit Kollegen anhand eines neurobiologischen Modells dar, wie sich die Anpassungsfähigkeit des Zentralnervensystems durch Inanspruchnahme unter Stress realisiert (Hüther, 2014; Huether, Doering, Rüger, Rüter & Schüssler, 1999; Hüther & Sachsse, 2007).⁶ Die Kernaussagen dieses Modells bestehen in Erklärungen stressbedingten Lernens und Vergessens sowie stressbedingter Entwicklung psychischer Beeinträchtigungen und Störungen aufgrund lebenslanger struktureller Veränderungen neuronaler Netzwerke der assoziativen Hirnrinde und damit verbundener, etablierter Funktionen. Eine neuroendokrine Stressreaktion erfolgt Huether et al. (1999) zufolge dann, wenn „demands either cannot or cannot adequately be responded to by the individual’s hitherto acquired behavioral, cognitive

⁵ Die Abgrenzung der Begriffe Affekt, Emotion, Gefühl und Stimmung ist umstritten, auch wenn es tendenzielle Übereinstimmungen gibt (vgl. Feldman Barrett, 2006, 2012; LeDoux, 2012; Russell, 2003, 2009; Russell & Feldman Barrett, 1999; Scherer, 2005). In der vorliegenden Arbeit wird der Begriff *emotionsbezogenes Erleben* verwendet, wenn das Verhalten oder Erleben mit einem dieser vier Begriffe benannt werden kann. Spezifische Begriffe werden je nach zitiertem Konzept verwendet.

⁶ Bei „Hüther“ und „Huether“ handelt es sich um denselben Autor Gerald Hüther. Bei der Zitation wurde die jeweils in der Veröffentlichung gewählte Schreibweise wiedergegeben.

or emotional strategies“ (S. 84). Bei ihrer Betrachtung berücksichtigen die Autoren die prozesshafte Interaktion zwischen stressauslösenden Stimuli (Stressoren) und psychophysiologischer Stressreaktion und definieren in diesem Sinne:

“The stress-reaction process (SRP) is triggered by a physiological and/or psychosocial, imagined or anticipated strain and it is characterized by a continuous interaction and feedback between a multi-phase cognitive and emotional appraisal of the strain and its meaning for the individual, and a sequence of physiological, cognitive, emotional and behavioral reactions that allow coping with the situation”
(Huether et al., 1999, S. 85).

In Anlehnung an Selye (1974) und Ursin und Olf (1993) beschreiben sie ein Kontinuum von kontrollierbaren bis hin zu unkontrollierbaren SRP. Einen kontrollierbaren SRP kann das betroffene Individuum durch eigene Anstrengung beenden. Kontrollierbare SRP gehen mit der neuronalen Stabilisierung erfolgreich angewandter Bewertungs- und Bewältigungsstrategien und einer zunehmenden Anpassung an die auslösenden Anforderungen (psychische Belastung) einher. Bei einem unkontrollierbaren SRP fehlen angemessene Bewertungs- und Bewältigungsstrategien. In unkontrollierbaren SRP werden durch neuroendokrine Prozesse unangemessene Denk- und Verhaltensweisen destabilisiert und gelöscht. Auf diese Weise wird Huether et al. (1999; vgl. Hüther, 2014) zufolge die Voraussetzung für eine umfassende Reorganisation zu Grunde liegender neuronaler Netzwerke vorbereitet. Foy, Kim, Shors und Thompson (2006) fassen dies wie folgt zusammen:⁷

„The key aspects are uncertainty and control: the less knowledge the individual has about a potentially harmful situation and the less control he or she feels can be exerted, the more stressful the situation is. ... We and other mammals appear to be driven by nature toward certainty” (Foy et al., 2006, S. 44).

Huether et al. (1999) beschreiben zwei neuroendokrine Achsen, die zeitversetzt am Stressreaktionsprozess beteiligt sind (vgl. DeRijk & de Kloet, 2006; Dickerson & Kemeny, 2004; Foy et al., 2006; Gunnar & Quevedo, 2007; Ströhle, 2003). Das Gewahrwerden eines realen oder vorgestellten Stimulus ist mit einem Aktivitätsmuster des Präfrontalkortex

⁷ Das Kontinuum der Kontrollierbarkeit von Stressreaktionsprozessen ist zu unterscheiden vom Konzept der Kontrollüberzeugung (Rotter, 1966). Ersteres bezieht sich auf die Fähigkeit, eine akute Situation durch das in dieser Situation gezeigte Verhalten erfolgreich bewältigen zu können. Letzteres stellt eine Persönlichkeitseigenschaft dar, die die situationsübergreifende Überzeugung abbildet, das Auftreten von Belohnungen bzw. Verstärkern durch eigene Anstrengung beeinflussen zu können. Ungeachtet dieser Unterscheidung wird auch in der psychologischen Literatur an unzähligen weiteren Stellen die Bedeutung der Kontrollierbarkeit persönlicher Belange für das Wohlbefinden und die psychische Gesundheit hervorgehoben. Neben der Kontrollüberzeugung seien beispielhaft Selbstwirksamkeit (Bandura, 1977) oder Kohärenzerleben (Antonovsky, 1979) genannt.

verbunden. Dieses wird mit Aktivitäten des limbischen Systems und insbesondere der Amygdala rückgekoppelt. Die subjektive Bewertung eines Stimulus als „novel, unexpected, challenging or threatening“ (Huether et al., 1999, S. 87) geht mit einer erhöhten unspezifischen neuronalen Aktivierung (*Arousal* bzw. *phasic arousal* [Posner, Russell & Peterson, 2005; Pribram & McGuinness, 1975; Sanders, 1983], s. u.) einher. In diesem Falle führen Projektionen zu tiefer liegenden Strukturen (locus coeruleus, Hirnstamm, Sympathisch-Adreno-Medulläres [SAM-] System) zur Ausschüttung von Noradrenalin, was die Aktivität der Amygdala sowie des Hypothalamus weiter anregt. Zugleich wird das Aktivitätsmuster des Präfrontalkortex durch Ausschüttung von Dopamin im mesolimbischen System angeregt (*tonic activation* [Gunnar & McGuinness, 1975; Posner et. al., 2005; Sanders, 1983]). Es kommt zu einer sich selbst verstärkenden Aktivierung der beteiligten Hirnareale (vgl. auch Gunnar & Quevedo, 2007). Diese Aktivierungsspirale kann Huether et al. (1999) zufolge im Falle eines kontrollierbaren Stimulus unterbrochen werden. Die unspezifische neuronale Aktivierung wechselt in ein spezifisches neuronales Aktivitätsmuster um, bei dem jene neuronalen Netze, die einer angemessenen Bewertungs- und Bewältigungsstrategie zugrunde liegen, aktiviert sind. Zahlreiche metabolische Prozesse bei den beteiligten Zellverbänden führen zu einer stärkeren Bahnung und Stabilisierung dieser neuronalen Netze. Mit jedem erfolgreichen Gebrauch angemessener Bewertungs- und Bewältigungsstrategien sind die zugrunde liegenden neuronalen Netze effizienter geknüpft.

Bei einem unkontrollierbaren SRP gelingt es nicht, die Aktivierungsspirale durch erfolgreiche Bewertungs- und Bewältigungsstrategien oder das Gewahrwerden anderer Stimuli zu unterbrechen. Die weitere Aktivierung führt zur Anregung sekretorischer Zellen des Hypothalamus und somit zur Aktivierung der Hypothalamus-Hypophysen-Nebennierenrinden-Achse (Hypothalamic Pituitary Adrenal [HPA-] System). Die Konzentration von Glukokortikoiden (beim Menschen Kortisol) im Blut nimmt zu und entsprechende Rezeptoren der Neuronen und Gliazellen sind vermehrt besetzt (vgl. auch DeRijk & de Kloet, 2006; Dickerson & Kemeny, 2004; Foy et al., 2006; Gunnar & Quevedo, 2007). Eine wichtige Funktion dieser zweiten Achse des Stressreaktionsprozesses wird in der neuroendokrinen Hemmung der beschriebenen Aktivierungsspirale gesehen. Diese ist Huether et al. (1999; vgl. Hüther, 2014) zufolge jedoch nicht umgehend wirksam, sondern realisiert sich zeitverzögert und längerfristig wirksam durch die Beeinflussung von Genexpressionen. Die damit verbundenen, vielfältigen strukturellen Veränderungen an Zellverbänden führen zu einer Destabilisierung oder gar Auflösung etablierter neuronaler Netzwerke. Auf diesen Netzwerken beruhende Bewertungs- und Verhaltensmuster werden verändert, destabilisiert oder gar

gelöscht (vgl. DeRijk & de Kloet, 2006; Foy et al., 2006; Gunnar & Quevedo, 2007; Sapolsky, 2003). Unter anhaltend unkontrollierbaren Bedingungen können sogar unbeteiligte Bewertungs- und Verhaltensmuster betroffen sein: „Thus, an organism exposed to an uncontrollable and unpredictable set of aversive stimuli shows not only a dramatic increase of adrenocortical activity while exposed to these conditions but also long-term deficits in other, unrelated test conditions“ (Foy et al., 2006, S. 45).

Da neuronale Anpassungsfähigkeit (neuronale Plastizität) und das damit verbundene Lernen und Vergessen während eines Stressreaktionsprozesses für das spätere Verständnis von Modellen zur Handlungs- und Selbstregulation (Hockey, 1997; Sanders, 1983) bedeutsam ist, soll es an dieser Stelle vertieft erläutert werden. Robert M. Sapolsky (2003) fasst diesbezüglich die Rolle von Glukokortikoiden (z. B. Kortisol) für die Aktivität und Anpassungsfähigkeit der hippocampalen Neuronen zusammen: Ein mäßig hoher Kortisol-Spiegel (Stimulation, z. B. durch eine anregende Umwelt oder durch Übung) ist mit einer erhöhten hippocampalen kognitiven und Gedächtnis-Leistung (Langzeitpotenzierung, Aufbau dendritischer Strukturen) sowie der Neubildung von Hippocampus-Neuronen verbunden. Dahingegen ist ein stark oder dauerhaft erhöhter Kortisol-Spiegel (unkontrollierbarer Stressreaktionsprozess) mit einer reduzierten hippocampalen kognitiven und Gedächtnis-Leistung (erniedrigte Anregbarkeit, Abbau dendritischer Strukturen) verbunden. Ein dauerhaft oder sehr stark erhöhter Kortisol-Spiegel kann zum Tode von Hippocampus-Neuronen führen. Hippocampale kognitive und Gedächtnis-Leistung folgen somit in Abhängigkeit des Glukokortikoid-Spiegels einem „inverse-U“ pattern“ (a. a. O., S. 1738; vgl. DeRijk & de Kloet, 2006). Foy et al. (2006) stellen darüber hinaus Geschlechtsunterschiede fest: Weibliche Individuen scheinen auf Grund hoher Östrogen-Konzentrationen weniger betroffen zu sein. Zum anderen stellen Foy et al. (2006) Stress, Kortisol-Konzentration und Altern in einen positiven Zusammenhang. In der Tat konnte festgestellt werden, dass das Fortschreiten der Demenz vom Alzheimer-Typ (F00; Dilling et al., 2011) positiv mit der Plasma-Konzentration von Kortisol in Zusammenhang steht (Csernansky et al., 2006; vgl. Teather, 2006).

Aus der Perspektive psychischer Belastung und Beanspruchung können stressauslösende Stimuli (Stressoren) als psychische Belastungsfaktoren und damit zusammenhängende Stressreaktionsprozesse als psychische Beanspruchung aufgefasst werden. Huether et al. (1999; vgl. Hüther, 2014) argumentieren schlüssig, dass sowohl kontrollierbare als auch unkontrollierbare SRP einen Doppelcharakter aufweisen: Beide können sowohl mit dem Aufbau angemessener Bewältigungsmöglichkeiten als auch mit Beeinträchtigungen der Gesundheit im Zusammenhang stehen. Sowohl kontrollierbare als auch unkontrollierbare SRP

sollten „of the right quality, of the right intensity, in the right context and at the right age“ (a. a. O., S. 91) auftreten, damit die ausgelösten strukturellen Veränderungen des Zentralnervensystems als angemessene Anpassungen an die ökologische Umwelt eines Individuums bewertet werden können. Ist dies nicht der Fall, muss mit verschiedenen körperlichen und psychischen Beeinträchtigungen und Störungen gerechnet werden (s. u.). Analog stellen Gunnar und Queveda (2007) fest: „It would not be an overstatement to say that the nervous system of mammals carries their singular epigenetic history and expresses it in a unique but lawful (i. e. predictable) ways“ (a. a. O., S. 162-163).

2.2.2 Die Arousal-Uhr

Arousal ist offenbar für die Koordination und Modulation neuronaler Aktivität wesentlich. Eine Reihe arbeitspsychologisch relevanter Modelle greift daher auf das Arousal-Konzept zurück (Hacker & Sachse, 2014; Hockey, 1997; Sanders, 1983; Thayer, 1967, 1997, 1996; Pribram & McGuinness, 1975). Eine Auswahl soll im Folgenden dargestellt werden. Zunächst werden jedoch die neurobiologischen Arousal-Systeme in Anlehnung an Barbara E. Jones (2003) und mit Ergänzungen aus Brown, Basheer, McKenna, Strecker und McCarley (2012) rekapituliert (siehe auch Tabelle 1).

Mit *generalized arousal* oder *genereller Aktiviertheit* ist zunächst eine hochfrequente Aktivität in der Hirnrinde (Gamma- und Theta-Wellen, elektroenzephalographisch messbar) und eine erhöhte Muskelspannung (elektromyographisch messbar) gemeint (Brown et al., 2012; Duffy, 1962; Hebb, 1955; Jones, 2003). Beides resultiert aus der neuronalen Aktivität bestimmter Kerne des Hirnstammes. In aufsteigenden neuronalen Bahnen freigesetztes Glutamat erregt Projektionen zwischen Thalamus und Hirnrinde sowie, am Thalamus vorbei, zwischen posteriorem Hypothalamus, basalem Vorderhirn und Hirnrinde. Von diesen Arousal-Systemen hinabführende neuronale Bahnen (Hirnstamm und Rückenmark) steigern oder modulieren die sensumotorische Empfindlichkeit und Aktivität (vgl. arousal bzw. phasic arousal bei Posner et al., 2005; Pribram & McGuinness, 1975; Sanders 1983). Zudem aktivieren sie das SAM- und das HPA-System (s. o.). Die Freisetzung von Acetylcholin aus Projektionen des Tegmentums und basalen Vorderhirns ist dagegen mit erhöhter Hirnaktivität (Bewusstsein, Aufmerksamkeit, Motivation), jedoch erniedrigter Muskelspannung bei belohnenden Aktivitäten im Wachzustand und im Rapid-Eye-Movement- (REM-) Schlaf verbunden. Sie unterstützen die Synchronisation kortikaler Aktivität. Die cholinergen Neuronen des basalen Vorderhirns werden darüber hinaus mit neuronaler Plastizität, Lernen und Gedächtnis sowie mit Schlaf nach längerer Wachheit (homöostatische Schlafantwort; Brown et al., 2012) in Verbindung gebracht. Cholinerge Neuronen des Hirnstammes kontrollieren u. a. die

Muskelspannung im Schlafzustand. Mit der Freisetzung von Serotonin aus den Raphe-Kernen des Mittelhirns geht ein entspannter, ruhiger und zufriedener Wachzustand einher. Serotonerge Projektionen verhindern aber auch Schlaf und sind bei Stressreaktionsprozessen aktiv. Letzteres steht vor allem mit der mentalen Simulation selbst- und personenbezogener, vergangener, gegenwärtiger oder zukünftiger Probleme zusammen (s. u.; Andrews-Hanna, 2012; Buckner, Andrews-Hanna & Schacter, 2008). Aus Projektionen in das mesolimbische System freigesetztes Dopamin induziert Belohnung und einen Zustand hoher Motivierung sowohl während des Wachzustandes als auch während des REM-Schlafs (vgl. tonic activation bei Pribram & McGuinness, 1975; Sanders 1983). Sowohl die Freisetzung von Noradrenalin (locus coeruleus) als auch von Histamin (posteriorer Hypothalamus) ist mit einem hochaktiven Wachzustand, z. B. während Stressreaktionsprozessen oder in unbekanntem Situationen, verbunden und verhindert REM- und Slow-Wave- (SW-) Schlaf. Orexine sind Neurotransmitter, die diffus verteilt im Gehirn wirken. Ihre Ausschüttung ist mit Erwachen, hoher Wachheit und Hunger verbunden. Orexine sind in der Lage, sowohl das SAM- als auch das HPA-System anzuregen. Während des SW-Schlafes werden die Arousal-Systeme von parallel verteilten Neuronen durch Ausschüttung von γ -Aminobuttersäure (γ -amino butyric acid; GABA) gehemmt. GABA wird außerdem mit der Synchronisation neuronaler Aktivität in Verbindung gebracht.

Die beschriebenen Arousal-Systeme sind nicht unabhängig voneinander aktiv, sondern sind auf komplexe Weise sowohl hemmend als auch erregend miteinander verschaltet. Mit Abbildung 1 soll ein Eindruck über diese Verschaltungen vermittelt werden, soweit sie sich aus den Schilderungen von Brown et al. (2012) sowie Jones (2003) ableiten lassen.

Psychische Belastung und Beanspruchung: Neurobiopsychologische Grundlagen

Tabelle 1. Substanzen der Arousal-Systeme und assoziierte Situationen bzw. Aktivitäten, Schlaf-Wach-Zustände und sensumotorische (SM) bzw. kortikale (CTX) Aktivität. Die Pfeile deuten an, ob die SM- oder CTX-Aktivität angeregt oder gehemmt wird, u. = unbestimmt. Unter der Überschrift „Einteilung“ wurden Zuordnungen zu Verhaltens-Kategorien vorgenommen (siehe weiter unten im Text). (Eigene Darstellung, vereinfacht nach den Angaben in Jones [2003] und Brown et al. [2012] und ergänzt um Kortisol [s. Löffler, 2008].)

Substanz	Assoziierte Situation bzw. Aktivität	Zustand	Einteilung	Aktivität	
				SM	CTX
Glutamat	Verarbeitung und Weiterleitung sensumotorischer und zentralnervöser Eingaben, Bewegung, Aufrechterhaltung des Wachzustandes, bewusste Kognition	Wachheit, REM-Schlaf	Wachen	↑	↑
Noradrenalin	Hoch aufregende Situation, Flucht- oder Kampf-Reaktion, emotionales Gedächtnis, Stressreaktion	Wachheit		↑	↑
Histamin	Aufrechterhaltung des Wachzustandes und Bewusstseins, neuartige/gefährliche Umgebungen, Erwartung von Nahrungsaufnahme	Wachheit	Bewältigen	u.	↑
Orexin	Aufrechterhaltung des Wachzustandes, Aufregung, Nahrungsaufnahme, Hunger, Metabolismus, Stressreaktion	Wachheit		u.	
Kortisol	Stressreaktion, Nahrungs-Entzug, Proteinbiosynthese-Hemmung, Energiebereitstellung	Tagesverlauf individuell	Aufgeben	↕	↕
Dopamin	Aufregung, Fortbewegung und Erkundung in Kombination mit positiver Verstärkung, Aufnahme von Nahrung und psychoaktiven Substanzen	Wachheit, REM-Schlaf	Konsu- mieren	↑	↑
Acetylcholin	Ruhe, Nahrungsaufnahme, positive Verstärkung, Lernen und Gedächtnis, Bewusstsein, Schlaf nach längerer Wachheit, Schlafkontrolle	Wachheit, REM-Schlaf		↓	↑
Serotonin	Ruhe, Zufriedenheit, Sattsein, Entspannung, rhythmische Bewegungsmuster, Stressreaktion, insbesondere in Zusammenhang mit mentaler Simulation von Problemen	ruhige Wachheit	mentales Simulieren	↓	↓
γ-Aminobuttersäure	Spezifische (Koordination) und allgemeine Hemmung sensumotorischer und kortikaler Aktivität (Schlaf)	SW-Schlaf, (Wachheit)	Schlafen	↓	↓

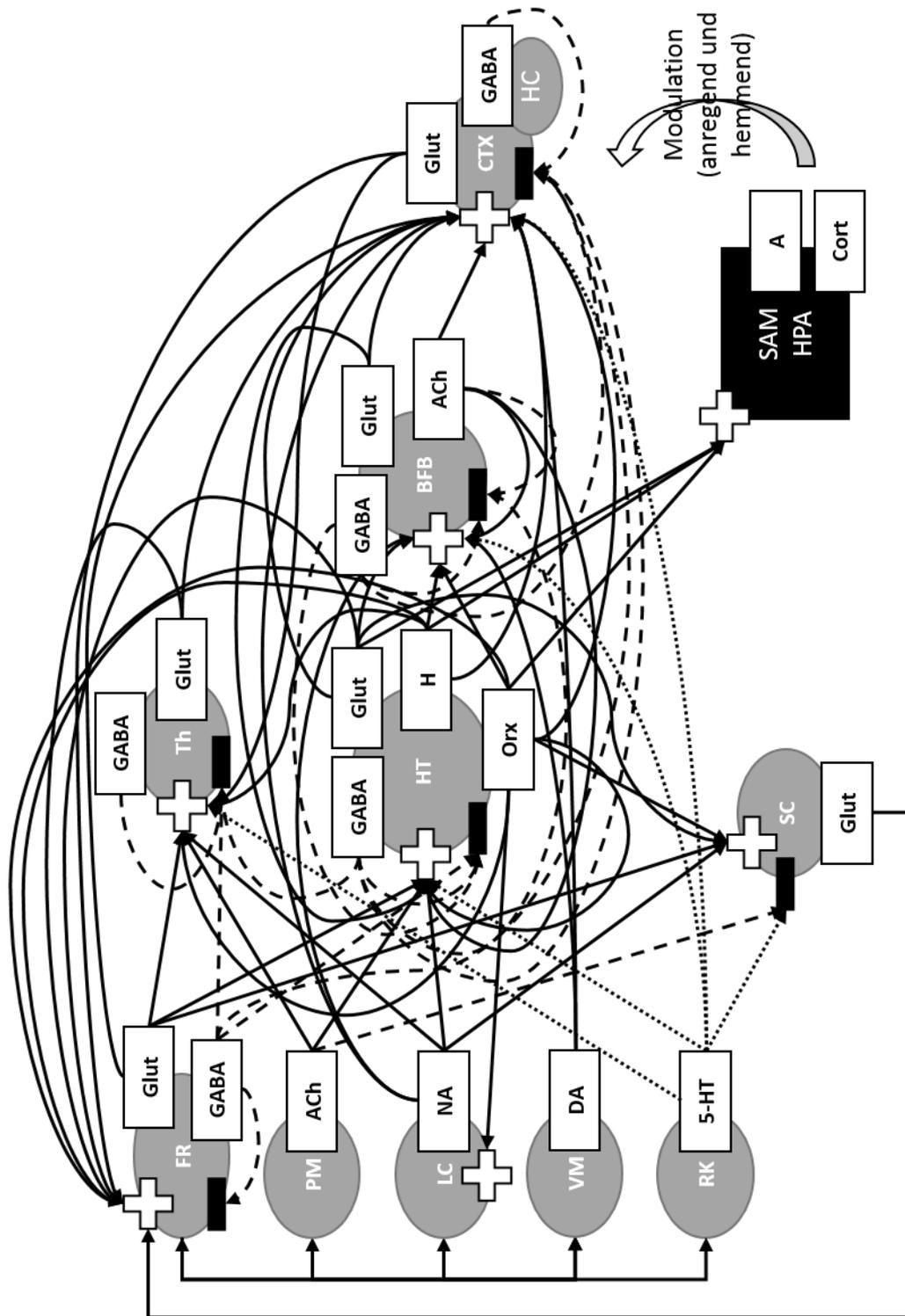


Abbildung 1. Arousal-Systeme (eigene Darstellung, vereinfacht nach den Angaben in Jones [2003] und Brown et al. [2012]). Graue Ovale stellen anatomische Strukturen des Gehirns dar: FR = formatio reticularis, PM = ponto-mesencephalon, LC = locus coeruleus, VM = ventral mesencephalon, RK = Raphe Kerne, Th = thalamus, HT = hypothalamus, BFB = basal forebrain, SC = spinal cord, CTX = cortex, HC = hippocampus. Weiße Kästen stellen die freigesetzten Neurotransmitter bzw. Neuromodulatoren dar: Glut = Glutamat, GABA = Gamma-Aminobuttersäure, ACh = Acetylcholin, NA = Noradrenalin, DA = Dopamin, 5-HT = 5-Hydroxytryptamin (Serotonin), H = Histamin, Orx = Orexin, A = Adrenalin, Cort = Kortisol. Durchgezogene Linien enden an einem weißen Plus-Symbol und stellen anregende Verbindungen dar. Gestrichelte Linien enden an einem schwarzen Minus-Symbol und stellen hemmende Verbindungen dar. Im schwarzen Kasten sind die beiden Stress-Achsen, das SAM- und das HPA-System abgebildet. Serotonin-Projektionen sind gepunktet dargestellt, da sie aktivierend und (z. T. indirekt) hemmend wirken.

Auch unter Rückgriff auf die zusammenfassende Darstellung in Tabelle 1 finden sich Systeme, die die generelle Aktiviertheit erhöhen (Glutamat, Acetylcholin, Dopamin, Noradrenalin, Histamin, Orexin) und solche, die sie hemmen (Serotonin, GABA). Darunter sind einige sowohl aktivierende als auch hemmende Systeme eher mit dem Konsum oder Aufsuchen appetitiver Reize und Zustände verbunden (Acetylcholin, Dopamin, Serotonin) und andere ausschließlich aktivierende Systeme eher mit der Bewältigung oder Vermeidung aversiver bzw. neuartiger Reize und Zustände (Noradrenalin, Histamin, Orexin). Glutamat ist unabhängig von der Qualität der Reize und Zustände mit der generellen Aktivierung verbunden. GABA ist ein Gegenspieler von Glutamat und ist sowohl spezifisch bei der Koordination der Aktivität als auch allgemein beim Schlaf und insbesondere beim SW-Schlaf mit der Hemmung verbunden. Arousal lässt sich daher vereinfachend in zwei Dimensionen beschreiben: Aktivierung mit den Polen aktiviert/wach - deaktiviert/schlafend und Valenz mit den Polen zufrieden/gut - unzufrieden/schlecht. In der Systematik fehlt jedoch ein spezifisches Arousal-System für geringe Aktivität und Unzufriedenheit. Müdigkeit und Erschöpfung sowie die Symptomatik des Burnout-Syndroms und der depressiven Episode (F32; Dilling et al., 2011) entsprechen diesem Muster. Chida und Steptoe (2009) zeigen in ihrer Meta-Analyse, dass diese Symptomatiken mit einem gestörten Kortisol-Haushalt (im Folgenden *Kortisol-Krise*) in Zusammenhang stehen. Daher soll die Kortisol-Krise die Einteilung der Arousal-Systeme ergänzen.

Die Arousal-Uhr

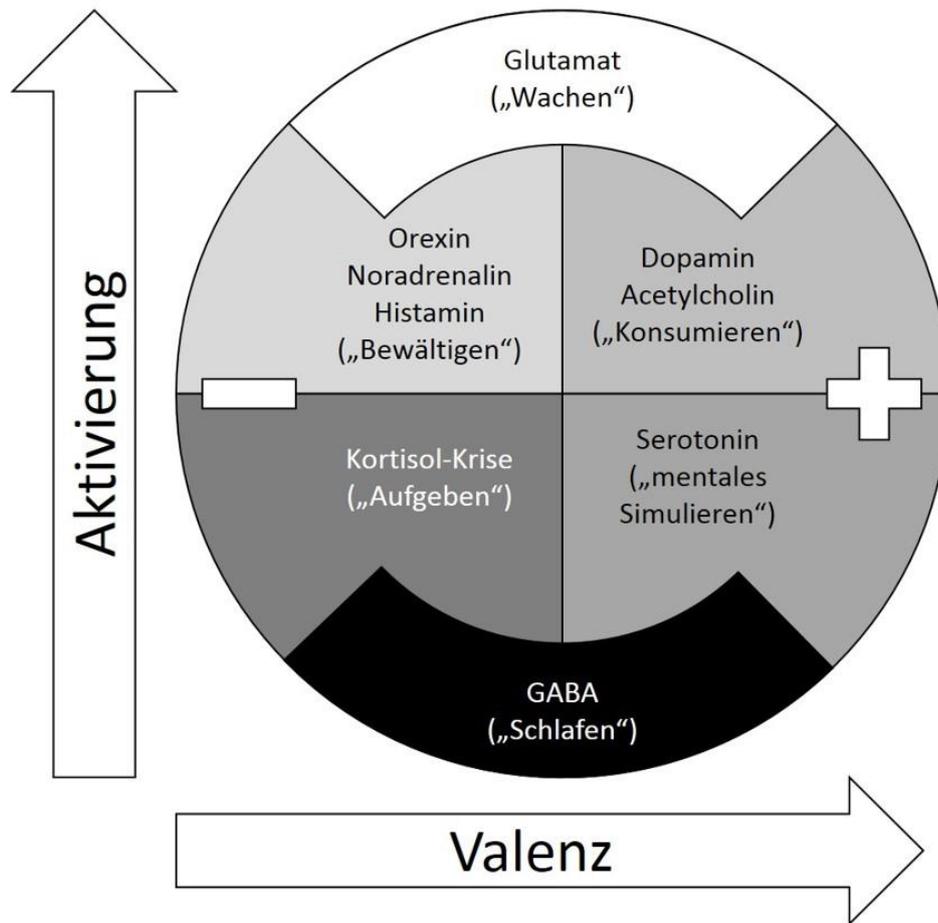


Abbildung 2. Die Arousal-Uhr. Einteilung der verschiedenen Arousal-Systeme in Abhängigkeit von Aktivität und Valenz. Die Begriffe in Klammern fassen die Hauptfunktionen der jeweiligen Arousal-Systeme umgangssprachlich zusammen.

In Abbildung 2 ist die vorgenommene Einteilung der Arousal-Systeme in Abhängigkeit von Aktivierung und Valenz als „Arousal-Uhr“ dargestellt. Diese Metapher wurde mit der Vorstellung gewählt, welche Aktivitäten wohl ein nahezu beliebiges gesundes Säugetier an einem Tag zu verrichten hat, ohne jedoch Tagesverläufe für Konzentrationsspiegel der Neuromodulatoren implizieren zu wollen.⁸ Es sei den Lesern überlassen, sich diese Vorstellung ebenfalls auszumalen. Es werden grob sechs Arousal-Systeme unterschieden. Diese sind mit ihren Neuromodulatoren sowie einem Beinamen in Klammern aufgeführt. Der Beiname fasst die jeweilige Hauptfunktion eines Arousal-Systems als Verhaltens-Kategorie zusammen. Diese werden in den folgenden Absätzen näher beschrieben.

⁸Beispielsweise weist jedoch der Kortisol-Spiegel eines gesunden Menschen einen charakteristischen Tagesverlauf (zirkadiane Rhythmik) auf: Morgens nach dem Aufwachen ist der Kortisol-Spiegel am höchsten, gegen Mittag gibt es eine relativ starke Ausschüttung von Kortisol und am Abend ist der Kortisol-Spiegel am geringsten. Darüber hinaus findet sich eine etwa stündliche Spitze bei der Ausschüttung von Kortisol. Kortisol wird hauptsächlich über Nacht biosynthetisiert (DeRijk & de Kloet, 2006).

Wachen: Das Glutamat-System steigert die neuronale Aktivität im Allgemeinen, was die sensumotorische Empfindlichkeit und Reaktionsfähigkeit sowie die Aufmerksamkeit einschließt (allgemeine Aktivierung). Seine Hauptfunktion besteht im Wachen über eintreffende Stimuli sowie im Entdecken appetitiver oder aversiver Stimuli. Dies schließt die Überwachung der Überlebens-Schaltkreise ein (LeDoux, 2012; s. o.). Es kann im Gesamtzusammenhang aller Arousal-Systeme gemeinsam mit dem Kortex als allgemeiner Zustandsmonitor aufgefasst werden. Hockey (1997) spricht in diesem Zusammenhang von *supervisory control* und führt einige Belege über den Einfluss auftretender Nebenanforderungen (z. B. dual task, Zeitdruck, Schlafdeprivation, Lärm, Zustands-Angst) auf die Effizienz und eingesetzte Strategie bei der Bewältigung verschiedener kognitiver Primäranforderungen an. Das Glutamat-System stimmt weitgehend überein mit dem System des phasic arousal bei Pribram und McGuinness (1975; vgl. Sanders, 1983).

Bewältigen: Ist ein neuer, unerwarteter herausfordernder oder aversiver Stimulus entdeckt (z. B. ein Beutetier oder ein Fressfeind), wird das Orexin/Noradrenalin/Histamin-System aktiviert und es kommt zu einem Stressreaktionsprozess (SRP). Angemessene Bewertungs- und Bewältigungsmuster werden gesucht. Abhängig davon, ob der SRP kontrollierbar ist oder nicht, kommt es zur Verstärkung oder Auflösung aktiver neuronaler Netzwerke. Im schlimmsten Falle können das HPA-System selbst oder andere neuronale Funktionen (z. B. Gedächtnis) und körperliche Strukturen (z. B. Fettdepots, Knochen, Muskeln, Immun-System) dauerhaft geschädigt werden (Kortisol-Krise) (Burke et al., 2005; Chida & Steptoe, 2009; DeRijk & de Kloet, 2006; Dickerson & Kemeny, 2004; Foy et al., 2006; Gunnar & Quevedo, 2007; Huether et al., 1999; Hüther, 2014; Hüther & Sachsse, 2007; Krishnan & Nestler, 2008; Kupfer et al., 2012; Sapolsky, 2003; Segerstrom & Miller, 2004; Sherwood Brown et al., 2004; Ströhle, 2003; Teather, 2006; Yehuda & Mostofsky, 2006).

Aufgeben: Das Kortisol-System bzw. HPA-System kann in diesem Zusammenhang als Anstrengungsmonitor aufgefasst werden, da seine Funktion in der Hemmung nicht zielführenden Energieverbrauchs durch das Orexin/Histamin/Noradrenalin- bzw. das dadurch innervierte SAM- und HPA-System besteht (DeRijk & de Kloet, 2006; Foy et al., 2006; Huether et al., 1999; Hüther, 2014). Es ist identisch mit dem *effort monitor* nach Hockey (1997). Auch Sanders (1983) nimmt in Anlehnung an Pribram und McGuinness (1975) ein *effort-system* an. Dort wird es u. A. mit dem Hippocampus identifiziert. Wie weiter oben dargelegt, folgt die hippocampale kognitive und Gedächtnis-Leistung in Abhängigkeit vom Kortisol-Spiegel einem „inverse-U“ pattern“ (Sapolsky, 2003, S. 1738; vgl. DeRijk & de Kloet, 2006). Über diese Abhängigkeit stehen effort monitor (Hockey, 1997) und effort-system (Pribram & McGuinness,

1975; Sanders, 1983) miteinander in Verbindung. Unter Berufung auf Befunde zu kognitiven Anforderungen unterscheidet Hockey (1997) automatisierte, aktive, anstrengende sowie passive Bewältigung (a. a. O., S. 81-82). Bei der automatisierten Bewältigung sind SAM- und HPA-System auf normalem Niveau aktiviert. Bei der aktiven Bewältigung ist zwar das SAM-, nicht jedoch das HPA-System verstärkt aktiviert. Bei der anstrengenden Bewältigung sind sowohl das SAM- als auch das HPA-System verstärkt aktiviert. Die passive Bewältigung beschreibt Hockey (1997) als zunehmende Abwendung von den geforderten Zielen bis hin zum vollständigen Aufgeben.

Der Zusammenhang zwischen Arousal (als Reaktion auf psychische Belastung, z. B. Schwierigkeit kognitiver Anforderungen) und Leistung hängt davon ab, wie Arousal aufgefasst wird. Wird Arousal als Stressreaktionsprozess (Aktivität von SAM- und HPA-System) bis hin zum vollständigen Aufgeben der Ziele aufgefasst (Hockey, 1997), so ist der Zusammenhang monoton fallend (vgl. Sanders, 1983). Wird Arousal hingegen als allgemeine kortikale und sensumotorische Aktivierung („general drive state“, Hebb, 1955, S. 240) aufgefasst, so ist der Zusammenhang konkav (umgekehrt U-förmig; sogenanntes „Yerkes-Dodson-Gesetz“, Duffy, 1962; Hebb, 1955; Malmö, 1959; Teigen, 1994; Yerkes & Dodson, 1908).

Konsumieren: Mit dem Dopamin/Acetylcholin-System sind vor allem appetitive Anreize und Verstärker und zugehöriges instrumentelles Verhalten verbunden. Posner et al. (2005) bringen das dopaminerge mesolimbische System mit dem Aufsuchen oder Konsumieren appetitiver Stimuli in Verbindung. Es kann zwischen einer Explorations- und Annäherungsphase sowie der eigentlichen Konsumtion unterschieden werden (Jones, 2003; LeDoux, 2012). Bei der Exploration und Annäherung wird die Herausforderung bzw. Möglichkeit durch instrumentelles Verhalten bewältigt, wobei der Anreiz handlungsleitend ist (z. B. Zubereitung der Nahrung). Das Dopamin-System verstärkt bereits dieses anreizgesteuerte Verhalten (Jones, 2003). Die eigentliche Konsumtion (z. B. Nahrung in den Mund führen) ist ein unkonditioniertes instrumentelles Verhalten (LeDoux, 2012). Während dieser Phase findet weitere Verstärkung statt. Das dabei aktivierte Acetylcholin-System ist darüber hinaus mit Lernprozessen und einer hohen Aktivität der Hirnrinde bei geringer Muskelspannung verbunden (Jones, 2003). Es kann vermutet werden, dass hier instrumentelles Verhalten und Stimuli bezüglich der Konsumtion assoziiert werden (z. B. Verarbeiten der Ereignisse des Tages) (LeDoux, 2012). Das Dopamin-System weist große Überschneidungen mit dem System für *tonic activation* bei Pribram und McGuinness (1975; vgl. Sanders, 1983) auf.

Mentales Simulieren: Gesättigt-Sein, Zufriedenheit und Entspannung einerseits und Stressreaktionsprozesse andererseits sind mit dem Serotonin-System verbunden. Auch rhythmische Bewegungen und Temperaturregulation stehen mit Serotonin-Aktivität in Verbindung. Dieses System verhindert einerseits SW- und REM-Schlaf, unterdrückt andererseits auch Theta und Gamma-Aktivität, die für aktive Wachheit typisch ist. Es wird vom Orexin/Noradrenalin/Histamin-System bzw. vom Dopamin/Acetylcholin-System, also bei gegebenen appetitiven oder aversiven Herausforderungen und Möglichkeiten gehemmt (Jones, 2003). Die kortikale Aktivität im Ruhezustand ist auf Grund ihrer Funktionalität von besonderer Bedeutung. Bei Säugetieren im Ruhezustand beobachtet man, dass die Aktivität rhythmisch zwischen bestimmten kortikalen Regionen schwankt (Fox et al., 2005). Wie weiter unten ausführlicher dargestellt wird, geht mit diesem Aktivitätsmuster die mentale Simulation von Problemen und Situationen einher. Diese Probleme und Situationen können sich auf die eigene Person, das persönliche soziale Umfeld, vergangene Erfahrungen oder Erwartungen für die Zukunft beziehen. Solche Simulationen bereiten effizientes und zielbezogenes zukünftiges Verhalten vor (vgl. Baumeister, Masicampo & Vohs, 2011).

Schlafen: Während des Schlafes verhindert die Kopplung des GABA-Systems mit den anderen Arousal-Systemen das Aufwachen. GABA wird auch mit der Hemmung des dopaminergen mesolimbischen Systems (Aufsuchen und Konsumieren appetitiver Stimuli) in Verbindung gebracht (Posner et al., 2005). Dennoch kann das Dopamin/Acetylcholin-System während des REM-Schlafes aktiv sein (Jones, 2003). Während des Wachzustandes übernimmt das GABA-System eine koordinierende Funktion, indem es für spezifische Hemmung in aktiven neuronalen Schaltkreisen sorgt.

2.2.3 Arousal und das Circumplex-Modell der Core Affects

Posner, Russell und Peterson (2005) schlagen über das *Circumplex-Modell* die Brücke von neurologischen Arousal-Systemen zu sprachlich mitgeteilten Stimmungen, Gefühlen bzw. Emotionen. Im Circumplex-Modell werden emotionale sprachliche Äußerungen (Selbst- oder Fremdeinschätzung) in ein zweidimensionales Koordinatensystem mit den Dimensionen *arousal* und *valence* eingeordnet (Abbildung 3). Die neurophysiologische Grundlage von arousal identifizieren sie in erster Linie mit der Aktivität der *formatio reticularis*. Die Auswirkungen ihrer Aktivität entsprechen der vertikalen Achse der Arousal-Uhr. Die neurophysiologische Grundlage von valence identifizieren sie in erster Linie mit dem dopaminergen mesolimbischen System und beziehen Hemmungen durch GABAerge und serotonerge Neuronen ein. Die Autoren gehen noch davon aus, dass die beiden genannten zentralnervösen Strukturen unabhängig voneinander das emotionsbezogene Erleben

beeinflussen. Wie man Abbildung 1 entnehmen kann, ist eine Unabhängigkeit jedoch keineswegs gegeben. Im Gegenteil: Die beteiligten Strukturen bilden Elemente eines komplexen Gesamtsystems und unterhalten zahlreiche aktivierende und hemmende Verbindungen (Brown et al., 2012; Cacioppo & Gardner, 1999; Jones, 2003). Ungeachtet dessen beobachtet man, dass Neuronen, die einen bestimmten Neurotransmitter ausschütten, vermehrt feuern, wenn bestimmte Verhaltensweisen (z. B. GABA beim Schlafen, Noradrenalin bei Flucht-oder-Kampf-Reaktionen, Dopamin bei Nahrungsaufnahme, Histamin beim Orientieren in neuen Umgebungen) ausgeführt werden (Tabelle 1; Brown et al., 2012; Cacioppo & Gardner, 1999; Jones, 2003). Eine vereinfachende Systematisierung dieser Zusammenhänge führt auch in der Arousal-Uhr zu zwei Dimensionen: Aktivität und Valenz (Abbildung 2). Mit der Herleitung der Arousal-Uhr wird jedoch die Behauptung, es gäbe für bestimmte Emotionen bestimmte neuronale Strukturen, relativiert. Lindquist, Wager, Kober, Bliss-Moreau und Feldman Barrett (2012) liefern in ihrem metaanalytischen Review empirische Evidenz gegen eine neuronale Lokalisierung von Basisemotionen, wobei ein endgültiges Votum noch aussteht (siehe z. B. die Kommentare zu Lindquist et al., 2012 im selben Heft oder Colibazzi et al., 2010; Gerber et al., 2008; Hamann, 2012). Der Herleitung der Arousal-Uhr zufolge würde es nicht ausreichen, die Aktivität neuronaler Regionen bei emotionsbezogener Informationsverarbeitung zu bestimmten (etwa durch Messung der Sauerstoffkonzentration im Blut bei bildgebenden Verfahren). Vielmehr müssten die an einer emotionsbezogenen Informationsverarbeitung beteiligten Neurotransmitter und Neuromodulatoren nach ihrer Art und Lokalität simultan quantifiziert werden. Denn bei der Auseinandersetzung eines Individuums mit einer Herausforderung oder Möglichkeit in der Umwelt (emotionale Episode) stellt sich ein Zustand ein, an dem alle Arousal-Systeme und alle Überlebens-Schaltkreise mehr oder weniger stark aktiviert beteiligt sind (vgl. global organismic state bei LeDoux [2012]). Je nach Erlebnisqualität und beherrschtem Vokabular kann dieser Zustand schließlich mit einem bestimmten emotionalen Begriff benannt werden (vgl. Feldman Barrett, 2006b, 2012).

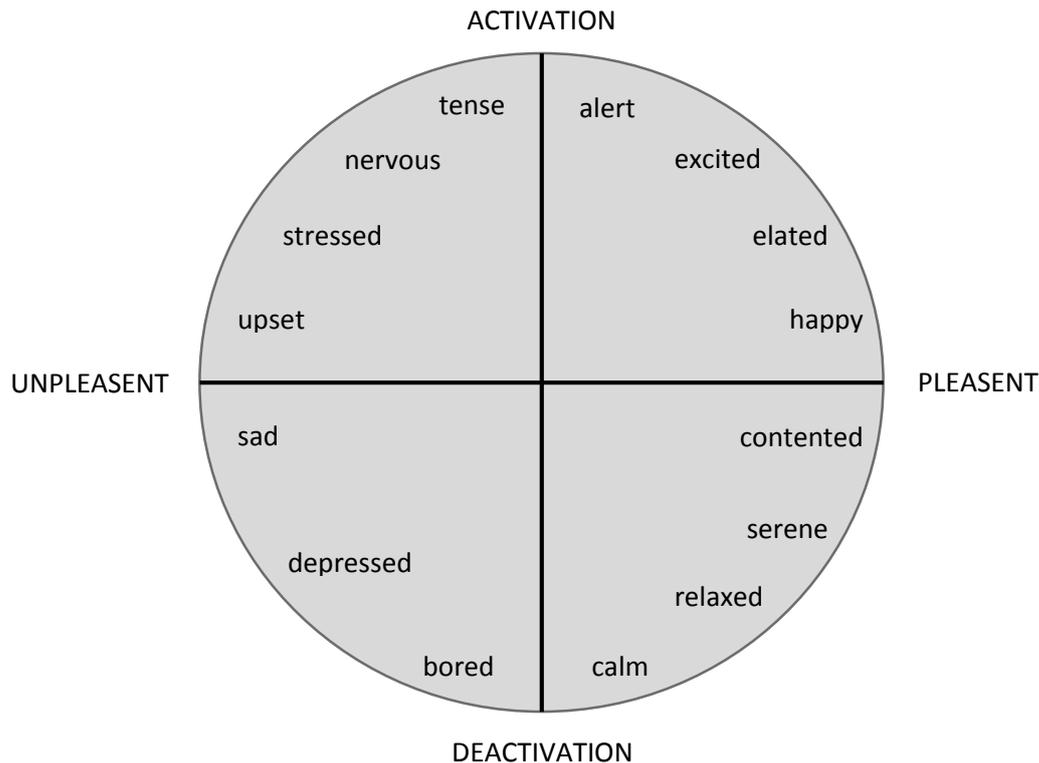


Abbildung 3. Circumplex-Modell nach Posner, Russell und Peterson (2005, S. 716, Figure 1)

Um das komplexe und ganzheitliche Erleben emotionaler Episoden von jenen emotionsbezogenen Zuständen abzugrenzen, die im Circumplex-Modell angeordnet werden, haben Russell und Feldman Barrett (Feldman Barrett, 2006b; Russell, 2003, 2009; Russell & Feldman Barrett, 1999) für letztere den Begriff *Core Affect* geprägt: “Core affect is a pre-conceptual primitive process, a neurophysiological state, accessible to consciousness as a simple non-reflective feeling: feeling good or bad, feeling lethargic or energised” (Russell, 2009, S. 1264). Core Affect ist frei und zeitvariabel fließend und kann daher am ehesten mit Stimmung (mood) gleichgesetzt werden (Russell, 2009). Core Affect stellt den Autoren zufolge eine einfache Form der Bedeutungs-Analyse (hilfreich, nicht hilfreich, belohnend, bedrohend) dar, die in den Sprachen aller menschlichen Kulturen ihre Entsprechung findet und unter Säugetieren homolog ist (vgl. Feldman Russell, 2006b; LeDoux, 2012). Eine Emotion manifestiert sich dagegen unter anderem in (a) Ausdruck in Gesicht und Stimme, (b) Aktivität des autonomen Nervensystems, (c) Verhalten (z. B. Annäherung vs. Vermeidung), (d) Erleben der Emotion, (e) Bewertung bzw. Einschätzung des Ereignisses/der Situation, (f) Zuordnung einer Ursache zur Emotion (Attribution) sowie in (g) Emotions-Regulation. Das Zusammenspiel all dieser Komponenten während einer emotionalen Episode nennt Russell (2003, 2009) *psychological construction*. Einen ähnlichen Zugang wählt Feldman Barrett (2006b) mit ihrem perzeptiv orientierten Ansatz der *conceptual acts*: „We experience an

instance of emotion in ourselves, or see it in others, when we conceptualize an ongoing, basic affective state via the process of categorization“ (a. a. O., S. 21). Alle genannten Komponenten sind, genauso wie Core Affect, durchgehend aktive Prozesse. Jedoch sind deren Korrelationen meist relativ schwach. Dies gilt auch dann, wenn eine bestimmte Emotion (z. B. Ärger) sehr intensiv erlebt wird (vgl. Feldman Barrett, 2006a, b; LeDoux, 2012). Daher sieht Russell (2003, 2009) für eine feste Zuordnung bestimmter Emotion zu bestimmten Mustern dieser Komponenten nur geringe Chancen. Er fordert jedoch, wie auch Feldman Barrett (2006a) auf Grund bereits erzielter Erkenntnisse dazu auf, die Untersuchung von Zusammenhängen verschiedener Manifestationen von Emotionen fortzusetzen. Russel (2003, 2009) ist davon überzeugt, dass Kategorien der Emotion, wie alle anderen Kategorien unseres Denkens, unscharf sind. Muster können prototypisch sein für eine bestimmte Kategorie, gerade an der Schwelle sein, zu einer Kategorie zu gehören oder zu keiner der etablierten Kategorien gehören. Dies entspricht der Annahme, dass Emotion auch ein sozial vermitteltes und erlerntes Konstrukt ist (Feldman Barrett, 2009).

Im Circumplex-Modell werden verschiedene dimensionale Modelle zu emotionalem Erleben zusammengefasst. Eines dieser Modelle stammt von Robert E. Thayer (1967). Seine Intention war es, Arousal in Anlehnung an die *Mood Adjective Checklist (MACL)* von Nowlis (1965) mittels Selbst-Bericht zu erfassen. Probanden wurden gebeten, mittels einer Liste von Adjektiven in Kombination mit einer vierstufigen Antwortskala ihr momentanes Befinden zu beschreiben (*Activation-Deactivation Adjective Check List, AD-ACL*). Thayer (1967) setzt diese Selbst-Berichte mit physiologischen Messungen zum Arousal, der Tageszeit und Stressinduktion in Verbindung. Er stellt fest, dass die Selbst-Berichte mit den weiteren erhobenen Variablen zum Arousal erwartungsgemäß zusammenhängen. 1967 unterscheidet er noch zwischen vier Arousal-Faktoren (*General Activation, High Activation, General Deactivation, Deactivation-Sleep*), die er auf einem hypothetischen Kontinuum angeordnet sieht. Später geht Thayer von zwei Arousal-Dimensionen aus (Thayer, 1978). Er behält die AD-ACL bei, fasst jedoch je zwei der vier Arousal-Faktoren zu einer Dimension zusammen: General Activation - Deactivation-Sleep wird zur Dimension *energy - sleep (Dimension A)*. Diese Dimension folgt einem circadianen Rhythmus und ist mit körperlicher und kognitiver Aktivität verbunden. Hohe Ausprägungen sind mit eher positiven Emotionen verbunden. High Activation - General Deactivation wird zur Dimension *tension - placidity (Dimension B)*. Diese Dimension assoziiert Thayer (1978) vorwiegend mit Stressreaktionsprozessen. Hohe Ausprägungen sind mit eher negativen Emotionen verbunden. Beide Dimensionen beginnen mit einem Zustand hohen energetischen Aufwands (*energy, tension*) und enden mit einem

Zustand niedrigen energetischen Aufwands (sleep, placidity). Dieser Umstand soll beobachtete Korrelationen zwischen diesen beiden Dimensionen begründen. Thayer (1978) argumentiert, dass Selbst-Berichte insofern den damals verfügbaren physiologischen Messungen voraus waren, weil die dimensionale Struktur der Selbst-Berichte anders als unidimensionale physiologische Messungen (z. B. Herzrate, Blutdruck, Muskeltonus) auch äußere Bedingungen wie die Tageszeit oder Stressinduktion widerspiegeln. Schon Nowlis und Nowlis (1956) stellen in ihren Studien über die Interaktion von Situation und Einnahme von Arzneimitteln fest, dass Selbst-Report von Stimmung (mood) mittels Adjektiv-Listen ihr brauchbarstes Instrument ist.

Russell (1979) fügt den ursprünglichen vier Arousal-Faktoren von Thayer (1967) durch Ergänzung der AD-ACL folgende Faktoren hinzu: *Pleasure, Displeasure, Arousal, Sleepiness, Dominance, Submissiveness* und *Depression*. Er ermittelt mit diesen Faktoren eine zweidimensionale Struktur emotionalen Erlebens. Eine Dimension heißt *pleasure-displeasure*, die zweite heißt *degree of arousal*. Russell (1979) stellt unabhängig von Thayer (1978) fest, dass dessen vier Faktoren eine zweidimensionale Struktur beschreiben. Allerdings scheint auch *Depression* dem Faktor *General Activation* gegenüber zu stehen. Weiter scheinen Thayers Dimensionen um etwa 45° gegenüber denen von Russell gedreht zu sein, mit *General Activation* im Quadranten *pleasure - high arousal*. Russell (1980) zeigt, dass das Circumplex-Modell sowohl den sprachlich-konzeptuellen Gebrauch emotionsbezogener Begriffe, als auch - weniger eindeutig - das introspektive emotionale Erleben widerspiegelt. Russell (1979, 1980) bezieht sich bei seinen Studien weniger auf biopsychologische Grundlagen sondern sehr viel stärker auf methodische Aspekte wie Antwort-Skalen, Item-Formate und Sortier-Aufgaben.

Einen ähnlichen Ansatz wie Russell (1979, 1980) verfolgen Randy J. Larsen und Edward Diener (1992) sowie Rainer Reisenzein (1994). Larsen und Diener (1992) beschreiben in ihrem theoriegeleiteten Circumplex-Modell ebenfalls zwei unabhängige Dimensionen *Activation* und *Pleasure*. Sie unterteilen die entstehende Ebene in acht Oktanten. Diesen acht Oktanten wurden insgesamt 48 emotionsbezogene Eigenschaftswörter zugeordnet. Reisenzein (1994) interessiert sich in diesem Zusammenhang vor allem für die Frage, wie unterschiedliche Qualitäten und Intensitäten von Emotionen erklärt werden können. Er kommt zu dem Schluss, dass die Qualität einer Emotion als Kombination von *Activation* und *Pleasure* aufgefasst werden kann. Die Intensität ergibt sich dahingehend aus dem Verhältnis von *Activation* zu *Pleasure*.

David Watson und Auke Tellegen (1985) gehen ebenfalls von einem Modell emotionalen Erlebens mit zwei unabhängigen Dimensionen aus: *Positive Affect (PA)* und *Negative Affect (NA)*. Personen mit hohem PA fühlen sich eher energiegeladen, konzentriert und engagiert.

Personen mit hohem NA fühlen sich eher verärgert, schuldig, ängstlich und nervös. Watson und Tellegen (1985) gehen ebenfalls von einem Circumplex-Modell aus, das bis auf Rotation um 45° und anschließende vertikale Spiegelung weitgehend mit jenem von Russell (1979, 1980) übereinstimmt. Watson, Clark und Tellegen (1988) legen entsprechend ihrer Konzeption den *Positive and Negative Affect Schedule (PANAS)* vor. Hierbei handelt es sich um eine Liste von zwanzig emotionsbezogenen Eigenschaftswörtern (jeweils zehn Items zu PA und NA), die auf einer fünfstufigen Skala im Selbst-Bericht bewertet werden sollen. Watson, Clark und Tellegen (1988) legen für den PANAS Validitäts- und Reliabilitätsmaße vor, die ihrer theoretischen Konzeption emotionalen Erlebens entsprechen.

Es gab weitere Beiträge zu dimensional Modellen emotionalen Erlebens (z. B. Wundt, 1912/1929; Schlosberg, 1941), auf die sich die Autorinnen und Autoren der genannten Konzepte bezogen. Gegen Ende des 20. Jahrhunderts konnten sämtliche genannten Konzepte im Circumplex-Modell integriert werden. Maßgeblich hierfür waren zum einen eine Weiterentwicklung statistischer Methoden (lineare Strukturgleichungsmodelle) und zum anderen eine multimethodale Herangehensweise. Durch lineare Modellierung und die Verwendung variierender Antwortformate konnten Green, Goldman und Salovey (1993) unsystematische und systematische Fehlerquellen in Bezug auf Positive und Negative Affect (Watson, Clark, & Tellegen, 1988; Watson, & Tellegen, 1985) kontrollieren. Watson, Tellegen und Clark (1999) konnten die Befunde von Green et al. (1993), Negative und Positive Affect würden eine bipolare Dimension aufspannen, durch eine Annäherung an das Circumplex-Modell widerlegen. Zugleich nahmen sie eine Erweiterung und Umbenennung ihrer Skalen in *Positive* und *Negative Activation* vor. Dadurch erschloss sich eine starke konzeptionelle Ähnlichkeit zu Thayers Modell von energetischer (*Energy*, entspricht *General Activation*) und angespannter Aktivierung (*Tension*, entspricht *High Activation*) (Thayer, 1996). Yik, Russell und Feldman Barrett (1999) bedienten sich schließlich linearer Strukturgleichungsmodelle und variierender Antwortformate, um vier dimensionale Modelle emotionalen Erlebens im Circumplex-Modell unter Kontrolle unsystematischer und systematischer Fehlerquellen zu integrieren (vgl. Feldman Barrett, & Russell, 1998): Energy und Tension von Thayer (1996), Negative und Positive Affect von Watson und Clark (1994; Watson, Clark, & Tellegen, 1988), Pleasure und Activation von Feldman Barrett und Russell (1998; *Current Mood Questionnaire, CMQ*) und schließlich das theoretische Konzept von Pleasure und Activation aus dem Review von Larsen und Diener (1992). Sie konnten empirisch belegen, dass die Dimensionen dieser vier Modelle bipolar sind und dass sie sich entsprechend Abbildung 4 im Circumplex-Modell anordnen lassen. Zuletzt konnten Yik, Russell und Steiger (2011) das Circumplex-Modell

weiter bestätigen und verfeinern, indem sie vorhandene Lücken im älteren Modell (Yik et al., 1999) durch Einbeziehung weiterer Modelle zum emotionalen Erleben schlossen. Darüber hinaus konnten sie verschiedene Persönlichkeitseigenschaften (traits) mit dem Circumplex-Modell in Beziehung setzen. Scherer (2005) konnte die Kreisfläche des Circumplex-Modells mit Begriffen füllen und so die Intensität bestimmter emotionsbezogener Begriffe abbilden.

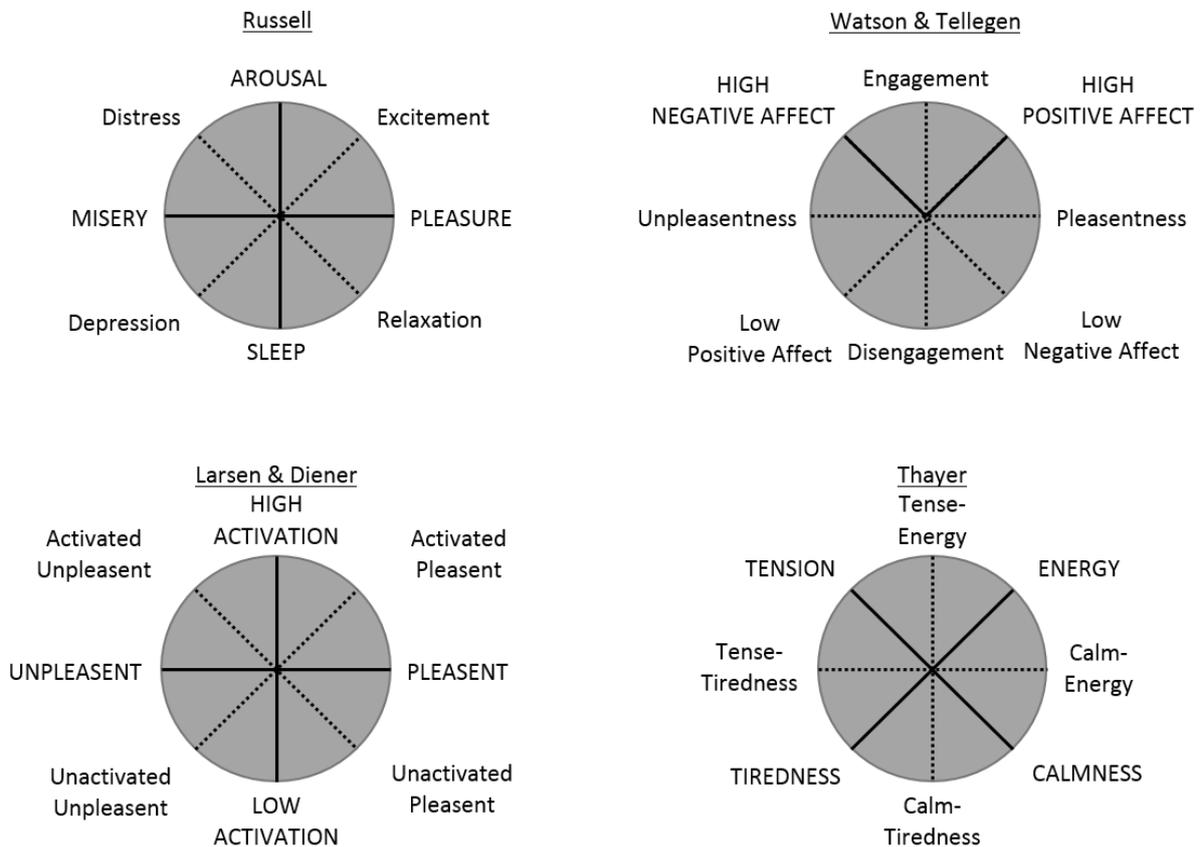


Abbildung 4. Synthese verschiedener dimensionaler Modelle emotionalen Erlebens im Circumplex-Modell (nach Yik, Russell, & Feldman Barrett, 1999, S. 601, Figure 1)

Das zweidimensionale Circumplex-Modell ist nicht unumstritten. Schimmack und Reisenzein (2002) wollen durch Feststellung einer Nullkorrelation zwischen den Residuen von Negative und Positive Affect nach Herauspartialisierung ihres Valence-Anteils die Notwendigkeit eines mindestens dreidimensionalen Modells belegt haben. Allerdings belegen sie mit ihrer Analyse nur das Offensichtliche: Wenn Negative und Positive Affect unabhängig voneinander variieren, dann variieren auch sämtliche Projektion dieser Variablen senkrecht zu einer vorgegebenen Geraden in derselben Ebene (z. B. Valence) unabhängig voneinander. Dies steht einem zweidimensionalen Modell genauso wenig entgegen wie einem beliebig höherdimensionalen Modell. Unabhängig hiervon können Schimmack und Grob (2000) belegen, dass ihr eigenes Verfahren zur Erfassung von Stimmungen am besten mit einem dreidimensionalen Modell zu beschreiben ist.

Bezieht man weitere Komponenten des emotionalen Erlebens über den Selbst-Bericht hinaus ein, so erhält man - durchaus erwartungsgemäß (Russell, 2009) - faktorenanalytisch weitere Dimensionen. So können Fontaine, Scherer, Roesch und Ellsworth (2007) unter Berücksichtigung von „(a) appraisals of events, (b) psychophysiological changes, (c) motor expressions, (d) action tendencies, (e) subjective experiences, and (f) emotion regulation“ (a. a. O., S. 1050) vier Dimensionen identifizieren. Hierzu baten sie belgische, englische und schweizerische Studierende im Rahmen einer Internet-Studie einzuschätzen, wie typisch 144 benannte Merkmale aus den Kategorien (a) bis (f) für 24 ausgewählte Emotionen sind. Fontaine et al. (2007) untersuchten demnach prototypische Repräsentationen ausgewählter Emotionen in einer selektiven Stichprobe. Nach einer Faktorenanalyse fügten sie den beiden Dimensionen *Valence* und *Arousal* noch *Potency* und *Unpredictability* hinzu. Für die Umschreibung der Dimensionen *Valence*, *Arousal* und *Potency* beziehen sich die Autoren auf Osgood, May und Miron (1975). *Unpredictability*, eine bis dahin im Zusammenhang mit Emotionen kaum untersuchte Dimension, repräsentiert Emotionen, die mit neuen Stimuli oder unbekanntem Situationen in Zusammenhang stehen.

Würde man die Arousal-Systeme (Abbildung 1; Brown et al., 2012; Jones, 2003) als neurophysiologische Grundlage der Stimmung bzw. der Core Affects zugrunde legen (z. B. Nowlis & Nowlis, 1956; Thayer, 1967, 1978; Posner, Russell & Peterson, 2005), wären gemäß obiger Ausführungen zu Arousal für ein einfaches Modell schon neun Dimensionen notwendig: Je eine Dimension für die mittleren Konzentrationen jeder der acht Neurotransmitter und eine weitere für die mittlere Kortisol-Konzentration. Der Zustand des Gesamtsystems müsste mit einer 9×9 -Matrix beschrieben werden, die den Übergang dieser Konzentrationen von Zeitpunkt t zu Zeitpunkt $t + \Delta t$ (z. B. nächster Beta- oder Gamma-Zyklus) modelliert. Diese Matrix wäre unter Berücksichtigung circadianer Rhythmen, neuronaler Potenzierung und der Neuroplastizität selbst zeitabhängig. Würde man in einem verfeinerten Modell alle in Abbildung 1 eingezeichneten Eingänge (Pfeilspitzen der eingezeichneten Pfade) berücksichtigen, so läge bereits ein knapp 100-dimensionales Problem vor. Die zeitabhängige Zustandsmatrix wäre entsprechend groß. Um die Zustandsmatrix selbst zu ermitteln, wäre das Eigenwertproblem zu lösen. D. h. man hätte bis zu neun bzw. bis zu knapp 100 paarweise linear unabhängige physiologischer Stimulierungen (Eigenvektoren) zu finden, unter denen die Konzentrationsverhältnisse an allen Eingängen beim nächsten Messzeitpunkt paarweise gleich blieben. Solche physiologischen Stimulierungen sind nicht standardisierbar, da die Verschaltungen und Reaktionsweisen von Arousal-Systemen individuell variieren. Ein auf physiologischer Messung basierendes mathematisches Modell, das die Vorhersage emotionalen

Empfindens in Abhängigkeit der physiologischen Stimulierung erlaubt bzw. die Dynamik emotionalen Empfindens exakt nachbildet, ist somit nicht möglich. Bei den Arousal-Systemen handelt es sich um ein nichtlineares dynamisches System (s. u.), für dessen Dynamik sich kein exakt lösbares zeitabhängiges Gleichungssystem angeben lässt. Ihr zeitliches Verhalten kann höchstens näherungsweise mittels numerischer Integration simuliert, jedoch nicht analytisch vorhergesagt werden (vgl. Strunk & Schiepek, 2006). Die zweidimensionale Arousal-Uhr ist eine starke Vereinfachung der Arousal-Systeme nach Abbildung 1. Die Darstellung in Abbildung 1 ist eine starke Vereinfachung der tatsächlichen Gegebenheiten in einem realen Gehirn. Erstens sind die Verbindungen in einem Gehirn sehr viel komplexer und variieren individuell (z. B. wirken Neurotransmitter an ihren unterschiedlichen Rezeptoren teils hemmend und teils erregend). Zweitens ist noch eine Vielzahl weiterer biochemischer Substanzen an der Neuromodulation beteiligt (s. u.). Unter der Arousal-Hypothese wäre es für verschiedene, vor allem praktische Fragestellungen, mehr als zu empfehlen, die zu erwartende Komplexität auf zwei, drei, vielleicht noch vier Dimensionen herunterzubrechen. Die Wahl der Basis bzw. des Koordinatensystems könnte an die Art der Fragestellung angepasst werden.

2.2.4 Emotion, Motivation und Verhalten

Feldman Barrett (2012) hebt als weitere bedeutsame Komponente des emotionalen Erlebens die selbstregulativen und sozialen Funktionen von Gefühlen/Emotionen hervor. Sie versteht Gefühle unter anderem als ein in einer sozialen Gemeinschaft geteiltes Kategoriensystem (vgl. Russell, 2003, 2009; Feldman Barrett, 1999, 2006b). Das Empfinden und Ausdrücken von Gefühlen hat ihrer Ansicht nach die Funktion (a) der aktuellen Situation eine Bedeutung zu geben, (b) eigene Verhaltensweisen, körperliche Zustände und Intentionen anderer Personen und sich selbst gegenüber zu kommunizieren und zu kommentieren sowie (c) das eigene und das Verhalten anderer Individuen zu steuern bzw. zu beeinflussen. Durch das Empfinden und Mitteilen von Gefühlen werden diese selbstregulativ und sozial wirksam und damit real. Formgebend ist hierbei unter anderem der nach sozialen/kulturellen Normen angemessene Ausdruck von Emotionen. Die Art und Weise des Gefühlsausdrucks und des Verständnisses von Gefühlen anderer Individuen muss zu großen Teilen sozial erlernt werden und ist somit von individuellen Erfahrungen abhängig. Dieser sozialen und kontextuellen Überformung von Gefühlen stehen sensumotorische Afferenzen und ein neurophysiologischer Zustand (im Sinne der Core Affects) gegenüber. Hiermit setzt Feldman Barrett (2012) ihre Überlegungen zum *conceptual act fort*. In diesem ist die emotionale Kategorisierung (*act of categorization*) des inneren Zustandes oder des beobachteten Verhaltens Anderer wesentlich: „The experience of feeling emotion occurs when a person categorizes his internal state. The experience of seeing

emotion in another person occurs when that person's behavior is categorized as emotional“ (Feldman Barrett, 2006b, S. 27). Kategorisierung setzt konzeptuelles Wissen voraus. Konzeptuelles Wissen (*conceptual knowledge*, Feldman Barrett, 2006b, S. 27) über eine Emotion baut sich Feldman Barrett (2006b) zufolge durch Erfahrungen auf, in denen eine gegebene Kombination von Situation und neurophysiologischem Zustand (Core Affect) mit der Benennung dieser Emotion verbunden wird. Dieses konzeptuelle Wissen wird in späteren, bekannten oder unbekanntem Situationen zur perceptiven Komponente für die Benennung einer Emotion (*top down*), während der Core Affect die wesentliche homöostatische Komponente darstellt (*bottom up*) (vgl. Cacioppo & Gardner, 1999). Feldman Barrett (2006b) zufolge tragen beide Komponenten und ihre Interaktion schließlich zur situativen Konzeptualisierung (*situated conceptualization*) einer Emotion bei (vgl. Cacioppo & Gardner, 1999; Russell, 2009). Da konzeptuelles Wissen über Emotionen nach diesem Ansatz stark von Erfahrungen abhängig ist, werden interindividuelle Unterschiede im emotionalen Erleben, Verhalten und insbesondere Ausdruck (z. B. in Feinheit, Fokus oder Verhaltenstendenz), aber auch in physiologischen Reaktionen verständlich (Cacioppo & Gardner, 1999; Feldman Barrett, 2006b, 2012).

Feldman Barrett (2006b) geht demnach davon aus, dass das Erleben einer Emotion aus dem Zusammenspiel mehrerer Komponenten hervorgeht. Klaus R. Scherer (2005) schlägt mit seinem *component process model* ein Konzept vor, dass die relative Aktivität verschiedener Komponenten beim Erleben einer Emotion hervorhebt. Seiner Ansicht nach sind Emotionen als physiologische Antworten auf Bewertungen externer und interner Stimuli zu verstehen. Bei einer Emotion werden die Zustände von fünf organismischen Subsystemen in synchronisierter und abhängiger Weise verändert: (a) Informationsverarbeitung (kognitive Komponente; Bewertung/Einschätzung), (b) Unterstützung (neurophysiologische Komponente; körperliche Symptome), (c) Ausführung (motivationale Komponente; Verhaltenstendenzen), (d) Verhalten (motorische Ausdrucks-Komponente; Gesichts- und Stimm-Ausdruck) sowie (e) Überwachung (subjektive Gefühls-Komponente; emotionales Erleben) (a. a. O., S. 698, Table 1). In seiner Sichtweise geschieht Emotion als schnelle, synchronisierte Anpassungsreaktion auf plötzliche Ereignisse, die als relevant eingeschätzt werden. Er möchte Gefühl bzw. Stimmung (als eine Komponente von Emotion) klar von einer Emotion bzw. einem emotionalen Ereignis abgegrenzt wissen. Aus der Perspektive psychischer Belastung und Beanspruchung fällt es allerdings schwer, die verwendeten Begrifflichkeiten von Scherer (2005) in klar voneinander abgegrenzten Naturphänomenen wiederzuerkennen. Wie nimmt man eine klare Unterteilung körperlicher Reaktionen in körperliche Symptome, emotionalem Erleben und kognitive Prozesse vor? Auf welche Weise kann man psychische Belastung konsensfähig in diskrete

Ereignisse unterteilen? Ab wann ist emotionales Erleben als Komponente psychischer Beanspruchung nicht mehr nur Gefühl, sondern schon Emotion (wann ist „emotional experiencing ... relatively high“ [a. a. O., S. 702])? Ein kontinuierliches Einströmen sensorischer und viszeraler Stimuli und eine kontinuierliche, zeitvariable Anpassung emotionalen Erlebens (Barrett, 2006b; Russell, 2009) ist mit dem Paradigma psychischer Belastung und Beanspruchung eher verträglich und schließt intensiv erlebte Emotionen im Sinne Scherers (2005) nicht aus.

LeDoux (2012) fasst bei allen theoretischen Auseinandersetzungen innerhalb der Emotionsforschung zusammen, dass der empirische gemeinsame Nenner die Introspektion ist und sein muss. Seiner Beobachtung nach werden Probanden letztlich immer um die Benennung von Gefühlen gebeten (vgl. Cacioppo & Gardner, 1999). Oder in den Begriffen von Feldman Barrett (2006b) ausgedrückt: Man fordert Probanden in entsprechenden Untersuchungen stets zu einem conceptual act auf. Solche Untersuchungen sind LeDoux' (2012) Ansicht nach nicht geeignet, um den Ausdruck einzelner Emotionen (sog. Basis-Emotionen; z. B. Ekman, 1972) substanziell auf physiologische Reaktionen zurückzuführen. Seine Auseinandersetzung mit der vorliegenden empirischen Evidenz gibt ihm Recht (vgl. Colibazzi et al., 2010; Gerber et al., 2008; Hamann, 2012; Lindquist et al., 2012) und mit dem Ansatz von Feldman Barrett (2006b, 2009, 2012) wird dies zumindest plausibel. Dennoch kommt Feldman Barrett (2006b) wie zuvor in ähnlicher Weise Nowlis und Nowlis (1956) und Thayer (1978), zu dem Schluss: “Verbal report, even with all of its failings, may be the only means of assessing the experience of emotion. If we want to know whether a person is experiencing an emotion, we have to ask them” (Feldman Barrett, 2006b, S. 24; vgl. Scherer, 2005). Umgekehrt spricht LeDoux (2012) den Gefühlen/Emotionen nicht ihre neurophysiologische Grundlage ab. Im Gegenteil: Seiner Ansicht nach verfügen die meisten Tiere und vor allem Säugetiere über sehr ähnliche neurologische Verschaltungen. Diese werden je nach Situation unterschiedlich stark aktiviert, initiieren spezifische Verhaltenstendenzen (Funktion) und werden während der Bewältigungsversuche des Individuums zweckmäßig verändert (Neuroplastizität, Lernen). LeDoux (2012) fordert konsequenter Weise von der Emotionsforschung, sich zukünftig im Sinne einer vergleichenden Verhaltenswissenschaft stärker diesen artübergreifenden neuronalen Verschaltungen und ihren Funktionen in emotionalen/motivationalen Situationen zuzuwenden.

Eine wesentliche Gemeinsamkeit der Positionen von Feldman Barrett (2006b), LeDoux (2012) und Scherer (2005) besteht in der begründeten Annahme, dass die affektive Bewertung einer gegebenen Situation mit bestimmten Motivationen und Verhaltenstendenzen einhergeht.

Feldman Barrett (2006b) schreibt: „The intensity of a core affective response ... results in a perceived urgency to act that is independent of the specific action taken (the specific action being tailored to the particular situation at hand)” (a. a. O., S. 31). Scherer (2005) drückt dies wie folgt aus: “Emotions prepare adaptive action tendencies and their motivational underpinnings. In this sense they have a strong effect on emotion consequent behavior, often interrupting ongoing behavior sequences and generating new goals and plans” (a. a. O., S. 702). Cacioppo und Gardner (1999) und LeDoux (2012) gehen hierfür von einem evolutionären Vorteil aus. Cacioppo und Gardner (1999) fassen Befunde und Konzepte zusammen, bei denen die Valenz-Dimension (etwas wird positiv oder negativ bewertet) ein artübergreifendes, übergeordnetes Bewertungsschema darstellt und Vermeidungs- und Annäherungs-Tendenzen steuert. LeDoux (2012) geht davon aus, dass die Aktivierung eines Überlebens-Schaltkreises mit einer affektiven Bewertung einhergeht und spezifische, jedoch durch Erfahrung modifizierbare und von der gegebenen Situation modulierte Verhaltenstendenzen motiviert und anstößt (s. o.).

Die Arousal-Systeme (s. Abbildungen 1 und Tabelle 1 weiter oben, Brown et al., 2012; Jones, 2003) sind mit verschiedenen Situations- und Verhaltenskategorien gekoppelt. Diese Tatsache führte weiter oben zu der theoretischen Ableitung der Arousal-Uhr (Abbildung 2). In der Arousal-Uhr werden zwei viel untersuchte Dimensionen des Circumplex-Modells (Aktivität und Valenz) mit sechs Verhaltens-Kategorien verknüpft. Mit der Arousal-Uhr wird somit die Hypothese aufgestellt, dass bestimmte Spektren der Core-Affects mit bestimmten Verhaltenskategorien verknüpft sind. Dieser theoretischen Position zufolge sollten sich Wohlbefinden, Gesundheit und Handlungsfähigkeit auf folgenden drei Ebenen systematisch bemerkbar machen: (a) Der biopsychologischen und insbesondere der neurophysiologischen Ebene, (b) der Ebene emotionalen Erlebens und (c) der Ebene der Interaktion des Individuums mit seiner Umwelt (Bewerten, Verhalten).

2.3 AROUSAL UND GESUNDHEIT

Der vorliegende Abschnitt ist von der Frage geleitet, wie eine gesunde Arousal-Dynamik gefördert werden kann. Hierzu ist zunächst eine Auseinandersetzung mit Beeinträchtigungen der Arousal-Dynamik hilfreich. Eine Arousal-Dysbalance liegt vor, wenn mindestens eines der beschriebenen Arousal-Systeme übermäßig aktiviert oder gehemmt ist. Die Arousal-Dynamik erfordert den Umsatz biochemischer Energie. Eine Erschöpfung körpereigener Energiereserven, wie sie beispielsweise bei anhaltenden unkontrollierbaren Stressreaktionsprozessen erfolgt, kann zu Arousal-Dysbalancen führen. Sie gipfelt in einer Erosion des

Körpers und des Selbst. Das Burnout-Syndrom wird in diesem Zusammenhang zunehmend als allgegenwärtiges Phänomen diskutiert. Dies sollte als Ausdruck einer wie auch immer begründbaren schwindenden Anpassungsfähigkeit der Mitglieder unserer Gesellschaft ernstgenommen werden. Für die Depression wird mit der Monoaminmangelhypothese eine weitere Arousal-Dysbalance diskutiert. Die Monoaminmangelhypothese der Depression besagt, dass diese unter anderem eine Ursache in einem Mangel an Serotonin oder Noradrenalin hat. Der heuristische psychopharmazeutische Behandlungsansatz für Depression setzt auf die Erhöhung dieser Monoamine. Einen anderen Ansatz verfolgen verhaltensorientierte Interventionen: Eine gesunde Arousal-Dynamik lässt sich durch das Erlernen angemessener persönlicher Bewertungs- und Bewältigungsstrategien sowie die Festigung eines insgesamt gesunden Lebensstils erreichen. Entsprechende Ansätze, wie beispielsweise Stressbewältigungstrainings, haben sich auch am Setting Arbeitsplatz etabliert und bewährt.

2.3.1 ...bis sich das Selbst zersetzt: Zusammenhang zwischen Stress und Energie

Der wichtigste erregende Neurotransmitter, Glutamat, geht über Zwischenprodukte aus der Verwertung zellulärer Energiespeicher, im Nervengewebe hauptsächlich Glucose, hervor. Die Synthese von Glutamat verläuft dabei, wie viele andere metabolische Prozesse auch, über den Citratzyklus. Für die Bildung von Glutamat erfolgt der Ausstieg aus dem Citratzyklus über α -Ketoglutarat. Glutamat wird unter Aufwendung weiterer chemischer Energie (Umsetzung von Nikotinamid-Adenin-Dinucleotid-Hydrid [NADH], NH_3 und H^+) durch die Glutamatdehydrogenase aus α -Ketoglutarat gebildet (Löffler, 2008, S. 157-165). Bei der Bildung von Glutamat wird somit chemische Energie investiert. Der wichtigste hemmende Neurotransmitter, γ -Aminobuttersäure (γ -amino butyric acid; GABA), ist ein direktes Abbauprodukt von Glutamat. Über die Glutamat-Decarboxylase wird Glutamat unter Energiefreisetzung zu GABA und CO_2 abgebaut. Sowohl GABA als auch Glutamat können über Zwischenschritte in den Citratzyklus zurückgeführt und dort unter Freisetzung (Glutamat \rightarrow GABA) oder Fixierung (GABA \rightarrow Glutamat) chemischer Energie ineinander überführt werden (Löffler, 2008). Glutamat und GABA sind somit nicht nur im Sinne der Verhaltenskategorien „Wachen“ und „Schlafen“, sondern auch im Hinblick auf den Energieverbrauch des Zentralnervensystems zwei bedeutende Gegenspieler. Die allgemeine Hemmung durch GABA kann als Folge einer hohen Aktivierung mittels Glutamat unter hoher Energieaufwendung betrachtet werden. Oder vereinfacht formuliert: Einer Phase hoher Ermüdung geht eine Phase starker Aktivität voraus. Dementsprechend konnte beobachtet werden, dass Schlaf u. a. als Reaktion auf die Ansammlung biochemischer Substanzen (z. B. Adenosin, Stickstoffmonoxid, Zytokine [Proteine zur Regulation von Zellwachstum und -differenzierung]) auftritt, die bei

wachaktivem Metabolismus anfallen (homöostatische Schlafreaktion; Brown et al., 2012). Beim Schlafen werden u. a. Energiereserven in Form von Adenosintriphosphat (ATP) aufgebaut (a. a. O.).

Dopamin geht unter Energieeinsatz über drei Schritte aus Phenylalanin, einer Aminosäure, hervor. Noradrenalin ist ein energiefixierendes Derivat von Dopamin, Adrenalin ist ein energiefixierendes Derivat von Noradrenalin. Diese Catecholamine werden im Nebennierenmark sowie im Nervengewebe (hier nur Noradrenalin und Dopamin) synthetisiert und in Granula bis zu ihrer Freisetzung zwischengespeichert. Die Freisetzung von Adrenalin und Noradrenalin erfolgt über neurale Stimulation in Notfallsituationen und „besteht in spezifischen Reaktionen der glatten Muskulatur, der Blutgefäße sowie der Frequenz und Kontraktionskraft des Herzens“ (Löffler, 2008, S. 348). Darüber hinaus werden Stoffwechselfvorgänge in Leber, Skelettmuskulatur und Fettgewebe angeregt, die unter anderem zu einem hohen Glucosespiegel führen und dem Arousal (z. B. erhöhte Glutamat-Synthese) zuträglich sind (SAM-System). Notfallsituationen stehen somit mit einem erhöhten Energieverbrauch in Zusammenhang. Bei nachlassender Stimulierung ihrer Freisetzung lässt die Wirkung von Noradrenalin und Adrenalin vor allem deswegen schnell nach, weil diese beiden Catecholamine sehr schnell abgebaut werden können (vgl. Löffler, 2008).

Bei einer anhaltenden Notfallsituation (unkontrollierbarer Stressreaktionsprozess, z. B. Nahrungs-Entzug) ist eine langfristige Energieversorgung notwendig, die weniger auf die energieintensiven Prozesse des sympathischen Systems zurückgreifen sollte. In einem solchen Falle setzt die Hypothalamus-Hypophysen-Nebennierenrinden-Achse (HPA-System) ein, um das Hormon Kortisol freizusetzen. Kortisol ist, wie Androgene (männliche Geschlechtshormone) und Östrogene (weibliche Geschlechtshormone), ein Derivat von Pregnenolon, das selbst wiederum ein Cholesterin-Derivat ist. Durch die Bildung von Kortisol aus Pregnenolon werden zwei energiereiche NADH-Einheiten frei (Löffler, 2008, S. 345). Die Kortisol-Freisetzung nimmt ihren Ausgang in der vielfältigen Innervierung des Hypothalamus und der damit einhergehenden Freisetzung des *corticotropin releasing hormone* (CRH). Dieses löst seinerseits die Freisetzung des *adrenocorticotropen Hormons* (ACTH) der Hypophyse aus, was schließlich in der Nebennierenrinde die Produktion und Freisetzung des Kortisols stimuliert. Das Kortisol selbst hemmt Hypothalamus und Hypophyse, so dass ein Gleichgewicht zwischen neuronal-adrenerger Anregung und hormoneller Hemmung entsteht (vgl. Jones, 2003; Löffler, 2008). Auch Kortisol bewirkt die Anregung von Stoffwechselfvorgängen, die Energie aus körpereigenen Energiereserven freisetzen (z. B. Abbau der Skelettmuskulatur durch Aktivierung Atrophie-induzierender Gene oder Förderung der

Gluconeogenese aus Amino- und Fettsäuren durch Biosynthese entsprechender Enzyme, vgl. Löffler, 2008, S. 189). Darüber hinaus wirkt es hemmend auf die Proteinbiosynthese, unterdrückt Immun-Antworten, erhöht Blutdruck und Herzzeitvolumen und ist entzündungshemmend. Gendefekte für die Expressierung verschiedener Kortisol-Rezeptoren können mit entsprechenden Erkrankungen (z. B. Bluthochdruck, Rheuma, Überempfindlichkeit des HPA-Systems, verschiedene Stoffwechselstörungen, Besonderheiten im Verhalten) einhergehen (DeRijk & de Kloet, 2006). Insbesondere der bei Nahrungsentzug herrschende Mangel an Glucose wäre für das Gehirn fatal, da das Nervengewebe nahezu ausschließlich diesen Ausgangsstoff als Energiequelle nutzen kann. Der durch Kortisol bewirkte Raubbau am eigenen Körper ermöglicht dem Individuum, Wochen ohne Nahrungszufuhr zu überleben und das Gehirn für eine angemessene Antwort mit ausreichend Glucose zu versorgen (Löffler, 2008). Ist ein Stressreaktionsprozess kontrollierbar, verhindert eine maßvolle Freisetzung von Kortisol das Überschießen einer adrenergen Reaktion auf eine Notfallsituation und führt zu einer Verstärkung funktionaler neuronaler Netzwerke. Ist er nicht kontrollierbar, so wird die adrenerge Reaktion dennoch gehemmt und es kommt zu einer Löschung aktivierter neuronaler Verschaltungen, die nicht (mehr) zu einer Bewältigung der gegebenen Situation beitragen (DeRijk & de Kloet, 2006; Huether et al., 1999; Hüther, 2014; Sapolsky, 2003). Dies entspricht dem Aufgeben unangemessener Bewertungs- und Bewältigungsmuster.

Glutamat-GABA-System, SAM-System und HPA-System sowie ihre Wechselwirkungen stehen unter der Maßgabe, die dem Individuum zur Verfügung stehenden körpereigenen Energiereserven unter verschiedenen Bedingungen für angemessene Antworten und Aktivitäten einzusetzen. Bewältigen, Haushalten, Aufgeben und Zersetzen stellen somit Verhaltens- bzw. Prozess-Kategorien dar, die auf einem Kontinuum des Energieverbrauchs angeordnet werden können (vgl. Duffy, 1962).

Hüther (2014) hebt in diesem Zusammenhang hervor, dass das Gehirn relativ schlecht zwischen vorgestellten und auf externen Stimuli beruhenden Situationen unterscheiden kann und die Bedeutungszuschreibung ohnehin eine Konstruktion ist (vgl. DeRijk & de Kloet, 2006; Foy et al., 2006). Das beständige gedankliche Verarbeiten unkontrollierbarer Stressreaktionsprozesse wird unter den Begriffen *Rumination* (unablässige negativ konnotierte Reflexion damit zusammenhängender vermeintlicher Ursachen, Symptome und Konsequenzen) und *Sich-Sorgen* (*worry*; unablässiges gedankliches Durcharbeiten eines bestehenden Problems mit unsicherem, aber wahrscheinlich negativem Ausgang) als bedeutendes Korrelat von Depression, Angststörungen und anderen psychischen Beeinträchtigungen bzw. Störungen untersucht (Nolen-Hoeksema, Wisco & Lyubomirsky,

2008). In Bezug auf die Arbeitswelt wird in diesem Zusammenhang zunehmend *Irritation* untersucht. Hierbei handelt es sich um „subjektiv wahrgenommene *emotionale* und *kognitive* Beanspruchung im Kontext der Erwerbsarbeit, hervorgerufen durch ein erlebtes Ungleichgewicht zwischen persönlichen Ressourcen und alltäglichen Belastungen“ (Mohr, Rigotti & Müller, 2005, S. 44; vgl. Mohr, Müller & Rigotti, 2005). Ein Item dort lautet beispielsweise: „Ich muss auch zu Hause an Schwierigkeiten bei der Arbeit denken“ (Mohr, Rigotti & Müller, 2005, S. 45). In der Arousal-Uhr sind solche Kognitionen der Verhaltens-Kategorie „mentales Simulieren“ zuzuordnen. Daher kann alleine die anhaltende gedankliche Auseinandersetzung im Zusammenhang mit einem unkontrollierbaren Stressreaktionsprozess über die beschriebenen Prozesse bis hin zu einer schleichenden materiellen Zersetzung des Selbst führen.

2.3.2 Burnout: Wenn die Seele bricht?

Wie fatal fehlgestaltete oder missverstandene Anforderungen bei der Arbeit für die psychische Gesundheit sein können, wird gegenwärtig allen Ortes unter dem Begriff *Burnout* diskutiert. Gert Kaluza (2011) beschreibt den Burnout-Prozess als Fortführung der oben beschriebenen Aktivierungsspirale. Lang andauernde hohe Anforderungen gekoppelt mit einem hohen Engagement führen zu Symptomen der Erschöpfung. Die Betroffenen neigen jedoch nicht zur Regeneration, sondern erhöhen ihr Engagement, um den weiterhin hohen Anforderungen gerecht zu werden. Dieser „Teufelskreis“ (a. a. O., S. 27) bzw. die allostatistische Belastung kulminiert in einem „psychovegetativen Erschöpfungssyndrom“ (a. a. O., S. 25; vgl. auch Gunnar & Queveda, 2007). In der Meta-Analyse von Chida und Steptoe (2008) finden sich dementsprechend Hinweise auf einen vergleichsweise niedrigen Kortisol-Spiegel am Morgen bei vergleichsweise hohen Messwerten für Burnout, Müdigkeit oder Erschöpfung.

So plausibel diese Beschreibung erscheinen mag, so metaphorisch muss sie angesichts mangelnden Konsenses bezüglich Definition, Symptomatik, Ätiologie und differenzialdiagnostischer Abgrenzbarkeit (vgl. Burisch, 2006; Känel, 2008; Korczak & Huber, 2012) zunächst auch bleiben. Korczak und Huber (2012) heben in diesem Zusammenhang ein Schema zur Diagnosestellung eines Burnout-Syndroms von Känel (2008) hervor. Burnout ist gemäß dieses Schemas eine psychosomatische bzw. funktionelle Krankheit mit dem Kardinalsymptom *anhaltende Erschöpfung* und steht in Verbindung mit Stressoren am Arbeitsplatz, Entfremdung und Zynismus gegenüber der Arbeit sowie reduzierter Leistungsfähigkeit (vgl. Känel, 2008, S. 480). Diese Konzeption lehnt stark am wohl gebräuchlichsten Messinstrument zu Burnout an, dem Maslach-Burnout-Inventory (MBI; Maslach, Jackson & Leiter, 1996). Andererseits stellt Burisch (2006) eine ganze Sammlung

unterschiedlichster Definitionsversuche für das Burnout-Syndrom zusammen. Dieser Mangel an Konsens macht verständlich, warum in der ICD-10 bislang nur eine Zusatzdiagnose für das Burnout-Syndrom aufgeführt ist. In Kapitel XXI („Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme von Gesundheitsdiensten führen“), Abschnitt Z73 („Probleme verbunden mit Schwierigkeiten bei der Lebensbewältigung“) findet sich der Schlüssel Z73.0: „Erschöpfungssyndrom (Burn-out-Syndrom)“ (Dilling et al., 2011, S. 415-419). „Damit hat das Burnout-Syndrom zwar *Krankheitswert*, ist aber *keine anerkannte somatische oder psychiatrische Krankheit*“ (Känel, 2008, S. 478).

Eine Definition von Burnout als „Symptomkomplex mit dem Kardinalsymptom Erschöpfung als Reaktion auf eine lange andauernde emotionale und interpersonale Belastung am Arbeitsplatz“ (a. a. O., S. 477) bleibt in nahezu allen Bestandteilen unbestimmt und damit beliebig. Allen Beschäftigten, die in ihrer Symptomatik in erster Linie erschöpft sind und an deren Arbeitsplatz eine wie auch immer geartete andauernde emotionale und interpersonale Belastung festgestellt wird, müsste Burnout attestiert werden. So stellen Korczak und Huber (2012) fest: „Anhand der analysierten Studienergebnisse ist ersichtlich, dass beinahe jedes eingesetzte Burn-out-Instrument ... in jeder Population ein Phänomen valide und verlässlich messen lässt ... vorausgesetzt, das Messinstrument wird der Population entsprechend angepasst und die Ergebnisse werden adäquat statistisch aufbereitet“ (a. a. O., S. 166-169). Trotz aller begrifflichen Schwierigkeiten, die mit der Eingrenzung des Burnout-Syndroms verbunden sind, verdeutlicht die Diskussion dieses Phänomens dessen Relevanz im Sinne der Arbeitssicherheit und des Gesundheitsschutzes. Das zugrunde liegende „komplizierte Wechselspiel zwischen Persönlichkeitseigenschaften und beruflichen Umweltfaktoren“ (a. a. O., S. 478) ist wohl noch in weiten Teilen unverstanden. Ungeachtet dessen sollte die wachsende öffentlich- und fachlich-mediale Präsenz des Burnout-Begriffs bzw. arbeitsbedingter Erschöpfung als Ausdruck schwindender Anpassungsfähigkeit der Beschäftigten an ihre Arbeitsumwelten ernst genommen werden. Dies umso mehr, je klarer das Verständnis des erosiven Effektes unkontrollierbarer Stressreaktionsprozesse auf den Körper wird.

2.3.3 Stress und Gesundheitsbeeinträchtigungen

Das Foltern von Menschen mit sogenannten nichtinvasiven, von Psychologen entwickelten Methoden, wie dies in Abu Ghraib und Bagram geschah (Mausfeld, 2009), ist Körperverletzung. Gleiches gilt nicht weniger für das Aufbauen einer dauerhaften Drohkulisse. Ein Beispiel besteht darin, Menschen dauerhaft den Zugang zum Arbeitsmarkt unter Inkaufnahme der Konsequenzen zu verwehren oder ihnen durch prekäre Arbeitsverhältnisse (z. B. Zeitverträge oder eine konkurrenzorientierte ‚hire and fire‘-Mentalität) damit zu drohen

(Paul & Moser, 2007; Rigotti & Mohr, 2011). Letzteres wird beispielsweise mit dem Burnout-Syndrom oder Depression/depressiver Verstimmung in Verbindung gebracht (a. a. O.). Mit ersterem stehen posttraumatische Belastungsstörungen (PTBS; F43.1, ICD 10; Dilling et al., 2011) in Verbindung. PTBS ist in der Arbeitswelt vor allem in Bezug auf Arbeitsunfälle relevant. Asmundson, Norton, Allerdings, Norton und Larsen (1998) stellten an einer Stichprobe von 139 verunfallten Beschäftigten 34,7 % vollständige und 18,2 % teilweise PTBS-Symptomatik fest. Ähnliche Ergebnisse legten Novara et al. (2009) vor. Neben Unfällen sind auch bewaffnete Raubüberfälle oder verbale oder physische Übergriffe, sowohl für Opfer als auch für Zeugen relevant (MacDonald, Colotla, Flamer & Karlinsky, 2003). Die Wahl des Begriffs „Körperverletzung“ bezieht sich zum einen auf die enge Verknüpfung von Stressreaktionsprozessen und insbesondere des Kortisol-Spiegels mit dem Immun-System. Je anhaltender stressauslösende Situationen sind, umso umfassender und stärker ist das Immun-System beeinträchtigt (Seegerstrom & Miller, 2004). Zum anderen finden sich Hinweise, dass Depression, Abbau des Hippocampus, Bauchverfettung, verringerte Knochendichte und ferner erhöhter Muskeltonus, Magengeschwüre und Diabetes über veränderte Kortisol-Spiegel miteinander in Beziehung stehen (DeRijk & de Kloet, 2006; Sherwood Brown, Varghese, & McEwen, 2004). In die Reihe der aufgeführten nichtinvasiven Körperverletzungen fällt ebenfalls soziale Unterdrückung oder soziale Abwertung. Auch diese sind mit einem hohen Kortisol-Spiegel und entsprechenden gesundheitlichen Beeinträchtigungen verbunden. Dies umso mehr, je unkontrollierbarer solche sozialen Konstellationen sind (Dickerson & Kemeny, 2004). Die erosive Wirkung einer gestörten Kortisol-Dynamik auf den ganzen Körper wurde bereits weiter oben näher erläutert.

Die Symptomatiken der depressiven Episode (vgl. Dilling et al., 2011, Abschnitt F32) sowie des Burnout-Syndroms (Burisch, 2006) entsprechen zu weiten Teilen dem Muster negativer Valenz und geringer Aktivierung in der Arousal-Uhr. Psychosoziale Modelle zur Entstehung depressiver Episoden und des Burnout-Syndroms lassen auf dauerhaft unkontrollierbare Stressreaktionsprozesse bzw. allostatistische Belastung als bedeutende Ursachen schließen (s. Beesdo-Baum & Wittchen, 2011; Burisch, 2006). Allostatistische Belastung und damit die anhaltende Aktivierung des SAM- und des HPA-Systems führen zu einer gestörten Kortisol-Dynamik (vgl. Foy et al., 2006; Huether et al., 1999; Hüther, 2014; Kaluza, 2011). Burke, Davis, Otte und More (2005) stellen in ihrer Metaanalyse ($N = 196$) entsprechend fest, dass Probanden mit klinischer Depression gegenüber nichtdepressiven Probanden wesentlich höhere Kortisol-Konzentrationen während der Erholung von induzierten Stress-Ereignissen ($d = 1,39$; $p < .001$) und am Abend ($d = 0,89$; $p < .0001$) und wesentlich niedrigere Kortisol-

Konzentrationen am Morgen ($d = -1,42; p < .0001$) aufweisen. Chida und Steptoe (2009) fassen in ihrer Metaanalyse zusammen, dass Menschen mit hoher Ermüdung, Erschöpfung oder einem Burnout-Syndrom niedrigere Kortisol-Konzentrationen am Morgen aufweisen. Umgekehrt stellen sie fest, dass eine überhöhte Kortisol-Konzentration am Morgen mit Stress im täglichen Leben und insbesondere bei der Arbeit in positivem Zusammenhang steht. Eine über die Wachperiode hinweg niedrige Ausschüttung von Kortisol steht ebenfalls mit Stress im täglichen Leben in Verbindung.

Posttraumatische Belastungsstörung (PTBS) wird ebenfalls mit der Aktivität des HPA-Systems und dem Kortisol-Spiegel im Tagesverlauf in Verbindung gebracht. Chida und Steptoe (2009) fassen hierzu zusammen, dass posttraumatische Belastungsstörung mit einer geringen Kortisol-Ausschüttung über den Tagesverlauf im Zusammenhang steht. Meewisse, Reitsma, de Vries, Gersons und Olf (2007) stellen fest, dass Plasma- und Serum-Spiegel sowie der gesamte Tagesverlauf (z. B. auch eine Messung am Abend) neben Geschlecht und den konkreten auslösenden Situationen näheren Aufschluss über den Zusammenhang zwischen PTBS und Kortisol-Spiegel geben können. Ein wesentlicher Prozess der PTBS ist die implizite Angst-Konditionierung. Während das explizite Gedächtnis hauptsächlich mit Langzeitpotenzierung im Hippocampus in Verbindung steht, findet die implizite Angst-Konditionierung hauptsächlich in der Amygdala statt (Sapolsky, 2003). Angst-Konditionierung beinhaltet „classical Pavlovian conditioning of automatic responses, procedural memory concerned with nonconscious skills and habits, and reflex pathways“ (a. a. O., S. 1738). Hippocampale Langzeitpotenzierung wird während Notfallsituationen unterdrückt, wohingegen Angst-Konditionierung in der Amygdala in einer solchen Situation gefördert wird. Dennoch werden kontextuelle deklarative, jedoch nicht sehr genaue Gedächtnisinhalte über die Umstände der Notfallsituation gebildet. Dies wird auf eine Wechselwirkung zwischen Hippocampus und Amygdala zurückgeführt. Wie Sapolsky (2003) zusammenfasst, sind auch hier Kortisol-Rezeptoren und erhöhte Kortisol-Konzentrationen maßgebend.

Hüther und Sachsse (2007) schlagen unter Rückgriff auf Befunde zur Entwicklung und Anpassungsfähigkeit des Zentralnervensystems Erklärungen für verschiedene psychische Syndrome („Bindungsstörung, Vertrauensstörung, Störungen der kognitiven Affektkontrolle, dissoziative Störungen“, a. a. O., S. 169) vor. Dabei sind vor allem frühkindliche Bindungserfahrungen und dabei ausgelöste Stressreaktionsprozesse für die Autoren zentral. Eine entsprechende Anpassung erweist sich zumeist unter den auslösenden Lebensbedingungen als funktional. „Sie kann aber auch häufiger, als uns lieb ist, und immer dann, wenn sich die bisherigen Lebensbedingungen nachhaltig ändern, dysfunktional werden“ (a. a. O., S. 172). Die

Autoren fassen zusammen: „Das neurobiologische Zentralorgan menschliches Gehirn ist in seiner Programmierung, Funktion und Fortentwicklung fast ausschließlich ein Sozialorgan“ (a. a. O., S. 179). Unabhängig hiervon fassen Gunnar und Queveda (2007) zusammen: „during human development social relationships play critical roles in regulating physiological stress reactions and protecting the developing brain from potentially deleterious effects of hormones and neurochemicals associated with stress reactions“ (a. a. O., S. 163). Dickerson und Kemeny (2004) stellen in diesem Zusammenhang fest, dass in akuten Laborsituationen die Merkmale *soziale Fremdbewertung* und *unkontrollierbare Anforderungen* gute Vorhersagekraft bezüglich der erhöhten Kortisol-Ausschüttung aufweisen.

In den Beiträgen in Yehuda und Mostofsky (2006) finden sich noch eine Reihe weiterer biopsychologisch begründbarer Verbindungen zwischen Stressreaktionsprozessen und Gesundheitsbeeinträchtigungen. Diese scheinen oftmals vermittelt zu sein über die - ebenfalls zum Teil biopsychologisch begründbare - Regulation von Konsum- und Risikoverhalten. Beiträge zu Essverhalten, Alkoholkonsum und Rauchen sollen beispielhaft herausgegriffen werden: (a) Stressreaktionsprozesse führen sowohl bei gesunden als auch bei Menschen mit Essstörungen zu einem veränderten Essverhalten. Diese Veränderung ist jedoch von Fall zu Fall unterschiedlich. Übergewichtigkeit und Adipositas (graduelle Abstufung der Körpervorfettung) scheinen unter anderem auf vermehrtes Essen während Stressreaktionsprozessen zurückzuführen sein (Dubnov & Berry, 2006). Ein intensives Bedürfnis, ein bestimmtes Nahrungsmittel einzunehmen (*food craving*, Gier nach Nahrungsmitteln), wird mit negativen Gefühlen und Stressreaktionsprozessen (z. B. alltäglicher Stress, negative Selbstbewertung, schädigende Kindheitserfahrungen, wiederholte Bemerkungen über den Körper oder das Essverhalten) in Verbindung gebracht. Letzteres führt bei Menschen mit Bulimia nervosa (F50.2; Dilling et al., 2011) häufig zu erhöhter Nahrungsaufnahme oder Essattacken und geringerem Bewältigungsverhalten (Dubnov & Berry, 2006; Rabinovitz, 2006). Zuverlässige Zusammenhänge zwischen Stress und Anorexia nervosa (F50.0; Dilling et al., 2011) konnten trotz ähnlich lautender Hinweise noch nicht belegt werden. Für Anorexia nervosa kann jedoch ein hyperaktives HPA-System als gesichertes Korrelat angesehen werden (Dubnov & Berry, 2006). (b) Alkohol wird häufig nach Stressreaktionsprozessen oder bei der Antizipation von Stressreaktionsprozessen konsumiert. Bei Menschen, bei denen Alkohol dämpfend auf den Stressreaktionsprozess wirkt, erhöht sich das Risiko zum Alkoholismus (F10.2; Dilling et al., 2011). Sowohl die verstärkenden als auch die beruhigenden Effekte von Alkohol scheinen auf eine Interaktion zwischen Alkohol und dem HPA-System zurückzuführen sein. Alkoholkonsum führt zu einem erhöhten Kortisol-Spiegel

mit allen oben beschriebenen erosiven Konsequenzen, jedoch einer schwindenden Bereitschaft zur funktionalen Stressreaktion. Die Antwortbereitschaft des HPA-Systems ist bei Kindern erhöht, deren Mütter vor der Geburt regelmäßig Alkohol konsumierten (Glavas & Weinberg, 2006). (c) Nikotin hat eine antidepressive Wirkung und kann erholsamen Schlaf (ausreichend REM-Schlaf) induzieren. Außerdem wirkt es sich positiv auf Gedächtnis und Aufmerksamkeit, insbesondere bei Alzheimer-Patienten aus. Rauchen kann z. T. als Selbstmedikation gegen Depression/depressive Verstimmung interpretiert werden. Eine Verwendung puren Nikotins als Medikament scheint aussichtsreich zu sein (Velázquez-Moctezuma & Drucker-Colín, 2006).

Die Autorinnen und Autoren unter (a) bis (c) heben jeweils hervor, dass neben dem Stressreaktionsprozess bzw. der Beteiligung des HPA-Systems und einer gestörten Kortisol-Dynamik weitere Faktoren relevant sind: Genetik, Familiengeschichte, vorgeburtliches Verhalten der Eltern und vor allem der Mutter, Bindungserfahrungen, erschütternde Lebensereignisse sowie die allgemeine soziale Situation seien als Beispiele genannt. Darüber hinaus machen sie jeweils deutlich, dass verschiedene neuromodulatorische Systeme (Noradrenalin, Dopamin, Acetylcholin, Glutamat, GABA, Serotonin, Opioide, weitere Hormone) auf komplexe Weise mit dem HPA-System wechselwirken. Ergänzend zu den bisher dargestellten Wirkungen von Stressreaktionsprozessen auf den Körper und die Leistungen und Funktionen des Gehirns sollte jedoch deutlich geworden sein: Eine unangemessene psychische Belastung am Arbeitsplatz (und selbstverständlich in anderen Settings) kann über die zum einen unspezifischen, zum anderen komplexen Einflüsse auf die Neuromodulation (psychische Beanspruchung) die körperliche und psychische Gesundheit und Leistungsfähigkeit in vielfältiger Weise beeinträchtigen.

2.3.4 Helfen Medikamente bei Depression? - Arousal als System

Um die in der Überschrift aufgeworfene Frage vorweg zu beantworten: Nicht immer, und falls doch, dann überraschend zeitverzögert. Im Übrigen helfen aber z. B. auch mehrfach ungesättigte Fettsäuren (Omega-3-Fettsäuren), manche Kräuter, Nikotin, Placebos und körperliche Aktivität (Geddes et al., 2003; Taylor et al., 2006; Neumann & Frasch, 2005; Puri, 2006; Velázquez-Moctezuma & Drucker-Colín, 2006; Wheatly, 2006; Yehuda, Rabinovitz & Mostofsky, 2006).

Die gegenwärtige Praxis der psychopharmakologischen Behandlung von Depression geht auf die so genannte Katecholamin-Hypothese zurück. Diese besagt, dass Depression mit einem Mangel an Noradrenalin einhergeht (vgl. Schildkraut, 1965). Diese Hypothese wurde später zur Monoaminmangelhypothese erweitert, da auch ein Mangel an Serotonin mit der depressiven

Episode im Zusammenhang zu stehen scheint (vgl. Coppen, 1967; Holsboer-Trachsler & Holsboer, 2008; Köhler, 2005, 2011). Tri- und tetrazyklische Antidepressiva sollen genauso, wie selektive Noradrenalin-Wiederaufnahmehemmer (selective noradrenaline reuptake inhibitor, SNRI), selektive Serotonin-Wiederaufnahmehemmer (SSRI) und Serotonin-Noradrenalin-Wiederaufnahmehemmer (SSNRI) die präsynaptische Wiederaufnahme von Noradrenalin und/oder Serotonin hemmen. Allerdings beeinflussen tri- und tetrazyklische Antidepressiva auch bzw. überwiegend (je nach Substanz) andere Neurotransmittersysteme und sind häufig mit vergleichsweise starken Nebenwirkungen verbunden. Die selektiven Wiederaufnahmehemmer sind verträglicher. Neuere Entwicklungen blockieren gezielt bestimmte Noradrenalin- oder Serotonin-Rezeptoren. Monoaminoxidasehemmer (MAOI, neuerdings MAO-A-I)⁹ sollen den Abbau der Monoamine im synaptischen Spalt verhindern und so deren dortige Präsenz erhöhen (Holsboer-Trachsler & Holsboer, 2008; Köhler, 2005, 2011). Wirkmechanismen der Antidepressiva sind Köhlers Ansicht nach häufig unklar und hypothetisch (Köhler, 2005, 2011). Er spricht dem Ansatz, Monoaminmangel mit Monoaminbereitstellung zu begegnen, eine „heuristische Fruchtbarkeit“ (Köhler, 2011, S. 242) zu. Seit dem Beitrag von Schildkraut (1965) scheint sich hier wenig geändert zu haben: „Although the hypothesis may not be directly testable by currently available experimental methods, this formulation is nonetheless of considerable heuristic value“ (a. a. O., S. 509). Holsboer-Trachsler und Holsboer (2008) ergänzen: „Es muss auch hinterfragt werden, ob sich aus dem Wirkmechanismus eines Medikaments immer auf den pathologischen Wirkmechanismus einer Erkrankung schließen lässt“ (a. a. O., S. 518). Erschwerend kommt hinzu, dass immer weitere biochemische Substanzen und Prozesse aufkommen, die mit Depression in Verbindung stehen. Die Bedeutung von Stressreaktionsprozessen und damit zusammenhängenden Substanzen für Depression, Panikstörung und posttraumatische Belastungsstörung hebt beispielsweise Ströhle (2003; vgl. Holsboer-Trachsler & Holsboer, 2008) hervor. Zu ihrer Behandlung beschreibt er unter anderem Möglichkeiten, auf neuronale Kortisol-Rezeptoren oder auf die Ausschüttung von Kortisol Einfluss zu nehmen. Er nimmt einen Ausschnitt der Aussagen im Review von Krishnan und Nestler (2008) vorweg. Diese beiden Autoren führen in Bezug auf Depressionen Evidenzen für weitere relevante biochemische Substanzen wie spezifische Transkriptions- und Wachstumsfaktoren und Zytokine (Proteine zur Regulation von Zellwachstum und Zelldifferenzierung) bis hin zu einzelnen Genen an (vgl. Holsboer-Trachsler & Holsboer, 2008; Kupfer, Frank, & Phillips,

⁹ Monoaminoxidase (MAO) ist ein Enzym, das in zwei Unterformen, A und B, vorkommt. MAO-A baut vornehmlich Adrenalin, Noradrenalin und Serotonin ab, während MAO-B Dopamin und Tyramin abbaut (Köhler, 2011).

2012). Auch sie berichten, dass diese neuen Erkenntnisse zur Entwicklung neuer Medikamente für Depression herangezogen werden. Darüber hinaus berichten sie über relevante epigenetische Phänomene. Dabei handelt es sich um Umwelterfahrungen die Gen-Transkriptionen und Gen-Funktionen ändern können, ohne zu Grunde liegende DNA-Sequenzen zu ändern. Als Beispiel führen sie das Ausmaß der Fürsorge durch die Mutter an. Solche Phänomene können Einfluss auf die Widerstandskraft (*resilience*) bzw. Verletzlichkeit (*vulnerability*) nehmen (vgl. Hüther, 2014; Hüther & Sachsse, 2007). Sie schlussfolgern “Collectively, these studies highlight the weaknesses of attempts to generate a ‘unified theory’ of depression” (Krishnan & Nestler, 2008, S. 898) und appellieren für eine interdisziplinäre Forschung in Bezug auf Depression.

Wie wirksam ist die psychopharmakologische Behandlung von Depressionen? Die Antwort hängt von der individuellen Ausformung (z. B. Gehemmtheit, Agitiertheit, Ängstlichkeit, Komorbidität) und den individuellen Entstehungsbedingungen dieser Störung ab (Holsboer-Trachsler & Holsboer, 2008). Rush et al. (2006) geben eine durchschnittliche Rate von 67 % ($N = 3\ 671$) mindestens einjähriger Symptomfreiheit bei Kombination von Psychotherapie und der Gabe von Antidepressiva an. D. h. 33 % der behandelten Patienten entwickeln während eines einjährigen Follow-Ups nach Remission erneut depressive Symptome. Zu ähnlichen Ergebnissen kommen Geddes et al. (2003) bezüglich der Rezidiv-Prophylaxe mit Antidepressiva nach Akutbehandlung. Bei der Gabe selektiver Serotonin-Wiederaufnahme-Hemmer beobachtet man bei wöchentlicher Messung und im Vergleich zu Placebos eine zunehmende Besserung der Symptomatik (Taylor, Freemantle, Geddes, & Bhagwager, 2006). Ein solcher verzögerter und nichtlinearer Verlauf wirft insofern noch offene Fragen auf, weil der Wirkmechanismus dieser Medikamente an den Synapsen nahezu umgehend nach Einnahme greift (a. a. O.). Die Wirksamkeit von Antidepressiva beschränkt sich jedoch nicht nur in beschriebenem Ausmaß auf affektive Störungen, sondern erstreckt sich andererseits auch auf Zwangsstörungen, Angststörungen, Essstörungen und Aufmerksamkeitsdefizit mit Hyperaktivitätssyndrom (Holsboer-Trachsler & Holsboer, 2008; Köhler, 2005, 2011).

Eine ähnliche Komplexität findet sich wohl bei der Behandlung mit Neuroleptika (Blockade von Dopamin- und Noradrenalinrezeptoren), Anxiolytika, Sedativa und Hypnotika (z. B. Erhöhung der Bindungsbereitschaft von GABA-Rezeptoren, Eingriff in das Serotonin-System, Blockade von Adrenalin- und Noradrenalinrezeptoren [β -Blocker]) sowie Antidementiva bzw. Nootropika (z. B. Hemmung des Abbaus von Acetylcholin [Acetylcholinesterasehemmer], Erhöhung der Dopamin-Konzentration) (Köhler, 2005, 2011). Dennoch ist die medikamentöse Behandlung eine von vielen wirksamen Möglichkeiten, eine gestörte Arousal-Dynamik und

Neuromodulation wieder zu normalisieren. In einigen Fällen (z. B. Manie, Schizophrenie) kann sie eine die Psychotherapie erst ermöglichende oder gar die einzige Möglichkeit der Behandlung sein (a. a. O.).

Betrachtet man noch einmal die komplexen Wechselwirkungen der Arousal-Systeme in Abbildung 1, so wird deutlich, dass eine „vereinheitlichte Theorie von Depression“, sollte es sie denn überhaupt geben, heuristisch-systemisch sein müsste. Die Elemente der Arousal-Systeme weisen essentielle sowohl anregende als auch hemmende Rückkopplungen auf, die zu einem nichtlinearen Antwortverhalten des Gesamtsystems führen. Letzteres muss schon allein auf Grund der Regel von Michaelis und Menten (1913) bzw. des Massenwirkungsgesetzes (Eckert & Müller, 2008) angenommen werden. Dieser Regel zufolge nimmt die Häufigkeit der Bindung von Neurotransmittern an ihre Rezeptoren konkav-monoton mit der Konzentration der Neurotransmitter zu. Sie nähert sich einer oberen Grenze, die durch die Anzahl bindungsfähiger Rezeptoren bestimmt ist. Der Arousal-Dynamik liegt somit ein nichtlineares dynamisches System zu Grunde (Strunk & Schiepek, 2006, S. 129; vgl. Duffy, 1962). Bei einer systemischen Betrachtung sind einfache Wenn-Dann-Beziehungen (Kausalbeziehungen) nur auf einer emergenten Ebene, z. B. in Begriffen der Psychologie, beschreibbar. Das hier betrachtete Beispiel könnte wie folgt mit emergenten Begriffen übersetzt werden: Wenn ein Mensch sehr lange unter unkontrollierbaren Stressreaktionen leidet, nimmt er psychischen und körperlichen Schaden (z. B. kognitive Beeinträchtigungen, affektive Störungen, Schlafstörungen, Störungen im Stoffwechsel und in den Immun-Funktionen). Wenn ein Mensch daraufhin mit Antidepressiva behandelt wird oder z. B. während einer Psychotherapie Möglichkeiten entwickelt, Stressoren zu bewältigen, kann sich sein Gesundheitszustand normalisieren. Die zugrunde liegenden Mechanismen auf der neurologischen, biochemischen Ebene verlaufen zwar deterministisch, sind jedoch im Einzelnen nicht vorhersagbar. Die auf emergenter Ebene beobachtbaren Kausalbeziehungen beruhen vielmehr darauf, dass sich das Gesamtsystem bei einer systemischen Störung (z. B. Gabe von Medikamenten, Vermitteln neuer Erfahrungen) in einem neuen Zustand einfindet. Davon sind jedoch alle Elemente des Gesamtsystems betroffen, auch wenn man nur einen einzigen Parameter (z. B. Serotonin-Konzentration, Präsenz von Stressoren) heuristisch manipuliert (vgl. Strunk & Schiepek, 2006). Eine solche Betrachtung macht die vielfältige, heterogene, z. T. eingeschränkte und verzögerte Wirkung von Antidepressiva zumindest anschaulich verständlich. Berücksichtigt man zusätzlich die in die Arousal-Dynamik eingreifenden sensorischen und somatischen Eingaben, so wird auch verständlich, warum Rückfälle innerhalb eines Jahres in 33 % der Fälle zu erwarten sind. Konnten die Bewertungs- und Bewältigungsmuster während der Therapie nicht angemessen

und funktional angepasst werden oder konnten die auslösenden Bedingungen nicht beseitigt werden, so ist mit einer Rückstellung des Gesamtsystems in den alten, stabilen, jedoch dysfunktionalen Zustand zu rechnen. Ein systemisches Verständnis des Gesamtorganismus liegt sicher auch den Gedanken von LeDoux (2012) zu Grunde, wenn er sich dafür ausspricht, *global organismic states* zu betrachten.

2.3.5 Verhaltensbezogene Förderung einer gesunden Arousal-Dynamik

Kontrollierbare Stressreaktionsprozesse (Huether et al., 1999; Hüther, 2014) und aktives bzw. anstrengendes Bewältigen von Anforderungen (Hockey, 1997) sind wesentliche Bestandteile der gesunden Auseinandersetzung eines Menschen mit seiner ökologischen und sozialen Umwelt. Auch unkontrollierbare Stressreaktionsprozesse können unter sich ändernden Bedingungen funktional sein (Huether et al., 1999; Hüther, 2014). Im Hinblick auf die körperliche und psychische Gesundheit sollten jedoch aversive Umstände vermieden werden, in denen der Mensch lernt, dass alle seine Bewältigungsversuche erfolglos bleiben. Solche Umstände führen zu Passivität („Aufgeben“), verminderter Anpassungsfähigkeit (das Erlernen instrumentellen Verhaltens ist erschwert) und körperlichen sowie psychischen Beeinträchtigungen (z. B. stressbedingte Erosion des Körpers, Depression). Dieses Reaktionsmuster auf anhaltend unkontrollierbare Umstände nennt Martin E. P. Seligman (1972) *erlernte Hilflosigkeit*. Er schlägt als Prävention und Therapie für erlernte Hilflosigkeit und Depression kontrollierbare Stressreaktionsprozesse vor: „Finally, successful psychotherapy in depression, like therapy for learned helplessness, may involve having patient come to believe that he can be effective in controlling the events that are important to him“ (a. O., S. 411).

Im Falle stressinduzierter gesundheitlicher Beeinträchtigungen lässt sich eine gesunde Arousal-Dynamik unter anderem über das Erlernen angemessener, funktionaler Bewertungs- und Bewältigungsmuster (Aufbau von Ressourcen) und die Festigung eines insgesamt gesunden Lebensstils bewirken. Diese optimistische Sichtweise leitet sich zum einen aus der Beobachtung ab, dass ein großer Teil nunmehr unangemessener Bewertungs- und Bewältigungsmuster ursprünglich als funktionale Anpassung erlernt wurde (Burisch, 2006; Foy et al., 2006; Gunnar & Quevedo, 2007; Huether et al., 1999; Hüther, 2014). Auch extreme Fälle erlernter Hilflosigkeit wurden - erlernt (Seligman, 1972). Optimistisch stimmt weiter die hohe Anpassungsfähigkeit aller beteiligten Elemente der Arousal-Systeme an neue ökologische und soziale Bedingungen (DeRijk & de Kloet, 2006; Foy et al., 2006; Gunnar & Quevedo, 2007; Huether et al., 1999; Hüther, 2014; Jones, 2003; Krishnan & Nestler, 2008; LeDoux, 2012; Sapolsky, 2003; Yehuda & Mostofsky, 2006). Das Erlernen neuer Bewertungs- und

Bewältigungsmuster erfordert sicher zunächst Anstrengung (Hockey, 1997). Es sollte aber auf lange Sicht auch zum Aufbau angemessener Kompetenzen in den Bereichen “Wachen”, „Bewältigen“, “Aufgeben”, “Konsumieren”, “mentales Simulieren/Entspannen” und “Schlafen” (siehe Arousal-Uhr) führen. Diese Kompetenzen sollten auch für die Zukunft eine funktionale Anpassung an die sich ändernde ökologische und soziale Umwelt erlauben.

Wesentlich für das zukünftig erfolgreiche “Bewältigen” ist das Erarbeiten, Planen und Üben neuer Bewertungs- und Bewältigungsmuster in kontrollierbaren Stressreaktionsprozessen mit verstärkendem Ausgang. Hierbei können je nach Problemstellung sachkundige Berater oder Therapeuten auf vielfältige Weise unterstützend begleiten (Faltermaier, 2005; Gunnar & Queveda, 2007; Huether et al., 1999; Renneberg & Hammelstein, 2006; Renneberg, Heidenreich & Noyon, 2009; Schwarzer, 2004; Seligman, 1972; Warschburger, 2009; Wilken, 2006; Wittchen & Hoyer, 2011). “Wachen” bezieht sich auf das funktionale Erkennen und Bewerten persönlich relevanter Herausforderungen und Möglichkeiten. Hiermit steht das “Aufgeben” als das rechtzeitige Erkennen persönlicher Leistungsgrenzen und das rechtzeitige Beenden nicht zielführenden Engagements in Verbindung. Hierfür haben sich im therapeutischen Kontext die kognitiven Verfahren und Selbstsicherheits-Trainings als wesentliche Bestandteile der Verhaltenstherapie etabliert (Helbig-Lang & Klose, 2011; Mühlig & Poldrack, 2011; Wilken, 2006). “Konsumieren” bezieht sich sowohl auf das Was (z. B. Ernährung, soziale Beziehungen, Kultur, Bildung, Genussmittel) als auch auf das Wie (instrumentelles Verhalten im Alltag, in der Freizeit und während der Arbeit). Auch für ihre gesundheitsförderliche Beeinflussung kommen Verfahren der Verhaltenstherapie in Frage (Bittner & Helbig-Lang, 2011; Helbig-Lang & Klose, 2011; Junge-Hoffmeister, 2011; Mühlig & Jacobi, 2011). Darüber hinaus finden sich spezifische primärpräventive Programme im Feld der Gesundheitspsychologie (Faltermaier, 2005; Renneberg & Hammelstein, 2006; Schwarzer, 2004), die auch am Setting Arbeitsplatz wirksam sind (Goldgruber & Ahrens, 2009). “Entspannen” setzt wahrgenommene Kontrollierbarkeit bestehender Herausforderungen und befriedigendes Konsumieren voraus und umfasst das Einstellen von Rumination und sorgenvollen Gedanken (Jones, 2003; Nolen-Hoeksema et al., 2008). Es dient unter anderem der Vorbereitung eines erholsamen Schlafs (Jones, 2003). Autogenes Training und progressive Muskelentspannung haben sich als wirksam unterstützende Verfahren etabliert (Ruhl, Hach & Wittchen, 2011).

Das Gehirn verbraucht im Wachzustand trotz seiner geringen Masse (etwa 2 % der Körpermasse, Brown et al., 2012) einen Großteil des gesamtorganismischen Energieumsatzes (etwa 20 %; a. a. O.). Restrukturierungen synaptischer Verbindungen während dem “Schlafen”

optimieren synaptischen Energie- und Raumverbrauch, sichern synaptische Plastizität und konsolidieren deklarative und prozedurale Gedächtnisinhalte. Durch das “Schlafen” werden demnach Erfahrungen, z. B. auf Grund neuer Bewertungs- und Bewältigungsmuster, gefestigt und effizienter verfügbar gemacht. Außerdem wird die Lernfähigkeit gesichert. Ein natürliches Verhältnis von SW- und REM-Schlaf ist hierfür Voraussetzung (Born & Plihal, 2000; Diekelmann & Born, 2010; Rasch & Born, 2013; Tononi & Cirelli, 2006). Auch das spezifische Immun-Gedächtnis wird während des Schlafens konsolidiert (Bildung spezifischer T-Helfer-Zellen) und für den zukünftigen Gebrauch optimiert (Rasch & Born, 2013). Darüber hinaus ist der Wiederaufbau von Energiereserven (in Form von Adenosintriphosphat [ATP]) eine wichtige homöostatische Funktion des Schlafs für das Gehirn (Brown et al., 2012). Schlaflosigkeit und -unterbrechung ist dagegen mit einer Reihe gesundheitlicher und kognitiver Defizite verbunden (a. a. O.). Zur Förderung eines ausgeglichenen, natürlichen Schlafs können o. g. Entspannungsverfahren angewandt werden. Diese können jedoch auch in umfassendere verhaltenstherapeutisch orientierte Trainings eingebunden werden (Müller & Paterok, 2010).

Als Beispiel einer primärpräventiven verhaltensbezogenen Intervention, die vor allem am Setting Arbeitsplatz zum Einsatz kommt, wird das Stressbewältigungstraining von Gert Kaluza (2011, 2012) vorgestellt. Das Ziel des Trainings formuliert Kaluza wie folgt: “Insgesamt sollen ein erweitertes Bewältigungsrepertoire sowie eine möglichst stressfreie Bewertung von Anforderungen und eigenen Kompetenzen die Teilnehmer zu einem flexiblen, situationsadäquaten Umgang mit alltäglichen Belastungen befähigen” (Kaluza, 2011, S. 60). Das Programm wird in Gruppen umgesetzt und umfasst 12 wöchentliche Sitzungen zu je 120 Minuten. Es ist nicht auf bestimmte Zielgruppen ausgelegt, sollte jedoch inhaltlich auf die spezifischen Bedürfnisse der jeweiligen Gruppe zugeschnitten werden. Theoretisch setzte es weitgehend auf den oben dargelegten Konzepten zu Stress auf. Erfolgreiches Stressmanagement unterteilt Kaluza (2011) in instrumentelles, mentales und regeneratives Stressmanagement. Er schlägt vor, folgende vier Trainingsmodule bei jeder Umsetzung durchzuführen (vgl. a. a. O., S. 62, Abb. 4.1 und S. 178-179, Tabelle 12.1): (a) Progressive Muskelentspannung (Entspannungstraining), (b) kognitive Umstrukturierung fälschlich vermuteter und z. T. stark internalisierter Erwartungen Dritter, die stressverstärkend wirken (Mentaltraining), (c) Aufbau von Fähigkeiten, Fertigkeiten und Strategien für erfolgreiches Bewältigen zukünftiger Herausforderungen (Problemlösetraining) und (d) Aufbau instrumentellen, verstärkenden Verhaltens im Alltag und der Freizeit (Genusstraining). Darüber hinaus gibt Kaluza (2011) Anleitungen zu Ergänzungsmodulen. Diese können je nach Bedarf der Gruppe optional und austauschbar eingesetzt werden. Die weitere Sensibilisierung für

persönlich relevante Herausforderungen und Möglichkeiten wird mit dem Modul Zukunftsvision vorgenommen. Weitere Strategien für das Bewältigen zukünftiger Herausforderungen sowie das Haushalten mit den eigenen Leistungsreserven werden in den Modulen Zeitplanung und Akutfallstrategie vermittelt. Spezielle Module zum Aufbau instrumentellen, verstärkenden Verhaltens heißen Sport und Bewegung und Soziales Netz. Zusätzlich gibt es die optionalen Exkurse Innere Achtsamkeit, Erholsamer Urlaub und Erholsamer Schlaf.

Dieses Training wurde unter anderem deshalb vorgestellt, weil es im Ganzen betrachtet mit sämtlichen Verhaltens-Kategorien der Arousal-Uhr korrespondiert. Einzelne Module lassen sich sicher auch eindeutig einer bestimmten Verhaltens-Kategorie der Arousal-Uhr zuordnen. Da dies jedoch nicht zuletzt wegen der zu Grunde liegenden Vernetzung und Dynamik der Arousal-Systeme willkürlich wäre, sei es den Leserinnen und Lesern überlassen, eine eigene Zuordnung zu versuchen. Das Stressbewältigungstraining von Gert Kaluza (2011, 2012) zeigt jedoch: (a) Die Arousal-Uhr deckt alle von diesem Autor als relevant betrachteten Verhaltens-Kategorien ab und (b) es ist möglich, Programme zur Verhaltensprävention zu entwickeln, die sämtliche für die Arousal-Dynamik relevanten Verhaltens-Kategorien berücksichtigen.

Als weitere Möglichkeit zur Förderung einer gesunden Arousal-Dynamik wäre die gesundheitsförderliche Gestaltung des Settings Arbeitsplatz (verhältnisbezogene Maßnahmen) zu nennen. Auch dabei kommt es darauf an, die Verhaltens-Kategorien der Arousal-Uhr zu berücksichtigen. Hierauf wird später zurückzukommen sein.

2.4 KOGNITION, EMOTION UND DIE MODULARE ARCHITEKTUR DES GEHIRNS

Bislang wurde dargestellt, dass sich psychische Beanspruchung als Inanspruchnahme der Überlebensschaltkreise bzw. Arousal-Systeme widerspiegelt. Die mit diesen Systemen zusammenhängenden Erlebensqualitäten (z. B. Core Affect) und Verhaltensweisen (z. B. Annäherung und Vermeidung, Aktivität und Ruhe) sind für die Interaktion eines Individuums mit seiner Umwelt überlebenswichtig, da sie dessen Homöostase sichern. Um jedoch auch langfristig planend und auf Erfahrung zurückgreifend mit der Umwelt interagieren zu können, bedarf es einer übergeordneten Kontrolle und Steuerung der beschriebenen Prozesse. Diese Funktionen übernimmt beim Menschen wie bei allen Säugetieren der Kortex. Ähnlich, wie die Arousal-Systeme, ist auch der Kortex als modulares Netzwerk bzw. nichtlineares dynamisches System aufzufassen. Überdies sind Kortex und Arousal-Systeme eng funktional und strukturell miteinander gekoppelt. Dies wurde bereits in Abbildung 1 angedeutet. So geben einerseits die Arousal-Systeme Informationen über den körperlichen Zustand und eine grundlegende

affektive Bewertung der gegenwärtigen Situation an den Kortex weiter. Andererseits ist der Kortex in der Lage, mittels kognitiver, integrativer Operationen die Reaktionen auf Arousal-Signale zu modulieren. Grundlage der folgenden Ausführungen sind Befunde über weiträumig verteilte kortikale Netzwerke: Dem Central Executive Network (zuständig für Problemlösung), dem Default Mode Network (zuständig für mentale Simulationen bezüglich der persönlichen sozialen Rolle in Vergangenheit, Gegenwart und Zukunft) sowie dem Salient Network (zuständig für das Entdecken von Konflikten, Handlungsfehlern oder Bedrohungen). Menon (2011) integriert diese drei Netzwerke im Tripel-Network-Modell. Im letzten Abschnitt dieses Kapitels wird dieses Modell um die Kopplung mit den Arousal-Systemen erweitert. Zuvor seien jedoch die drei beteiligten Netzwerke sowie ihre Funktionen (erster Abschnitt), auch im Zusammenhang mit psychischer Gesundheit (zweiter Abschnitt), im Einzelnen betrachtet.

2.4.1 Drei weiträumig verteilte kortikale Netzwerke

Nach gegenwärtigem Kenntnisstand kann der Kortex in weiträumig verteilte, untereinander weitgehend separierte funktionale Netzwerke eingeteilt werden. Um solche Netzwerke zu bestimmen und voneinander abzugrenzen, werden bildgebende Verfahren (zumeist funktionelle Magnet-Resonanz-Tomografie, fMRT) eingesetzt. Die Probanden werden gebeten, sich möglichst nicht zu bewegen und an möglichst nichts Bestimmtes zu denken, ohne jedoch einzuschlafen (resting state-fMRT, rs-fMRT). Räumlich getrennte Kortex-Areale, die zu demselben Netzwerk gehören, weisen unter diesen Umständen eine positive Korrelation ihrer Aktivitäten über die Zeit hinweg auf und sind über weiße Substanz (lange, myelinisierte Axonenbündel) miteinander verbunden (Bressler & Menon, 2010; Menon, 2011; van den Heuvel & Hulshoff Pol, 2010). Menon (2011) nennt 14 unterschiedliche funktionale Netzwerke, weist jedoch auch darauf hin, dass die Unterteilung von der zeitlichen und räumlichen Auflösung der Apparatur sowie von den verwendeten Verfahren (Vorannahmen und Auswertung) abhängig ist (vgl. Bressler & Menon, 2010; van den Heuvel & Hulshoff Pol, 2010). Die unterschiedlichen Unterteilungen konvergieren jedoch in dem Sinne, dass die verschiedenen funktionalen Netzwerke hierarchisch organisiert zu sein scheinen. Daher werden bei rs-fMRT globale kortikale Aktivitäten subtrahiert, die mit viszeralen Aktivitäten wie Atmung oder mit motorischen Aktivitäten wie Bewegung einhergehen (a. a. O.).

In Bezug auf kognitive Funktionen zeichnet sich ein Konsens ab, bei dem drei funktionale kortikale Netzwerke unterschieden werden: Das Central Executive Network (CEN), das Default Mode Network (DMN) und das Salient Network (SN) (Bressler & Menon, 2010; Menon, 2011; Vincent, Kahn, Snyder, Raichle & Buckner, 2008). Das CEN ist über Teile des parietalen sensorischen, des zentralen motorischen, des medialen Temporalkortex sowie des

Präfrontalkortex verteilt. Es interagiert über Thalamus, Hypothalamus, Hippocampus und die Amygdala mit den Arousal-Systemen (Miller & Cohen, 2001; Miller, Freedman & Wallis, 2002; Vincent et al., 2008). Als zentrales Teilnetzwerk (so genannter Hub) des CEN erweist sich der dorsolaterale Präfrontalkortex (Miller & Cohen, 2001; Miller et al., 2002). Seine Aktivität ist mit der Bewältigung von Anforderungen verbunden, die eine Bindung der Aufmerksamkeit an externale Stimuli voraussetzen. Seine Funktionen bestehen in der Koordination und Aufrechterhaltung instrumenteller, zielbezogener Interaktion mit der Umwelt. Er enkodiert besonders schnell Verhaltens-Kontingenzen, Regeln über erfolversprechende Verhaltensweisen in Abhängigkeit der gegebenen Situation sowie Kategorisierungs-Regeln. Er ist darüber hinaus in der Lage, mit der Erwartung von Belohnung verbundene Ziele auch über lange Zeiträume und komplexe Tätigkeiten hinweg aufrecht zu erhalten und gegenüber konkurrierenden Zielen zu priorisieren. Durch seine Aktivität werden automatisierte Verhaltensweisen überformt (a. a. O.). Die Kontrolle des Verhaltens über den Präfrontalkortex erfordert jedoch zusätzlichen Energieaufwand und wird vermieden, wenn automatisierte Verhaltensweisen ebenfalls zur Zielerreichung dienlich sind (Miller et al., 2002).

Das CEN steht dem Konzept der Central Executive von Baddeley (2012) nahe. Allerdings werden episodische Kognitionen hauptsächlich im Default Mode Network (DMN) verarbeitet (Andrews-Hanna, 2012; Buckner et al., 2008; Spreng, Mar & Kim, 2008). Das DMN erhielt seinen Namen vor dem Hintergrund seiner Entdeckung. Um eine Grundlinie der Hirnaktivitäten zu erhalten, vor dessen Hintergrund sich anforderungsbezogene Hirnaktivitäten besser abgrenzen lassen, wurde erstmals rs-fMRT eingesetzt. Ziel war es, jene Aktivitäten des Kortex zu messen, die bei der Bewältigung konkreter kognitiver Anforderungen gehemmt werden (Raichle et al., 2001). Es stellt sich heraus, dass sich das DMN räumlich vom CEN abgrenzen lässt (Vincent et al., 2008). Mehr noch: Die Aktivitäten von CEN und DMN sind antikorreliert. D. h. je aktiver das CEN desto weniger aktiv das DMN und umgekehrt. Im Ruhezustand fluktuiert die Aktivität dieser beiden Netzwerke gegenläufig (Fox et al., 2005). Mediale Bestandteile des DMN sind der mediale Präfrontalkortex, der mediale parietale Kortex, der posteriore cinguläre Kortex sowie der retrospleniale Kortex. Ggf. ist auch der Precuneus zum DMN zu zählen. Weitere, nicht zum CEN gehörende Areale des DMN finden sich im parietalen, temporalen und frontalen Kortex. Über den medialen temporalen Kortex besteht eine Verbindung zum Hippocampus und umliegenden Kortex-Arealen (medial temporal lobe [MTL]-Teilnetzwerk). Als Hubs (zentrale integrative Teilnetzwerke) stellen sich der mediale Präfrontalkortex sowie der posteriore cinguläre Kortex heraus. Insbesondere der dorsale Anteil des medialen Präfrontalkortex ist hier hervorzuheben (dMPFC-Teilnetzwerk) (Andrews-

Hanna, 2012; Bressler & Menon, 2010; Buckner et al., 2008; Menon, 2011; Raichle et al., 2001; Spreng et al., 2008; Vincent et al., 2008). MTL-Teilnetzwerk, dMPFC-Teilnetzwerk und Hubs sind untereinander vernetzt, weisen jedoch inhaltliche Unterschiede bei Kognitionen auf. Das MTL-Teilnetzwerk ist mit episodischem und autobiografischem Gedächtnis, episodischen Gedanken über die Zukunft, Erinnern kontextueller Assoziationen, konzeptueller und semantischer Verarbeitung, (u. a. bildlicher) Vorstellung/Imagination sowie räumlicher Navigation verbunden. Das dMPFC-Teilnetzwerk wird mit Metakognition (eigene und die Anderer; Theory of Mind), moralischen Entscheidungen, dem Verständnis und der Begründung sozialer Vorgänge sowie konzeptueller Verarbeitung in Verbindung gebracht. Die DMN-Hubs sind zu integrativer Verarbeitung in der Lage und stehen mit selbstbezogener Kognition und Selbstreflexion, Metakognition (eigene und die ähnlicher/nahestehender Anderer), autobiografischem Gedächtnis, episodischen Gedanken über die Zukunft, moralischen Entscheidungen sowie der Repräsentation und Vorwegnahme von Konsequenzen hinsichtlich ihrer Valenz in Verbindung (Andrews-Hanna, 2012; Buckner et al., 2008; Spreng et al., 2008). In Verbindung mit der oben geschilderten Fluktuation zwischen CEN und DMN (Fox et al., 2005) spricht vieles dafür, dass ein Mensch im Ruhezustand vergangene Episoden, Gegenwart und Zukunftserwartungen erinnert bzw. simuliert und dabei nach möglichen instrumentellen Verhaltensweisen sucht. Mit solchen selbstbezogenen und sozialen mentalen Operationen ist der Mensch auch während der Bewältigung externaler Anforderungen beschäftigt (Andrews-Hanna, 2012; Buckner et al., 2008; Spreng et al., 2008). Man schätzt, dass nur 5 % des Metabolismus im Gehirn zu motorischem, instrumentellem Verhalten führt (van den Heuvel & Hulshoff Pol, 2010). Der ökologische Nutzen dieses mentalen Energieaufwands kann sicher darin gesehen werden, dass nicht zielführendes motorisches Verhalten durch Simulation schon vor der Ausführung ausgeschlossen und auf diese Weise mit Energiereserven hausgehalten werden kann.

Auch, wenn sich die Aktivität von CEN und DMN nicht prinzipiell ausschließen (Andrews-Hanna, 2012; Buckner et al., 2008; Fox et al., 2005) stellt sich doch die Frage, wie es zur verstärkten Aktivierung des CEN bei herausfordernden externalen Anforderungen kommt (Miller & Cohen, 2001; Miller et al., 2002). Das Salient Network (SN) scheint für die Entdeckung solcher Anforderungen zentral zu sein und für die Umschaltung von DMN auf CEN zu sorgen (Bressler & Menon, 2010; Menon, 2011; Menon & Uddin, 2010). Das SN ist topologisch zwischen CEN und DMN lokalisiert. Seine Hauptbestandteile sind die anteriore Insula sowie der anteriore cinguläre Kortex. Diese Areale stehen über posterioren cingulären Kortex, Thalamus, limbischem System und Hirnstamm in enger Verbindung mit den Arousal-

Systemen (a. a. O.; Critchley, 2009; Vincent et al., 2008). Die Funktion des SN besteht in der Zuteilung von Aufmerksamkeit und begrenzten kortikalen Ressourcen und Funktionen zu jenen externalen und internalen Stimuli, die in der gegenwärtigen Situation besonders bedeutsam für die Aufrechterhaltung der Homöostase sind. Hierzu integriert das SN viszerale und sensumotorische Signale und greift auf die Dimensionen der Core Affects der Arousal-Systeme (Aktivität und Valenz) zurück (a. a. O.; Berntson et al., 2011; Critchley, 2009; Ochsner & Gross, 2006). Der anteriore cinguläre Kortex kann als Konflikt-Monitor betrachtet werden. Er wird aktiviert, wenn konflikthafte Anforderungen bestehen. Solche Anforderungen bestehen, wenn in der gegenwärtigen Situation bislang zielführendes, automatisiertes Verhalten nicht mehr zum erwünschten Ziel führt oder verschiedene zielführende Verhaltensweisen in Konkurrenz stehen. Bei solchen Anforderungen wird der dorsolaterale Präfrontalkortex (graduell) hinzugeschaltet, um den bestehenden Konflikt aufzulösen. Darüber hinaus ist der anteriore cinguläre Kortex über seine Verschaltungen im limbischen System und mit dem Hirnstamm in der Lage, eine schnelle motorische Korrektur von Fehlverhalten zu unterstützen (Botvinick, Cohen & Carter, 2004; Bush, Luu & Posner, 2000; Critchley, 2009; Kerns et al., 2004; Carter & van Veen, 2007). Die Aktivität des anterioren cingulären Kortex bei konflikthafter Anforderungen geht mit erhöhtem sympathischen Arousal (kardiovaskuläre und elektrodermale Aktivität, Pupillenerweiterung) einher. Darüber hinaus besteht eine Interaktion zwischen anteriorem cingulärem Kortex und autonomem Arousal bei der Verarbeitung emotionaler Reize (z. B. Fotos von Menschen, die Freude, Trauer, Wut oder Ekel ausdrücken; Critchley, 2009). Anteriore Insula und Amygdala tragen gemeinsam zu einer angemessenen Bewertung bewusster Repräsentationen im Hinblick auf Aktivität und Valenz bei (Brentson et al., 2011; Critchley, 2009). In der anterioren Insula (vor allem rechtsseitig) laufen viszerale Afferenzen über den körperlichen Zustand (z. B. homöostatisches Ungleichgewicht, autonomes Arousal, Schmerz, Entzündung) zusammen. Deren Repräsentationen sind die Grundlage für bewusste Motive und Emotionen und beeinflussen die Verarbeitung emotionaler Reize, die Empathiefähigkeit sowie die kognitive Leistungsfähigkeit (Critchley, 2009). Zusammenfassend kann das SN als Zustandsmonitor aufgefasst werden, der im Bedarfsfalle zusätzliche kortikale Funktionen und Ressourcen für die Aufrechterhaltung der organismischen Homöostase bzw. die Erreichung belohnender Ziele in den Dienst stellt. Es bleibt noch anzumerken, dass nicht nur das SN, sondern auch das CEN und das DMN z. B. über Thalamus, Hypothalamus, Hippocampus oder Amygdala mit den Arousal-Systemen in Verbindung stehen (Miller & Cohen, 2001; Andrews-Hanna, 2012). Weiter oben wurde darüber hinaus darauf

hingewiesen, dass bereits die mentale Vorstellung stressinduzierender Situationen, z. B. durch Simulation im DMN (Rumination, Irritation), die Arousal-Systeme beeinflussen kann.

2.4.2 Kortikale Funktionalität bei psychischen und Verhaltensstörungen

Ein vertieftes Verständnis der drei kortikalen Netzwerke, CEN, DMN und SN, ergibt sich aus deren empirisch belegten Beziehungen zu psychischen Störungen (Broyd et al., 2009; Buckner et al., 2008; van den Heuvel & Hulshoff Pol, 2010; Menon, 2011; Zhang & Raichle, 2010). Dysfunktionen im Central Executive Network (CEN) stehen mit Depression (F32), Schizophrenie (F20), frontotemporaler Demenz, Alzheimer Demenz (F00) und Autismus (F84) in Zusammenhang (Menon, 2011; Zhang & Raichle, 2010). Dysfunktionen des CEN können sich nach Menon (2011) äußern in (a) schwacher Konnektivität und somit Integrität des CEN (z. B. bei Schizophrenie), (b) abnormaler Einbeziehung kortikaler Netzwerke, die normalerweise nicht zum CEN gehören und (c) herabgesetztem Zugriff auf bedeutsame, zielbezogene Stimuli. Bei letzterem spielt auch das Salient Network (SN) eine bedeutsame Rolle. Die meisten Befunde weisen auf eine geringe Aktivität des dorsolateralen Präfrontalkortex sowie des posterioren parietalen Kortex bei den genannten Störungen hin (Menon, 2011). Bei Schizophrenie zeigen sich erhöhte Konnektivitäten zwischen dorsolateralem Präfrontalkortex und den Basalganglien sowie geringe Konnektivitäten zwischen ventralem Präfrontalkortex und Amygdala (Zhang & Raichle, 2010). Die Konnektivität zwischen CEN und Default Mode Network (DMN) ist bei Schizophrenie erhöht (Broyd et al., 2008).

Dysfunktionen im Default Mode Network (DMN) stehen mit Depression, Alzheimer Demenz, leichter Intelligenzminderung (F70), Schizophrenie, Epilepsie (G40), Autismus, Angststörungen (F41) und dem Aufmerksamkeits-Defizit-Hyperaktivität-Syndrom (F90) in Zusammenhang (Broyd et al., 2008; Menon, 2011; Zhang & Raichle, 2010). Nach Broyd et al. (2008) können sich Dysfunktionen des DMN zeigen in (a) Interferenzen während der Aufgabenbewältigung, (b) gestörter Antikorrelation zwischen CEN- und DMN-Aktivitäten, (c) veränderter Konnektivität und somit der Integrität des DMN sowie (d) veränderter Aktivität des DMN. Länge und Schwere einer depressiven Episode stehen mit erhöhter Aktivität des ventralen anterioren cingulären Kortex und anliegendem ventromedialen Präfrontalkortex in Zusammenhang. Dies erklärt erhöhte Rumination und Selbstreflexion (Broyd et al., 2008; Menon, 2011; Zhang & Raichle, 2010). Bei Alzheimer Demenz steht dagegen eine geringe Aktivität und Konnektivität des parietalen cingulären Kortex, des medialen Temporallappens sowie des Hippocampus im Vordergrund, was Defizite beim episodischen und autobiografischen Gedächtnis erklärt (a. a. O.). Durch die Stärke der Deaktivierung des DMN

lassen sich Alzheimer Demenz und leichte Intelligenzminderung voneinander trennen (a. a. O.). Schizophrenie ist durch eine veränderte funktionale und anatomische Konnektivität innerhalb des DMN und insbesondere des medialen Präfrontalkortex sowie veränderter Morphologie und grauer Substanz gekennzeichnet. Zusätzlich zeigen sich eine erhöhte Konnektivität zum CEN und eine erhöhte Antikorrelation zwischen CEN- und DMN-Aktivität. Hiermit wird die Positivsymptomatik (Halluzinationen, Wahn) bei Schizophrenie erklärt (a. a. O.). Darüber hinaus ist der parietale cinguläre Kortex eingeschränkt mit dem Kleinhirn verbunden (Zhang & Raichle, 2010). Epileptische Anfälle und einhergehende Bewusstlosigkeit werden auf abnormale Aktivitätsmuster auf Grund defizitärer Konnektivität im DMN (insbesondere zwischen Hippocampus und parietalem cingulären Kortex) zurückgeführt. Die Aktivitätsmuster variieren jedoch nach Typ der Erkrankung (Broyd et al., 2008; Menon, 2011; Zhang & Raichle, 2010). Bei Autisten gelingt die Deaktivierung des DMN während zielbezogener Kognitionen nicht in ausreichendem Maße. Außerdem zeigt sich im DMN eine geringe funktionale Konnektivität. Bei medialem Temporallappen, Hippocampus, Amygdala und posteriorem medialen Kortex findet sich vergleichsweise wenig graue Substanz. Parietaler cingulärer Kortex und medialer Präfrontalkortex sind hypoaktiv, wodurch Defizite bei der Verarbeitung sozialer Informationen erklärt werden (a. a. O.). Eine verringerte Konnektivität zwischen parietalem cingulärem Kortex und medialem Präfrontalkortex findet sich beim Aufmerksamkeits-Defizit-Hyperaktivität-Syndrom (Broyd et al., 2008; Zhang & Raichle, 2010).

Dysfunktionen im Salient Network (SN) stehen mit Angststörungen, Depression, Schmerzempfinden, Sucht (F1x.2), Schizophrenie, frontotemporaler Demenz, Autismus und dem Aufmerksamkeits-Defizit-Hyperaktivität-Syndrom in Zusammenhang (Menon, 2011; Zhang & Raichle, 2010). Nach Menon (2011) sind diese und andere psychische und Verhaltensstörungen auf Abnormalitäten im SN bezüglich (a) der Entdeckung bedeutsamer Stimuli, (b) der Kontrolle über das Umschalten von DMN auf CEN und der Zuweisung kognitiver Funktionen (z. B. Arbeitsgedächtnis und Aufmerksamkeit), (c) der Interaktion zwischen anteriorer und posteriorer Insula bei der Zuweisung physiologischer Antwortbereitschaft sowie (d) der Kopplung zwischen anteriorer Insula und anteriorem cingulären Kortex für den schnellen Zugriff auf motorische Kontrolle zurückzuführen. Menon (2011) fasst zusammen: "The consequence of abnormalities at any of these levels is deficient context-dependent engagement and disengagement of cognitive systems important for attending to salient external stimuli or internal mental events." Angststörungen gehen mit einer erhöhten Aktivität der anterioren Insula einher, was die Erwartung aversiver körperlicher Zustände, vermehrte sorgenvolle Gedanken und Vermeidungsverhalten erklärt. Diese

Zusammenhänge sind klinisch besonders bedeutsam, da eine hohe Komorbidität von Angst mit Depression, Phobien, posttraumatischer Belastungsstörung und Autismus besteht (Menon, 2011). Bei Depression ist die Konnektivität zwischen anteriorem cingulärem Kortex, Amygdala, Thalamus und weiteren Regionen des limbischen Systems reduziert. Hierauf wird mangelnde Reaktivität und Aktivität von Depressiven zurückgeführt (Zhang & Raichle, 2010). Schmerzempfinden und Empathie für Schmerz sind zentrale Funktionen des SN. Diese können durch Überaktivität oder veränderte Konnektivität im SN unangemessen ausgeprägt sein, z. B. bei chronischem Schmerz (Menon, 2011). Bei Sucht finden sich abnormale Aktivitäten und Konnektivitäten in den meisten Teilnetzwerken des SN (anteriore Insula, Amygdala und subkortikale dopaminerge Netze) (a. a. O.). Schwere und Symptomatik von Schizophrenie (Halluzination, Wahn und Realitätsverlust) korrelieren mit reduzierten Volumina in anteriorer Insula und anteriorem cingulären Kortex. Halluzinationen, und möglicherweise unangemessene Beachtung bedeutsamer Stimuli, stehen mit geringer Aktivität der anterioren Insula in Zusammenhang. Erhöht ist die Konnektivität zwischen anteriorem cingulärem Kortex und limbischem System, womit Aufmerksamkeitsdefizite bei Schizophrenie erklärt werden (Broyd et al., 2008; Menon, 2011; Zhang & Raichle, 2010). Frontotemporale Demenz wird mit einer Degeneration der anterioren Insula und einem damit einhergehenden Verlust, angemessen auf bedeutsame Stimuli zu reagieren, in Zusammenhang gebracht (Menon, 2011). Autisten weisen bei der Verarbeitung sozialer Informationen eine Hypoaktivität der anterioren Insula auf (Menon, 2011; Zhang & Raichle, 2010). Beim Aufmerksamkeits-Defizit-Hyperaktivitätssyndrom finden sich verstärkte Verbindungen zwischen dem anterioren cingulären Kortex, der anterioren Insula, dem Thalamus und dem Kleinhirn. Die Aktivität des parietalen cingulären Kortex (Teil des DMN) korreliert positiv mit diesem Netzwerk, worauf das Aufmerksamkeitsdefizit z. T. zurückgeführt wird (Broyd et al., 2008; Zhang & Raichle, 2010).

2.4.3 Doubleloop-Perturbation-Modell neuronaler Aktivität

Menon (2011) stellt auf Grundlage der Befunde zur Organisation des Kortex sowie damit zusammenhängender Erkenntnisse über psychische und Verhaltensstörungen das Tripel-Network-Modell der kognitiven Aktivität vor. In diesem Modell steht das Salient Network (SN) im Zentrum und ist für die Entdeckung bedeutsamer internaler und externaler Stimuli sowie die angemessene Zuweisung kognitiver Funktionen zuständig. Eingaben erhält das SN von sensumotorischen, viszeralen und limbischen Afferenzen (bottom-up- bzw. Stimulusgesteuerte Aktivität), dem Central Executive Network (CEN) (top-down- bzw. instrumentelle, zielbezogene Aktivität) sowie dem Default Mode Network (DMN) (selbstbezogene und soziale Informationsverarbeitung). Es verfügt über Möglichkeiten, schnell in die Steuerung

motorischer Aktivitäten eingreifen zu können, z. B. bei fehlerhaftem Verhalten. Die angemessene Zuweisung kognitiver Funktionen nimmt das SN durch Beeinflussung der Aktivitäten von CEN und DMN vor. Liegt eine Situation vor, in der zielbezogene Aktivität erforderlich ist, so schaltet das SN von DMN auf CEN um.

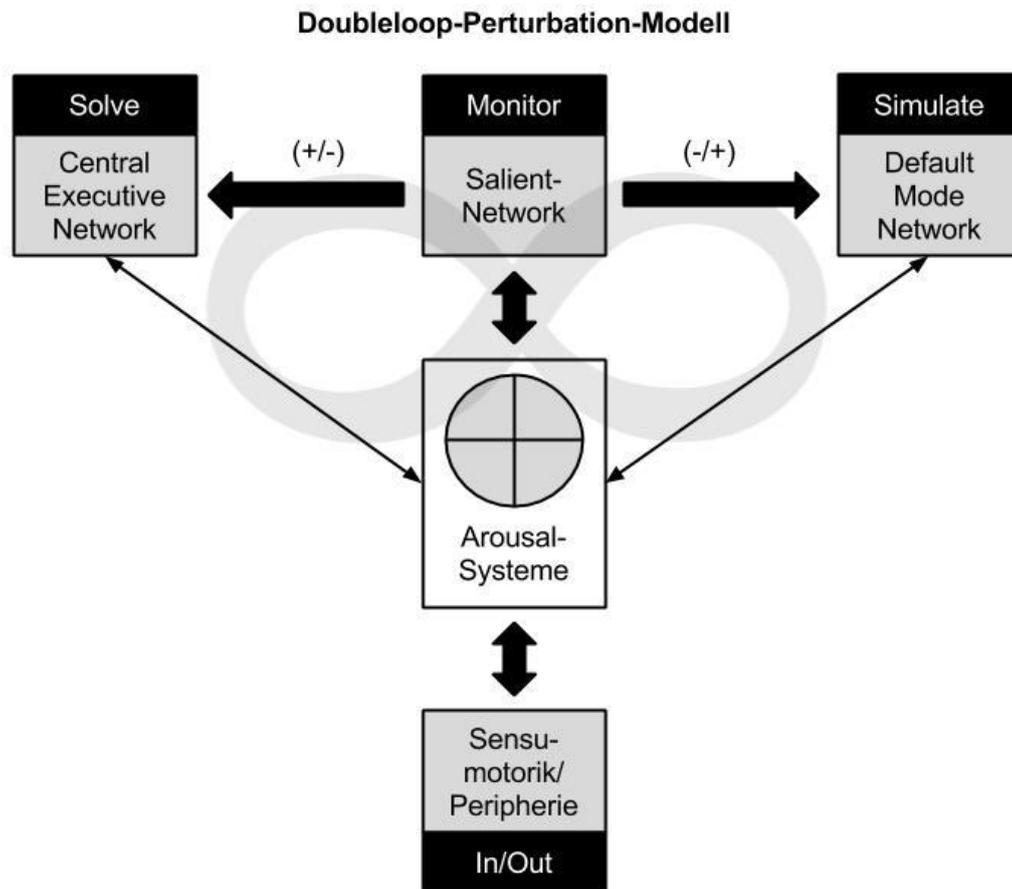


Abbildung 5. Doubleloop-Perturbation-Modell neuronaler Aktivität. Eigene Darstellung in Anlehnung an Menon (2011).

Mit dem in Abbildung 5 vorgeschlagenen Doubleloop-Perturbation-Modell (∞ -Modell) wird die Vorstellung von Menon (2011) weitgehend übernommen. Es wird jedoch eine wichtige Differenzierung vorgenommen. Im Zentrum des ∞ -Modells stehen die Arousal-Systeme, die der Anatomie folgend zwischen sensumotorischen und viszeralen auf- und absteigenden Bahnen und Kortex geschaltet sind. Mit dieser Erweiterung lässt sich insbesondere die alternierende Aktivitäten von CEN und DMN (Fox et al., 2005) im Ruhezustand erklären. Im zufriedenen, gesättigten Ruhezustand (Serotonin bzw. mentale Simulation in der Arousal-Uhr; vgl. Brown et al., 2012; Jones, 2003) ist zunächst keine instrumentelle Aktivität notwendig. Das SN sorgt für eine Deaktivierung des CEN und eine Aktivierung des DMN (Menon, 2011). Aktivität des DMN geht mit Simulationen selbstbezogener und sozialer Situationen, erlebter Episoden oder der eigenen Zukunft einher (Andrews-Hanna, 2012; Buckner et al., 2008; Spreng et al., 2008; Zhang & Raichle, 2010). Diese Simulationen werden, genau wie sensumotorische

und viszerale Afferenzen, durch die Arousal-Systeme präkognitiv bewertet (vgl. Hüther et al., 1999; Hüther, 2014). Das SN verarbeitet diese limbischen Eingaben und schaltet von DMN auf CEN, wenn es diese Eingaben als konfliktbehaftet oder homöostatisch bedeutsam bewertet. Im CEN wird nun instrumentelles Verhalten zur Lösung dieses Konflikts (Strategien, Pläne, Vorgehensweisen) generiert (vgl. Berntson et al., 2012; Botvinick et al., 2004; Bush et al., 2000; Carter & van Veen, 2007; Critchley, 2009; Kerns et al., 2004; Menon, 2011). Kann eine erfolgreiche Lösung für das simulierte Problem generiert werden, so ist keine weitere instrumentelle Aktivität des CEN notwendig und das SN schaltet wieder um auf das DMN. Im Ruhezustand findet nach dem ∞ -Modell somit eine beständige Fluktuation zwischen CEN- und DMN-Aktivität statt (Doubleloop; Fox et al., 2005). Ihre Funktion besteht in der Lösung simulierter sozialer und selbstbezogener Konflikte oder der Generierung von Strategien und Plänen zur Realisierung simulierter Zukunftsvisionen. Darüber hinaus findet durch die vorab stattfindende Generierung instrumenteller Verhaltensweisen im Ruhezustand eine Vorbereitung für eine energetisch effiziente, zielführende Gestaltung der Umwelt statt (van den Heuvel & Hulshoff Pol, 2010; Zhang & Raichle, 2010). Baumeister et al. (2011) fassen eindrucksvoll die Bedeutung bewusster Prozesse in der Doubleloop für zukünftiges zielführendes Verhalten zusammen. Sie unterstreichen den Einfluss von (a) mentaler Simulation und Übung, (b) Antizipation, Planung und Intentionsbildung, (c) Durchspielen, Interpretieren und Reflektieren vergangener Ereignisse, (d) Begründung, Entscheidung und Problemlösung, (e) mentaler Simulation aus der Perspektive Anderer, (f) Selbstachtung und Selbstbestätigung, (g) Framing und Zielsetzung sowie (h) Kommunikation und gegenseitigem Verständnis auf das zukünftige Verhalten. So betrachtet, ist die Quelle psychischer Beanspruchung hauptsächlich in der Fähigkeit zur Simulation, basierend auf Erfahrungen (als Echo vergangener psychischer Belastung), zu suchen.

Psychische Belastung stellt im ∞ -Modell eine potenzielle systemische Störung der Doubleloop dar. Psychische Belastung wird über sensumotorische und viszerale Afferenzen an die Arousal-Systeme weitergeleitet (Brown et al., 2012; Jones 2003), wo sie präkognitiv bewertet wird. Über die Arousal-Systeme verfügbare angemessene Reaktions- und Verhaltensmuster bzw. -gewohnheiten können umgehend zur Bewältigung der Belastung eingesetzt werden. Die Inanspruchnahme kognitiver Funktionen ist in einem solchen Falle nicht notwendig. Es findet automatisierte Bewältigung der Belastung statt. Das Gehirn alterniert weiterhin zwischen CEN- und DMN-Aktivität, die nicht zwingend mit der gegebenen Belastung in inhaltlichem Zusammenhang steht. Wird in der psychischen Belastung jedoch ein Konflikt oder eine für die Homöostase bedeutsame Situation erkannt, wird über das SN das

CEN hinzugeschaltet. Kognitive Funktionen (Aufmerksamkeit, Arbeitsgedächtnis, Zugriff auf das Langzeitgedächtnis, Planung) zur Generierung instrumentellen Verhaltens werden für die Bewältigung der Belastung in den Dienst gestellt (vgl. Berntson et al., 2012; Botvinick et al., 2004; Bush et al., 2000; Carter & van Veen, 2007; Critchley, 2009; Kerns et al., 2004; Menon, 2011). Das DMN wird deaktiviert (Menon, 2011). Die Doubleloop des ∞ -Modells ist gestört (Perturbation) und es findet kontrollierte Bewältigung der Belastung statt. Wie erfolgreich die Suche nach angemessenem Bewältigungsverhalten ist, entscheidet nun darüber, wie kontrollierbar der angestoßene Stressreaktionsprozess ist und wieviel und welche Form von Energie zur Bewältigung mobilisiert wird (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther et al., 1999; Hüther, 2014). Auch die in der Belastung entdeckten motivationalen Anreize sind für die Aufrechterhaltung der Bewältigungsbemühungen entscheidend (Brown et al., 2012; Jones, 2003; LeDoux, 2012). Die Kontrolle bei der Bewältigung der Belastung muss jedoch nicht vollständig sein. Vielmehr wird davon ausgegangen, dass das Ausmaß der Kontrolle mit der Stärke der Aktivierung des SN, speziell des anterioren cingulären Kortex, zusammenhängt. Bei geringer aufmerksamer Kontrolle kann das DMN weiterhin aktiv sein und mit den CEN-Aktivitäten interferieren. Dies kann die kognitive Leistungsfähigkeit unter Umständen beeinträchtigen, erlaubt aber andererseits die parallele Simulation (Andrews-Hanna, 2012; Botvinick et al., 2004; Carter & van Veen, 2007; Fox et al., 2005; Kerns et al., 2004; Menon & Uddin, 2010; Ochsner & Gross, 2005).

Gegenwärtig werden Kognitionen über das Selbst bzw. über die vergangene, gegenwärtige und zukünftige soziale Rolle der eigenen Person hauptsächlich mit Aktivitäten des DMN gleichgesetzt (vgl. Andrews-Hanna, 2012; Buckner et al., 2008; Heatherton, 2010; Lord, Diefendorff, Schmidt & Hall, 2011; Spreng et al., 2008). Das ∞ -Modell legt nahe, dass das Selbst durch beständige Selbst-Reflexion (Simulation sozialer, selbstbezüglicher Situationen mit einhergehender Problemlösung) hinsichtlich Energieeffizienz modifiziert und stabilisiert wird. Störungen der Doubleloop durch sensumotorische An- oder Hinweisreize oder viszerale Signale über die Homöostase stellen Herausforderungen und Möglichkeiten dar (LeDoux, 2012), in denen simulierte Problemlösungen an der Realität geprüft werden können (Baumeister et al., 2011). Je erfolgreicher und müheloser dies gelingt, desto weniger Modifikationen des vorgesehenen Verhaltens durch umgehende Korrektur (Zuschaltung von CEN-Funktionen über das SN; Menon, 2011) sind notwendig. Automatisiertes Verhalten ist besonders energieeffizient und verlangt bei erfolgreicher Ausführung auch keinerlei Modifikation („System 1“ bei Kahneman, 2011). Kommt es jedoch während der Bewältigung zu Konflikten (Aktivierung des SN), muss unter Energieeinsatz eine kontrollierte Korrektur des

(automatisch) vorgesehen Verhaltens vorgenommen werden („System 2“ bei Kahneman, 2011). Damit werden auch automatisch die zur simulierten Problemlösung eingesetzten (sozial-) kognitiven Konzepte und Verhaltensmuster geschwächt (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther et al., 1999; Sapolsky, 2003). Durch Speicherung dieser konflikthaften Episode sowie ggf. der erfolgreich eingesetzten Verhaltensweisen (Hüther et al., 1999; Sapolsky, 2003) steht nun neues Erfahrungs-Material für Simulationen und beständige Selbst-Reflexion in der Doubleloop bereit. Das Selbst sowie die (sozial-) kognitiven Konzepte über die Welt können an diese Erfahrung angepasst werden, das Selbst kann sich entwickeln. Ohne Störungen (psychische Belastung) ist jedoch mit keinerlei Entwicklung des Selbst zu rechnen.

3 SELBSTORGANISATION IN ARBEITSSYSTEMEN

Menschen sind, wie alle Organismen, offene Systeme, die dem zweiten thermodynamischen Hauptsatz unterliegen (Penzlin, 2014; Strunk & Schiepek, 2006). Das heißt sie unterliegen der allgemeinen Tendenz hin zu einer völligen Auflösung übergeordneter Strukturen. Zugleich stehen sie in einer energetischen und stofflichen Austauschbeziehung mit ihrer Umwelt. Energie und Stoff werden derart in organismische Strukturen umgesetzt, dass ein Organismus trotz beständigen mikroskopischen Wandels über die Lebensspanne als kontinuierlich bestehende makroskopische Einheit in artspezifischer Weise mit der Umwelt in Interaktion tritt. Hierbei sind die mikroskopischen metabolischen Prozesse nur bedingt bzw. nur in einem statistischen Sinne vorhersagbar und abhängig von Fluktuationen des Energie- und Stoffdurchsatzes. Dieser Umstand erfordert Anpassungsfähigkeit und garantiert Individualität und innerartliche Varianz. Unter dem Begriff der Selbstorganisation werden Phänomene zusammengefasst, bei denen es trotz des zweiten thermodynamischen Hauptsatzes, also trotz mikroskopischer Fluktuationen zu makroskopisch klassifizierbaren Ordnungsstrukturen kommt (vgl. Strunk & Schiepek, 2006). Im vorliegenden Zusammenhang wurde der Begriff der Selbstorganisation gewählt, weil es um die Frage geht, wie ein Mensch trotz mikroskopischer Fluktuationen und individueller Anpassung als Mensch erkennbar und als Persönlichkeit identifizierbar bleibt. Hierbei liegt der Fokus auf der Begründung von Verhaltensweisen, die ein Mensch ohne besondere Aufforderung oder Umweltbedingungen, also „intrinsisch motiviert“, ausführen würde (vgl. Deci & Ryan, 2000).

Primäre Verhaltensweisen dienen der Sicherstellung der organismischen Integrität und Funktionsfähigkeit bzw. der Homöostase. Bei diesen Verhaltensweisen geht es darum, den Zerfall organismischer Strukturen durch geeigneten Energie- und Stoffdurchsatz auszugleichen (Penzlin, 2014). Die effektive Gestaltung der Umwelt im Sinne der effizienten Erfüllung primärer Bedürfnisse ist eine weitere Klasse von Verhaltensweisen, die vor allem für den Menschen (aber nicht nur für ihn) typisch sind. Verhaltensweisen dieser Klasse werden unter dem Begriff der Kompetenz zusammengefasst (White, 1959; Deci & Ryan, 2000). Menschen neigen darüber hinaus zur Ausbildung sozialer Beziehungen, die von Empathie und Fürsorge getragen sind. Daher wird dem Menschen ein Bedürfnis nach sozialer Teilhabe und Zugehörigkeit zugeschrieben (Baumeister, 2000; Baumeister & Leary, 1995; Deci & Ryan, 2000; Heatherton, 2011). Auf Grundlage dieser Trias intrinsischer Bedürfnisse - Sicherstellung der Homöostase, Gestaltung der Umwelt und Bildung sozialer Beziehungen - bildet sich die Dynamik menschlicher Gesellschaften heraus. Energie- und Stoffumsatz wird zu einem gesellschaftlichen Anliegen, die individuelle Kompetenz zur Gestaltung der Umwelt wird für

dieses Anliegen in gesellschaftlichen Dienst gestellt und die Möglichkeiten zur sozialen Teilhabe und Zugehörigkeit sind abhängig von den Kompetenzen, die ein Individuum in eine soziale Gruppe einbringen kann. Auf diese Weise in einen gesellschaftlichen Kontext eingebunden ergeben sich für den einzelnen Menschen soziale Möglichkeiten und Grenzen. Möchte er die Vorteile in Anspruch nehmen, die mit der Zugehörigkeit zu einer sozialen Gruppe einhergehen, so muss er auch deren Erwartungen erfüllen. Hiermit ergibt sich das Spannungsfeld von Autonomie und Selbstkontrolle: Autonomie bedeutet die uneingeschränkte Möglichkeit der Erfüllung persönlicher Bedürfnisse in der persönlich bevorzugten Art und Weise. Selbstkontrolle ist erforderlich, wenn die eigenen Bedürfnisse oder die persönlich bevorzugte Art und Weise ihrer Erfüllung mit den (durchaus wandelbaren) Erwartungen der sozialen Gruppe kollidieren. In diesem Falle ist unter Mehraufwand von Energie alternatives, gesellschaftskonformes Verhalten zu entwickeln und umzusetzen. Das Spannungsfeld von Autonomie und Selbstkontrolle lässt sich über die gesellschaftliche Umwelt hinaus auf die Anforderungen aus der Umwelt insgesamt verallgemeinern (vgl. Baumeister, 2010; Baumeister, Bratslavsky, Muraven & Tice, 1998; Baumeister & Leary, 1995; Deci & Ryan, 2000; Gollwitzer, 1999; Heatherton, 2011; Kuhl, 1996; Lord, Diefendorff, Schmidt & Hall, 2010).

Im vorliegenden Kapitel wird dargelegt, dass das ∞ -Modell plausible Erklärungen und einige hypothetische Schlussfolgerungen für die Trias intrinsischer Bedürfnisse anbietet. Außerdem können Schlussfolgerungen für die menschengerechte Gestaltung von Arbeitssystemen abgeleitet werden. Einige einschlägige arbeitspsychologische Konzepte und Theorien werden in diesen theoretischen Rahmen eingeordnet. Dabei wird es nicht um eine umfassende Zusammenstellung etablierter arbeitspsychologischer Erkenntnisse gehen. Vielmehr stehen arbeitspsychologische Bewertungskriterien für die persönlichkeits- und gesundheitsförderliche Arbeitsgestaltung im Vordergrund. Ulich (2011; Ulich & Wülser, 2012) nennt (a) Ganzheitlichkeit, (b) Anforderungsvielfalt, (c) Möglichkeiten der sozialen Interaktion, (d) Autonomie, (e) Lern- und Entwicklungsmöglichkeiten, (f) Zeitelastizität und stressfreie Regulierbarkeit sowie (g) Sinnhaftigkeit. Diese werden anhand einiger konkreter und weit verbreiteter Theorien und Konzepte illustriert. Die Auswahl der Beispiele orientiert sich dabei an arbeitspsychologischen Begrifflichkeiten, die auch für den empirischen Teil der vorliegenden Arbeit relevant sind.

3.1 SICHERSTELLUNG DER HOMÖOSTASE

Die biologische Integrität und Funktionsfähigkeit muss als Grundvoraussetzung und prioritäre Zielstellung aller Handlungen von Organismen betrachtet werden. Grundvoraussetzung deshalb, weil dies die Sicherstellung von Energie- und Stoffreserven für das Handeln einschließt. Prioritäre Zielstellung deshalb, weil jeder Organismus daraufhin ausgerichtet ist (Penzlin, 2014). Können homöostatische Krisen nicht bewältigt werden, beschleunigt dies den Tod. Handlungen, die über die Sicherstellung der Homöostase hinausgehen, wie zum Beispiel soziale Handlungen, treten bei solchen Krisen in den Hintergrund. Gesellschaften und Arbeitssysteme wären pathogen und selbstzerstörerisch, wenn ihre Gestaltung dieses ihr eigne organismische Fundament nicht berücksichtigen würde.

3.1.1 Theoretische Grundlegung

Die Sicherstellung der Homöostase umfasst Verhaltensweisen, die biologische Integrität und Funktionsfähigkeit aufrechterhalten. Meist handelt es sich um Verhaltensweisen, die mit dem Energieaustausch (Atmung, Temperaturregulierung), dem Stoffwechsel (Wasser- und Salzhushalt, Nahrung und Verdauung, organismische und zelluläre Auf- und Abbauprozesse) und der Abwehr von Noxen (Gifte, Krankheitserreger) und physischen Gefahren (z. B. direkte inner- oder überartliche Bedrohungen) bzw. mit Stressreaktionsprozessen in Zusammenhang stehen (Hüther, 2014; LeDoux, 2012; Penzlin, 2014). Mit diesen Verhaltensweisen sind die Arousal-Systeme auf sehr direkte Weise verbunden. Ungleichgewichte oder Krisen der Homöostase werden in den Arousal-Systemen abgebildet. Diese initiieren die Generierung von Verhaltensweisen, die in der gegebenen Situation zur (Re-) Stabilisierung der Homöostase geeignet sind (Brown et al., 2010; Hüther, 2014; Jones, 2003; LeDoux, 2012; Penzlin, 2014). Drängende Krisen, die eine Gefahr der Integrität und Funktionsfähigkeit signalisieren, haben vor allen anderen psychophysiologischen Prozessen Priorität. Das heißt in solchen Situationen werden nicht zielführende Verhaltensweisen unterbrochen; es wird keine Rücksicht mehr auf gesellschaftliche Verpflichtungen und Konventionen genommen. "... in the human being who is missing everything in life in an extreme fashion, it is most likely that the major motivation would be the physiological needs rather than any others" (Maslow, 1943, S. 373). Je drängender und aversiver eine Krise erlebt wird, desto eher werden basale (primäre, unconditionierte) Verhaltensweisen initiiert, die zu einer direkten Linderung der Krise beitragen (vgl. Hüther, 2014).

Die Bewältigung homöostatischer Krisen und einhergehenden Stressreaktionsprozessen ist mit hohem Energieaufwand und möglicherweise anderweitigen Schädigungen von Organsystemen verbunden (Hüther et al., 1999; Sapolsky, 2003; Yehuda & Mostofsky, 2006).

Daher ist es für einen Organismus günstiger, die Sicherstellung der Homöostase präventiv vorzunehmen. Das heißt, der Organismus sollte sich besten Falls in einem physiologischen Zustand befinden, in dem homöostatische Krisen unwahrscheinlich sind. Dementsprechend verfügen nicht zuletzt Säugetiere über Funktionen, die auf Speicherung, optimale Nutzung und Wiederverwendung von Stoffen und Energieträgern ausgerichtet sind (Penzlin, 2014; Rasch & Born, 2013). Mechanisches, Immun- und kognitives Gedächtnis stellen das Ergebnis der Anpassungsfähigkeit eines Organismus dar und sichern die Verfügbarkeit erfahrungsgemäß erfolgreicher Verhaltensweisen und Reaktionen. Deren Aufbau sowie die Homöostase sind an einen ausgewogenen circadianen Rhythmus gebunden (Rasch & Born, 2013).

Mit der Arousal-Uhr werden Verhaltensweisen kategorisiert, die in erster Linie mit dem homöostatischen Zustand in Verbindung stehen: (a) Wachen umfasst die allgemeine Aufmerksamkeit und Handlungsbereitschaft. Möglichkeiten und Herausforderungen werden in einem wachsamem Zustand überhaupt erst entdeckt (Brown et al., 2012; Jones, 2003; LeDoux, 2012). (b) Bewältigen umfasst die Bereitstellung und Kanalisierung von Energie zur Bewältigung einer Bedrohung oder Herausforderung. Die Generierung neuartigen Bewältigungsverhaltens sowie erhöhte Lernbereitschaft sind hiermit assoziiert (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther et al., 1999; LeDoux, 2012; Sapolsky, 2003). (c) Konsumieren bedeutet die mentale Vorwegnahme und physische Ausführung instrumenteller Verhaltensweisen zur Erfüllung homöostatischer Bedürfnisse (z. B. Nahrungs- oder Flüssigkeitsaufnahme, Sexualverhalten). Erfolgreiche Verhaltensweisen werden durch dopaminerge Belohnung verstärkt (Brown et al., 2012; Jones, 2003; LeDoux, 2012). (d) Ein ruhiger und sattsamer Zustand geht mit der Simulation eigenen Verhaltens und Reflexionen über die eigene Person in sozialen Kontexten der Vergangenheit, Gegenwart und Zukunft einher (Andrews-Hanna, 2012; Buckner et al., 2008; Menon, 2011). Diese Kognitionen bereiten energetisch effizientes und wünschenswertes Verhalten in der Zukunft vor (Baumeister et al., 2011). (e) Schlafen ist die Voraussetzung für die Festigung von kognitivem und Immun-Gedächtnis, sichert die Lernfähigkeit und effiziente Funktionsfähigkeit des Gehirns und ermöglicht die Wiederherstellung von Energie- und Stoffreserven (Brown et al., 2012; Rasch & Born, 2013). (f) Aufgeben bedeutet, Energie- und Stoffreserven nicht bei Verhaltensweisen aufzubreuchen, die nicht zielführend sind. Dies kann kontrolliert erfolgen, indem ein Engagement bei vorhersehbarer Erfolglosigkeit bewusst zurückgenommen wird. Im Kortisol-System besteht jedoch auch ein biologischer Mechanismus, der nicht zielführendes Engagement mit schädigender Wirkung eindämmt (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther et al., 1999; Sapolsky, 2003).

3.1.2 Arbeitspsychologische Konsequenzen

Sicherstellung der Homöostase erfordert Raum und Zeit, die aufgezählten Verhaltensweisen in erforderlichem Maße ausführen zu können. Außerdem erfordert sie persönliche Aufmerksamkeit gegenüber grundlegenden biologischen Bedürfnissen. Da die Arousal-Uhr in einem engen Zusammenhang mit dem Circumplex-Modell der Core Affects steht (s. o.), erscheint der Zugang zum eigenen emotionalen Zustand hierfür essenziell (s. o.; vgl. LeDoux, 2012). Emotionale Zustände spiegeln die (psychische) Beanspruchung der Arousal-Systeme wider. Ihre Erfassung kann daher für die personenbezogene Analyse von Arbeitssystemen sehr aufschlussreich sein (Schüpbach, 2013; Schüpbach & Zölch, 2007; Warr, 1994; Wieland & Hammes, 2014b). Der über die Arousal-Systeme vermittelte homöostatische und damit einhergehende emotionale Zustand stellt einen inneren Belastungszustand dar, der die Generierung und Ausführung sichernder und lindernder Verhaltensweisen beansprucht. Für die Gestaltung von Arbeitssystemen bedeutet dies, dass die homöostatischen Bedürfnisse zu berücksichtigen sind. Den Beschäftigten sind Räume, Zeiten und Zugänge zu Ressourcen zuzusichern, sodass sie diese Bedürfnisse in ausreichendem Maße erfüllen können (Allmer, 1996; Rau, 2011; Ulich & Wiese, 2011; Warr, 1994). Dies umfasst eine ausreichende materielle Versorgung, eine angemessene Arbeitszeit-, Pausen- und Urlaubsplanung sowie die Gestaltung geeigneter Rückzugsräume (z. B. sanitäre Anlagen, Pausen- und Ruheräume, Sozialräume) (Allmer, 1996; Rau, 2011; Ulich & Wiese, 2011). Stressreaktionsprozesse¹⁰ stellen neben der Optimierung instrumentellen Verhaltens eine zweite wichtige Säule zum Aufbau kompetenten Verhaltens dar (s. u.). Sie sollten jedoch immer kontrollierbar sein, da sie andernfalls Schlaf verhindern und mit kognitiven Einbußen, Leistungseinbußen und gesundheitlichen Beeinträchtigungen einhergehen (Allmer, 1996; Brown et al., 2012; DeRijk & de Kloet, 2006; Foy et al., 2006; Gunnar & Quevedo, 2007; Hockey, 1997; Hüther et al., 1999; Jones, 2003; Kaluza, 2011; Krajewski & Wieland, 2003; Sanders, 1984; Schönflug, 1987; Rasch & Born, 2013; Rau, 2011; Ulich & Wiese, 2011; Warr, 1994; Wieland-Eckelmann, 1992; Wieland-Eckelmann, Allmer, Kallus & Otto, 1994; Yehuda & Mostofsky, 2006). Während die aufgezählten Maßnahmen eher auf die Verhältnisse bzw. Bedingungen am Arbeitsplatz abzielen, können verhaltensbezogene Maßnahmen einen zusätzlichen Beitrag leisten. Elaborierte Stressmanagementtrainings (Greif, 2011; Kaluza, 2011, 2012) sind geeignet, individuelle Kompetenzen für die Sicherstellung der Homöostase und insbesondere der

¹⁰ Stressreaktionsprozesse sind hier im Sinne von Anpassungsprozessen zu verstehen, die sich in Folge neuartiger, unerwarteter oder herausfordernder Situationen einstellen (Hüther et al., 1999).

Energie- und Stoffreserven zu entwickeln. Solche Maßnahmen können sinnvoll in ein ganzheitliches betriebliches Gesundheitsmanagement eingebettet werden.

3.2 AUFBAU VON KOMPETENZ

Je effektiver ein Mensch seine Umwelt im Sinne seiner eigenen Bedürfnisse gestalten kann, desto besser gelingt ihm die Sicherstellung seiner Homöostase. Daher erscheint es biologisch sinnvoll, die eigenen Fähigkeiten zur Gestaltung der Umwelt, also die eigenen Kompetenzen, mehr und mehr auszubauen. In der Interaktion des Menschen mit der Umwelt werden sowohl das Verständnis von der Umwelt als auch die eigenen Fähigkeiten zur Umwelt-Gestaltung entwickelt. Diese Anpassungsfähigkeit des Menschen wird in Arbeitssystemen zur Erzeugung von Produkten und Dienstleistungen für Menschen eingesetzt. Die natürliche Neigung des Menschen, die eigene Kompetenz in interessanten Situationen auszubauen und zu optimieren, kann bei der Arbeitsgestaltung für den einzelnen Menschen und ein Unternehmen wachstumsorientiert berücksichtigt werden. Außerdem kann die Qualität wirtschaftlicher Erzeugnisse nur davon profitieren, wenn Arbeitssysteme die Beschäftigten als Experten der eigenen Bedürfnisse und als Experten des Konsums begreifen.

3.2.1 Theoretische Grundlegung

White (1959) beschreibt bei Kindern und Tieren Spielverhalten, Exploration und Manipulation und bei Erwachsenen darüber hinaus das Engagement für neuartige und aufregende Erlebnisse. Diese Verhaltensweisen sind seiner Ansicht nach nicht direkt auf homöostatische Bedürfnisse und ein konsumptives Ziel hin ausgerichtet. Vielmehr werden diese Verhaltensweisen intrinsisch ausgelöst. D. h. in geeigneten Situationen (s. u.) werden sie allein durch Prozesse des Zentralnervensystems bzw. selbst-determiniert ausgelöst und wirken in sich belohnend. Deci und Ryan (2000) übernehmen in ihrer Self Determination Theory (SDT) diese Sichtweise in Bezug auf das Bedürfnis nach Kompetenz nahezu vollständig. White (1959) schreibt in Anlehnung an Hebb (1949), dass Menschen aller Altersklassen ihre Aufmerksamkeit vor allem auf Situationen richten, die ein gewisses Ausmaß an "difference-in-sameness" bieten. Situationen, die das "effectance motive" auslösen, sind demnach vertraut und bieten einige interessante neue Aspekte. In solchen Situationen lernen Menschen jene Effekte, die das generierte variable und neue Verhalten auslösen. So bauen sie ihre Kompetenz zur effektiven Gestaltung ihrer Umwelt nach und nach aus. Mit dieser Konzeption geht White den für die persönliche Kompetenz-Entwicklung entscheidenden, infinitesimalen Schritt über jene Ansicht Allports (1937), "a person likes to do what he can do well" (a. a. O., S. 150), hinaus. In Bezug auf Yerkes und Dodson (1906) und Hebb (1949) schreibt White weiter, dass die

Anpassung an neue Umwelten am besten bei mittlerer Herausforderung an die bereits vorhandenen Kompetenzen gelingt. Hiermit lässt sich eine Brücke zur Stressforschung schlagen. Die beschriebenen Situationen entsprechen in ihrem Charakter jenen, die kontrollierbare Stressreaktionsprozesse und damit einhergehendes Lernen sowie positive Emotionen (z. B. Stolz, Erleichterung) auslösen (Gunnar & Quevedo, 2007; Hüther, 2014; Hüther et al., 1999; Sapolsky, 2003).

Die erlernten Episoden und Verhaltensweisen werden im Gedächtnis gespeichert und stehen nunmehr dem Default Mode System (DMN) zur Verfügung. Unter Verwendung des ∞ -Modells lassen sich weitere Schlussfolgerungen ziehen. Die neuen Gedächtnisinhalte können nun in Simulationen des DMN einbezogen werden. Es besteht daher die Möglichkeit, vertraute Konzepte und Situationen mit neuen Aspekten mental zu kombinieren und ggf. entstehende Konflikte im Rahmen der Doubleloop mental zu bewältigen. Tritt diese simulierte Situation in der Realität ein, kann das simulierte Bewältigungsverhalten an der Realität geprüft und ggf. modifiziert werden. In letzterem Fall entstehen neue Gedächtnisinhalte, die eine Modifikation vorhandener Konzepte erforderlich machen. Dies kann wiederum durch Simulation und Realitäts-Check erfolgen. Das ∞ -Modell legt daher nahe, dass der Mensch fähig ist, erworbenes Wissen durch Simulation auf ähnliche Situationen zu generalisieren. Um die Angemessenheit dieser Generalisierung bzw. seine Kompetenz zu prüfen, bedarf er jedoch des bestätigenden oder korrigierenden Feedbacks aus der Umwelt.

Diese theoretischen Betrachtungen können einige Aspekte der Kompetenz-Entwicklung näher beschreiben. White (1959) fasst zusammen, dass Jugendliche und Erwachsene zwar weiterhin interessante Situationen aufsuchen, dieses Verhalten jedoch mit dem Alter abnimmt. Er erklärt dies mit absorbierenden Anforderungen in der Schule bzw. im Beruf. Eine weiterer Erklärungs-Ansatz kann aus den obigen Ausführungen abgeleitet werden: Mit zunehmendem Wissen nimmt auch die Möglichkeit der mentalen Simulation von Situationen zu, sodass es weniger notwendig wird, entsprechende Situationen tatsächlich aufzusuchen. White (1959) schreibt außerdem, dass Kinder zunächst sehr langsam und dann zunehmend schneller lernen. Auch hier kann eine Simulations-Hypothese aufgestellt werden: Mit zunehmendem Wissen können Effekte des eigenen Verhaltens in der Umwelt mental vorweggenommen werden und simulierte Bewältigungsverhalten haben bei der Ausführung in der Realität zunehmend bestätigenden oder korrigierenden Charakter. Strittig bei dem Kompetenz-Konzept von White (1959) sowie Deci und Ryan (2000) ist die Bedeutung externaler Stimuli für die Auslösung des "effectance motive". Die Autoren gehen davon aus, dass interessante Situationen ("difference-in-sameness") aufgesucht werden und entsprechendes Verhalten auslösen. Sie schreiben dabei

dem zentralen Nervensystem die Rolle der energetischen Quelle zu, ohne dies genauer zu spezifizieren. Geht man von einer Simulations-Hypothese aus, so könnten auch solche Situationen auslösenden Charakter haben oder gar aktiv hergestellt und aufgesucht werden, deren Bewältigung bereits zu einem gewissen Grade erfolgreich mental simuliert wurde. Der Anreiz in einer derartigen Situation würde somit in der Möglichkeit bestehen, die eigene Kompetenzerwartung zu prüfen bzw. zu bestätigen. Interessanter Weise lässt sich das "effectance motive" unter dieser Betrachtung entgegen den Annahmen von White (1959) sowie Deci und Ryan (2000) im Sinne einer kybernetischen Annäherungs-Schleife (Lord et al., 2010) formulieren: Ausgehend von sozialpsychologischen Befunden kann angenommen werden, dass Menschen sich im Allgemeinen selbst als besonders kompetente Bewältiger herausfordernder Situationen simulieren (Baumeister, 2010; s. u.). In der Interaktion mit der Umwelt versuchen sie dann, die Diskrepanz zwischen überhöhtem Selbstbild und dem Feedback aus den Effekten des Verhaltens in der Umwelt auf verschiedene Weisen zu minimieren, wobei Kompetenz-Aufbau nur eine von vielen Möglichkeiten darstellt. In der Regel neigen sie jedoch auch bei noch bestehenden Diskrepanzen weiterhin zu selbstwertdienlichen Verzerrungen bzw. Simulationen (a. a. O.). Das überhöhte Selbstbild kann somit als zeitlich vorweglaufender Standard für das Wachstum von Kompetenz betrachtet werden. "What a man can be, he must be" (Maslow, 1943, S. 382).

White (1959) schreibt, dass der Aufbau von Kompetenzen ein fortwährender Prozess ist, der nur gelegentlich von homöostatischen Krisen oder Bedürfnissen unterbrochen wird. Da Kompetenz-Aufbau in der Arousal-Uhr mit dem Quadranten "Bewältigen" (kontrollierbare Stressreaktionsprozesse) sowie allgemeiner Wachsamkeit verknüpft ist, muss außerdem von einem erhöhten Energiebedarf und Metabolismus (Anpassungsleistung) ausgegangen werden (Gunnar & Quevedo, 2007; Hüther et al., 1999; Hüther, 2014; Sapolsky, 2003; Yehuda & Mostofsky, 2006). Dies bedeutet, dass der Kompetenz-Aufbau einen wachsamem Arousal-Zustand sowie aufgefüllte Energie- und Stoffreserven voraussetzt. Überfordernde Situationen bzw. unkontrollierbare Stressreaktionsprozesse (Kortisol-Krise in der Arousal-Uhr) sind dem Kompetenz-Aufbau genauso wenig dienlich, wie vollkommen vertraute Umgebungen ohne interessante Anreicherungen (Gunnar & Quevedo, 2007; Deci & Ryan, 2000; Hüther et al., 1999; Sapolsky, 2003; White, 1959). Es stehen somit beim Kompetenz-Aufbau drei Arten von Störungen der Doubleloop (∞ -Modell, s. o.) in Betracht: (a) Externale Stimuli als Anreize zur Auslösung des "effectance motive" oder als Möglichkeit zur Prüfung der eigenen Kompetenzerwartung. (b) Die Modifikation vorhandener mentaler Konzepte durch Feedback über die Effekte des eigenen Verhaltens in der Umwelt. (c) Die weiter oben bereits beschriebene

Notwendigkeit zur Erfüllung homöostatischer Bedürfnisse. Homöostatische Bedürfnisse können daher mit Prozessen des Kompetenz-Aufbaus störend interferieren. Andererseits stehen diese Arten von Störungen insofern in einem funktionalen Zusammenhang, da mit zunehmender Kompetenz zur effektiven Interaktion mit der Umwelt eine zunehmend angemessene Erfüllung homöostatischer Bedürfnisse in verschiedensten Situationen gelingt (vgl. White, 1959).

3.2.2 Arbeitspsychologische Konsequenzen

Bereits White (1959), aber auch Deci und Ryan (2000) stellen gesundheitliche Beeinträchtigungen als Folge einer Deprivation des Bedürfnisses nach Kompetenz heraus. Die Simulations-Hypothese kann hierzu einen erklärenden Beitrag leisten: Eine Quelle für Arousal-Aktivitäten ist das Default Mode Network (DMN), dessen Aktivität mit Simulationen selbstbezogener oder sozialer Situationen, Erfahrungen oder Erwartungen über die Zukunft einhergeht. Es ist unter anderem bei automatisierten und rhythmischen (kurzzyklischen) Tätigkeiten aktiv (s. o.). Zahlreiche einzelne Arbeitstätigkeiten verlaufen bei hoher Kompetenz automatisiert ab (Hacker & Sachse, 2014; Lord, Diefendorff, Schmidt & Hall, 2010). Nach dem ∞ -Modell befindet sich das mentale Gesamtsystem währenddessen in der Doubleloop, ist also mit der Lösung simulierter Probleme befasst. Die einsetzenden Simulationen beziehen sich auf Grund vorhandener Hinweisreize vermutlich vornehmlich auf die eigene Arbeitssituation. Darüber hinaus wird selbstwertschädigenden Formen der mentalen Simulation (z. B. negative Denkmuster bei Depression oder sorgenvolle Gedanken bei Angststörungen, s. o.) unter solchen Bedingungen unbegrenzter Raum gelassen. Fordert eine Arbeitstätigkeit ausschließlich automatisierte Routinen ohne Aussicht auf eine bessere Zukunft, so besteht Gefahr, dass sie zu einem Großteil von Stressreaktionen (z. B. Ausschüttung von Noradrenalin und Stressempfinden in Folge simulierter Probleme) bis hin zu unkontrollierbaren simulierten Stressreaktionsprozessen mit allen gesundheitlichen Folgen begleitet wird (Hüther, 2014; Hüther & Sachsse, 2007). Einsetzende Simulations-Prozesse können sich außerdem negativ auf die Leistung auswirken, da sie mit der Ausführung weitgehend beherrschter Tätigkeiten interferieren (Fox et al., 2005). Eine Strategie bei automatisierten oder kurzzyklischen Tätigkeiten kann somit darin bestehen, die Aufmerksamkeit willentlich auf diese zu lenken und Simulationen zu unterdrücken. Hacker und Richter (1980) definieren gerade Monotonie als Folge von Tätigkeiten "die erforderliche Aufgabenzuwendung mit unzureichenden kognitiven Auseinandersetzungsmöglichkeiten verbinden" (a. a. O., S. 43). Die Autoren verbinden Monotonie mit folgender Symptomatik: "Aktuell psychophysiologische Symptome herabgesetzter Aktivität; Unzufriedenheit mit Tätigkeit; Befindensbeeinträchtigungen mit

körperlicher Symptomatik" (a. a. O.). Bei automatisierten oder kurzzyklischen Tätigkeiten muss also unabhängig von der Strategie (Rumination oder Aufmerksamkeit fokussieren) mit gesundheitsschädigenden Folgen gerechnet werden. Dieser Sachverhalt wurde vor allem während der Hochphase der wissenschaftlichen Betriebsführung deutlich (vgl. Ulich, 2011). Solche Teufelskreise können z. B. durch vollständige Tätigkeiten (Hacker & Sachse, 2014) vermieden werden, da hier aufgabenbezogene Herausforderungen gestellt werden. Bei hierarchisch vollständigen Tätigkeiten können während der Ausführung eher automatisierter Teiltätigkeiten Probleme anspruchsvollerer Teiltätigkeiten mental simuliert und gelöst werden, was selbstwertschädigende Simulationen verhindert und zugleich zu einer effizienteren Bewältigung von Arbeitsaufträgen führt (vgl. Hacker & Sachse, 2014). Bereits die mit der mentalen Problemlösung einhergehende Aussicht auf Erfolg und Vorbereitung der Ausführung wird von Dopaminausschüttung und positiven Emotionen begleitet und wirkt motivierend (Brown et al., 2012; Jones, 2003; LeDoux, 2012; Lord et al., 2010).

Um Arbeitssysteme kompetenzförderlich zu gestalten, sind auf Grund der Voraussetzungen bezüglich der Arousal-Systeme zunächst jene Punkte zu berücksichtigen, die weiter oben bezüglich der Sicherstellung der Homöostase beschrieben wurden. Insbesondere dürfen die Energie- und Stoffreserven nicht durch unkontrollierbare Stressreaktionsprozesse erschöpft werden (s. o.). Die Gestaltung des Arbeitsplatzes und der Arbeitsaufgaben stellen die *extrinsischen* Quellen der psychischen Belastung dar, die über die Kompetenzförderlichkeit entscheiden. Sie sollten nach dem Prinzip "difference-in-sameness" gestaltet sein. Das heißt, sie sollten an vorhandene Kompetenzen anknüpfen und um interessante Herausforderungen angereichert sein (White, 1959; Deci & Ryan, 2000). Interessant bedeutet hierbei, dass die Beschäftigten diese Herausforderungen prinzipiell freiwillig, ohne zusätzliche extrinsische Anreize (z. B. Aufforderungen oder Gratifikationen) aufsuchen würden (Deci & Ryan, 2000). Für die Gestaltung von Arbeitsplätzen und Arbeitsaufgaben ergibt sich hieraus die Option, Beschäftigten Gestaltungsfreiheiten bezüglich kompetenzförderlicher Arbeitsaufgaben einzuräumen (Wieland, 2014a; Wieland & Hammes, 2014a). Das heißt, das Arbeitssystem hält Elemente für Arbeitsaufgaben bereit, denen sich die Beschäftigten bei der Anreicherung ihrer eigenen Arbeitsaufgaben bedienen können (vgl. Sonntag & Stegmaier, 2007). Ulich (2011) stellt in diesem Zusammenhang die kompetenzförderlichen Effekte von differenzieller Arbeitsgestaltung und einer Abkehr vom Gedanken eines "one best way" der Arbeitsgestaltung heraus. Individuelle Unterschiede in Bezug auf Kompetenzen und ihre Entwicklung sollten bei der Arbeitsgestaltung Berücksichtigung finden und führen zu verbesserter Qualifikation und erhöhter Produktivität. Auch Hacker (Hacker & Sachse, 2014) beschreibt Lernpotenzial als

Verhältnis individuell vorhandener Kompetenzen und herausfordernden Anforderungen in der Arbeit. Bieten Arbeitsplätze Lernpotenzial, dann wirkt sich dies günstig auf die allgemeine und kognitive Leistungsfähigkeit, die psychische und körperliche Gesundheit, das Wohlbefinden sowie die persönliche Entwicklung aus und kann Alterungsprozesse verzögern. "Anforderungsarme Arbeitsprozesse *ohne* Lernpotenziale für Hinzulernen, Erhaltungslernen und ohne tätigkeits- und arbeitsprozessinterne Ansatzmöglichkeiten von 'intrinsischen' Lernmotiven führen [hingegen] auf verschiedenen Wegen zur *Verringerung* von Leistungsvoraussetzungen" (Hacker & Sachse, 2014, S. 495).

Die mentalen Simulationen eigenen kompetenten Verhaltens stellen die entsprechenden *intrinsischen* Quellen der psychischen Belastung dar. Es wurde dargestellt, dass diese Simulationen den Aufbau von Kompetenzen zeitlich vorwegnehmen. Weiter wurde argumentiert, dass diese Vorwegnahme die Herstellung und das Aufsuchen herausfordernder Situationen bedingen können. Voraussetzung für die arbeitsbezogene Beanspruchung dieser Simulationsprozesse ist ein Arbeitsauftrag, der im Rahmen des ∞ -Modells eine systemische Störung darstellt. Die Ausrichtung der Simulationsprozesse auf einen Arbeitsauftrag führt zur Interpretation desselben auf dem Hintergrund individueller, tätigkeitsbezogener Erfahrungen bzw. Kompetenzen. Zielvorstellungen sowie möglichst effiziente Pläne und Strategien über die Art und Weise der Zielerreichung (Operative Abbildsysteme; Hacker & Sachse, 2014) werden auf diesem Hintergrund entwickelt - ein Prozess, der in der Arbeitspsychologie als Redefinition der Aufgabe bezeichnet wird (Hacker & Sachse, 2014; Hackman, 1970; Ulich, 2011). Die Herstellung eines übereinstimmenden Verständnisses eines Arbeitsauftrages bei Auftraggeber und Auftragnehmer stellt daher u. U. eine wesentliche Gestaltungsdimension dar.

Beim Aufbau von Kompetenzen, aber auch beim Prozess der Zielerreichung sind Beschäftigte auf Feedback über die Effekte ihres Verhaltens in der Umwelt angewiesen. Sie müssen Informationen darüber erhalten, inwiefern sie die gestellten Anforderungen "gemeistert" (im Sinne von Bandura, 1977) bzw. bewältigt haben. Bei der Gestaltung von Arbeitsaufgaben ist daher darauf zu achten, dass die Beschäftigten jederzeit Feedback über diese Effekte erhalten können (Hacker & Sachse, 2014). Die Selbstwirksamkeitserwartung oder Kompetenzerwartung (self-efficacy; Bandura, 1977) stellt ein geeignetes Konstrukt dar, die Überzeugung der Beschäftigten im Hinblick auf ihre effektive Interaktion mit dem Arbeitssystem zu analysieren (Abele, Stief & Andrä, 2000; Latham & Pinder, 2005; Schyns & von Collani, 2002). Dies stellt sich auch für die gesundheitsbezogene Selbstwirksamkeit bzw. Gesundheitskompetenz trotz der spezifischen Verhaltens-Domäne "Gesundheit" als gültig heraus (Wieland & Hammes, 2009).

Eine weitere Option zur Gestaltung kompetenzförderlicher Arbeitssysteme besteht in der Berücksichtigung und Beeinflussung des individuellen kompetenzbezogenen Selbstbildes. Hiermit sind Führungsaufgaben angesprochen. Das Führungssystem (direkte Führungskraft oder selbstorganisierte Arbeitsgruppe) kann im Mitarbeitergespräch gemeinsam mit den einzelnen Beschäftigten den erwünschten Kompetenz-Aufbau vereinbaren und daraufhin die Gestaltung der Arbeit und der Arbeitsaufträge sowie geeignete Qualifikationsmaßnahmen initiieren (Felfe, 2009; Nerdinger, 2013; Sonntag & Stegmaier, 2007).

Arbeitssysteme, die in der beschriebenen Weise kompetenzförderlich gestaltet sind, weisen eine Reihe von arbeitspsychologischen Merkmalen auf, deren präventive, gesundheits- und persönlichkeitsförderliche Wirkung nachgewiesen ist: Sie sind *vollständig*, da sie insbesondere durch den herausfordernden Anteil das Central Executive Network über das Salient Network beanspruchen und somit intellektuelle Anforderungen stellen. Durch ihre Anknüpfung an vorhandene Kompetenzen erlauben sie jedoch auch den Rückgriff auf etablierte klassifizierende, algorithmische und konzeptionelle Denkprozesse (perzeptiv-begriffliche Anforderungen) sowie auf beherrschte sensumotorische Routinen (Hacker & Sachse, 2014; vgl. Sonntag & Stegmaier, 2007). Können Beschäftigte ihre Arbeitsaufgaben entsprechend ihrer Kompetenzen herausfordernd (mit-) gestalten, dann kann von angemessenen Anforderungen bei hoher Kontrolle ausgegangen werden. Untersuchungen zum Job-Demand-Control-Modell von Karasek (1979; Karasek & Theorell, 1990) belegen, dass solche Arbeitsaufgaben ("active jobs") mit einem aktiven Lebensstil und einem guten Gesundheitszustand einhergehen (a. a. O.; de Jonge, Dollard, Dormann, Le Blanc, & Houtman, 2000; Lerner, Levine, Malspeis & D'Agostino, 1994; Peter, 2010; Richter, Buruck, Nebel & Wolf, 2011; Semmer & Udris, 2007; Siegrist & Dragano, 2008; Taris & Feij, 2004; Ulich & Wülser, 2012; Wieland, 2010b). Ulich (2011; Ulich & Wülser, 2012) zählt darüber hinaus noch die Merkmale *Anforderungsvielfalt*, *Lern- und Entwicklungsmöglichkeiten* sowie *Zeitelastizität und stressfreie Regulierbarkeit* auf, die durch kompetenzförderliche Arbeitsgestaltung zum großen Teil abgedeckt sind. Mit den weiteren Merkmalen gesundheits- und persönlichkeitsförderlicher Arbeitsgestaltung, *Möglichkeit der sozialen Interaktion*, *Autonomie* und *Sinnhaftigkeit* (a. a. O.), sind das Bedürfnis nach sozialer Teilhabe und Zugehörigkeit sowie die autonome Natur des Menschen angesprochen.

3.3 SOZIALE TEILHABE UND ZUGEHÖRIGKEIT

Es ist allseitig bekannt: Der Mensch ist ein soziales Wesen. Soziale Interaktionen sowie die Teilhabe an sozialen Gruppen sind wesentliche Determinanten der individuellen Persönlichkeit.

Gesunde soziale Beziehungen sind getragen von gegenseitiger Empathie und Fürsorge. Die Teilhabe an einer sozialen Gruppe setzt aber auch den persönlichen Beitrag zu dieser Gruppe durch den Einsatz vorhandener Kompetenzen voraus. Diese Sachverhalte gelten im Rahmen von Familien, Freundeskreisen, Vereinen und Staaten genauso, wie für Beschäftigte in Unternehmen. Bei der Gestaltung von Arbeitssystemen gilt es daher, einerseits Möglichkeiten für die Pflege privater sozialer Beziehungen einzuräumen (Life-Domain-Balance). Andererseits ist es für ein Unternehmen erfolgversprechend, wenn Kommunikation und Kooperation auf der natürlichen Bildung gesunder sozialer Beziehungen gründet.

3.3.1 Theoretische Grundlegung

Bei sozialer Isolation, so schreibt bereits Maslow (1943), "... the person will feel keenly, as never before, the absence of friends, or a sweetheart, or a wife, or children. He will hunger for affectionate relations with people in general, namely, for a place in his group, and he will strive with great intensity to achieve this goal" (a. a. O., S. 381). Deci und Ryan (2000) beschreiben neben einem Bedürfnis nach Kompetenz ebenfalls ein Bedürfnis nach sozialer Teilhabe und Zugehörigkeit. Auch dieses Bedürfnis wird ihrer Ansicht nach intrinsisch, also ohne besondere äußere Umstände oder Anreize ausgelöst. In Übereinstimmung mit Baumeister und Leary (1995) vermuten sie den evolutiven Vorteil dieses angeborenen Bedürfnisses in materiellem Austausch (z. B. Nahrung, Produkte) sowie professioneller Spezialisierung und Kooperation in sozialen Gruppen (vgl. auch Baumeister, 2010; Heatherton, 2011). Baumeister und Leary (1995) spezifizieren auf empirischer Grundlage, welche Qualitäten und Quantitäten soziale Beziehungen zur Erfüllung dieses Bedürfnisses aufweisen müssen: "First, there is a need for frequent, affectively pleasant interactions with a few other people, and, second, these interactions must take place in the context of a temporally stable and enduring framework of affective concern for each other's welfare" (a. a. O., S. 497). Das heißt, es kommt weniger auf eine Vielzahl flüchtiger Bekanntschaften an, sondern vielmehr auf wenige tiefe Beziehungen, die durch häufige, angenehme Kontakte und gegenseitige Empathie und Fürsorge gekennzeichnet sind. Sie beschreiben, dass soziale Beziehungen relativ schnell und einfach, d. h. ohne besondere Umstände entstehen. Das Erleben einer gemeinsamen äußeren Gefahr oder Herausforderung kann die Bildung sozialer Beziehungen jedoch befördern. Außerdem schildern sie, dass es zu einem relativ starken Widerstand gegen die Aufhebung bestehender sozialer Beziehungen kommt. Dies sogar dann, wenn diese Beziehungen nicht besonders intensiv oder sogar unvorteilhaft für eine Person sind. Wird eine wichtige soziale Beziehung (z. B. Ehe- oder Lebenspartner, nahestehendes Familienmitglied, intime Freundschaft) aufgelöst, so wird sie in der Regel in dem Sinn ersetzt, dass das Bedürfnis nach sozialen Beziehungen

wieder ausreichend erfüllt wird (Baumeister & Leary, 1995; Deci & Ryan, 2000). Die Deprivation des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit sowie die (drohende) Auflösung wichtiger sozialer Beziehungen sind mit Stressreaktionsprozessen und langfristig mit Beeinträchtigungen der körperlichen und psychischen Gesundheit verbunden. Baumeister und Leary (1995) zählen Herzinfarkte, Tuberkulose, Krebs, Immunhemmung, Hypercortisolismus und psychische und Verhaltensstörungen auf und berichten darüber hinaus über Zusammenhänge mit Kriminalität und Suizidalität. Auch aus der Stressforschung ist bekannt, dass problematische soziale Beziehungen zu den bedeutendsten Auslösern von Stressreaktionsprozessen und entsprechenden Gesundheitsbeeinträchtigungen zählen (Gunnar & Quevedo, 2007; Hüther & Sachsse, 2007).

Heatherton (2011) beschreibt die neurophysiologische Grundlage des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit ganz im Sinne des ∞ -Modells. Selbstbezogene und soziale Kognitionen werden mit Strukturen des Default Mode Networks (DMN) assoziiert. Hierunter fasst er zum einen die Selbstwahrnehmung, zum anderen Mentalizing (Theory of Mind). Die Selbstwahrnehmung ist notwendig, um über das eigene Verhalten und dessen Wirkung reflektieren zu können sowie es an eigene soziale Normen angleichen zu können (vgl. auch Baumeister, 2010). Mit Mentalizing ist die Überzeugung verbunden, dass auch andere Menschen zu sozialen Kognitionen und Emotionen befähigt sind. Außerdem ist hiermit die Fähigkeit verbunden, diese Kognitionen und Emotionen Anderer zu verstehen und vorherzusagen. Insbesondere gilt dies für die Urteile anderer über das eigene Verhalten. Heatherton (2011) zufolge ist diese Fähigkeit die Voraussetzung für ein harmonisches Zusammenleben und eine gelingende Kooperation. Funktionierende soziale Beziehungen gehen mit positiven sozialen Emotionen (z. B. Zuneigung, Bewunderung, Stolz, Liebe) einher (vgl. Baumeister & Leary, 1995). Ein möglicher Ausschluss aus einer sozialen Gruppe ist mit negativen sozialen Emotionen (z. B. Schuld, Scham, Eifersucht, Neid) verbunden (vgl. Baumeister & Leary, 1995) und führt zur Aktivität des Salient Network (SN) sowie von Strukturen des Central Executive Network (CEN). Die Hinzuschaltung exekutiver Netzwerke ermöglicht die Hemmung sozial unerwünschter Emotionen, Gedanken und Verhaltensweisen (z. B. Vorurteile, Gewalt, Opportunismus, Unehrlichkeit) sowie die Generierung sozial erwünschter Emotionen, Gedanken und Verhaltensweisen (z. B. Unvoreingenommenheit, Höflichkeit, Freigiebigkeit).

Baumeister (2010) schreibt dem Selbst in dem von Heatherton (2011) beschriebenen Prozess eine vermittelnde Rolle zu: "[The self] is an interface between the animal body and the social system" (a. a. O., S. 139). "The self ... is formed in interactions and relationships with people.

Moreover, the self functions to create and sustain relationships, to fulfill important roles, and to keep a favored position in the social system" (a. a. O., S. 142). Im vorliegenden Zusammenhang sind vor allem seine Zusammenfassungen zu Selbstvertrauen und Selbstdarstellung zu berücksichtigen. Selbstvertrauen beschreibt er als inneren Maßstab, der Auskunft darüber gibt, inwieweit man von anderen Menschen sozial akzeptiert wird. Selbstdarstellung bezieht sich auf die Fähigkeit, sich gegenüber andern Menschen auf bestimmte Weise zu präsentieren. Darüber hinaus neigt der Mensch zu selbstwertdienlichen Verzerrungen. Das heißt, er generiert ein positives Bild über sich Selbst und seine Rolle im sozialen Umfeld. Diese inneren Prozesse erlauben ihm, sich gegenüber anderen Menschen auf erwünschte Weise zu präsentieren: "Inner processes serve interpersonal functions" (a. a. O., S. 155). An anderer Stelle (Baumeister et al., 2011) beschreibt er gemeinsam mit Kollegen die Bedeutung mentaler selbstbezogener und sozialer Simulationen für zukünftig erfolgreiches und effizientes Verhalten. Somit kann - in Analogie zur Argumentation bezüglich des Aufbaus von Kompetenzen weiter oben - auch das Bedürfnis nach sozialer Teilhabe und Zugehörigkeit im Sinne einer kybernetischen Annäherungsschleife formuliert werden: Gemäß des ∞ -Modells wird in der Doubleloop ein positives Bild des Selbst sowie der eigenen sozialen Rolle als Standard generiert. Aus der Interaktion mit dem sozialen Umfeld erhält man ein Feedback (soziale Emotionen). Erfahrene Diskrepanzen (negative soziale Emotionen) werden durch Anpassung des Verhaltens sowie der selbstbezogenen und sozialen Simulationen abgebaut. Das positive Selbstbild im sozialen Kontext kann somit als zeitlich vorweglaufender Standard für das Wachstum der sozialen Teilhabe und Zugehörigkeit angesehen werden. Again: "What a man can be, he must be" (Maslow, 1943, S. 382).

Die Bedürfnisse nach Kompetenz sowie nach sozialer Teilhabe und Zugehörigkeit weisen nicht nur eine formale und neurologische Analogie, sondern sogar eine enge Verschränkung auf. Die Akzeptanz eines Menschen durch eine soziale Gruppe basiert auf deren Bewertung seiner Aktivitäten. Baumeister und Leary (1995) formulieren dies so: "The need to belong should stimulate goal-directed activity designed to satisfy it" (a. a. O., S. 500). Zielbezogene Aktivitäten können sich auf die soziale Unterstützung der Angehörigen und Freunde (Empathie und Fürsorge), die Übernahme von Aufgaben im Haushalt der Familie oder die Übernahme von Verantwortung im Sportverein beziehen. Sie können sich aber auch auf Arbeitsaufgaben innerhalb einer Arbeitsgruppe oder den gesellschaftlichen Nutzen der erzeugten Produkte oder Dienstleistungen beziehen. Um sozial akzeptiert zu werden, sind demnach Kompetenzen erforderlich, die zur Herstellung eines sozialen Beitrages bzw. eines gesellschaftlichen Nutzens befähigen (Hacker & Sachse, 2014). Hiermit wird verständlich, dass sich zumindest im

Bildungsweg und Berufsleben Kompetenz-Aufbau und soziales Umfeld eines Menschen gegenseitig bedingen. Für solche Interaktionen und ihre Wechselwirkungen mit dem Freizeitverhalten über die Lebensspanne stehen einige empirische Befunde bereit (vgl. Rosenstiel, 2007; Semmer & Udris, 2007; Sonntag & Stegmaier, 2007). Auf Grund der Kongruenz zwischen beiden Bedürfnissen liegt es nahe, die Fähigkeit zur effektiven, d. h. Bedürfnis-erfüllenden Gestaltung sozialer Beziehungen (unter Einsatz von Mentalizing und angemessener Beachtung sozialer Emotionen) als soziale Kompetenz zu bezeichnen (vgl. Kanning, 2002).

3.3.2 Arbeitspsychologische Konsequenzen

Raum und Zeit sollte nicht nur für die Sicherstellung der Homöostase sondern auch für die Erfüllung des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit gewährt werden. Dies bezieht sich vor allem auf ausreichende Möglichkeiten, gegenseitige Empathie und Fürsorge in der Familie, im Freundeskreis und im privaten sozialen Umfeld zu erleben. Ulich (2011, Ulich & Wiese, 2011; Ulich & Wülser, 2012) schreibt in diesem Zusammenhang von Life Domain Balance. Dienstreisen, Arbeitszeiten und Urlaubsplanung sind so einzurichten, dass Arbeit und andere Lebensdomänen optimal vereinbar sind (vgl. Klump & Gemmiti, 2011; Rau, 2011; Ulich & Wiese, 2011). Auch Angebote zusätzlicher Sozialeinrichtungen und -leistungen (z. B. Betriebskindergarten, Tagesmütter, Sozialberatung) sind als betriebliche Maßnahmen denkbar (Hämming & Bauer, 2010; Ulich & Wiese, 2011). Maßnahmen, die geeignet sind, die Präsenzzeit am Arbeitsplatz zu erhöhen, bergen jedoch das Risiko, Möglichkeiten zum Erleben gegenseitiger Empathie und Fürsorge zwischen sich nahestehenden Menschen einzuschränken. Das Bedürfnis nach sozialer Teilhabe und Zugehörigkeit kann nur zum Teil durch informale Beziehungen am Arbeitsplatz (z. B. Interessengruppen, Freundschaften, romantische Beziehungen) kompensiert und ausgelebt werden, auch wenn sie sich in jeder Organisation relativ schnell und leicht entwickeln (White, 1959; Deci & Ryan, 2000; Rosenstiel, 2007). Auf Grund ihres Potenzials zur Bedürfnis-Erfüllung sollten informale Beziehungen am Arbeitsplatz dennoch eher gefördert, denn durch formale Strukturen (z. B. Hierarchie, Funktionsteilung) verhindert werden (Neuberger, 2002; Rosenstiel, 2007).

Bei Kenntnis des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit erscheinen die Beobachtungen und Schlussfolgerungen aus den Hawthorne- und Tavistock-Studien (zusammenfassend bei Ulich, 2011) wie Selbstverständlichkeiten: Soziale Beziehungen bilden sich ohne weiteres Zutun in Abhängigkeit eingesetzter Technik, sie bieten Sicherheit und Unterstützung und sie beeinflussen Zufriedenheit und Leistung der einzelnen Beschäftigten sowie der Gruppe als Ganze. Diese Erkenntnisse führten zum Konzept der soziotechnischen

Systemgestaltung und sind in arbeitspsychologische Konzepte der Gruppenarbeit eingeflossen. Auf Grund nachgewiesener Erfolge ist das Konzept der teilautonomen Arbeitsgruppe besonders hervorzuheben (Ulich, 2011): Relativ unabhängigen Arbeitseinheiten werden vollständige Aufträge als Kernaufgaben übertragen. Sämtliche Teilaufgaben sollten einen inhaltlichen Gesamtzusammenhang aufweisen und innerhalb der Arbeitseinheit zu Kooperations-Erfordernissen führen. Das Feedback über das Arbeitsergebnis, das mit Gruppenentlohnung gekoppelt sein kann, sollte auf die Leistung der gesamten Arbeitsgruppe, weniger jedoch auf Leistungen Einzelner zurückzuführen sein. Die Autonomie der Arbeitsgruppe bezieht sich zunächst auf die interne Selbstorganisation bezüglich der Arbeitsteilung sowie der Art und Weise der Auftragserfüllung. Hiermit kann die Arbeitsgruppe flexibel auf Störungen und Schwankungen reagieren. Die Autonomie kann sich jedoch auch auf Entscheidungen über die Annahme von Arbeitsaufträgen, die Zusammensetzung der Arbeitsgruppe (Personalfragen), die Anschaffung und den Einsatz von Arbeitsmitteln bzw. Technik, die Entlohnung der einzelnen Gruppenmitglieder und andere strategische und organisatorische Entscheidungen erstrecken. Ulich (2011) führt zahlreiche Praxisbeispiele an, in denen die Einführung teilautonomer Arbeitsgruppen zu erhöhter Leistung, Zufriedenheit und Motivation sowie verbesserter Qualifikation führten. Die Produktivität betreffender Unternehmen stieg zum Teil erheblich. In teilautonomen Arbeitsgruppen leistet jedes Gruppenmitglied seinen sozial wertvollen Beitrag (Kompetenz) bei der Bewältigung einer gemeinsamen Herausforderung (Arbeitsauftrag an die Gruppe) und im Sinne allseitigen Gewinns (Anerkennung der Leistung, Lohn, soziale Beziehungen, Kompetenz-Aufbau etc.). Ihre Erfolge scheinen zu einem gewissen Teil auf das Bedürfnis nach sozialer Teilhabe und Zugehörigkeit zurückzuführen zu sein.

Die Einführung teilautonomer Arbeitsgruppen führt u. a. zu einer Verflachung der hierarchischen Organisation (Ulich, 2011). Führungskräften kommt in einer teilautonomen Gruppe weniger die direkte Weisungsbefugnis gegenüber den Gruppenmitgliedern sondern vielmehr die Vertretung der Gruppe gegenüber außenstehenden Interessenvertretern zu (sog. Grenzregulation; Ulich, 2011). Einige teilautonome Arbeitsgruppen weisen überhaupt keine formale Führungskraft auf, sondern treffen sämtliche Entscheidungen selbstorganisiert. Gibt es eine formale Führungskraft mit Weisungsbefugnis, so stellt sich die Frage, wie die Führungskraft-Geführten-Beziehung persönlichkeits- und gesundheitsförderlich und unter Berücksichtigung von Unternehmenszielen gestaltet werden kann. Inzwischen liegen einige Überblicksarbeiten vor, die belegen, dass u. a. eine Kombination hoch ausgeprägter mitarbeiter- und aufgabenorientierter sowie nicht-autoritärer Führung positiv mit Führungserfolg (Leistung, Zufriedenheit, Gesundheit) in Zusammenhang steht (Gregersen, Kuhnert, Zimmer & Nienhaus,

2011; Neyberg, Bernin & Theorell, 2005; Judge, Piccolo & Ilies, 2004; Rowolt & Heinitz, 2008; Stadler & Spieß, 2005; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Darüber hinaus ist eine partizipative Unternehmenskultur mit vielen Möglichkeiten der Mitbestimmung für die Beschäftigten positiv mit Gesundheit assoziiert (Neyberg et al., 2005). "Mitarbeiterorientierung steht hier im Wesentlichen für die Wertschätzung, Achtung und Sorge für den Mitarbeiter sowie die Offenheit und Zugänglichkeit des Vorgesetzten, während die Leistungsorientierung [bzw. Aufgabenorientierung] ein Verhalten beschreibt, das durch die Klärung von Zielen und Beaufsichtigung (Kontrolle) durch den Vorgesetzten sowie die Aktivierung der Leistungsmotivation der Mitarbeiter gekennzeichnet ist" (Wegge & Rosenstiel, 2007, S. 490). Demnach spricht mitarbeiterorientierte Führung das Bedürfnis nach sozialer Teilhabe und Zugehörigkeit an, während aufgabenorientierte Führung auf den zielbezogenen Einsatz der Kompetenzen der Geführten abzielt. Wieland, Winizuk und Hammes (2009) nennen eine Kombination beider Führungsstile "ganzheitliche Führung" und zeigen, dass dieser Führungsstil in Kombination mit persönlichkeits- und gesundheitsförderlich gestalteten Arbeitsplätzen mit einem besonders guten Gesundheitszustand in Zusammenhang steht. Felfe (2009) hebt über die genannten Führungsstile hinaus den kompetenten Einsatz von Führungsinstrumenten besonders hervor. Er betont, dass Mitarbeitergespräche, Zielvereinbarungen sowie Feedback- und Qualifizierungsgespräche zwischen Führungskraft und Geführten jeweils als Verhandlungen zu gestalten sind. Die Interessen des Unternehmens, vermittelt über die Führungskraft, sowie die Interessen der Geführten sollen auf diese Weise optimal aufeinander abgestimmt werden.

In einer sozialen Gemeinschaft unterliegt sämtliches beobachtbares Verhalten einer sozialen Bewertung. Die mental vorweggenommene oder erfahrene soziale Bewertung verschiedener Verhaltensalternativen ist für ein Individuum handlungsleitend (Baumeister, 2010; Baumeister & Leary, 1995; Heatherton, 2011). Die Arbeitstätigkeit und der Arbeitsinhalt stellen dabei eine zentrale Kategorie dar. „Arbeitstätigkeiten sind auch im Grenzfall kooperationsloser Einzeltätigkeiten gesellschaftliche Vorgänge. Sie ergeben sich aus der gesellschaftlichen Arbeitsteilung und werden wegen der Erzeugung eines Ergebnisses für andere (gegen Lohn) stets bewertet. Diese Bewertung wird vom arbeitenden Menschen vorweggenommen und reguliert in dieser Vorwegnahme seine Tätigkeit in ausschlaggebenden Umfang mit“ (Hacker & Sache, 2014, S. 38). Diese Bewertungen erfolgen durch unterschiedlichste Personenkreise. In der Arbeit sind es die Kolleginnen und Kollegen sowie die Vorgesetzten und ggf. auch die Geschäftspartner und Kunden, die Ihre Zufriedenheit mit der erbrachten Leistung über das Ausmaß an Anerkennung und Wertschätzung, aber auch über die Höhe der monetären

Gegenleistung ausdrücken. Außerhalb der Arbeit sind das Unternehmen, für das man arbeitet, die Arbeitstätigkeit sowie der dadurch ermöglichte Lebensstil ebenfalls Gegenstände sozialer Bewertung. Sinnhafte Aufgaben geben "das Gefühl, an der Erstellung gesellschaftlich nützlicher Produkte beteiligt zu sein" (Ulich & Wülser, 2012, S. 253) und gehen mit einer positiven sozialen Bewertung der eigenen Arbeit einher. Jemand, der einer sinnhaften Arbeitstätigkeit nachgeht, kann daher von einer "Übereinstimmung individueller und gesellschaftlicher Interessen" (a. a. O.) ausgehen. Bei der Gestaltung von Arbeit ist daher neben der unternehmensinternen Anerkennung und Wertschätzung der Arbeitsleistungen (z. B. durch Führungskräfte) auch auf die gesellschaftliche Bewertung der jeweiligen Tätigkeit zu achten und diese den Beschäftigten ggf. zugänglich zu machen bzw. zu verdeutlichen. Semmer und Udris (2007) fassen unter Rückgriff auf Forschungen zur Arbeitslosigkeit fünf psychosoziale Funktionen der Erwerbsarbeit zusammen: (a) Aktivität und Kompetenz, (b) Zeitstrukturierung, (c) Kooperation und Kontakt, (d) soziale Anerkennung und (e) persönliche Identität. Wären Sinnhaftigkeit und soziale Anerkennung der Arbeitstätigkeit nicht gegeben oder nicht erkennbar und wäre sie zusätzlich von routinierter und sozial isolierter Einzeltätigkeit geprägt, so wäre sie ihrer psychosozialen Funktionen fast vollständig beraubt. Eine solche Form der Arbeit käme psychosozial der Arbeitslosigkeit (mit all ihren Folgen) schon sehr nahe.

Der individuellen Bewertung sozialer Gegenleistungen für erbrachte Arbeitsleistungen kommt im Modell der beruflichen Gratifikationskrisen von Siegrist (1996a, b, 2002) eine besondere Bedeutung zu. Gratifikationen können Siegrist zufolge im Lohn/Gehalt, in Wertschätzung sowie in Arbeitssicherheit, Aufstiegschancen und qualifikationsgerechter Beschäftigung bestehen. Diesen Gratifikationen steht die persönliche Verausgabung zur Erfüllung beruflicher Anforderungen und Verpflichtungen gegenüber. Kommen Beschäftigte zu dem Schluss, dass Verausgabungen und Gratifikationen nicht im erwarteten Verhältnis stehen (Verletzung der sozialen Reziprozität), dann kommt es zu Gratifikationskrisen (vgl. Richter et al., 2011; Ulich & Wülser, 2012; Wieland, 2010b). Empirische Befunde belegen, dass diese Krisen mit körperlichen (Herz-Kreislauf-Erkrankungen) und psychischen Beeinträchtigungen und Erkrankungen (Depression, Burnout, Alkoholabhängigkeit u. a.) positiv im Zusammenhang stehen (Überblick bei: Siegrist & Dragano, 2008; Van Vegchel, De Jonge, Bosma & Schaufeli, 2005). Persönlichkeits- und gesundheitsförderliche Arbeitssysteme sollten daher über ein gerechtes, transparentes und möglicher Weise sogar partizipatives Gratifikationssystem verfügen, das den Abgleich der persönlichen Erwartungen mit den sozial-situativen Möglichkeiten erleichtert.

3.4 DAS SPANNUNGSFELD VON AUTONOMIE UND SELBSTKONTROLLE

Besonders wohl fühlen sich Menschen vermutlich, wenn sie alle ihre persönlichen Bedürfnisse uneingeschränkt in der persönlich bevorzugten Art und Weise erfüllen können. Im Umkehrschluss bedeutet dies, sich nicht an äußere oder verinnerlichte physisch-soziale Bedingungen anpassen zu müssen. Allein die Tatsache, dass man für die Erfüllung persönlicher Bedürfnisse auf Kompetenzen sowie auf tragfähige soziale Beziehungen angewiesen ist, führt jedoch zu der Erfordernis, sich in gewissem Ausmaß physisch-sozialen Bedingungen anzupassen. D. h. man kann seine persönlichen Bedürfnisse nicht immer bzw. nicht immer in der persönlich bevorzugten Art und Weise umfassend erfüllen. Auf diese Weise entsteht ein Spannungsfeld zwischen Autonomie und Selbstkontrolle. Selbstkontrolle, also die gezielte Anpassung des eigenen Verhaltens an die äußeren Bedingungen, ist deshalb schwierig und begrenzt, weil dafür zusätzliche Energie- und Leistungsressourcen aufgewendet werden müssen. Daher neigt man gemeinhin dazu, die persönlich bevorzugte Art und Weise der Bedürfniserfüllung (Autonomie) gegenüber Selbstkontrolle vorzuziehen. Autonomie in der Arbeit bedeutet, Arbeitsaufträge in der persönlich bevorzugten Art und Weise zu erfüllen. Dies schont persönliche Ressourcen, schützt die Gesundheit und führt unter Berücksichtigung der Bedürfnisse nach Kompetenz sowie sozialer Teilhabe und Zugehörigkeit zu persönlicher Entwicklung. Um Arbeitsaufträge erfüllen zu können, müssen jedoch zunächst die erforderlichen Qualifikationen erworben sowie die sozialen Normen und Werte der Organisation verinnerlicht werden. Darüber hinaus kann es zu unvorhergesehenen Störungen in der Arbeit kommen, die ebenfalls bewältigt werden müssen. Damit sind aufgabenbezogene Quellen für Anforderungen an die Selbstkontrolle umrissen. Um persönliche Ressourcen zu schonen sowie die Gesundheit von Beschäftigten zu schützen, wird im Folgenden gefordert, *vermeidbare* Selbstkontrollanforderungen in der Arbeit durch geeignete Maßnahmen der Arbeitsgestaltung zu minimieren sowie Möglichkeiten zur autonomen Erfüllung von Arbeitsaufträgen soweit es geht zu erweitern. Es wird sich zeigen, dass hierfür bereits einige gut ausgearbeitete arbeitspsychologische Konzepte (Tätigkeitsspielraum, Ursachen von Fehlhandlungen, Regulationsbehinderungen, Führungsverhalten) vorliegen.

3.4.1 Theoretische Grundlegung

Neben dem Bedürfnis nach Kompetenz und dem Bedürfnis nach sozialer Teilhabe und Zugehörigkeit postulieren Deci und Ryan (2000) auch ein Bedürfnis nach Autonomie. Sie spannen ein Kontinuum auf, in dem Fremdkontrolle als der Autonomie gegenüberliegender Pol abgebildet wird. Fremdkontrolliertes Verhalten basiert auf extrinsischer Motivierung. D. h. andere Personen kontrollieren die Konsequenzen des gezeigten Verhaltens. Autonomie ist Deci

und Ryan (2000) zufolge dann möglich, wenn die dieser Kontrolle zu Grunde liegenden sozialen Regeln, Normen und Werte weitest möglich in das persönliche Wertesystem integriert werden konnten. In diesem Fall, so die Autoren, sind autonomes und intrinsisches Verhalten für die betreffende Person kaum mehr zu unterscheiden. Sie handelt ohne weiteres Zutun entsprechend der vorherrschenden sozialen Normen. Problematisch an dieser Konzeption der Autonomie ist, dass (a) Autonomie als Maß der Erfüllung des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit interpretiert werden kann (Autonomie ist Identifikation mit den Gruppennormen und erzeugt ausschließlich gruppenkonformes Verhalten), (b) die scharfe Trennung oder konsequente Identifikation von Autonomie und intrinsischer Motivation vermieden wird und (c) soziale Regeln, Normen und Werte als lokale und dynamische Rahmenbedingungen für das Verhalten kaum als verlässliches Fundament für Autonomie geeignet scheinen.

In der vorliegenden Arbeit wird die Auffassung vertreten, dass der Mensch nicht ein Bedürfnis nach Autonomie *hat*, sondern grundsätzlich ein autonomes Wesen *ist* (vgl. Allport, 1937). Der Mensch handelt unter den mental repräsentierten Möglichkeiten der Bedürfnis-Erfüllung in der durch Erfahrung vermittelten, persönlich bevorzugten Art und Weise. Ein Säugling schreit in Abwesenheit der Mutter nicht, weil es noch nicht autonom ist, sondern weil dies die persönlich bevorzugte (in diesem Falle noch die einzig beherrschte) Art und Weise ist, die Erfüllung des Bedürfnisses nach der Nähe zur Mutter herbeizuführen. Allgemein schränkt die mentale Repräsentation der physisch-sozialen Rahmenbedingungen (extern oder internalisiert) jedoch die Möglichkeiten der Bedürfnis-Erfüllung ein. Sind die momentan vorherrschenden Bedürfnisse oder die persönlichen Präferenzen bezüglich der Art und Weise ihrer Erfüllung nicht mit der mentalen Repräsentation der physisch-sozialen Situation vereinbar, dann müssen angemessenere Repräsentationen und/oder Verhaltensweisen generiert und ggf. ausgeführt werden. Diese Anpassung des internen Zustandes und der Verhaltensweisen an die gegebenen physisch-sozialen Rahmenbedingungen und Anforderungen wird mit Selbstregulation bezeichnet (Baumeister, 2010; Heatherton, 2011; Lord et al., 2010). Hierbei ist es nützlich, zwischen automatisierter und kontrollierter Selbstregulation zu unterscheiden. Kontrollierte Selbstregulation erfordert die bewusste Planung und Steuerung des eigenen Verhaltens unter Einsatz kognitiver Funktionen. Sie wird mit Selbstkontrolle bezeichnet (vgl. Gollwitzer, 1999; Kuhl, 1996). Weiter unten wird dargestellt, dass Selbstkontrolle mit dem energetisch aufwändigen Hinzuschalten des Central Executive Network einhergeht. Außerdem wird dargestellt, dass Selbstkontrolle eine Voraussetzung für das Erlernen und Festigen angemessener Verhaltensweisen in neuartigen

Situationen ist. Biopsychologisch stellt das Spannungsfeld von Autonomie und Selbstkontrolle somit ein energetisch-strukturelles dar. Energetisch aufwändige Selbstkontrolle ist nicht notwendig, wenn bereits die inneren (strukturellen) Voraussetzungen für ein angemessenes und zugleich persönlich bevorzugtes Verhalten in einer gegebenen Situation gelegt sind (a. a. O.). Insofern kommen wir wieder auf Deci und Ryan (2000) zurück: Sind angemessene Verhaltensweisen in bestimmten physisch-sozialen Situation durch Lernprozesse als persönlich bevorzugte Verhaltensweisen zur Bedürfnis-Erfüllung internalisiert (und wenn dies auch "nur" der Sicherstellung der sozialen Zugehörigkeit dient), so ist dies als autonomes Verhalten zu bezeichnen. Die ökologische Begründung autonomen Verhaltens ist jedoch eine energetische. Und den Ausgangspunkt für die Entwicklung autonomen Verhaltens bilden physisch-soziale Herausforderungen für die Bedürfnis-Erfüllung (systemische Störungen), die nicht auf soziale Fremdkontrolle beschränkt sind.

Solange bei den mental repräsentierten Möglichkeiten der Bedürfniserfüllung automatisierte oder persönlich bevorzugte (primäre, erlernte oder durch Simulation vorgebahnte) Verhaltensweisen eingesetzt werden können, die zur tatsächlichen Bedürfniserfüllung führen, kann das Individuum autonom und effizient handeln. Auslöser für Selbstkontrollprozesse ist eine mentale Repräsentation der gegenwärtigen Situation, die im Konflikt steht mit aktuellen Bedürfnissen oder der persönlich bevorzugten Art und Weise ihrer Erfüllung. In einer solchen Situation kann das Individuum schwerlich Anreize zur Erfüllung der aktuellen Bedürfnisse erkennen oder schwerlich persönlich bevorzugte Verhaltensweisen abrufen, die zur Bedürfniserfüllung geeignet erscheinen (vgl. Gollwitzer, 1999; Kuhl, 1996). Implizit sind Konflikte zwischen inhaltlich verschiedenen und insbesondere hierarchisch unterscheidbaren (sog. primären und sekundären) Bedürfnissen enthalten. Derartige Konflikte machen sich biopsychologisch auf Ebene der Arousal-Systeme sowie auf kortikaler Ebene bemerkbar. Auf Ebene der Arousal-Systeme werden sensumotorische Aufmerksamkeit und Reaktionsbereitschaft gesteigert (Brown et al., 2012; Hüther et al., 1999; Jones, 2003). Stressreaktionsprozesse und die Mobilisierung von Energie werden initiiert (a. a. O.). Es bilden sich Core Affects mit hoher Aktivität und negativer Valenz (hohe Anspannung bei Scherer, 2009). Auf kortikaler Ebene wird das Salient Network über den anterioren cingulären Kortex (Konflikt-Monitoring) und die Insula (Integration viszeraler, sensorischer, motorischer und emotionaler Information) aktiviert (Botvinick et al., 2004; Brentson et al., 2011; Bush et al., 2000; Carter & van Veen, 2007; Critshley, 2009; Kerns et al., 2004; Ochsner & Gross, 2005). Dies führt zu einer Hemmung des Default Mode Network und einer Aktivierung des Central Executive Network (Fox et al., 2005; Menon, 2011; van den Heuvel & Hulshoff Pol, 2010;

Vincent et al., 2008). Die Doubleloop (simulierte Problemlösung) wird gestört und das CEN wird für die Generierung angemessener Verhaltensweisen beansprucht (vgl. Menon, 2011; Miller & Cohen, 2001; Miller et al., 2002). Das heißt, mentale Repräsentationen der gegenwärtigen Situation werden mittels Suche nach Möglichkeiten zur Bedürfniserfüllung (Anreize) angepasst und es werden Verhaltensweisen generiert, die unter diesen neu entdeckten Anreizen zu einer erfolgreichen Bedürfniserfüllung führen sollen. Dies kann zuweilen durch eine Doubleloop erfolgen, in der die aktuelle Situation mit Erfahrungen aus der Vergangenheit verglichen wird, in denen ähnliche Situationen erfolgreich bewältigt werden konnten (Nutzung des autobiografischen und episodischen Gedächtnisses; vgl. Baumeister et al., 2011; Andrews-Hanna, 2012; Heatherton, 2011; Lord et al., 2010). Bei äußerlicher motorischer Ruhe kann so ein angemessenes Verhaltensprogramm für die gegenwärtige Situation generiert werden. Bezüglich der Arousal-Systeme ist bekannt, dass eine erhöhte Ausschüttung von Serotonin u. a. mit erhöhter kortikaler Aktivität bei motorischer Ruhe in stressreichen Situationen einhergeht (Brown et al., 2012; Jones, 2003). Führen diese Prozesse bei aller Anstrengung nicht zur Generierung angemessener Verhaltensweisen, so schaukelt sich der Stressreaktionsprozess ggf. bis hin zu einem unkontrollierbaren Stressreaktionsprozess mit allen Konsequenzen auf (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther, 2014; Hüther et al., 1999; Sapolsky, 2003). Insbesondere werden die Energieressourcen für Selbstkontrollprozesse stark ausgeschöpft („ego-depletion“; vgl. Baumeister et al., 1998). Das Individuum verharrt in einer Lage geringer Aktivität mit negativer Valenz (geringe Kontrolle bei Scherer, 2009). Kuhl (2001) spricht bei Menschen, die sich generell schlecht aus solchen Gedankenspiralen lösen können, treffend von Lageorientierung. Können angemessene Verhaltensweisen generiert oder abgerufen werden, so sind sie instrumentell im Dienste der Bedürfniserfüllung und wirken u. a. durch Dopaminausschüttung motivierend (Hüther et al., 1999; Jones, 2003; LeDoux, 2012; Lord et al., 2010). Es entstehen Core Affekts mit hoher Aktiviertheit und positiver Valenz (hohe Kontrolle bei Scherer, 2009). Die Verhaltensweisen werden umgesetzt und ihre Effekte in der aktuellen Situation bezüglich neuer Möglichkeiten zur Bedürfniserfüllung oder bezüglich tatsächlich stattgefundenener Bedürfniserfüllung evaluiert (LeDoux, 2012). Im Sinne der Bedürfniserfüllung erfolgreich umgesetzte Verhaltensweisen werden verstärkt und sind in zukünftigen, ähnlichen Situationen leichter abrufbar (Lernprozesse; Gunnar & Quevedo, 2007; LeDoux, 2012; Hüther et al., 1999; Hüther, 2014; Sapolsky, 2003). Durch zunehmende Erfahrungen mit ähnlichen Situationen kommt es zu einer Internalisierung der angemessenen Verhaltensweisen. In sozialen Kontexten betrifft dies insbesondere die Internalisierung sozialer Werte, Normen und Verhaltensregeln bzw. sozialer Pflichten (vgl. Carver et al., 2000; Deci &

Ryan, 2000; Higgins, 1996; Lord et al., 2010). Feedback über soziale Effekte der Selbstregulation wird über soziale Emotionen vermittelt (Baumeister, 2010; Baumeister & Leary, 1995; Heatherton, 2011; Lord et al., 2010). Verbleiben nach der Verhaltensaufführung weiterhin Konflikte, so hält der Selbstkontrollprozess an. Können in der neuen Situation automatisierte oder persönlich bevorzugte Verhaltensweisen zur Bedürfniserfüllung eingesetzt werden, so ist Selbstkontrolle nicht länger erforderlich. Sie ist erst bei der erneuten Entdeckung eines Konfliktes wieder notwendig. Bis dahin kann das Individuum autonom handeln.

Weiter oben wurde die beschriebene Konzeption in Anlehnung an Heatherton (2011) bezüglich des Bedürfnisses nach sozialer Teilhabe und Zugehörigkeit dargestellt. Hier wurde eine Verallgemeinerung für die gesamte Bedürfnis-Triade (Sicherstellung der Homöostase, Bedürfnis nach Kompetenz und Bedürfnis nach sozialer Teilhabe und Zugehörigkeit) sowie eine Integration im Sinne des ∞ -Modells vorgenommen. Zudem ist die beschriebene Konzeption mit kybernetischen Vorstellungen über Selbstregulationsprozesse vereinbar (Baumeister, 2010; Carver, Sutton & Scheier, 2000; Hacker & Sachse, 2014; Lord et al., 2010; Miller et al., 1960). Nach diesen Vorstellungen können Selbstregulationsprozesse wie folgt als negative Feedbackschleife beschrieben werden (vgl. Hacker & Sachse, 2014; Lord et al., 2010; Miller et al., 1960): Ein kognitiv repräsentiertes Ziel (Standard) sowie Informationen über die Situation und über das Ergebnis eigenen Verhaltens (Feedback) werden mittels eines Comparators gegeneinander abgeglichen. Wird eine Diskrepanz zwischen Standard und Ergebnis festgestellt, so werden das Verhalten (z. B. Anstrengung) oder Kognitionen (z. B. Interpretationen des Ziels, des Ergebnisses oder der Diskrepanz) zum Zwecke der Beseitigung der Diskrepanz verändert, Selbstkontrollprozesse setzen ein. Angepasst an die gegebene Situation wird schließlich das intendierte Verhalten ausgeführt. Informationen über das Ergebnis dieses Verhaltens werden wiederum dem Comparator als Feedback zugeführt und die Feedbackschleife schließt sich. Sie wird solange aufrechterhalten, bis die Diskrepanz beseitigt ist. Solche Schleifen können insofern hierarchisch verschachtelt sein, dass verhaltensnahe Prozesse zum Gelingen übergeordneter, abstrakterer Ziele beitragen (vgl. Hacker & Sachse, 2014; Lord et al., 2010; Miller et al., 1960). Im Rahmen des ∞ -Modells bestehen Ziele in der Erfüllung von Bedürfnissen. Der Comparator ist mit den Arousal-Systemen und dem Salient Network gleichzusetzen und steht in enger Verbindung mit emotionaler Evaluation und Aktivierung (vgl. auch Baumeister, 2010; Baumeister & Leary, 1995; Carver et al., 2000; Heatherton, 2011; LeDoux, 2012; Lord et al., 2010). Die Veränderung von Kognitionen und Verhalten erfolgt im Central Executive Network bzw. einer problembezogenen Doubleloop. Weiter oben wurde beschrieben, dass sogar so genannte Wachstumsbedürfnisse (vgl. Maslow,

1943; White, 1959) bezüglich der Kompetenz und der sozialen Teilhabe und Zugehörigkeit in kybernetischer Weise aufgefasst werden können: In der Doubleloop werden effiziente Verhaltensweisen für zukünftige Situationen generiert (vgl. Andrews-Hanna, 2012; Baumeister et al., 2011), die in Situationen mit entsprechenden Anreizen umgesetzt werden und so zur Selbstaktualisierung beitragen. Gollwitzer (1999) hebt in diesem Zusammenhang die unterstützende Funktion von Implementierungs-Intentionen zur Realisierung persönlicher Ziele hervor. Es handelt sich dabei um Instruktionen der Form: „Whenever situation x arises, I will initiate the goal-directed response y!“ (a. a. O., S. 493).

Carver, Sutton und Scheier (2000) gehen davon aus, dass es neben negativen Feedbackschleifen auch positive Feedbackschleifen geben kann. Negative Feedbackschleifen dienen dazu, Diskrepanzen zwischen dem Ist-Zustand und dem avisierten Ziel zu verringern. Bei den Zielen handelt es sich um Annäherungsziele. Gelingt die Zielerreichung, so geht dies mit Begeisterung ("Elation"; a. a. O., S. 745) einher. Misslingt sie, dann folgt Depressivität ("Depression", a. a. O.). Bei positiven Feedbackschleifen soll eine bestimmte Situation vermieden werden, also der Abstand zu dieser Situation möglichst vergrößert werden. Entsprechende Ziele heißen Vermeidungsziele. Gelingt hier die Zielerreichung, dann erleben die Betroffenen Erleichterung ("Relief", a. a. O.). Gelingt sie nicht, dann folgt Angst ("Anxiety", a. a. O.). Problematisch an Vermeidungszielen ist nach Carver et al. (2000) die Tatsache, dass der Abstand zur zu vermeidenden Situation theoretisch unendlich wachsen müsse. Sie gehen jedoch davon aus, dass neue Umweltreize Anreizcharakter bekommen können, sodass die betreffende Person im Reizumfeld zwischen Annäherungs- und Vermeidungszielen alterniert. Sie unterscheiden Personen, die eher appetitive Reize aufsuchen von Personen, die eher aversive Reize vermeiden und weisen ihnen Extraversion respektive Neurotizismus als Persönlichkeitseigenschaften zu. In diesem Zusammenhang nehmen sie Bezug auf die Feldtheorie Lewins, der Verhalten beschreibt als "Behavior = Function of person and environment = Function of live-space ($Be = F[P, E] = F[L Sp]$)" (Lewin, 1939, S. 878).

Die situationsbedingte bzw. problembezogene Störung der Doubleloop und insbesondere die verstärkte Beanspruchung des Central Executive Systems gehen bei Selbstkontrollprozessen mit einem erhöhten Energieverbrauch einher (Hüther et al., 1999; Miller & Cohen, 2001; Miller et al., 2002). Allein diese Tatsache beschränkt die individuellen Kapazitäten zur Selbstkontrolle, da die körperlich verfügbare Energie begrenzt ist (vgl. „ego-depletion“ bei Baumeister et al., 1998). Daher wurde weiter oben in Übereinstimmung mit Hockey (1997) und Sanders (1982) ein Anstrengungs- bzw. Effort-Monitor postuliert, der über Kortisol-Ausschüttung das Ausmaß der Anstrengung und Leistung in Abhängigkeit von den

Anforderungen nach einem konkaven Muster (umgekehrt U-förmig) moduliert (DeRijk & de Kloet, 2006; Foy et al., 2006; Hüther et al., 1999; Sapolsky, 2003). Darüber hinaus sind verschiedene kognitive Funktionen über unterschiedliche kortikale Netzwerke verteilt (Menon, 2011; van den Heuvel & Hulshoff Pol, 2010; Vincent et al., 2008). Jedes dieser Netzwerke verfügt über je eine begrenzte Zahl an Neuronen und Rezeptoren für Neurotransmitter und Neuromodulatoren die dessen Kapazität begrenzen. Selbst ein einzelnes Neuron, dessen Kapazitäten z. B. durch maximale Besetzung der Rezeptoren ausgeschöpft sind, kann daher zu einem "Flaschenhals" für eine folgende Beanspruchung der Funktionen seines Netzwerkes werden (vgl. Penzlin, 2014). Die zahlreichen Möglichkeiten über den Verlauf der Leistung in Abhängigkeit der Beanspruchung von Leistungsfunktionen, wie sie z. B. Wieland-Eckelmann (1992) beschreibt, kann daher entsprechend des Massenwirkungsgesetzes wie folgt zusammengefasst werden: Der eine Leistung limitierende Faktor ist jene beanspruchte Leistungsfunktion (funktionales neuronales Netzwerk) mit der geringsten verfügbaren Kapazität zur Signalverarbeitung und -weiterleitung. In Bezug auf die Selbstkontrolle bedeutet diese Feststellung, dass der Selbstkontrolle sowohl energetische als auch strukturelle Grenzen gesetzt sind (vgl. Baumeister et al., 1998; Heatherton, 2011; Lord et al., 2010; Wieland-Eckelmann, 1992). Sind diese Grenzen erreicht, dann ist das Individuum bald nur noch begrenzt oder gar nicht mehr fähig, angemessene Verhaltensweisen in unbekanntem oder herausfordernden Situationen zu generieren. Diese Ausschöpfung der Kapazitäten kann mehrere Gründe haben: (a) allgemeine energetische Ressourcen (Glukose, Glutamat) gehen durch Aktivität zur Neige (allgemeine Erschöpfung, Müdigkeit), (b) die Situation lässt sich trotz Anstrengung nicht im Sinne der Bedürfniserfüllung gestalten, (c) das Individuum kann trotz mentaler Anstrengung keine geeigneten Verhaltensweisen generieren, (d) es treten beständig neue und möglicherweise konfligierende Bedürfnisse durch An- oder Hinweisreize (external oder internal) auf (Baumeister et al., 1998; Gollwitzer, 1999; Heatherton, 2011; Kuhl, 1996; LeDoux, 2012; Lord et al., 2010). Betroffene zeigen weniger Engagement oder geben ihr Ziel vollständig auf (Baumeister et al., 1998; Hockey, 1997; Lord et al., 2010; Wieland-Eckelmann, 1992). Kapazitäten bestimmter Leistungsfunktionen oder allgemeine energetische Ressourcen müssen regeneriert (Tätigkeitswechsel, Pausen, Schlaf) und ggf. durch metabolische Prozesse (Energie- und Stoffzuführung, Neuroplastizität) auf- und ausgebaut werden (DeRijk & de Kloet, 2006; Foy et al., 2006; Gunnar & Quevedo, 2007; Hüther, 2014; Hüther et al., 1999; Hockey, 1997; Hacker & Richter, 1984; Lord et al., 2010; Penzlin, 2014; Sanders, 1982; Sapolsky, 2003; Rasch & Born, 2013; Wieland-Eckelmann, 1992; Yehuda & Mostofsky, 2006). Überfordern die Anforderungen an die Selbstkontrolle die verfügbaren

Kapazitäten, dann können verschiedene Strategien verfolgt werden. Diese werden im folgenden Abschnitt näher betrachtet.

3.4.2 Arbeitspsychologische Konsequenzen

Arbeitsaufträge sowie physische und soziale Bedingungen am Arbeitsplatz stellen in ihrer Gesamtheit die psychische Belastung dar und beanspruchen psychische Leistungsfunktionen. Im Zusammenhang mit dem Spannungsfeld von Autonomie und Selbstkontrolle stellt sich die Frage, unter welchen Bedingungen Arbeitsaufträge (a) in der persönlich bevorzugten Art und Weise bewältigt werden können (Autonomie) bzw. (b) zu ihrer Erfüllung Selbstkontrolle erfordern. Im vorhergehenden Abschnitt wurde dargelegt, dass Selbstkontrolle zu einer erhöhten Inanspruchnahme begrenzter Energiereserven führt und ihr deshalb Grenzen gesetzt sind. Neben diesem Kosten-Aspekt gibt es jedoch auch einen Nutzen-Aspekt von Selbstkontrolle. Selbstkontrolle führt zur Entwicklung angemessener Verhaltens- und Bewertungsmuster in der gegebenen physisch-sozialen (Arbeits-) Situation. Durch wiederholten Einsatz dieser Muster können diese verinnerlicht werden und sich als bevorzugte Art und Weise der Bewältigung von Arbeitsaufträgen etablieren. Ein Übermaß an Selbstkontrollanforderungen geht mit der Gefahr einer Beeinträchtigung von Gesundheit und Leistungsfähigkeit einher. Eine derartige Überbeanspruchung sollte daher vermieden werden. Andererseits lassen sich in einem Unternehmen oder einer Organisationseinheit als soziotechnischem System Störungen und Schwankungen nicht vermeiden (Ulich, 2011). Diese gehen mit Veränderungen der psychischen Belastung einzelner Beschäftigter einher und schränken deren Möglichkeiten ein, jederzeit in der persönlich bevorzugten Art und Weise ihrer Arbeit nachzugehen. – Sie erfordern Selbstkontrolle. Angesichts dieses Kosten-Nutzen-Dualismus der Selbstkontrolle kann das Ziel der Arbeitsgestaltung zwar in einer Minimierung der Anforderungen zur Selbstkontrolle, nicht jedoch in ihrer vollständigen Beseitigung bestehen. Umgekehrt kann das Ziel zwar in einer Maximierung von Autonomie, nicht jedoch in völlig uneingeschränkter Autonomie bestehen. D. h. dort, wo sich am Arbeitsplatz Anforderungen zur Selbstkontrolle vermeiden lassen, sollten sie alleine schon zur Schonung von Ressourcen sowie zum Gesundheitsschutz vermieden werden. Diese Forderung widerspricht nicht der Forderung nach Lern- und Entwicklungspotenzialen in der Arbeit. Letztere lassen sich, wie weiter oben beschrieben, über Aufgaben realisieren, die die Bedürfnisse zum Aufbau von Kompetenz (z. B. vollständige, vielfältige und sinnhafte Aufgaben) sowie zur sozialen Teilhabe und Zugehörigkeit (z. B. in teilautonomen Arbeitsgruppen) erfüllen. Das Aufsuchen neuer Herausforderungen in der Arbeit sowie der

Aufbau unterstützender sozialer Beziehungen erfolgen intrinsisch, d. h. ohne besondere äußere Veranlassungen, vorausgesetzt, der Arbeitsplatz ist entsprechend anregend gestaltet.

Der anwendungsbezogene Schwerpunkt der vorliegenden Arbeit liegt in der betrieblichen Gesundheitsförderung bzw. dem betrieblichen Gesundheitsmanagement. Sowohl in den rechtlichen Normen (z. B. Arbeitsschutzgesetz) als auch in der Arbeitspsychologie hat sich die Einsicht durchgesetzt, dass verhältnisbezogene Maßnahmen Vorrang vor verhaltensbezogenen Maßnahmen haben sollten (Bamberg, Ducki & Metz, 2011a; Faller, 2010; Hacker & Sachse, 2014). In der Arbeitspsychologie spricht man vom „Primat der Aufgabe“ (Ulich, 2011; Ulich & Wülser, 2012). D. h. die Gestaltung der Arbeitsaufgabe ist der primäre Ansatzpunkt für gesundheitsförderliche Interventionen. Maßnahmen, die am Verhalten der Beschäftigten ansetzen (z. B. Handlungsanweisungen oder Qualifizierung) haben in diesem Kontext unterstützenden bzw. ergänzenden Wert. Daher wird in der folgenden Diskussion arbeitspsychologischer Konsequenzen auf Themen wie Personalauswahl (Passung zwischen Anforderungen und Qualifikation) oder Sozialisation in Organisationen (Angleichung organisationaler und persönlicher Normen und Werte) verzichtet (Rosenstiel, 2007; Rosenstiel, Molt & Rüttinger, 2005; Semmer & Udris, 2007). Hier könnte etwa postuliert werden, dass eine geringe Passung von Person und Arbeitsplatz bzw. Organisation zu erhöhten Anforderungen an die Selbstkontrolle einhergehen. Darüber hinaus wird auf besondere Anforderungen in bestimmten Ausschnitten des Arbeitsmarktes, wie beispielsweise im Dienstleistungssektor, in Call Centern oder in der Zeitarbeit, verzichtet. Beispielsweise stellen die Emotionsarbeit in Dienstleistungsberufen (Nerdinger, 2012) und insbesondere in Call Centern (Scherrer, 2001, 2002) oder ein häufiger Wechsel des Arbeitsplatzes in der Zeitarbeit (Wieland, 2001; Wieland & Krajewski, 2002, 2007) besondere Anforderungen an die Selbstkontrolle dar. Im Folgenden liegt der Fokus auf Merkmalen der Arbeitsgestaltung sowie der Qualität sozialer innerbetrieblicher Beziehungen.

Zunächst sei ein Blick auf Gestaltungsmerkmale geworfen, die zu einer erhöhten Autonomie führen. Wesentlich sind hier Freiheitsgrade bezüglich der Art und Weise, in der eine Person einen Arbeitsauftrag ausführt. Hacker und Sachse (2014) fassen diese Freiheitsgrade unter dem Begriff Tätigkeitsspielraum zusammen und definieren ihn als „die Gesamtheit der Entscheidungsmöglichkeiten für selbstständige Zielsetzungen im auftragsgerechten Handeln im Sinne der Beeinflussbarkeit („*control*“) der eigenen Tätigkeit und ihrer Ausführungsbedingungen auf der Grundlage von *Durchschaubarkeit* der Arbeitssituation und *Vorhersehbarkeit* von Anforderungen“ (Hacker & Sachse, 2014, S. 86). Ulich (2011) unterteilt den Tätigkeitsspielraum unter Verwendung der hierarchischen Beziehung von Tätigkeit, Handlung

und Operation (vgl. Hacker & Sachse, 2014). Er unterscheidet Handlungs-, Gestaltungs- und Entscheidungsspielraum. Handlungsspielraum bezieht sich in Anlehnung an Hacker (1978) „auf Verfahrenswahl, Mitteleinsatz und zeitliche Organisation von Aufgabenbestandteilen“ (Hacker, 1978, S. 72, nach Ulich, 2011, S. 187). Er ist der Flexibilität bei der Durchführung aufgabenbezogener Operationen zugeordnet. „Der *Gestaltungsspielraum* wird durch die Möglichkeit zur selbstständigen *Gestaltung* von Vorgehensweisen nach eigenen Zielsetzungen bestimmt“ (Ulich, 2011, S. 187). Diese Art von Freiheitsgraden ist demnach eng verknüpft mit der Variabilität bei der Organisation aufgabenbezogener Handlungen. Der Entscheidungsspielraum bezieht sich dagegen auf Möglichkeiten, als Person oder Gruppe Einfluss auf die „*Festlegung* bzw. Abgrenzung von Tätigkeiten oder Aufgaben“ (a. a. O.) nehmen zu können. Das Ausmaß des Entscheidungsspielraums bezeichnet Ulich (2011) als Entscheidungskompetenz sowie als Autonomie. Auch, wenn diese Konzeption in der Definition von Hacker und Sachse (2014) z. T. aufgeht, verdeutlicht Ulich (2005), dass aufgabenbezogene Freiheitsgrade auf mindestens diesen drei Ebenen gestaltet werden können. Sie tragen zur Autonomie bei, da sie auf verschiedenen Ebenen erlauben, Arbeitsaufgaben in gewissem Ausmaß in der persönlich bevorzugten Art und Weise auszuführen (Handlungs- und Gestaltungsspielraum) oder gar selbst auszuwählen und zu gestalten (Entscheidungsspielraum). Im Entscheidungsspielraum ist eine wesentliche Erweiterung im Vergleich zur Definition von Hacker und Sachse (2014) zu erkennen, denn Hacker und Sachse (2014) setzen bereits einen Auftrag als gegeben voraus, indem sie von „auftragsgerechtem Handeln“ sprechen. Wieland (2014) verdeutlicht darüber hinaus, dass diese Freiheitsgrade durch geeignete organisationale Strukturen und Strukturen auf Abteilungs- bzw. Gruppenebene unterstützt oder überhaupt erst ermöglicht werden. Er stützt sich dabei auf den soziotechnischen Systemansatz, demzufolge technisches und soziales Teilsystem einer Organisation durch eine geeignete Gestaltung von Organisationseinheiten und deren Aufgaben optimal auf einander abgestimmt werden sollten (vgl. Ulich, 2011). Hiermit sind auch eine menschengerechte Funktionsteilung zwischen Mensch und Maschine („Technik als Option“, Ulich, 2011) sowie das Konzept teilautonomer Arbeitsgruppen (s. o.) angesprochen.

Nun sei der Blick solchen Arbeitsplatzmerkmalen zugewandt, die zu vermeidbaren Selbstkontrollanforderungen führen. Als Merkmale der Arbeitsgestaltung werden Ursachen von Fehlhandlungen sowie Regulationsbehinderungen besprochen. Für die Vermeidung von Fehlhandlungen durch Beschäftigte ist Selbstkontrolle genauso notwendig, wie für den Umgang mit Arbeitsplatzmerkmalen, die unabhängig von der eigentlichen Arbeitsaufgabe zur Inanspruchnahme psychischer Leistungsfunktionen führen und die Aufgabenerfüllung

erschweren oder behindern (Regulationsbehinderungen). Vermeidbare Selbstkontrolle ist ebenfalls notwendig, wenn die Führungskraft autoritär führt und somit stark in die Art und Weise der Aufgabenerfüllung eingreift.

Im Rahmen des ∞ -Modells wurde dargelegt, dass das Salient Network u. a. dann das Central Executive Network hinzuschaltet, wenn Fehler in der Handlungsausführung entdeckt werden oder bislang erfolgreiche Handlungs- und Bewertungsmuster nicht mit den äußeren Bedingungen vereinbar sind (Botvinick, Cohen & Carter, 2004; Bush, Luu & Posner, 2000; Critchley, 2009; Kerns et al., 2004; Carter & van Veen, 2007). In einem solchen Moment setzt Selbstkontrolle ein, um mögliche Fehlhandlungen und damit zusammenhängende Unfälle und Sachschäden zu vermeiden. Dieser Prozess erfordert zusätzliche Energie und belegt kognitive Ressourcen und sollte daher möglichst über eine geeignete Aufgabengestaltung vermieden werden. Denn auch eine gut gemeinte Korrektur des eigenen Handelns muss nicht zwingend zur Vermeidung einer Fehlhandlung führen. Hacker und Sachse (2014) widmen sich daher in ihren Ausführungen zu Fehlhandlungen vor allem den in der Arbeitsgestaltung liegenden Ursachen von Fehlhandlungen und stellen als wesentliche Ursache den Mangel an oder die unangemessene Nutzung von Informationen heraus. Um Fehlhandlungen zu vermeiden, sollten Arbeitsmittel so gestaltet sein, dass alle handlungsrelevanten Informationen in menschengerechter Weise zur Verfügung gestellt werden. Hacker und Sachse (2014) führen aus, dass dabei allgemeines Wissen über Wahrnehmungsschwellen, Gedächtnisleistungen, kontingente Erwartungen (z. B. „Übersehen“ einer Information, weil sie auf Grund von Lerneffekten nicht erwartet wird) oder kognitive Interferenzen (z. B. mögliche Verwechslungen von Informationen auf Grund von Ähnlichkeit) berücksichtigt werden sollten. Verhaltensbezogene Maßnahmen, wie beispielsweise Fehlertrainings, dienen ihrer Meinung nach vor allem der Unterstützung verhältnisbezogener Maßnahmen zur Fehler- und Unfallvermeidung.

Regulationsbehinderungen behindern Beschäftigte bei der Erfüllung ihrer Arbeitsaufträge und erfordern zusätzlichen Handlungsaufwand, insbesondere, da die Beschäftigten noch keine effizienten Strategien im Umgang mit ihnen entwickelt haben (Greiner, Leitner, Weber, Hennes & Volpert, 1987). Unter Regulationsbehinderungen fallen informatorische (z. B. fehlende oder schwer zugängliche Informationen, s. o.) oder motorische Erschwerungen (z. B. schwer zugängliche Bedienelemente einer Maschine), aber auch Unterbrechungen, die auf Störungen durch Personen, Funktionsstörungen (z. B. Maschinenausfall) oder Blockierungen (z. B. fehlendes Arbeitsmaterial) zurückgehen. Diese Regulationshindernisse erfordern, wenn überhaupt möglich, kurzfristige Reaktionen. Außerdem sind aufgabenunspezifische (allgemeine, störende Umgebungsbedingungen) und aufgabenspezifische (monotone

Bedingungen, Zeitdruck) Regulationsüberforderungen hinzuzuzählen, die zu einer Verminderung der Regulationsfähigkeit führen (vgl. Lüders, 1999; Ulich, 2011). In der vorliegenden Arbeit wird die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann, Saßmannshausen, Rose & Schwarz, 1999) für die Erfassung von Regulationsbehinderungen verwendet. Hier werden auch unklare Arbeitsaufträge als Regulationsbehinderungen aufgefasst. Bei Semmer (1984) wird dieser Aspekt als Regulationsunsicherheit bezeichnet und stellt eine Unterkategorie von Regulationsproblemen dar (vgl. auch Frese & Zapf, 1994; Semmer, Zapf & Dunckel, 1999). Die hier aufgeführten Regulationsbehinderungen müssen von Beschäftigten zusätzlich zur eigentlichen Arbeitsaufgabe bewältigt werden, obwohl sie nicht Bestandteil derselben sind. Damit erfordern sie eine Anpassung der Art und Weise, wie eine Aufgabe bewältigt werden kann, also (zusätzliche, eigentlich unnötige) Selbstkontrolle. Regulationsbehinderungen können zum größten Teil durch technische Maßnahmen der Arbeitsgestaltung vermieden werden. Dies wäre zum einen zu empfehlen, um die unnötige Inanspruchnahme psychischer Leistungsfunktionen zu vermeiden. Zum anderen deshalb, weil ein Zusammenhang von Regulationsbehinderungen und gesundheitlichen Beeinträchtigungen belegt ist (Greiner et al., 1987; Lüders, 1999; Semmer et al., 1999; Ulich, 2011; Wieland-Eckelmann et al., 1999).

Weiter oben wurde bereits dargestellt, dass die Führungskraft durch eine Kombination mitarbeiter- und aufgabenorientierter Führung („ganzheitliche Führung“; Wieland et al., 2009) zu einer Erfüllung der Bedürfnisse nach Kompetenz und sozialer Teilhabe und Zugehörigkeit beiträgt. Darüber hinaus wurde herausgestellt, dass nicht-autoritär geführt werden sollte. Autoritäre Führung ist gekennzeichnet durch „Strikte Anweisung und Kontrolle, Leiter [Führungskraft] entscheidet allein, wenig Kommunikation zwischen Gruppenmitgliedern“ (Felfe, 2009, S. 28, Tabelle 2). In der vorliegenden Arbeit wird autoritäre Führung über den Fragebogen zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007) erhoben. In diesem Fragebogen umfasst dieses Konstrukt neben den genannten bevormundenden Verhaltensweisen auch solche, die auf der persönlichen Beziehungsebene als beleidigend oder erniedrigend erlebt werden können (z. B. „Persönlichen Ärger oder Ärger mit dem Vorgesetzten lässt er/sie an uns aus“ oder „Unser/e Vorgesetzte/r kritisiert uns Mitarbeiter auch in Gegenwart anderer“). Gregersen et al. (2011) sprechen in diesem Zusammenhang von „Führungsverhalten als Risikofaktor bzw. Stressor“ (a. a. O., S. 7) und fassen zusammen: „Diese *führungsbezogenen Stressoren* wirkten negativ auf die Arbeitszufriedenheit und den Erschöpfungs- sowie allgemeinen Gesundheitszustand der Mitarbeiter. Sie können zudem zu einer Erhöhung der Fehlzeiten und Langzeitabwesenheiten sowie einer Verstärkung von

Wechselabsichten führen“ (a. a. O.; vgl. auch Stadler & Spieß, 2005). Da die bevormundenden Verhaltensweisen (Kontrolle, Anweisung, Vorgaben) prinzipiell die Art und Weise der Erfüllung von Arbeitsaufträgen einschränken, führen sie zwar zu Selbstkontrollanforderungen, sind aber zumindest von Seiten der Beschäftigten vorhersehbar. Die darüber hinausgehenden beleidigenden und erniedrigenden Verhaltensweisen stellen die Beziehung zwischen Führungskraft und betroffenen Beschäftigten in Frage. Sie lösen negative soziale Emotionen (Baumeister & Leary, 1995) aus und führen über diesen Weg zu weiteren, eher unvorhersehbaren Selbstkontrollanforderungen. Besteht das Ziel darin, vermeidbare Selbstkontrollanforderungen zu minimieren sowie die Gesundheit von Beschäftigten zu schützen, so ist autoritäre Führung im hier umschriebenen Sinne möglichst zu vermeiden.

3.5 UNTERNEHMENSKULTUR ALS KONTEXT FÜR SELBSTORGANISATION

Um grundlegende Bedürfnisse erfüllen zu können, so können die vorhergehenden Argumentationen zusammengefasst werden, sind entsprechende Freiheiten (Maslow, 1943), Spielräume (Hacker & Sachse, 2014; Ulich, 2011) oder Gestaltungsfreiheiten (Wieland, 2014a; Wieland & Hammes, 2014a) zu garantieren. Beschränkungen dieser Freiheiten bedeuten bereits Gefahren für entsprechende Bedürfnis-Erfüllung und führen zu Widerständen (Maslow, 1943). Freiheiten reichen jedoch nicht aus. Es bedarf auch der entsprechenden Hinweis- und Anreize sowie der Möglichkeiten und Herausforderungen für Bedürfnis-Erfüllung und persönliche Entwicklung (Gunnar & Quevedo, 2007; Hüther & Sachsse, 2007; LeDoux, 2012; Nerdinger, 2013). Sofern in Unternehmen gemeinsame Ziele verfolgt werden, bedarf es darüber hinaus des Vertrauens in die Belegschaft. Vertrauen darin, dass die Beschäftigten bei einer anregend gestalteten Arbeitsumgebung (Bedürfnis nach Kompetenz) und einer von Empathie und Fürsorge geprägten Unternehmenskultur (Bedürfnis nach sozialer Teilhabe und Zugehörigkeit) ohne weiteres Zutun zu ihrer persönlichen und der unternehmerischen Entwicklung gleichermaßen beitragen werden (s. o.; Baumeister, 2010; Baumeister & Leary, 1995; Deci & Ryan, 2000; Hacker & Sachse, 2014; Ulich, 2011; White, 1953). Hierbei sollten die anregenden Elemente der Arbeitsumgebung selbstverständlich mit den Unternehmenszielen in Zusammenhang stehen. Und ebenso selbstverständlich sollten Empathie und Fürsorge bzw. soziale Beiträge im Unternehmen auf dem Prinzip der Reziprozität (oder Gerechtigkeit), des gegenseitigen Gebens und Nehmens beruhen. Genau hier liegt die Quelle des Spannungsfeldes von Autonomie und Selbstkontrolle (s. o.). Was sind die Unternehmensziele, Ziele der Arbeitseinheit, Ziele der einzelnen Beschäftigten? Welche Beiträge können, wollen und dürfen die einzelnen Beschäftigten zum Unternehmenserfolg einbringen? Nach welchen Regeln und

Normen und in welcher Art und Weise werden Beiträge zum Unternehmenserfolg bewertet und belohnt? Klarheit hierüber, und somit auch über die individuellen Anforderungen an die Selbstkontrolle, verschaffen nur Transparenz und der partizipative Dialog. D. h. es bedarf Möglichkeiten der Mitbestimmung, bei denen die Interessen der Unternehmensmitglieder verschiedener Funktionen und Hierarchien ausgetauscht, hinterfragt, konsensuell integriert und in Gestaltungsmaßnahmen umgesetzt werden können (organisationale Demokratie; Weber, 1999). Neben den gesetzlich garantierten Möglichkeiten der Mitbestimmung kann diese durch gestalterische Maßnahmen zusätzlich befördert werden. Führungsverhalten und Führungsinstrumente sowie Gruppenarbeit sind geeignete Möglichkeiten, partizipative Dialoge zu etablieren (Felfe, 2009; Wegge & Rosenstiel, 2007; Ulich, 2011). Die Einführung von Qualitäts- oder Gesundheitszirkeln sowie die Durchführung vollständiger Gefährdungsbeurteilungen (im Sinne des Arbeitsschutzgesetzes) und betrieblichem Eingliederungsmanagement zählen ebenso zu diesen Möglichkeiten (Badura, Walter & Hehlmann, 2010; Bamberg, Ducki & Metz, 2011; Faller, 2010; Ulich & Wülser, 2012). Auf den partizipativen, funktions- und hierarchieübergreifenden Dialog mit Gestaltungsabsicht zielt am umfassendsten das Konzept der Organisationsentwicklung im Sinne einer lernenden Organisation ab (Bamberg et al., 2011; Gebert, 2007; Rosenstiel, 2007; Rosenstiel, Molt & Rüttinger, 2005). Eine persönlichkeits- und gesundheitsförderliche Unternehmenskultur weist somit neben der wirtschaftlichen Gewinnorientierung (z. B. Balderjahn & Specht, 2011) mindestens folgende Qualitäten auf: Empathie, Fürsorge, Gerechtigkeit, Transparenz, Partizipation und nicht zuletzt Vertrauen. Zahlreiche Fallbeispiele belegen, dass häufig selbst im Sinne gesteigerter Produktivität erfolgreich eingeführte partizipative Strukturen und Prozesse partialisierten Formen der Arbeit weichen müssen. Ulich (2011) führt Beispiele auf, aus denen als Begründung die vermeintlich bessere Kontrollier- und Berechenbarkeit solcher Arbeitsformen hervorgeht. D. h. Entscheidungsträger verzichten häufig zugunsten vereinfachter Steuerbarkeit auf einen möglichen zusätzlichen Gewinn und nehmen hierfür eine Erosion der Qualifikation, Leistungsfähigkeit und Gesundheit der Belegschaft in Kauf.

4 BEGRÜNDUNG UND FORMULIERUNG DES EMPIRISCH-PRAKTISCHEN UNTERSUCHUNGSANLIEGENS

Die bisherigen theoretischen Ausführungen der vorliegenden Arbeit erfolgten mit dem Ziel, psychische Beanspruchung, ihre Ursachen (psychische Belastung) sowie ihre langfristigen Folgen (Leistung und Gesundheit) aus einer biopsychologischen Perspektive zu spezifizieren. Es wurde dargestellt, dass psychische Beanspruchung eine untrennbare Kombination von Emotion, Motivation bzw. Bedürfnissen und Kognition sowie körperlichen Reaktionen darstellt. Langfristige Folgen, wie psychische und körperliche Beeinträchtigungen und Erkrankungen (Kosten), aber auch die Ausbildung von Qualifikation und angemessenen Bewertungs- und Bewältigungsmustern (Nutzen) stellen eine körperliche Widerspiegelung von Erfahrungen aus der Interaktion mit der physischen und sozialen Umwelt dar. Entscheidend für die Bilanz von Kosten und Nutzen ist, welche systemischen Störungen (Belastung) auf den Organismus in seinem gegenwärtigen Zustand einwirken. In Bezug auf die Kosten sind es unkontrollierbare Stressreaktionsprozesse sowie vermeidbare Selbstkontrollanforderungen, die zu Beeinträchtigungen bis hin zu organismischen Erosionen führen. Es wurden Überlegungen darüber angestellt, wie solchen Prozessen durch eine persönlichkeits- und gesundheitsförderliche Gestaltung von Arbeitssystemen vorgebeugt werden kann. Hierzu wurde von intrinsischen Bedürfnissen (Homöostase, Kompetenz, soziale Teilhabe und Zugehörigkeit) ausgegangen. Das Maß an Selbstkontrollanforderungen, das mit Schwierigkeiten der Bedürfnis-Erfüllung aufgrund einschränkender Rahmenbedingungen einhergeht, wurde als Weiche zwischen Nutzen und Kosten psychischer Beanspruchung identifiziert. Die Schlussfolgerung für eine persönlichkeits- und gesundheitsförderliche Gestaltung von Arbeitssystemen lautet zusammengefasst, Selbstkontrollanforderungen durch das Angebot an Möglichkeiten und angemessene Herausforderungen zur Bedürfnis-Erfüllung auf einem individuell kontrollierbaren Niveau zu halten (Beanspruchungsoptimalität). D. h. diese Anforderungen müssen an die energetischen und strukturellen Kapazitäten, die Bedürfnisse sowie die Anpassungsfähigkeiten der Beschäftigten angepasst sein. Einschlägige arbeitspsychologische Bewertungskriterien, Konzepte und Merkmale von Arbeitssystemen konnten in diesen theoretischen Rahmen eingebracht werden. Im Folgenden wird dargestellt, dass diese theoretischen Grundlagen Ausgangspunkt für die empirischen Untersuchungen der vorliegenden Arbeit sind. Auch hier wird die psychische Beanspruchung im Zentrum stehen und ihr Verhältnis zu psychischer Belastung und langfristigen Beanspruchungsfolgen untersucht.

4.1 DAS WUPPERTALER KONZEPT

Die Arbeitsgruppe um Rainer Wieland legt in ihren Praxis- und Forschungsprojekten ein Wirkungsmodell (Fünf x Fünf-Wirkungsmodell; Hammes, Wieland & Winizuk, 2009; Wieland, 2006b, 2008, 2009, 2010a, 2013, 2014a; Wieland & Görg, 2009; Wieland & Hammes, 2014; Wieland, Krajewski & Memmou, 2004; Wieland, Winizuk & Hammes, 2009) sowie ein zugehöriges Inventar (Wuppertaler Inventar zur ressourcenorientierten Arbeits- und Organisationsgestaltung [WITAL]; Wieland, 2014b) für Mitarbeiterbefragungen zugrunde, in dem das soziotechnische System sowie die Unternehmenskultur als Kontext arbeitsbezogener Prozesse definiert werden. Bei der Analyse wird auf die partizipative Unternehmenskultur besonderes Augenmerk gelegt (Görg & Wieland, 2008, 2009; Wieland & Görg, 2009). Belastung, Beanspruchung und langfristige Beanspruchungsfolgen vollziehen sich in diesem soziotechnischen und kulturellen Kontext. Belastungsfaktoren werden in drei Merkmalskategorien unterteilt: (a) Arbeitsaufgabe und Arbeitsplatz, (b) Führung und Zusammenarbeit sowie (c) Personenmerkmale. Arbeitsaufgabe und Arbeitsplatz werden im Hinblick auf Tätigkeitsspielräume (drei Items), Arbeitsanforderungen (im Sinne vollständiger Tätigkeit; drei Items), Regulationsbehinderungen (fünf Items), Leistungs- und Zeitvorgaben (zwei Items) sowie Kommunikations- und Kooperationsanforderungen (zwei Items) in Bezug auf drei Schnittstellen (Arbeitsaufgabe, Zusammenarbeit/Kommunikation und Computer/Maschine) analysiert. Hierfür kommt die Synthetische Beanspruchungs- und Arbeitsanalyse zum Einsatz (SynBA; Wieland-Eckelmann, Saßmannshausen, Rose & Schwarz, 1999). Die Items der SynBA sind bedingungsbezogen formuliert (z. B. "Die Arbeit besteht aus kurzen, sich wiederholenden Tätigkeiten.") und sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 5 = "Trifft vollständig zu") durch die befragten Beschäftigten zu bewerten. Führung und Zusammenarbeit wird unter den Aspekten der partizipativen Unternehmenskultur (s. o.; drei Items), ganzheitlicher Führung (sechs Items), autoritärer Führung (vier Items), Qualität der Zusammenarbeit (drei Items) sowie Anerkennung und Wertschätzung (zwei Items) abgefragt (Fragebogen zu Führung und Zusammenarbeit [FFZ]; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Die Items (z. B. "Unser/e Vorgesetzte/r lässt einen wissen, wie gut man seine Arbeit getan hat.") sind ebenfalls auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 5 = "Trifft immer zu") zu bewerten. Als Personenmerkmal wird gesundheitsbezogene Selbstwirksamkeit (Gesundheitskompetenz) erhoben (Gesundheitskompetenz-Fragebogen [GKF]; Wieland, 2010a; Wieland & Hammes, 2009, 2010). Die Items orientieren sich am Fragebogen zur allgemeinen Selbstwirksamkeit von Schwarzer und Jerusalem (1999) (z. B. "Wenn ich Pläne bezüglich meiner Gesundheit mache, dann kann ich

diese auch umsetzen.") und sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 5 = "Trifft vollständig zu"). Die erhobenen Belastungsfaktoren werden im Sinne des Belastungs-Beanspruchungs-Paradigmas als verursachende und gestaltbare Merkmale bzw. Qualifikationen betrachtet. Beanspruchung ergibt sich dem Modell und Paradigma zufolge aus dem unmittelbaren Zusammenwirken der Belastungsfaktoren auf die Person und wird als Prozessvariable aufgefasst. Mit dem WITAL wird sie in Form von Emotionen und Befindenzuständen während der Arbeit gemessen (Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchung während der Arbeit [EEB]; Scherrer, 2002; Wieland, 2004, 2006a; Wieland, Krajewski & Memmou, 2006; Wieland, Metz & Richter, 2001). Insbesondere funktionale Beanspruchung (z. B. "energiegeladen", "leistungsbereit"), dysfunktionale Beanspruchung (z. B. "nervös", "körperlich verspannt") und Kontrollerleben ("einflussreich") haben sich als gute Prädiktoren für die Klassifikation gesundheits- und persönlichkeitsförderlicher Arbeit und ihrer Folgen erwiesen. Diese Befindenzustände werden auch alleinstehend mit dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB; Hammes & Wieland, 2012a, b; Wieland, 2013; Wieland & Hammes, 2014b) erhoben. Die Eigenschaftswörter sind auf einer siebenstufigen Skala (1 = "kaum" bis 7 = "außerordentlich") dahingehend zu bewerten, ob sie die Intensität des allgemeinen Befindens während der Arbeit treffend beschreiben. Die langfristigen Beanspruchungsfolgen stellen die Ergebnisvariablen dar. Im Wirkungsmodell werden hier Leistungs- und Gesundheitsindikatoren subsummiert. Das Inventar weist unter anderem einen Fragebogen zu Häufigkeiten körperlicher Beschwerden (Magen-Darm, Muskel-Skelett, Herz-Kreislauf, unspezifische Beschwerden, allergische Reaktionen) (HkB; Baumgart et al., 2002 in Anlehnung an Fahrenberg, 1994; Manz, 1998) sowie Fragen zu Absentismus und Präsentismus auf. Die Häufigkeiten körperlicher Beschwerden (z. B. "Wie oft haben Sie Rückenschmerzen?" oder "Wie oft haben sie Konzentrationsstörungen?") werden in Bezug auf die letzten zwölf Monate auf einer fünfstufigen Ordinalskala (0 = "praktisch nie", 1 = "etwa 2 x im Jahr", 2 = "etwa 2 x im Monat", 3 = "etwa 3 x die Woche" und 4 = "fast täglich") abgefragt. Die dargestellten Fragebögen bilden einen großen Teil der Merkmale persönlichkeits- und gesundheitsförderlicher Arbeit ab, wie sie weiter oben theoretisch erörtert und begründet wurden. Hier wurden ausschließlich jene Fragebögen des WITAL beschrieben, die für den empirischen Teil der vorliegenden Arbeit relevant sind, die aber zugleich auch den Kern des WITAL darstellen. Je nach Anwendungsfeld und Problemstellung wird das WITAL um geeignete Fragebögen ergänzt. So wurden im "Projekt Arbeit Rücken Gesundheit (PAKT)" zusätzliche Fragen zur Rückengesundheit und zur Beanspruchung des Bewegungs- und Stützapparates aufgenommen (Hammes et al., 2010). In

verschiedenen Studien konnte belegt werden, dass die Wirkungszusammenhänge der aufgezählten Variablen erwartungsgemäß ausfallen (Hammes et al. 2010; Hammes & Wieland, 2012a, b; Hammes, Wieland & Winizuk, 2009; Wieland, 2004, 2008, 2013, 2014a; Wieland & Hammes, 2014b; Wieland, Krajewski & Memmou, 2006; Wieland, Metz & Richter, 2001; Wieland, Winizuk & Hammes, 2009). In Abbildung 6 findet sich eine Darstellung des Fünf x Fünf-Wirkungsmodells mit einer Zuordnung der genannten Instrumente zu den entsprechenden Merkmalsbereichen.

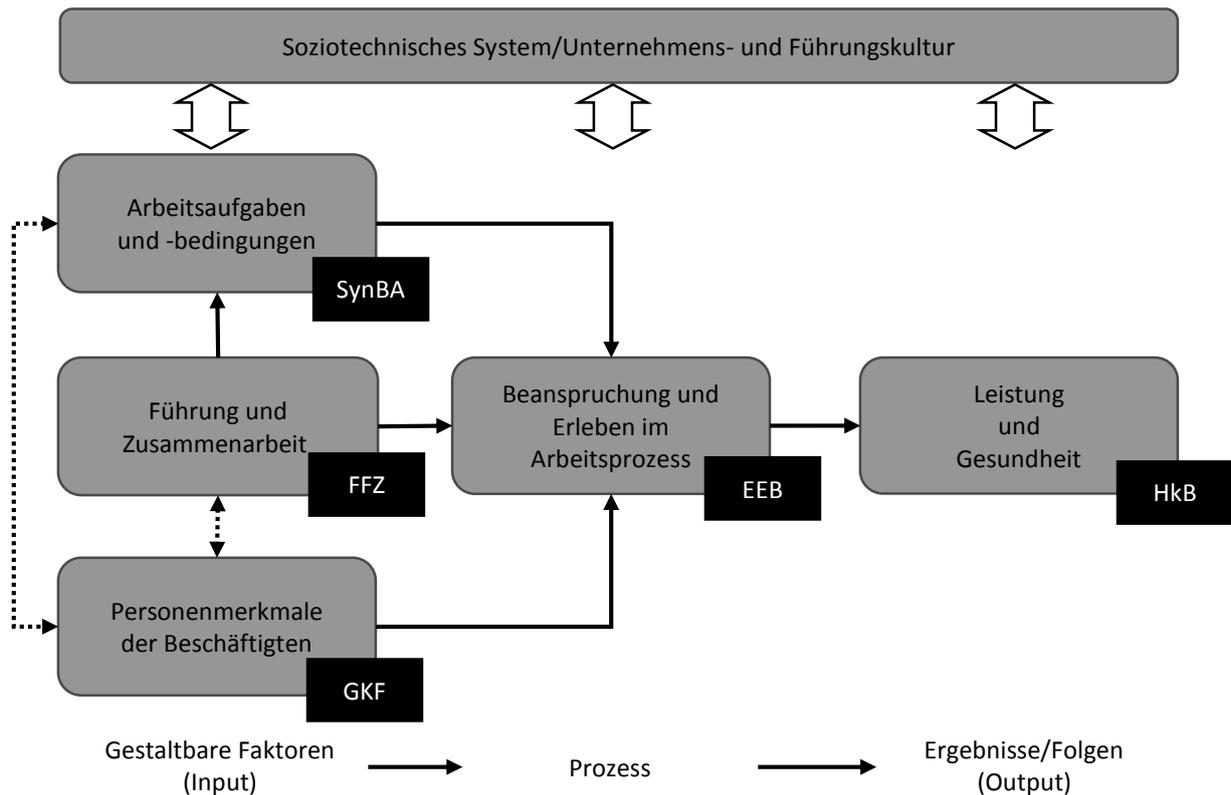


Abbildung 6. Fünf x Fünf-Wirkungsmodell und eingesetzte Instrumente. Dargestellt ist das Fünf x Fünf-Wirkungsmodell in Anlehnung an Wieland (2008, 2009, 2013, 2014a). Die schwarzen Kästen repräsentieren die je Merkmalsbereich eingesetzten Instrumente. SynBA = Synthetische Beanspruchungs- und Arbeitsanalyse (Wieland et al., 1999), FFZ = Fragebogen zu Führung und Zusammenarbeit (Scherrer & Wieland, 2006; Wieland & Scherrer, 2007), GKF = Gesundheitskompetenz-Fragebogen (Wieland, 2010a, Wieland & Hammes, 2009, 2010), EEB = Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben während der Arbeit (Scherrer, 2002; Wieland, 2004, 2006a; Wieland, Krajewski & Memmou, 2006; Wieland, Metz & Richter, 2001), HkB = Häufigkeit körperlicher Beschwerden (Baumgart et al., 2002).

4.2 HINTERGRUND UND STRUKTUR DER EEB

Im Zentrum des empirischen Teils der vorliegenden Arbeit stehen die Eigenschaftswörter der EEB als Indikatoren der psychischen Beanspruchung. Die Entwicklung der EEB folgt der Linie der "mental-workload Forschung" sowie dimensionaler Modelle von Emotionen, die weiter oben in einen engen Zusammenhang mit den Arousal-Systemen gestellt wurden. Die EEB basiert im Wesentlichen auf einer deutschen Übersetzung und Anpassung der Activation Deactivation Adjective Check List (AD ACL) von Thayer (1967; s. Wieland-Eckelmann, 1992;

Wieland-Eckelmann & Bösel, 1987). In seinem Werk "Kognition, Emotion und psychische Beanspruchung" verwendet Wieland dieses Verfahren in einer Trait- (stabiles Persönlichkeitsmerkmal) und einer State-Version (gegenwärtiger individueller Zustand) (Wieland-Eckelmann, 1992). Die Trait-Version ("Fragebogen zur Erfassung des habituellen Aktiviertheitsniveaus (GA-HA-Trait)", a. a. O., S. 365-372) unterscheidet sich von der State-Version ("GA-HA-State") nur durch eine entsprechende Instruktion. Wieland-Eckelmann (1992) unterscheidet zwei Klassen von Beanspruchungszuständen: "psychomentale Funktionsbeanspruchung" und "emotionale Selbstbeanspruchung". Psychomentale Funktionsbeanspruchung ist "die Folge der Inanspruchnahme aufgabenbezogener psychischer Leistungsfunktionen ..., deren Regulationsfunktion darin besteht, externe Ist-Zustände in selbst- und/oder fremdgesetzte Soll-Zustände zu überführen" (Wieland-Eckelmann, 1992, S. 102). Beanspruchungszustände dieser Klasse können mittels der Dimension "Energetische Aktiviertheit" (GA-Skala; Wieland-Eckelmann & Bösel, 1987) (z. B. "energiegeladen", "leistungsbereit") erfasst werden. Emotionale Selbstbeanspruchung ist die "Inanspruchnahme psychophysischer Selbstregulationsfunktionen als Folge emotionaler Anforderungen/ Belastungen, die angestrebte interne Systemzustände (persönliche Ziele, Bedürfnisse, Leistungsstandards) gefährden" (Wieland-Eckelmann & Bösel, 1992, S. 102). Beanspruchungszustände dieser Klasse können mittels der Dimension "Ängstliche Anspannung" (HA-Skala; Wieland-Eckelmann & Bösel, 1987) (z. B. "nervös", "aufgeregt") erfasst werden. Dem Konzept von Wieland-Eckelmann (1992) zufolge variiert die psychomentale Funktionsbeanspruchung aufgrund externer aufgabenbezogener Anforderungen, wohingegen die emotionale Selbstbeanspruchung mit Bedrohungen angestrebter interner Systemzustände variiert. Letztere stellen seiner Ansicht nach eine zusätzlich zu regulierende Anforderung dar (a. a. O.). Sieht man von der angstbezogenen Tönung des Konzepts der emotionalen Selbstbeanspruchung ab, so ist es sehr gut mit der Vorstellung mental erzeugter Emotionen durch Simulationen und erfahrungsgeleiteter Bewertungsmuster vereinbar, wie es im ∞ -Modell postuliert wird. Das ∞ -Modell sagt jedoch auch voraus, dass psychomentale Funktionen zur Lösung simulierter Probleme eingesetzt werden (dies stellt ebenfalls eine Selbstbeanspruchung dar) und dass Emotionen, moderiert durch die Arousal-Systeme, auch externe Quellen (systemische Störungen, psychische Belastung) haben können. Solche externe Quellen sowie langfristige Beanspruchungsfolgen werden im Rahmen der vorliegenden Arbeit mit den Eigenschaftswörtern der EEB in Beziehung gesetzt.

Eine erste Annäherung an die aktuelle EEB erfolgt im Zusammenhang mit der Entwicklung der Synthetischen Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999), die ebenfalls fester Bestandteil des WITAL ist. GA-HA-State konnte in mehreren Studien derart modifiziert werden, dass sie die Wirkungen der mit der SynBA erfassten psychischen Belastung im Sinne psychischer Beanspruchung abbildet. In diesem Zusammenhang erweisen sich vier Dimensionen als zentral: Mentale, motivationale, emotionale und physische Beanspruchung (Baggen et al., 1993; Wieland, 1999; Wieland-Eckelmann, 1992; Wieland-Eckelmann et al., 1991). Die Studien weisen die so genannte "Doppelrolle" psychischer Belastung und Beanspruchung nach: Arbeitsplätze, die im Sinne arbeitspsychologischer Kriterien gut gestaltet sind (z. B. vielfältige und vollständige Tätigkeiten mit entsprechenden Handlungsspielräumen), führen zu funktionalen Beanspruchungsmustern (hohe mentale und motivationale und geringe emotionale und physische Beanspruchung). Bei schlecht gestalteten Arbeitsplätzen kehrt sich das Beanspruchungsmuster in ein dysfunktionales um. Diese Erkenntnis führt zur Forderung nach einer beanspruchungsoptimalen Gestaltung von Arbeitsplätzen sowie der Definition der Beanspruchungsbilanz als Kennwert für die Beanspruchungsoptimalität von Arbeitsplätzen. Belastung soll nicht reduziert oder beseitigt, sondern so gestaltet werden, dass die Beschäftigten ein optimales Beanspruchungsmuster (hohe funktionale und geringe dysfunktionale Beanspruchung) aufweisen (Wieland, 1999; Wieland-Eckelmann, 1996). SynBA wird entsprechend um ein Zusatzmodul zur Erfassung psychischer Beanspruchung ergänzt (Wieland-Eckelmann et al., 1999).

In der Folgezeit erfährt dieses Zusatzmodul zwei Erweiterungen. Wieland, Metz und Richter (2001) erweitern es im Rahmen des Projektes "Gesundheit und Sicherheit in neuen Arbeits- und Organisationsformen" (1997-2001) um die Dimensionen Monotonie (vgl. Hacker & Richter, 1984), Ärger (vgl. Schwenkmezger, 1990) und Kontrollerleben. Im selben Projekt fügt Scherrer (2002) die Dimensionen positives und negatives Befinden hinzu. Dies erfolgt in Anlehnung an die beiden Dimensionen Positive und Negative Affect bei Watson, Clark und Tellegen (1988). Beide Erweiterungen erfolgen, um die besonderen Anforderungen von Beschäftigten in Call Centern abbilden zu können. Sie erweisen sich in späteren Projekten auch für andere Branchen und betriebliche Handlungsfelder als relevant (Wieland, 2004, 2008; Wieland & Görg, 2009; Wieland, Krajewski & Memmou, 2004). Scherrer (2002) stellt in ihrer Dissertation ausführlich dar, dass sich die insgesamt neun theoriegeleiteten Dimensionen der EEB faktorenanalytisch auf drei Dimensionen reduzieren lassen: Funktionale Beanspruchung, dysfunktionale Beanspruchung sowie Kontrollerleben. Auch diese übergeordnete Struktur kann

mehrfach repliziert werden und ist Grundlage für die Operationalisierung der Beanspruchungsbilanz (Wieland, 2004, 2008; Wieland et al., 2004). Inzwischen liegt die EEB auch in einer Kurzform, dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB), vor (Wieland & Hammes, 2014). Dieses Instrument umfasst die ursprünglich vier Dimensionen des Zusatzmoduls psychische Beanspruchung der SynBA (mentale, motivationale, emotionale und physische Beanspruchung) sowie das Kontrollerleben. Mentale und motivationale Beanspruchung repräsentieren hier die funktionale Beanspruchung, emotionale und physische Beanspruchung repräsentieren die dysfunktionale Beanspruchung. Es ist daher ebenfalls möglich, durch Gegenüberstellung funktionaler und dysfunktionaler Beanspruchung eine Beanspruchungsbilanz zu bilden. Wieland und Hammes (2014) können zeigen, dass verschiedene Charakteristika eines Arbeitssystems (Arbeitsaufgabe und -situation, Führung und Zusammenarbeit, Gesundheitsindikatoren) systematisch und in erwarteter Weise mit bestimmten Beanspruchungsmustern (funktionale und dysfunktionale Beanspruchung, Beanspruchungsbilanz und Kontrollerleben, operationalisiert mit dem WSIB) einhergehen. Demnach liegen eine umfassende (25 Items) und eine kurze Version (neun Items) der EEB vor, die psychische Belastung (Charakteristika eines Arbeitssystems) im Sinne psychischer Beanspruchungsmuster systematisch abbilden können. In Tabelle 2 sind sämtliche EEB-Items der aktuellen Fassung (Wieland, 2014b) sowie ihre Zuordnungen zu den theoriegeleiteten Dimensionen bzw. faktorenanalytischen Komponenten aufgelistet.

Begründung und Formulierung des empirisch-praktischen Untersuchungsanliegens

Tabelle 2. Struktur der Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben während der Arbeit (EEB; Scherrer, 2002; Wieland, 2014b; Wieland & Hammes, 2014; Wieland, Metz & Richter, 2001).

EEB-Item	Theoriegeleitete Dimension	Faktorenanalytische Komponente
konzentriert*	Mentale Beanspruchung	Funktionales Beanspruchungserleben
aufmerksam*		
leistungsbereit*	Motivationale Beanspruchung	
energiegeladen*		
einflussreich*	Kontrollerleben	Positives Kontrollerleben
selbstsicher		
beschwingt		
vergnügt	Positives Gefühl	
lustig		
fröhlich		
zufrieden		
leicht		
nervös*	Nervöse Anspannung/ Emotionale Beanspruchung	Dysfunktionales Beanspruchungserleben
aufgeregt*		
entspannt ^{a)}		
körperlich verspannt*	Körperliche Beschwerden	
körperlich unwohl*		
zornig	Ärger	
aggressiv		
gelangweilt	Monotonie	
ängstlich	Negatives Gefühl	
betrübt		
deprimiert		
bedrückt		
ärgerlich		

*Dieses Item gehört auch zum Wuppertaler Screening Instrument psychische Beanspruchung (WSIB; Wieland & Hammes, 2014). ^{a)}Dieses Item wird für die Skalen-Bildung umgepolt.

Für die vorliegende Arbeit ist die Struktur der EEB vor allem für die Interpretation der generierten Klassifikatoren von Interesse. Jeder Klassifikator nutzt nur zwischen drei und fünf Items der EEB bzw. des WSIB. Diese Itemkombinationen werden je Kriterium algorithmisch

und datenbasiert ausgewählt (Merkmals-Selektion). Die beschriebenen Dimensionen der EEB werden herangezogen, um eine inhaltliche Interpretation dieser Merkmals-Selektionen in Zusammenhang mit den entsprechenden Kriterien vorzunehmen.

4.3 FORMULIERUNG DES FORSCHUNGSANLIEGENS

Im Rahmen der vorliegenden Arbeit werden die Eigenschaftswörter der EEB als Prädiktoren bzw. Features zur Vorhersage der oben aufgezählten weiteren Merkmale (psychische Belastung und langfristige Beanspruchungsfolgen) betrachtet. Mittels Verfahren der Feature-Selektion und statistischen Mustererkennung werden jene Kombinationen von Eigenschaftswörtern ermittelt, die besonders hohe Klassifikationsraten für diese Merkmale aufweisen. Die Merkmale werden hierzu dichotomisiert ("günstig/ungünstig" bzw. "hoch/niedrig"). In Anknüpfung an vorhergehende Untersuchungen (Hammes, Wieland & Winizuk, 2009; Hammes & Wieland, 2012a, b, 2014; Hammes, Wieland & Krajewski, 2011; Kieseler & Hammes, 2012; Wieland & Hammes, 2014b) besteht das Ziel darin, die EEB als ökonomisches und praktikables Screening-Instrument für psychische Belastung, psychische Beanspruchung und langfristige Beanspruchungsfolgen zu stärken. Hiermit soll der bestehende Bedarf an geeigneten Instrumenten für die Praxis bedient werden. Im Einzelnen werden folgende Fragestellungen verfolgt: (a) Welche Eigenschaftswörter der EEB sind für die Einschätzung psychischer Belastung und langfristiger Folgen beizubehalten? (b) Wie kann die EEB bei Vorgabe von einer Teilmenge der Merkmale geeignet gekürzt werden? (c) Welche Vorhersagegenauigkeiten lassen sich mit Hilfe der Eigenschaftswörter der EEB in Bezug auf die einzelnen Merkmale maximal erreichen? (d) Welche praktischen Empfehlungen können auf Grund der Befunde bezüglich der Fragen (a) bis (c) abgeleitet werden? (e) Wie lassen sich die Befunde bezüglich der Fragen (a) bis (d) in ein praktikables Instrument überführen? (f) Wie muss die EEB ggf. verbessert bzw. erweitert werden, um für alle Merkmale eine hinreichende Klassifikationsrate zu erzielen? Bevor diesen Fragestellungen nachgegangen wird, erfolgt zunächst eine theoretische Darstellung und Begründung der gewählten methodischen Verfahren.

5 KLASSIFIKATOR-INDUKTION: METHODISCHES VORGEHEN UND THEORETISCHE BEGRÜNDUNG

Im vorliegenden Kapitel steht das methodische Vorgehen des Autors im Vordergrund. Dieses Vorgehen rankt sich um Methoden der Klassifikator-Induktion. Damit ist gemeint, dass aus gegebenen Daten eine Rechenvorschrift ermittelt wird, die einzelne Instanzen (z. B. Angaben von Personen auf einem Fragebogen) möglichst zuverlässig in ein vorgegebenes System von Klassen (z. B. eine Diagnose trifft zu oder trifft nicht zu) einordnet (Kotsiantis, Zaharakis, & Pintelas, 2006). In der psychologischen Forschung gehören solche Verfahren zum methodischen Standard. Binärlogistische Regression kann beispielsweise als Klassifikator-Induktion aufgefasst werden, da von dichotomisierten (zweiwertigen) Daten verschiedener Variablen auf die Zugehörigkeit zu einer von zwei Klassen geschlossen wird. Lineare Regression geht hier sogar weiter, da der Wertebereich eines Regressionsmodells als kontinuierliches Klassensystem aufgefasst werden kann (Hastie et al., 2009). Ungeachtet dessen werden die statistischen Standardmethoden in der psychologischen Forschung zu selten als Klassifikator-Induktoren - oder allgemeiner: als Modell-Induktoren - aufgefasst und noch seltener in einen entsprechenden methodischen Rahmen eingebettet. Kaum ein psychologisches Modell wird nach Induktion mit Hilfe von Testdaten geprüft. Man beschränkt sich hier gerne auf die Angabe von Fit-Indizes. Kaum ein psychologisches Modell wird im Hinblick auf die Korrektur von Vorhersagefehlern kalibriert. Zwar werden häufig Fehlerschätzungen genannt, diese fließen jedoch selten zurück in die Modell-Anpassung. Dass derart kritische Prüfungen und Verfeinerungen von datenbasierten Modellen aber einen Mehrwert bieten, beweisen derzeitige gesellschaftliche Debatten über Massendatenspeicherung und Big Data. Die Gestalter unserer Zukunft, allen voran Technologiegiganten wie Amazon, Apple, Facebook, Google und Microsoft, verdienen jährlich mehrere Milliarden US-Dollar mit Klassifikatoren. Automatisiertes Fahren und Industrie 4.0 sind nur auf Grundlage guter Klassifikatoren möglich. Behörden, Versicherer, Banken, Börsenhändler, aber auch Naturwissenschaftler und Ingenieure vertrauen in zunehmendem Maße geprüften und stetig verbesserten datenbasierten Vorhersagemodellen. Und mit diesen Modellen wird in großem Ausmaß etwas grundständig psychologisches relativ zuverlässig vorhergesagt: Das Erleben und Verhalten von Menschen in bestimmten Lebensumständen. Diese Vorhersagen werden genutzt, um - wie auch immer zu bewertende - gewinnbringende kulturelle Fakten zu schaffen. Die gesellschaftlichen Debatten beziehen sich freilich auf den eingegangenen faustischen Handel, den man erst erkennt, wenn man den Versuchungen dieser neuen Technologien längst erliegt. Alleine, um hier den technologischen und methodologischen Anschluss zu sichern und die Menschen bei diesen

gesellschaftlichen Entwicklungen angemessen praktisch begleiten zu können, scheint es für eine stärkere Öffnung der psychologischen Forschung in Richtung moderner wissenschaftlicher Datenverarbeitung (Klassifikation, Big Data etc.) höchste Zeit zu sein.

In der vorliegenden Arbeit wird am Beispiel eines praxisrelevanten arbeitspsychologischen Problems dargestellt, wie solche Methoden in der psychologischen Forschung angewendet werden können. Doch das Feld der Klassifikator-Induktion ist weit. Das Ziel der Arbeit kann nicht darin bestehen, einen vollständigen Überblick über alle Techniken der Klassifikator-Induktion zu liefern. Hierfür sei auf einschlägige Fach- und Lehrbücher verwiesen (z. B. Clarke, Fokoué & Zhang, 2009; Hastie et al., 2009; Maimon & Rokach, 2005). Im Folgenden werden nur jene theoretischen Grundlagen und speziellen Techniken der Klassifikator-Induktion beschrieben, die für das Verständnis der gewählten Vorgehensweise relevant sind. Um der Leserschaft das Verständnis der Vorgehensweise zusätzlich zu erleichtern, werden die theoretischen Grundlagen im vorliegenden Methoden-Teil dort beschrieben, wo sie gebraucht werden. Aus demselben Grunde wird auf einem gesonderten theoretischen Vorspann zur Klassifikator-Induktion verzichtet.

Für eine genaue Darstellung des methodischen Vorgehens und der theoretischen Begründung ist eine mathematische und algorithmische Formalisierung unverzichtbar. Diese Formalisierung beschränkt sich im Folgenden auf das Notwendigste und wird erläutert. Dennoch kann nicht davon ausgegangen werden, dass die Leserschaft mit allen Notationen vertraut ist. Daher findet sich in Anhang A1 eine Zusammenstellung von verwendeten mathematischen und algorithmischen Notationen mit Erläuterungen, die das Verständnis hoffentlich erleichtern.

Der Aufbau des Methoden-Teils orientiert sich an der chronologischen Reihenfolge, in der die einzelnen Analyseschritte erfolgen: (1) Auswahl geeigneter Stichproben, (2) Auswahl einer geeigneten Analyse- und Entwicklungsumgebung, (3) Aufbereitung der Daten, (4) deskriptiv-statistische Analyse, (5) grundlegendes Design der eingesetzten Klassifikator-Induktoren inklusive Kreuzvalidierung und Erläuterung verwendeter Gütemaße, (6) Beschreibung der verschiedenen Experimentalreihen zur Klassifikator-Induktion inklusive Beschreibung der speziellen Spezifikationen (Art der Klassifikatoren und Vorgehensweisen bei der Merkmals-Selektion), (7) Finale Auswahl und Kalibrierung geeigneter Klassifikatoren mittels Bootstrapping, (8) Validierung und Anwendung der finalen Klassifikatoren und schließlich (9) Darstellung der Ergebnisse. Diese Analyseschritte entsprechen weitgehend den Empfehlungen,

wie sie der Literatur zur Klassifikator-Induktion entnommen werden können (Jain et al., 2000; Kotsiantis et al., 2006; Krizek, 2008; Maimon & Rokach, 2005).

5.1 AUSWAHL GEEIGNETER STICHPROBEN

Für die vorliegende Arbeit werden Stichproben für drei unterschiedliche Zwecke verwendet. Vier Stichproben werden für die Induktion von Klassifikatoren verwendet, zwei Stichproben werden verwendet, um die resultierenden Klassifikatoren zu validieren (Guyon & Elisseeff, 2003) und eine weitere Stichprobe wird zur Demonstration der praktischen Anwendbarkeit der in dieser Arbeit präsentierten Methoden eingesetzt. Diesen Stichproben ist gemeinsam, dass sie jeweils im Rahmen von Interventionsprojekten zu betrieblicher Gesundheitsförderung erhoben wurden. Damit ist die Auswahl der zu Grunde liegenden Datenbasis bereits ein erster Schritt für die Sicherstellung des praktischen Nutzens hier präsentierter Forschungsergebnisse.

5.1.1 Induktionsstichproben

Zur Klassifikator-Induktion werden Daten aus vier Mitarbeiterbefragungen verwendet, die im Rahmen von Interventionsprojekten zur betrieblichen Gesundheitsförderung durchgeführt wurden. Die Eignung als Induktionsstichprobe bemisst sich in erster Linie daran, ob die vollständige Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchung während der Arbeit (EEB) zum Einsatz kam. Darüber hinaus muss mindestens eines der folgenden Instrumente im eingesetzten Fragebogen-Inventar enthalten sein: (a) die Mensch-Aufgaben-Schnittstelle der Synthetischen Beanspruchungs- und Arbeitsanalyse (SynBA), (b) der Fragebogen zu Führung und Zusammenarbeit (FFZ), (c) der Gesundheitskompetenz-Fragebogen (GKF), (d) der Fragebogen zu Häufigkeiten körperlicher Beschwerden (HkB) oder (e) Fragen zu Absentismus- und Präsentismustagen. Die im Folgenden beschriebenen Stichproben erfüllen diese Bedingungen.

Versicherer: Bei einem Versicherungsunternehmen konnten $N = 889$ Datensätze erhoben werden. 370 der befragten Personen waren männlich, 495 Personen weiblich. Das Alter wurde in Kategorien abgefragt, der Median liegt in der Altersgruppe der 31- bis 40-Jährigen. Die Befragten übten verschiedene Verwaltungs- und Bürotätigkeiten aus. 127 der befragten Personen gaben an, Führungsverantwortung zu tragen. Es wurden alle oben aufgeführten Instrumente eingesetzt.

Diverse: Im Rahmen eines überbetrieblichen Gender-Projektes konnten $N = 300$ Datensätze erhoben werden. 154 hiervon stammten von männlichen und 137 von weiblichen Befragten. Das Alter wurde in Kategorien abgefragt, der Median liegt in der Altersgruppe der 41- bis 50-

Jährigen. Die Befragten übten diverse, nicht näher beschriebene Tätigkeiten aus. 13 der befragten Personen gaben an, Führungsverantwortung zu tragen. Es wurden alle oben aufgeführten Instrumente eingesetzt.

Entsorgung: An einem Projekt zur Förderung der Rückengesundheit nahmen bundesweit elf kommunale Abfallentsorgungs- und Straßenreinigungsunternehmen teil. Im Rahmen dieses Projektes konnten $N = 634$ Datensätze erhoben werden. 543 männliche Befragte überwogen stark gegenüber 30 weiblichen Befragten. Das Alter wurde in Kategorien abgefragt, der Median liegt in der Altersgruppe der 40- bis 44-Jährigen. Es wurden hauptsächlich Abfallentsorger und Straßenreiniger befragt. 93 der befragten Personen gaben an, Führungsverantwortung zu tragen. Mit Ausnahme der Fragen zu Absentismus- und Präsentismustagen wurden alle oben aufgeführten Instrumente eingesetzt. An der Erhebung dieser Stichprobe war der Autor beteiligt.

Informationstechnologie: Bei einem Unternehmen aus der IT-Branche konnten $N = 303$ Datensätze erhoben werden. 209 der Befragten waren männlich, 93 weiblich. Das Alter sollte in Jahren angegeben werden, der Median liegt bei 38 Jahren. In dem Projekt wurden vordergründig Beschäftigte mit Bildschirmarbeitsplätzen befragt. 48 der Befragten gaben an, Führungsverantwortung zu tragen. Von den oben aufgeführten Instrumenten wurde neben der vollständigen EEB ausschließlich die SynBA eingesetzt.

Die genannten Stichprobenumfänge beziehen sich auf die Stände vor Bereinigung und Aufbereitung der Daten. Beispielsweise gehen eventuelle Unstimmigkeiten nach Addition der Geschlechter-Häufigkeiten auf fehlende oder ungültige Angaben zurück. Bei der späteren deskriptiv-statistischen Auswertung der einzelnen Arbeitsplatzmerkmale und Gesundheitsindikatoren (Kriterien) kann es auf Grund fehlender oder ungültiger Werte ebenfalls zu teils starken Schwankungen der Stichprobenumfänge kommen.

5.1.2 Validierungsstichproben

Wasserwirtschaft: Als eine Validierungsstichprobe liegen Daten einer Mitarbeiterbefragung aus einem öffentlich-rechtlichen Verband für Wasserwirtschaft vor. Die Mitarbeiterbefragung wurde im Rahmen eines Projektes zum demografischen Wandel durchgeführt. Die Stichprobe umfasst $N = 159$ Datensätze. Unter den Befragten waren 120 Personen männlich und 36 Personen weiblich. Das Alter wurde in Kategorien erfragt, der Median liegt in der Gruppe der 45- bis 49-Jährigen. Befragt wurden technische Fachkräfte verschiedener Standorte des Unternehmens sowie Angestellte der Verwaltung. 41 Befragte gaben an, Führungsverantwortung zu tragen. Der eingesetzte Fragebogen enthielt die Instrumente SynBA, FFZ,

GKF und HkB sowie Fragen nach Absentismus- und Präsentismustagen. Aus der EEB wurden nur zwölf Items eingesetzt. In dieser Auswahl sind jedoch die neun Items des WSIB enthalten. An der Erhebung dieser Stichprobe war der Autor beteiligt.

Technologieunternehmen: Eine weitere Validierungsstichprobe stammt aus einer Mitarbeiterbefragung in einem transnationalen Hochtechnologieunternehmen mit Sitz in Deutschland. Diese Stichprobe umfasst $N = 160$ Datensätze. 38 Personen gaben weibliches, 120 Personen männliches Geschlecht an. Das Alter wurde in Kategorien erfasst, der Median liegt in der Gruppe der 41 bis 50-Jährigen. Teilgenommen haben Beschäftigte aus den Bereichen Fertigung, Einkauf, Technik und Entwicklung. 29 Personen gaben an, Führungsverantwortung zu tragen, 131 waren ohne Führungsverantwortung. Mit Ausnahme von Absentismus- und Präsentismustagen wurden alle oben aufgeführten Kriterien erhoben. Bei der Befragung wurde ausschließlich das WSIB (nicht die vollständige EEB) eingesetzt. Diese Stichprobe ist unterteilt in 11 Bereiche mit Stichprobenumfängen zwischen 7 und 22 Teilnehmern. Für jeden dieser Bereiche können separate Profile bestimmt und an Hand tatsächlicher Messungen validiert werden. An der Erhebung dieser Stichprobe war der Autor beteiligt.

Die Daten aus den beiden Validierungsstichproben eignen sich, um Klassifikatoren bezüglich des WSIB zur Vorhersage von 16 bzw. allen 18 Kriterien zu validieren (Guyon & Elisseff, 2003). Eine Stichprobe, mit der die Klassifikatoren bezüglich der vollständigen EEB validiert werden könnten, steht leider nicht zur Verfügung.

5.1.3 Anwendungsstichprobe

Für die Demonstration der Anwendungsmöglichkeiten vorliegender Forschungsbefunde kann auf eine weitere Stichprobe aus dem oben genannten transnationalen Hochtechnologieunternehmen zurückgegriffen werden. In diesem Unternehmen wird im Rahmen eines kontinuierlichen Projekts des betrieblichen Gesundheitsmanagements ausschließlich das WSIB eingesetzt. Es wird dort genutzt, um für einzelne Unternehmensbereiche den Bedarf für Maßnahmen der betrieblichen Gesundheitsförderung zu ermitteln und entsprechende Interventionen zu evaluieren. Aus diesem Unternehmen liegen $N = 551$ vollständige Datensätze zur Bedarfseinschätzung vor. Sie sind nach Unternehmensbereichen differenziert, sodass Vergleiche zwischen Funktionen (z. B. Produktion [$N = 151$] und Führungskräfte [$N = 36$]) möglich sind. An der Erhebung dieser Stichprobe war der Autor beteiligt. Ausgehend von diesen Daten können durch Klassifikation Profile für Personengruppen (z. B. Unternehmensbereiche oder Funktionen) erstellt werden, die die Ausprägungen der 18 Kriterien vorhersagen. Diese Profile können auf ihre praktische Anwendbarkeit und

Plausibilität geprüft werden. Da ausschließlich das WSIB eingesetzt wird, stehen weitergehende Informationen über die vorliegende Stichprobe (z. B. Geschlechts- oder Altersverteilung) nicht zur Verfügung.

5.2 EINGESETZTE ANALYSE- UND ENTWICKLUNGSUMGEBUNG

Das dieser Arbeit zu Grunde liegende Projekt erforderte neben der Anwendung herkömmlicher statistischer Analyseverfahren zum größten Teil die Entwicklung bzw. Programmierung eigener Analyse- und Auswertungsverfahren, basierend auf Algorithmen, die in der Literatur beschrieben werden. Daher war es notwendig, eine Analyse- und Entwicklungsumgebung einzusetzen, die neben Standardverfahren größtmögliche Freiheit bei der Entwicklung eigener Verfahren erlaubt. Eine geeignete Umgebung wurde in SageMathCloud™ gefunden. Diese Plattform bietet eine Integration des Mathematiksystems Sage (Informationen hierzu auf <http://www.sagemath.org>), des Statistiksystems R (Informationen hierzu auf <http://www.r-project.org>) sowie der Programmiersprache Python, die u. A. eine Einbindung von Mathematik-, Statistik- und Grafik-Bibliotheken erlaubt und auf die effiziente Verarbeitung strukturierter Daten ausgelegt ist. Die Benutzer-Schnittstelle von SageMathCloud™ ist internetbasiert. D. h. die Umgebung wird über das Internet (<http://cloud.sagemath.org>) bedient und die Berechnungen finden auf leistungsfähigen Servern statt. Darüber hinaus bietet SageMathCloud™ für jedes Projekt eine vollständige Dateiverwaltung und Konsolen-Bedienung sowie Zusatzleistungen wie etwa Aufgabenlisten, Benutzerverwaltung u. v. m. Die Planung, Verwaltung, Datenhaltung und Durchführung des Projektes konnte so in einer einzigen Umgebung vorgenommen werden. Neben der Verfügbarkeit einer reichhaltigen Umgebung und leistungsfähiger Server war die ständige Verfügbarkeit des vollständigen Projekts über das Internet (auch über minder leistungsfähige Client-Computer mit unterschiedlichen Betriebssystemen) eine zusätzliche Erleichterung bei der Durchführung. Sämtliche im Folgenden vorgetragenen Befunde wurden in dieser Umgebung unter Verwendung eigenständig programmierter Algorithmen ermittelt.

5.3 AUFBEREITUNG DER DATEN

Um einen Klassifikator induzieren und anwenden zu können, ist die Aufbereitung der entsprechenden Daten ein bedeutender vorbereitender Schritt. Die Daten müssen in einheitlicher Weise vom Klassifikator-Induktor verarbeitet werden können und es müssen Regeln für den Umgang mit fehlenden oder ungültigen Daten festgelegt werden (Jain et al., 2000; Kotsiantis, Kanellopoulos & Pintelas, 2006; Kotsiantis, Zaharakis & Pintelas, 2006;

Maimon & Rokach, 2005). Die Anwendungsstichprobe liegt als Datenbank-Tabelle vor, die als Datei im CSV-Format (Comma Separated Values) exportiert wurde. Die weiteren Daten liegen als SPSS-Dateien, getrennt nach sechs Stichproben (Induktions- und Validierungsstichproben), vor. Bei der Aufbereitung der Daten steht zunächst die vereinheitlichte Kodierung der Rohwerte im Zentrum. Mit Rohwerten sind die Angaben der Befragten zu den einzelnen Items der weiter oben aufgeführten Instrumente (EEB, SynBA, FFZ, GKF, HkB sowie Absentismus- und Präsentismustage) gemeint. Die vorliegenden SPSS-Dateien werden auf übereinstimmende Items geprüft, die Items werden einheitlich mit Variablennamen versehen. Anschließend werden die Items von SynBA, FFZ, GKF und HkB einheitlich nach bestehenden Auswertungsvorschriften zu den einzelnen Merkmalen bzw. Skalen aggregiert. Schließlich wird für jedes Merkmal bzw. jede Skala eine eigene Daten-Datei im CSV-Format angelegt. Diese CSV-Dateien enthalten eine laufende Nummer für die Befragten, eine laufende Nummer für die Herkunfts-Stichprobe, Rohwerte für alle 25 EEB-Items sowie das entsprechende aggregierte Merkmal. Für die Generierung von Vorhersage-Modellen werden die Daten der vier Induktionsstichproben (Versicherer, Diverse, Entsorgung und Informationstechnologie) zusammengeführt. An anderer Stelle (Hammes et al., 2009; Wieland & Hammes, 2014; Wieland et al., 2009) konnte bereits gezeigt werden, dass die Zusammenlegung und gemeinsame Analyse dieser Daten statistisch vertretbar ist.

Für die Validierung der Vorhersage von Erwartungswerten werden die Daten der Validierungsstichproben (Wasserwirtschaft und Technologieunternehmen) nach demselben Vorgehen separat in CSV-Dateien gespeichert. Diese Dateien enthalten nur jene neun EEB-Items, die dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB) entsprechen, sowie die jeweiligen aggregierten Merkmale.

Für die Demonstration der Anwendungsmöglichkeiten wird eine weitere CSV-Datei angelegt (s. o.). Hierin sind ausschließlich Rohwerte der neun WSIB-Items sowie eine laufende Nummer für Bereiche aus der Anwendungsstichprobe (Technologieunternehmen) enthalten.

Allen CSV-Dateien ist gemeinsam, dass ausschließlich Instanzen (Datensätze von befragten Personen) mit durchgehend gültigen Werten aufgenommen wurden. D. h. diese Dateien enthalten keine als fehlend oder ungültig gekennzeichneten Werte.

5.4 EINTEILUNG IN ZWEI KLASSEN: DICHOTOMISIERUNG

Ziel der vorliegenden Arbeit ist die Ermittlung von Verfahren, mit denen man, ausgehend von den EEB-Items, auf 18 Merkmale der Arbeitssituation und des Gesundheitszustandes von

Beschäftigten in Betrieben schließen kann. Diese 18 Merkmale sollen dabei in zwei Klassen (0 = niedrige Werte, 1 = hohe Werte) vorhergesagt werden. Es stellt sich die Frage, warum eine solche Vereinfachung der Datenlage vorgenommen wird und warum ausschließlich auf den einfachsten Fall von zwei Klassen eingegangen wird. Bevor das methodische Vorgehen bei der Dichotomisierung näher beschrieben wird, sollen daher zunächst diese Entscheidungen theoretisch begründet werden.

5.4.1 Theoretische Begründung für die Dichotomisierung der Daten

Für die vorliegende Arbeit ist der Beitrag von Hughes (1968) über die mittlere Vorhersagegenauigkeit von Verfahren zur statistischen Mustererkennung von besonderer Relevanz. Hughes (1968) gibt mathematisch exakte Antworten und mathematisch fundierte Schätzungen zu der Frage nach der Anzahl zu wählender Klassen, der Anzahl zu wählender Prädiktoren sowie der zu erwartenden Korrektklassifikation (Vorhersagegenauigkeit; Anteil an Instanzen, der durch ein gegebenes Verfahren korrekt den vorgegebenen Klassen zugeordnet wird) bei gegebener Anzahl von Instanzen (Umfang einer Stichprobe). Hierbei setzt er explizit keinerlei Vorbedingungen bezüglich Art und Form der Verteilung von Prädiktor-Messwerten voraus. Er setzt auch explizit keine Unabhängigkeit der Prädiktoren voraus, da diese Bedingung bei realen Daten kaum erfüllt ist. Seine Resultate sind demnach für jede Art und Form der Verteilung gültig.

Ein wichtiges Resultat von Hughes (1968) besteht in der Feststellung, dass mit zwei gleich verteilten Klassen 0 und 1 bei unendlich vielen Messungen die höchste theoretisch mögliche Korrektklassifikation κ zu erreichen ist, denn diese berechnet sich nach der Formel

$$\kappa_p = 1 - p(1 - p),$$

mit p : Wahrscheinlichkeit für Klasse 0. Das Maximum dieser Funktion liegt bei $p = 1/2$.¹¹ Dies entspricht zwei gleich verteilten Klassen. In diesem Falle beträgt die Korrektklassifikation $\kappa_{1/2} = .750$. Für mehr als zwei gleichverteilte Klassen muss obige Formel rekursiv angewendet werden. Bezeichnen wir mit $\kappa(c)$ die Korrektklassifikation bei c gleichverteilten Klassen. Dann ist die Korrektklassifikation bei $c + 1$ gleichverteilten Klassen gegeben durch

$$\kappa(c + 1) = \frac{\kappa_{1/(c+1)}}{c + 1} + \frac{\kappa_{1/(c+1)} \kappa(c) c}{c + 1}.$$

Dieser Ansatz ergibt sich aus folgender Argumentation: Bei den $c + 1$ gegebenen Klassen $\{0..c\}$ fasse man zunächst die Klassen $\{1..c\}$ zu Klasse 1' zusammen. Klasse 0 hat dann die

¹¹ Für $p \in \{0, 1\}$ ist $\kappa_p = 1$. Hierbei gehören alle Instanzen derselben Klasse an.

Wahrscheinlichkeit $1/(c + 1)$, Klasse 1' die Wahrscheinlichkeit $c/(c + 1)$. Nach Hughes (1968) gelingt die korrekte Klassifikation auf diese beiden Klassen mit

$$\kappa_{1/(c+1)} = \kappa_{c/(c+1)} = 1 - \frac{1}{c+1} \left(1 - \frac{1}{c+1}\right) = \frac{(c+1)^2 - (c+1) + 1}{(c+1)^2}.$$

Die Korrektklassifikation der Klassen $\{1..c\}$ gelingt schließlich mit $\kappa(c)$ nach Induktionsvoraussetzung. Mit diesem Ansatz lässt sich durch vollständige Induktion beweisen (siehe Anhang A1), dass

$$\kappa(c) = \frac{1}{c} \sum_{i \in \{1..c\}} \prod_{j \in \{i..c\}} \kappa_{1/j} = \frac{1}{c} \sum_{i \in \{1..c\}} \prod_{j \in \{i..c\}} \frac{j^2 - j + 1}{j^2}.$$

Im Falle von drei gleichverteilten Klassen ist die theoretisch bestmögliche Korrektklassifikation mit $\kappa(3) = .648$ bereits um gut 10 % geringer als bei zwei gleichverteilten Klassen. Bei vier gleichverteilten Klassen sinkt die theoretisch bestmögliche Korrektklassifikation nochmals um 5 % auf $\kappa(4) = .598$. Damit wird deutlich, dass die höchste Korrektklassifikation bei der Vorhersage von zwei gleichverteilten Klassen zu erwarten ist.

Ein weiteres wichtiges Resultat von Hughes (1968) bezieht sich auf die optimale Messkomplexität bei endlich vielen Messungen. Mit Messkomplexität meint er die notwendige Anzahl von Zellen, in denen Häufigkeiten der verschiedenen möglichen Messwerte der verwendeten Prädiktoren abgelegt werden können. Zunächst stellt er fest, dass die maximal mögliche Korrektklassifikation bei *endlich* vielen Messungen unterhalb jener theoretisch maximal möglichen Korrektklassifikation für unendlich viele Messungen nach obiger Formel liegt. Weiter stellt er fest, dass es bei endlich vielen Messungen eine optimale Messkomplexität gibt. Bei dieser Messkomplexität wird die maximal mögliche Korrektklassifikation erreicht. Liegt die Messkomplexität ober- oder unterhalb dieser optimalen Messkomplexität, dann wird die maximal mögliche Korrektklassifikation immer unterboten. Geht man von 1 000 Messungen aus (in dieser Größenordnung liegen die Umfänge der vorliegenden Stichproben), so liegt die optimale Messkomplexität bei 23 Zellen. Ein EEB-Item hat die Messkomplexität 7, da Häufigkeiten der sieben möglichen Messwerte (Antwortmöglichkeiten) abgebildet werden müssen. Zwei Items haben die Messkomplexität $7^2 = 49$ und drei die Messkomplexität $7^3 = 343$. Im ersten Fall wird die optimale Messkomplexität unterboten, im zweiten und dritten überboten. Die maximal mögliche Korrektklassifikation kann also in allen drei Fällen nicht erzielt werden. Daher werden die sieben Antwortkategorien der EEB-Items für einige Experimentalreihen auf die zwei Werte 0 und 1 abgebildet (s. u.). Damit lässt sich die Messkomplexität von m kombinierten EEB-Items als 2^m angeben. Bei vier EEB-Items betrüge die Messkomplexität 16

und läge noch unterhalb der optimalen Messkomplexität. Bei fünf EEB-Items betrüge die Messkomplexität schon 32 und läge oberhalb der optimalen Messkomplexität. Es ist also zu erwarten, dass maximal vier bis fünf auf diese Weise rekodierte EEB-Items für eine akzeptable Korrektklassifikation zu kombinieren sind.

5.4.2 Methodisches Vorgehen bei der Dichotomisierung

Die Dichotomisierung der Daten gehört zu der vorbereitenden Aufbereitung der Daten (Jain et al., 2000; Kotsiantis, Kanellopoulos & Pintelas, 2006; Kotsiantis, Zaharakis & Pintelas, 2006; Maimon & Rokach, 2005). Die Dichotomisierung der 18 vorherzusagenden Merkmale (Kriterien) erfolgt durch Mediansplit. Für die Dichotomisierung eines der 18 Kriterien werden alle vorliegenden N Instanzen (Einträge in der entsprechenden CSV-Datei, s. o.) zunächst der Rangreihe dieses Kriteriums nach aufsteigend sortiert. Anschließend werden diese Instanzen in einer Schleife durchlaufen, während ein Schleifenzähler i , beginnend bei Null, hochzählt. Einer neuen Variablen wird je Instanz eine 0 zugewiesen, solange $i \leq N/2$. Für die verbleibenden Instanzen wird diese Variable auf 1 gesetzt. Auf diese Weise werden je Kriterium zwei Klassen mit dem Umfang von $N/2$ Instanzen generiert. Klasse 0 enthält Instanzen mit niedrigen Werten im jeweiligen Kriterium, Klasse 1 Instanzen mit hohen Werten.

Um eine hinreichende Zellenbesetzung bei der Kombination von EEB-Items zu gewährleisten, sowie Anforderungen bestimmter Verfahren (z. B. binärlogistische Regression, Messkomplexität bei statistischer Mustererkennung, s. o.) zu genügen, müssen auch die EEB-Items dichotomisiert werden. Diese Dichotomisierung wird jeweils spezifisch für die 18 Kriterien vorgenommen, indem Rekodierungs-Vorschriften von den sieben Antwortkategorien auf die zwei Werte 0 und 1 erzeugt werden. Grundlage sind bedingte Wahrscheinlichkeiten. Bei gegebenem Kriterium und gegebenem EEB-Item werden für jede Antwortkategorie $a \in \{1..7\}$ die Wahrscheinlichkeiten $p(a|0)$ bzw. $p(a|1)$ unter der Bedingung, dass eine Instanz zu Klasse 0 bzw. 1 gehört, ermittelt. Die Antwortkategorie a bzw. der Rohwert des betreffenden Items einer Instanz wird in den Wert 0 rekodiert, wenn $p(a|0) > p(a|1)$. Andernfalls wird in den Wert 1 rekodiert. Sämtliche Rekodierungen dieser Art sind in Anhang A2 aufgeführt.

Sowohl die dichotomisierten Daten als auch die ermittelten Rekodierungs-Vorschriften werden für die spätere Verwendung in separaten strukturierten Dateien abgelegt.

5.5 DESKRIPTIV-STATISTISCHE ANALYSE

Die 25 Items der EEB werden auf Grund ihrer Bedeutung als Einzelmerkmale in der vorliegenden Arbeit jeweils separat deskriptiv-statistisch ausgewertet. Für jedes Item werden

zunächst alle vorliegenden Instanzen (alle vorliegenden Angaben von Personen zu diesem Item) zusammengetragen. Anschließend werden die Anzahl der Instanzen, das arithmetische Mittel sowie die Standardabweichung bestimmt. Darüber hinaus wird für jede der sieben Antwortkategorien die relative Häufigkeit bestimmt. Somit liegt die Verteilung der Antworten auf die Antwortkategorien vor. Diese Daten werden für jedes Item in einem Histogramm visualisiert. Aus diesen Histogrammen kann der Modus (häufigste Antwortkategorie) sowie der Median (Antwortkategorie, bei der die Stichprobe in Hälften unterteilt werden kann) abgelesen werden. Die resultierenden 25 Histogramme werden dem Rang des Mittelwertes folgend aufsteigend in einer Abbildung zusammengeführt. Diese Rangreihe ist weitgehend identisch mit der Rangreihe der Schiefe. Der Abbildung können daher Antworttendenzen entnommen und interpretiert werden.

Für die deskriptiv-statistische Analyse der 18 vorherzusagenden Merkmale (Kriterien) werden die aggregierten Werte aus den oben genannten CSV-Dateien verwendet. Es werden Anzahl der Instanzen, Modus, Median, arithmetisches Mittel, Standardabweichung sowie Minimum und Maximum bestimmt. Darüber hinaus wird für die beiden Klassen 0 und 1 (s. o.) der jeweilige Klassenmittelwert des Kriteriums ($\mu^{(0)}$ und $\mu^{(1)}$) bestimmt. Diese Klassenmittelwerte werden später für die Vorhersage von Erwartungswerten benötigt. Neben den genannten univariaten deskriptiven Kennwerten werden Spearman-Rang-Korrelationen zwischen den 25 EEB-Items und den 18 Kriterien angegeben. Sämtliche deskriptiv-statistischen Kennwerte werden für den späteren Abruf in anderen Verfahren sowie für die Verschriftlichung in einer zentralen CSV-Datei gesammelt.

5.6 THEORETISCHE GRUNDLAGEN DER KLASSIFIKATOR-INDUKTION

Bei der Klassifikator-Induktion geht es darum, auf der Basis von gegebenen Daten eine Berechnungsvorschrift zu finden (zu induzieren), die einen möglichst verlässlichen Schluss von bestimmten Prädiktoren (z. B. Angaben zu EEB-Items) auf bestimmte klassierte Kriterien (z. B. Gesundheitskompetenz niedrig oder hoch) erlaubt. Im vorliegenden Abschnitt wird dieses Ziel der Klassifikator-Induktion genauer definiert und anschließend das generelle Vorgehen bei der Klassifikator-Induktion beschrieben. Die Herausforderungen der Klassifikator-Induktion werden zum Teil an Hand historischer Entwicklungen beschrieben, da hierdurch Verbindungen zur psychologischen Forschungspraxis deutlich werden. Außerdem kann so begründet werden, warum in der vorliegenden Arbeit Merkmals-Selektion und statistische Mustererkennung unter dem übergeordneten Begriff der Klassifikator-Induktion behandelt werden.

5.6.1 Herausforderungen an die Klassifikator-Induktion

Die Entwicklung maschineller Klassifikatoren war schon immer inspiriert von den kognitiven Leistungen biologischer Organismen und insbesondere des Menschen (Marill & Green, 1960, 1963). Das Perceptron (Rosenblatt, 1960) stellt einen Prototypen dar, der das Schema *Informationsaufnahme (Messung) – Verarbeitung – Reaktion (Klassifikation)* nachempfunden. Die Klassifikator-Induktion ist daher ein Teilgebiet der Erforschung und Anwendung künstlicher Intelligenz (Carbonell, Michalski & Mitchell, 1983). Blicken wir zurück auf die vorhergehenden Kapitel der vorliegenden Arbeit, so vermag die maschinelle Simulation menschlicher Verarbeitung von Informationen höchstens näherungsweise oder in Ausschnitten zu gelingen. Detaillierte Kenntnisse zu Grunde liegender biologischer Prozesse (Forschungsstand) sind die eine, technische Möglichkeiten ihrer Übersetzung in maschinelle Prozesse (Sensoren, Aktoren, Prozessoren, mathematische Modellierung) die andere Limitation. Es ist jedoch davon auszugehen, dass der Wille zur Umsetzung des technisch Machbaren stets besteht. Das Human Brain Project (www.humanbrainproject.eu) wurde unter anderem zu diesem Zwecke von der Europäischen Kommission ausgerufen.

Merkmals-Selektion beschäftigt sich mit der Frage, welche Informationen für eine Klassifikation zwingend notwendig bzw. relevant sind, und welche vernachlässigt werden können. In den Anfängen der maschinellen bzw. Rechner-gestützten Klassifikation ging es häufig darum, physische Objekte an Hand bestimmter Merkmale bestimmten Klassen zuzuordnen. Ein Problem bestand lange Zeit im maschinellen Lesen von Texten. Wurden Normschriften verwendet, so konnte das Problem mit Hilfe von Schablonen physikalisch gelöst werden. Bei handschriftlichen Texten funktionierte das nicht. Die Herausforderung kann man sich bei der Unterscheidung der Zeichen "B" und "8" wohl recht gut vorstellen. Bei diesem und ähnlichen Problemen wurden die physikalischen Merkmale, die für die Unterscheidung relevant zu sein schienen, auf Grundlage des Vorwissens und der Erfahrung der Forscher extrahiert (Duda, 1970; Kanal, 1974). Dieses Beispiel verdeutlicht bereits, wie wesentlich Merkmals-Selektion für die Entwicklung guter Klassifikatoren ist. Es verdeutlicht jedoch auch, dass Merkmals-Selektion ursprünglich nicht maschinell vorgenommen wurde. Dies änderte sich, als für die zu klassifizierenden Objekte Beschreibungen bzw. Repräsentationen vorlagen, die maschinell verarbeitet werden konnten. Solche Repräsentationen sind zuweilen recht umfangreich, d. h. sie beschreiben die Objekte in einem hochdimensionalen Raum. Eine Herangehensweise zur Bewältigung solcher Datenmengen besteht in der Dimensionsreduktion. Verfahren wie die Diskriminanzanalyse, die Hauptkomponentenanalyse oder die Faktorenanalyse führten zu Beschreibungen von Objekten, die für die Entwicklung von

Klassifikatoren handhabbar waren (Clarke et al., 2009; Hastie et al., 2009; Jain et al., 2000; Kanal, 1974; Nagy, 1968). Solche Verfahren gehören auch zu den statistischen Standardverfahren der psychologischen Forschung. Seit den 1990-er Jahren wird im Detail untersucht, wie man Zusammenhänge zwischen verschiedenen Merkmalen für die Merkmals-Selektion nutzbar machen kann. Anders, als beispielsweise bei der Hauptkomponentenanalyse, besteht das Ziel nicht darin, korrelierte Merkmale zu einem einzigen Merkmal zusammenzufassen, sondern von korrelierten Merkmalen nur die relevanten Merkmale für die Klassifikation zu nutzen und redundante (mit relevanten hoch korrelierende) Merkmale auszuschließen (Blum & Langley, 1997; Caruna & Freitag, 1994; Clarke et al., 2009; Dash & Liu, 1997; Guyon & Elisseeff, 2003; Hua, Xiong, Lowey, Suh & Dougherty, 2005; Kohavi & John, 1997; Kohavi & Sommerfield, 1995; Krizek, 2008; Ng, 1998). Schrittweise lineare Regression kann als Beispiel für diese Vorgehensweise herangezogen werden (Breiman & Spector, 1992; Clarke et al., 2009; Hastie et al., 2009; Miller, 1984). Dieses Verfahren endet mit einem Satz von Merkmalen, die zwar einen hohen Zusammenhang mit der zu erklärenden Variable aufweisen, untereinander jedoch möglichst gering korrelieren. Verfahren zur Merkmals-Selektion können im Hinblick darauf untersucht werden, inwiefern sie tatsächlich relevante Merkmale auswählen und weniger relevante bzw. redundante ausschließen (Blum & Langley, 1997; Caruna & Freitag, 1994; Clarke et al., 2009; Dash & Liu, 1997; Guyon & Elisseeff, 2003; Kanal & Chandrasekaran, 1971; Kohavi & Sommerfield, 1995; Kudo & Sklansky, 2000; Krizek, 2008; Molina, Belanche & Nebot, 2002). Dass die Bestimmung relevanter Merkmale jedoch keine einfache Aufgabe darstellt, belegt bereits die Tatsache, dass zwei relevante Einzelmerkmale nicht zwingend ein für die Klassifikation geeignetes Paar von Merkmalen darstellen (Cover, 1974; Cover & Van Campenhout, 1977; Elashoff, Elashoff & Coldman, 1967; Guyon & Elisseeff, 2003; Toussaint, 1971).

Bereits im Zusammenhang mit der Dichotomisierung der Daten weiter oben wurde auf die Zusammenhänge zwischen Messkomplexität (Anzahl ausgewählter Merkmale), Umfang verwendeter Stichproben und Korrektklassifikation hingewiesen (Cover, 1965; Hughes, 1968). Diese Zusammenhänge haben sich auch in der Folgezeit mehrfach bestätigt (Blum & Langley, 1997; Clarke et al., 2009; Hua et al., 2005; Jain & Zongker, 1997; Kanal & Chandrasekaran, 1971; Kohavi & John, 1997; Kohavi & Sommerfield, 1995; Kudo & Sklansky, 2000; Maimon & Rokach, 2005; Ng, 1998; Raudys, 2006; Reunanen, 2003; Siedlecki & Sklansky, 1989). Werden zu viele Merkmale ausgewählt, dann ist der resultierende Klassifikator zu sehr an die verwendeten Daten (Trainingsdaten) angepasst und kann unbekannte Fälle nur unzureichend korrekt klassifizieren. Werden zu wenige Merkmale ausgewählt, dann gelingt die Trennung

verschiedener Klassen nicht in ausreichendem Maße, das Modell ist zu einfach. Die Korrektklassifikation eines Klassifikators ist umso stärker von diesem Zusammenhang beeinflusst, je geringer der Umfang zur Verfügung stehender Daten ist (Blum & Langley, 1997; Clarke et al., 2009; Hua et al., 2005; Kohavi & John, 1997; Ng, 1998).

Nicht nur die Trainingsdaten haben einen Einfluss darauf, welche Merkmale zu einer guten Klassifikation führen, sondern selbstverständlich auch die Art des zu induzierenden Klassifikators (Clarke et al., 2009; Doumpos & Slappa, 2005; Hua et al., 2005; Kohavi & John, 1997; Raudys, 2006). Maschinelle Klassifikation gehört zu den Verfahren der statistischen Mustererkennung. Bei statistischer Mustererkennung geht es darum, systematische Muster in den Beschreibungen von Instanzen (z. B. physischen Objekten) aufzufinden und neue Instanzen an Hand ihrer Beschreibungen solchen Mustern zuzuordnen (Kanal, 1974). Bei der Klassifikation geht man von der Existenz eines diskreten Klassensystems aus, sowie davon, dass jede Instanz genau einer Klasse angehört. Es sind solche Muster in den Beschreibungen der Instanzen aufzufinden, die möglichst fehlerfrei die Klassenzugehörigkeit der Instanzen widerspiegeln. Es wird zwischen supervidierter und nicht-supervidierter Klassifikation unterschieden (Carbonell et al., 1983; Fu, 1994; Jain et al., 2000; Kotsiantis, Zaharakis & Pintelas, 2006; Maimon & Rokach, 2005; Nagy, 1968). Bei supervidierter Klassifikation ist die Klassenzugehörigkeit der Instanzen bekannt und der Klassifikator wird darauf trainiert, die korrekte Klassenzuordnung von Instanzen an Hand ihrer Beschreibungen vorzunehmen. Bei nicht-supervidierter Klassifikation wird das Klassensystem durch die Unterscheidung systematischer Muster in den Beschreibungen der Instanzen erst entwickelt und daraufhin die Klassenzuordnung vorgenommen. In der vorliegenden Arbeit ist die Klassenzugehörigkeit der Instanzen stets bekannt, weshalb im Folgenden nur der Fall supervidierter Klassifikation vertieft wird.

Bereits in den Anfängen der Erforschung maschineller Klassifikation beobachtete man ein Phänomen, das man mit dem Begriff des optimistischen Bias belegte (Kanal, 1974; Kanal & Chandrasekaran, 1971; Nagy, 1968). Als wichtigstes Gütemaß für Klassifikatoren gilt die Korrektklassifikation (Vorhersagegenauigkeit) von Instanzen, die nicht in die Entwicklung des Klassifikators eingegangen sind (Test- oder Anwendungsdaten) (Dash & Liu, 1997; Duda, 1970; Kohavi, 1995). Schätzt man diese Korrektklassifikation alleine auf Grundlage der verwendeten Trainingsdaten, so wird sie in den meisten Fällen stark überschätzt. Daher hat sich die holdout-Methode durchgesetzt. Damit ist gemeint, dass man die zur Verfügung stehenden Daten in Trainings- und Testdaten unterteilt. Der Klassifikator wird mit den Trainingsdaten trainiert und die Korrektklassifikation wird mit Hilfe der Testdaten geschätzt (Kanal, 1974;

Kanal & Chandrasekaran, 1971; Kohavi, 1995). Eine entscheidende Frage besteht darin, in welchem Verhältnis ein Datensatz in Trainings- und Testdaten unterteilt wird. Diese Frage ist umso entscheidender, je weniger Daten zur Verfügung stehen. Wird der Trainingsdatensatz zu gering angesetzt, dann ist der Klassifikator nicht zur Generalisierung fähig. Wird er zu hoch angesetzt, dann ist der Klassifikator zu stark an die Trainingsdaten angepasst, was sich ebenfalls ungünstig auf die Fähigkeit zur Generalisierung auswirkt. Wird andererseits der Testdatensatz zu gering angesetzt, kommt es zu einer hohen Varianz bei der Klassenzuordnung und der entsprechend geschätzten Korrektklassifikation kann nur eingeschränkt vertraut werden (John, Kohavi & Pfleger, 1994; Kohavi, 1995; Kohavi & John, 1996; Kohavi & Sommerfield, 1995; Kohavi & Wolpert, 1996; Narendra & Fukunaga, 1977; Stone, 1974). Bezüglich des Umfanges der Testdaten hat sich folgende Heuristik als praktikabel erwiesen: Sind d Merkmale ausgewählt, so sollten je Klasse mindestens $10d$ Instanzen im Testdatensatz enthalten sein (Jain et al., 2000; Nagy, 1968). Rechnen wir, wie weiter oben ausgeführt wurde, mit etwa vier Merkmalen bei zwei Klassen so sollten die Testdatensätze mindestens 80 Instanzen umfassen.

Offenbar, so wird deutlich, sind Merkmals-Selektion und statistische Mustererkennung bzw. Klassifikation eng miteinander gekoppelt. Die Entwicklung eines guten Klassifikators gelingt nur bei einer geeigneten Auswahl von Merkmalen. Klassifikation und Merkmal-Selektion sind beide abhängig von Umfang und Qualität eingehender Daten. Aus diesem Grunde wird in der vorliegenden Arbeit der Begriff der Klassifikator-Induktion bevorzugt. Ein Klassifikator-Induktor führt sowohl die Merkmals-Selektion als auch das Training des Klassifikators durch. Die Merkmals-Selektion erfolgt durch einen Selektions-Algorithmus. Dieser generiert nach vorgegebenen Operationen verschiedene Sätze von Merkmalen. Mit jedem dieser Merkmals-Sätze wird ein Unterprogramm zum Trainieren eines entsprechenden Klassifikators aufgerufen. Die Kombination aus Merkmals-Satz und Klassifikator, die auf den Trainingsdaten die beste Korrektklassifikation erzielt, wird als Ergebnis der Klassifikator-Induktion ausgegeben. Da die Merkmals-Selektion in einer Schleife das Training des Klassifikators umklammert, wird ein solches Verfahren wrapper-Verfahren genannt (Blum & Langley, 1997; Dash & Liu, 1997; Guyon & Elisseeff, 2003; Kohavi & John, 1997; Kohavi & Sommerfield, 1995). Der Test dieser Kombination erfolgt im Anschluss außerhalb dieser Schleife. Weiter unten wird beschrieben, dass durch Festlegung verschiedener Trainings- und Testdatensätze (Kreuzvalidierung) verschiedene Kombinationen von Merkmals-Satz und Klassifikator induziert werden können. Kreuzvalidierung erfolgt in einer weiteren äußeren Schleife. In dieser Schleife können die unterschiedlichen Klassifikatoren hinsichtlich ihrer Korrektklassifikationen auf den Testdaten validiert und daraufhin untereinander verglichen werden (Blum & Langley, 1997; Clarke et al.,

2009; Dash & Liu, 1997; Hastie et al., 2009; Jain et al., 2000; Kohavi, 1995; Kohavi & John, 1997; Kohavi & Sommerfield, 1995; Krizek, 2008; Stone, 1974).

5.6.2 Formale Beschreibung eines Klassifikator-Induktors

Mit der folgenden formal-mathematischen Beschreibung eines Klassifikator-Induktors werden Symbole eingeführt, die im weiteren Verlauf konsistent genutzt werden. Die Formalisierung orientiert sich weitgehend an üblichen Beschreibungen des Klassifikations-Problems (z. B. Devroye, 1988). Wir bezeichnen mit X die Datenmatrix der Prädiktor-Werte je Instanz. D. h. der Eintrag x_{ij} ist der Wert von Instanz (oder Person) $i \in \{1..N\}$ für Prädiktor (oder in unserem Falle für EEB-Item) $j \in \{1..M\}$. Außerdem bezeichnen wir mit \mathbf{y} die Liste der vorherzusagenden Kriterien-Werte. $\mathbf{y} \in \mathcal{C}^N$ ist ein N -Tupel mit Einträgen aus der Menge der Klassen \mathcal{C} (in unserem Falle ist $\mathcal{C} = \{0, 1\}$). D. h. y_i ist die tatsächliche Klasse, der Instanz i angehört (z. B. Gesundheitskompetenz in der Klasse 0 = niedrig oder 1 = hoch). Weiter nehmen wir eine Einteilung der Instanzen in Trainingsdaten $T \subseteq \{1..N\}$ und Testdaten $t = \{1..N\} \setminus T$ vor. Ziel der Klassifikator-Induktion \mathcal{I} ist das Auffinden einer zweistelligen Abbildung f (Klassifikator) mit Tupeln aus Elementen der Klassen \mathcal{C} als Bildmenge, sowie einer Prädiktor-Teilmenge $S \subseteq \{1..M\}$ (Merkmals-Selektion), sodass eine dreistellige Abbildung (Zielfunktion) κ maximiert (oder minimiert) wird. $\mathcal{I}(X, \mathbf{y}, T) = (f, S)$ ist das Ergebnis der Klassifikator-Induktion unter Verwendung der Trainingsdaten. $f(X, S)$ ist das Ergebnis des resultierenden Klassifikators f (die Vorhersage) in Anwendung auf die Datenmatrix X und unter Verwendung der selektierten Merkmale S . Sei nun $U \subseteq \{1..N\}$ eine beliebige Teilmenge der Instanzen. $\kappa(f(X, S), \mathbf{y}, U)$ ist das Ergebnis der Zielfunktion, das aus einem Vergleich der Vorhersage mit den tatsächlichen Kriterien-Werten für die Instanzen U resultiert. Im Falle von $U = T$ wird die Zielfunktion κ für die Trainingsdaten ausgewertet. Im Falle von $U = t$ wird sie für die Testdaten ausgewertet. Die Unterscheidung dieser Fälle ist bedeutsam. Während der Klassifikator-Induktion wird häufig an Hand der Trainingsdaten geprüft, ob die Testfunktion nach einer Iteration ein besseres Ergebnis erzielt, als zuvor. Problematisch hieran ist eine gewisse Zirkularität: Die Prüfung erfolgt an Hand derselben Daten, wie die Induktion. Eine Folge kann eine Überanpassung des Klassifikators f an die Trainingsdaten und eine mangelnde Vorhersagekraft für ungesehene Daten sein (siehe oben). Nach der Klassifikator-Induktion kann mit Hilfe der Testdaten eine tatsächliche Validierung des Klassifikators f an Hand der Auswertung der Zielfunktion κ vorgenommen werden, denn die Testdaten sind nicht in die Induktion des Klassifikators eingeflossen.

5.6.3 Allgemeiner Aufbau eines Klassifikator-Induktors

Ein Klassifikator-Induktor \mathcal{I} weist eine innere Struktur auf, bei der verschiedene Teilprozesse in einer Schleife durchlaufen werden. Diese Teilprozesse sind (a) ein Generator g für die Generierung einer Prädiktor-Teilmenge S , (b) ein Lerner (eine vierstellige Abbildung) ℓ für die Spezifizierung von Parametern des Klassifikators f in Abhängigkeit der Trainingsdaten T und der Prädiktor-Teilmenge S : $f = \ell(X, \mathbf{y}, T, S)$, (c) die Auswertung der Zielfunktion κ auf den Trainingsdaten T , (d) eine Abbruchbedingung θ für das Verlassen der Induktions-Schleife sowie ggf. (e) die Auswertung der Zielfunktion κ auf den Testdaten t . Bei Start des Induktors wird eine erste Prädiktor-Teilmenge generiert. Diese Prädiktor-Teilmenge wird zusammen mit den Trainingsdaten an den Lerner übergeben. Dieser erzeugt einen Klassifikator mit spezifischen Parametern. Durch Auswertung der Zielfunktion auf den Trainingsdaten kann die Güte des erzeugten Klassifikators bestimmt werden. Schließlich wird die Abbruchbedingung geprüft. Sie kann den Wert F (für falsch) oder W (für wahr) annehmen. Nimmt sie den Wert F an, so beginnt die Induktions-Schleife bei der Generierung einer neuen Prädiktor-Teilmenge erneut. Nimmt sie den Wert W an, so wird die Induktions-Schleife verlassen. Resultat sind der zuletzt spezifizierte Klassifikator sowie die zuletzt generierte Prädiktor-Teilmenge. Ggf. kann mittels Auswertung der Zielfunktion auf den Testdaten eine Validierung dieses Klassifikators auf ungesehenen Daten vorgenommen werden (vgl. Blum & Langley, 1997; Dash & Liu, 1997; Duda, 1970; Jain et al., 2000; Kanal, 1974; Kotsiantis, Zaharakis & Pintelas, 2006; Krizek, 2008; Langlay, 1994; Liu & Yu, 2005; Maimon & Rokach, 2005). Mit Abbildung 7 wird der generelle Aufbau eines Klassifikator-Induktors veranschaulicht.

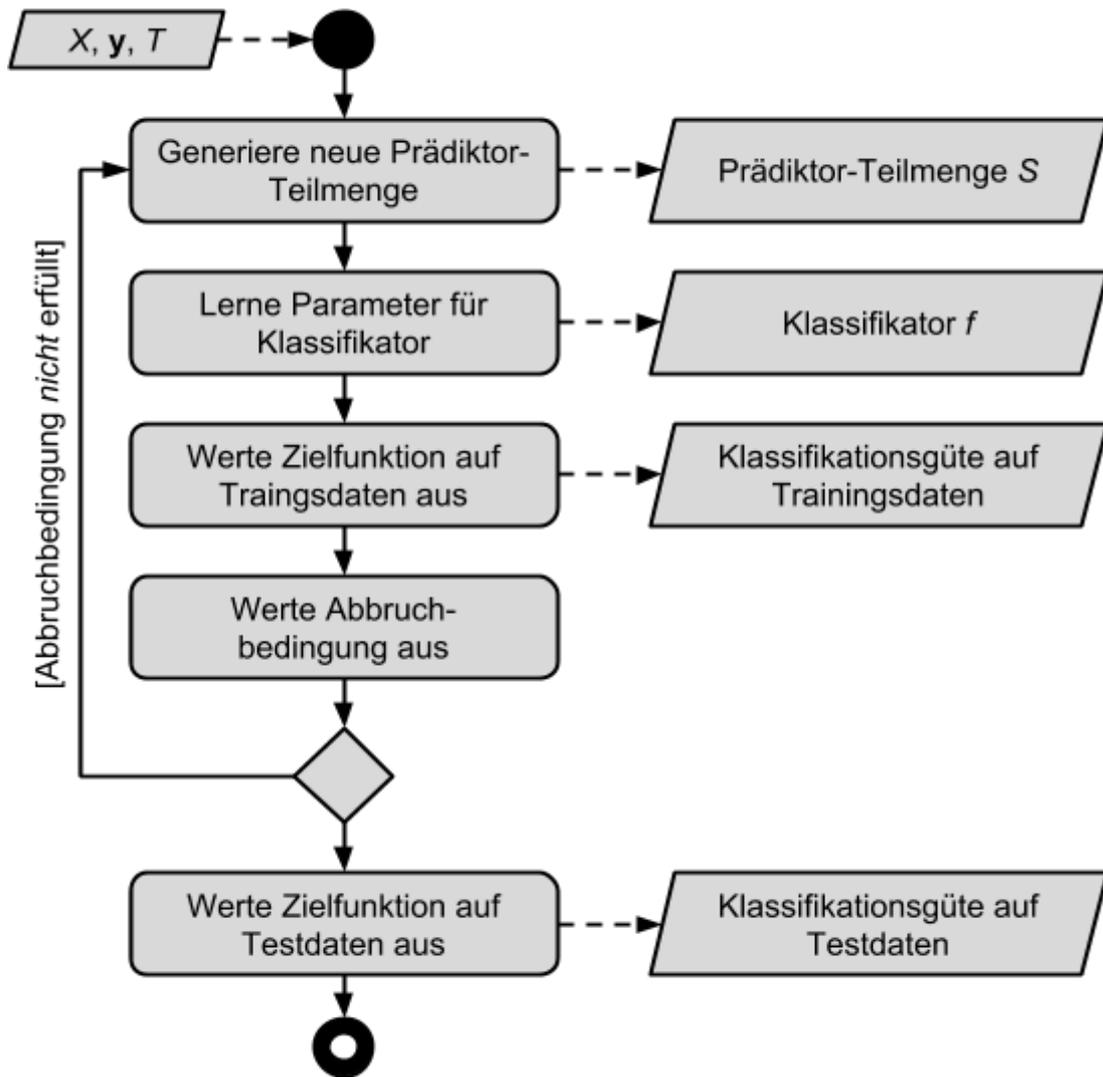


Abbildung 7. Grundlegender Aufbau eines Klassifikator-Induktors (Flussdiagramm) (eigene Darstellung).

Die verschiedenen Teilprozesse der Klassifikator-Induktion können auf verschiedenste Weisen spezifiziert werden. So gibt es verschiedene Selektions- bzw. Such-Algorithmen zur Generierung von Prädiktor-Teilmengen (Merkmals-Selektion), verschiedene Arten von Klassifikatoren und entsprechenden Lernern, verschiedene mögliche Zielfunktionen sowie unterschiedliche Möglichkeiten, die Abbruchbedingung zu definieren. Aus der Literatur kann man sich jeweils an einem reichhaltigen Angebot bedienen und verschiedene Algorithmen nach Art der Problemstellung kombinieren (Clarke et al., 2009; Hastie et al., 2009; Jain et al., 2000; Kotsiantis, Zaharakis & Pintelas, 2006; Krizek, 2008; Maimon & Rokach, 2005; Somol, Pudil, Novovikova & Paclik, 1999; Yusta, 2009). Darüber hinaus kann man Algorithmen an die eigene Problemstellung anpassen. In der Literatur werden Empfehlungen angegeben, bei welcher Art von Problemstellung welche Algorithmen am erfolgversprechendsten sind. Regeln zur Anwendung bestimmter Algorithmen, die beispielsweise auf Grund mathematischer Erkenntnisse abgeleitet worden wären, existieren jedoch nicht. Gäbe es solche Regeln, so könnten sie ebenfalls als Algorithmus implementiert werden und eine sachlich begründete

Entscheidung durch die Designer von Klassifikator-Induktoren wäre nicht mehr nötig. Die Designer sind hier auf ihr Verständnis der unterschiedlichen Algorithmen sowie eine genaue Analyse der eigenen Problemstellung angewiesen und müssen ggf. verschiedene Möglichkeiten der Spezifikation ausprobieren. Insofern ist es angebracht und durchaus üblich, von Klassifikationsexperimenten zu sprechen, wenn verschiedene Verfahren (unabhängige Variable) im Hinblick auf ihre Klassifikationsgüte (abhängige Variable) verglichen werden. In weiteren Abschnitten werden die für diese Arbeit gewählten Algorithmen beschrieben und die Auswahl begründet.

5.7 ALLGEMEINE SPEZIFIKATIONEN VERWENDETER KLASSIFIKATOR-INDUKTOREN

In sämtlichen Reihen von Klassifikationsexperimenten wird in der vorliegenden Arbeit zur Einschätzung der Klassifikationsgüte die Korrektklassifikation (Vorhersagegenauigkeit) als Zielfunktion verwendet. Daneben werden für jede Spezifikation des Klassifikator-Induktors Kontingenztafeln ausgewertet. Sämtliche Spezifikationen der Klassifikator-Induktoren werden 10-fach kreuzvalidiert, d. h. zehn Mal mit verschiedenen Trainings- und Testdaten im Hinblick auf ihre Klassifikationsgüte geprüft. Da bei Zugrundelegung verschiedener Trainingsdaten im Rahmen der Kreuzvalidierung mit verschieden ausgewählten Prädiktor-Teilmengen zu rechnen ist, wird außerdem die Stabilität bzw. Übereinstimmung der ausgewählten Prädiktor-Teilmengen untersucht.

5.7.1 Beurteilung der Klassifikationsgüte

In der vorliegenden Arbeit wird als Zielfunktion die Korrektklassifikation κ gewählt (Fu, 1994; Hughes, 1968; Kohavi, 1995; Kohavi & Wolpert, 1996; Kotsiantis, Zaharakis & Pintelas, 2006; Maimon & Rokach, 2005). Mit Korrektklassifikation (Vorhersagegenauigkeit) ist der Anteil der Instanzen U gemeint, der durch den Klassifikator f korrekt den Klassen 0 und 1 zugewiesen wird. Dieser Anteil ist definiert durch die Menge $K(f(X, S), \mathbf{y}, U) := \{i \in U \mid f(X, S)_i = y_i\} \subseteq U$. Bezeichnet man mit $\text{card } A$ die Anzahl an Elementen in der Menge A (Kardinalität der Menge A), dann ist die Korrektklassifikation gegeben durch

$$\kappa(f(X, S), \mathbf{y}, U) := \frac{\text{card } K(f(X, S), \mathbf{y}, U)}{\text{card } U}.$$

Sie kann Werte zwischen 0 (keine Übereinstimmung zwischen Vorhersage und tatsächlichen Kriterien-Werten) und 1 (vollständige Übereinstimmung) annehmen. Ziel ist eine möglichst hohe Korrektklassifikation. $\kappa_{\text{Training}} := \kappa(f(X, S), \mathbf{y}, T)$ wird im Folgenden als Trainings-Korrektklassifikation bezeichnet, $\kappa_{\text{Test}} := \kappa(f(X, S), \mathbf{y}, t)$ als Test-Korrektklassifikation.

Trainings- und Test-Korrektklassifikation sollten sich möglichst nicht statistisch bedeutsam voneinander unterscheiden. Ist die Trainings-Korrektklassifikation bedeutend höher als die Test-Korrektklassifikation, so muss von einer Überanpassung des Klassifikators an die Trainings-Daten ausgegangen werden. Die Test-Korrektklassifikation ließe sich in diesem Falle vermutlich durch einen weniger an die Trainings-Daten angepassten Klassifikator steigern. Ist die Trainings-Korrektklassifikation bedeutend niedriger als die Test-Korrektklassifikation, dann muss von einer mangelnden Fähigkeit der Generalisierung des Klassifikators ausgegangen werden. Eine hohe Test-Korrektklassifikation kann in diesem Falle ein reines Zufallsergebnis sein. Sowohl Trainings- als auch Test-Korrektklassifikation sollten oberhalb der Korrektklassifikation $\kappa_{\text{konst.}}$ bei konstanter Zuordnung liegen. Konstante Zuordnung bedeutet, dass alle Instanzen derselben Klasse, z. B. der Klasse 0, zugeordnet werden. Auf Grund der Dichotomisierung der Daten (s. o.) muss diese bis auf minimale Rundungsfehler .500 betragen. Trainings- und Test-Korrektklassifikation können auf der anderen Seite höchstens die theoretisch maximal mögliche Korrektklassifikation κ_{Hughes} nach Hughes (1968) annehmen, sollten dieser aber möglichst nahe kommen. Hughes (1968) hat einen Algorithmus für die Berechnung dieser Korrektklassifikation in Abhängigkeit von Messkomplexität und Anzahl Messungen (Instanzen) angegeben. Dieser wurde für das vorliegende Projekt implementiert; entsprechende Angaben finden sich bei einigen Beurteilungen der Klassifikationsgüte von Klassifikator-Induktoren.

Sind zwei verschiedene Korrektklassifikationen κ_a und κ_b auf Grundlage von Fallzahlen N_a und N_b gegeben, so lassen sie sich mittels Wald-Test auf statistisch bedeutsame Unterschiede hin untersuchen. Zunächst berechnet man hierzu den Erwartungswert, der sich als Differenz $\langle \kappa_a - \kappa_b \rangle = \kappa_a - \kappa_b$ darstellt. Weiterhin benötigt man die Standardabweichung. Diese lässt sich mittels der Formel

$$\sigma(\kappa_a - \kappa_b) = \sqrt{\frac{\kappa_a(1 - \kappa_a)}{N_a} + \frac{\kappa_b(1 - \kappa_b)}{N_b}}$$

berechnen. Der Quotient $\tau := \langle \kappa_a - \kappa_b \rangle / \sigma(\kappa_a - \kappa_b)$ weist eine t-Verteilung auf, sodass sich die statistische Bedeutsamkeit dieses Unterschiedes in einem entsprechenden p -Wert ausdrücken lässt. Ganz entsprechend weist bei Betrachtung des Differenz-Betrages $|\kappa_a - \kappa_b|$ die Größe $X^2 := (|\kappa_a - \kappa_b| / \sigma(|\kappa_a - \kappa_b|))^2$ eine χ^2 -Verteilung auf (Agresti, 2002, 2007). SageMathCloud™ bietet über die Integration des Statistiksystems R Funktionen, um entsprechende p -Werte numerisch zu bestimmen. Hiermit gibt es zwei Möglichkeiten, Trainings- und Test-Korrektklassifikation sowie Korrektklassifikation bei konstanter

Zuordnung und theoretisch maximal mögliche Korrektklassifikation miteinander zu vergleichen. Entsprechende Angaben finden sich bei der Beurteilung der Klassifikationsgüte der unterschiedlichen Klassifikator-Induktoren.

Neben der Betrachtung der Korrektklassifikation lässt sich für die Beurteilung der Klassifikationsgüte eines Klassifikators auch die Kontingenztafel auswerten. Die Kontingenztafel ist eine 2 x 2-Matrix, in der die Anzahlen von korrekt bzw. fälschlicher Weise den Klasse 0 bzw. 1 zugewiesenen Instanzen eingetragen werden. Zusätzlich werden Randsummen angegeben. In Abbildung 8 ist das Schema veranschaulicht.

		Vorhersage		Summe
		1	0	
Messwerte	1	a	b	$a + b$
	0	c	d	$c + d$
Summe		$a + c$	$b + d$	$a + b + c + d$

Abbildung 8. Schematische Darstellung einer 2 x 2-Kontingenztafel.

Die Korrektklassifikation lässt sich mit Hilfe der in Abbildung 8 angegebenen Variablen mittels der Formel $\kappa = (a + d)/(a + b + c + d)$ berechnen. Gebräuchlicher bei der Auswertung einer Kontingenztafel ist jedoch die Schätzung eines Zusammenhangsmaßes sowie von χ^2 . In der vorliegenden Arbeit wird für jeden Klassifikator bzw. jeden Klassifikator-Induktor Cramers φ und eine Schätzung von χ^2 mit entsprechendem p -Wert angegeben. Cramers φ ist gegeben durch die Formel

$$\varphi = \frac{a d - b c}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$$

Die Schätzung von χ^2 lässt sich mit Hilfe der Formel

$$X^2 = (a + b + c + d)\varphi^2$$

bestimmen. Diese Größe weist eine χ^2 -Verteilung auf, sodass sich ein entsprechender p -Wert ermitteln lässt (Agresti, 2002, 2007).

5.7.2 Kreuzvalidierung

Weiter oben wurde beschrieben, dass bei der Klassifikator-Induktion häufig eine Unterteilung der Instanzen in Trainings- und Testdaten vorgenommen wird, um die Klassifikationsgüte eines Klassifikators auch bezüglich ungesehener Daten einschätzen zu können. Bei einer gegebenen Unterteilung mit z. B. 50 % Trainings- und 50 % Testdaten könnte man zwei Durchläufe der Klassifikator-Induktion vornehmen. Zunächst induziert man den Klassifikator mit den Trainingsdaten und validiert ihn mit den Testdaten. Anschließend vertauscht man die Rollen von Trainings- und Testdaten, führt also nun die Induktion mit den Testdaten und die Validierung mit den Trainingsdaten durch. Die beiden erhaltenen Ergebnisse kann man schließlich miteinander vergleichen. Dies wäre ein einfacher Fall von Kreuzvalidierung (Efron, 1983; Hastie et al., 2009; Kohavi, 1995; Kohavi & John, 1997; Mosteller & Tuckey, 1968; Stone, 1974). Kreuzvalidierung bedeutet, dass jede verfügbare Instanz in der einen oder anderen Form mindestens einmal zu den Trainingsdaten und mindestens einmal zu den Testdaten gehört. Leave-one-out Kreuzvalidierung ist eine besondere Variante, bei der alle Instanzen bis auf eine Validierungsinstanz den Trainingsdaten zugewiesen werden. Die Validierung wird ausschließlich mit der Validierungsinstanz durchgeführt. Führt man dies in einer Schleife über alle Instanzen durch, so wird jede Instanz genau einmal als Validierungsinstanz eingesetzt (Mosteller & Tuckey, 1968).

Für den hier vorliegenden Fall vergleichsweise weniger möglicher Prädiktoren (25 EEB-Items) und moderater Umfänge von Instanzen (Stichproben mit Umfängen der Größenordnung 1 000) wird in der Literatur die 10-fach Kreuzvalidierung empfohlen. Hierbei werden zehn Validierungsdurchläufe bzw. Klassifikator-Induktionen mit jeweils 90 % der Instanzen durchgeführt. Die verbleibenden 10 % der Instanzen werden jeweils für die Validierung des resultierenden Klassifikators verwendet (Mosteller & Tuckey, 1968). Auch hier wird darauf geachtet, dass jede Instanz genau einmal für Validierungszwecke eingesetzt wird. Mit dieser Festlegung ist auch die Heuristik erfüllt, nach der bei vier ausgewählten Merkmalen und zwei Klassen mindestens 80 Instanzen in einem Testdatensatz enthalten sein sollten (Jain et al., 2000; Nagy, 1968).

Für die vorliegende Arbeit wird die 10-fach Kreuzvalidierung wie folgt realisiert: Wenn die Instanzen geladen werden, wird ihnen eine laufende Nummer zugewiesen. Zugleich liegen sie jedoch nach Stichproben geordnet in den CSV-Dateien vor. Daher wird die Modulo-Division der laufenden Nummer für die Zuweisung einer Instanz zu einer der zehn Validierungsstichproben genutzt. Hat eine Instanz die laufende Nummer i , so wird sie der Validierungsstichprobe $i \bmod 10$ zugeordnet. Hierbei ist $i \bmod 10$ der Rest nach ganzzahliger

Division von i durch 10, also z. B. $137 \bmod 10 = 7$. Die zehn Validierungsstichproben $V_a, a \in \{0..9\}$ sind daher definiert durch $V_a := \{i \in \{1..N\} | i \bmod 10 = a\}$. Auf dieser Grundlage werden für jeden der zehn Validierungsdurchgänge die Trainingsdaten mit $T_a := \{1..N\} \setminus V_a$ und die Testdaten mit $t_a := V_a$ festgelegt. Auf diese Weise wird sichergestellt, dass jede Instanz genau einer Validierungsstichprobe angehört und dass zugleich sämtliche Stichproben gleichmäßig und disjunkt auf die zehn Validierungsstichproben verteilt werden.

Da der Klassifikator-Induktor bei jedem Validierungsdurchgang andere Daten für die Induktion nutzt, ist mit unterschiedlichen Klassifikatoren und Prädiktor-Teilmengen je Validierungsdurchgang zu rechnen (Kalousis, Prados & Hilario, 2005; Krizek, 2008; Kuncheva, 2007; Somol & Novovicova, 2010). Dies wird bei Kreuzvalidierung durch das Subskript a bei Klassifikator f_a und Prädiktor-Teilmenge S_a gekennzeichnet. Auch die Korrektklassifikationen werden sich je Validierungsdurchgang unterscheiden: $\kappa_{\text{Training}, a} := \kappa(f_a(X, S_a), \mathbf{y}, T_a)$ sind nun zehn verschiedene Trainings-Korrektklassifikationen und $\kappa_{\text{Test}, a} := \kappa(f_a(X, S_a), \mathbf{y}, t_a)$ zehn verschiedene Test-Korrektklassifikationen ($a \in \{0..9\}$). Bei Kreuzvalidierung nennt man als Maß für die Klassifikationsgüte eines Klassifikator-Induktors die Cross-Validation-Korrektklassifikation $\kappa_{\text{CV}} := \langle \kappa_{\text{Test}, a} \rangle_{a \in \{0..9\}}$, also den Mittelwert der verschiedenen Test-Korrektklassifikationen. Gleiches gilt für die Trainings-Korrektklassifikation κ_{Training} . Die Kontingenztafel Q_{CV} bei Kreuzvalidierung ist die Matrizen-Summe der Kontingenztafeln Q_a aller Validierungsdurchgänge: $Q_{\text{CV}} := \sum_{a \in \{0..9\}} Q_a$. Kreuzvalidierung führt also zu einer über die Validierungsdurchgänge hinweg gemittelten Einschätzung der Klassifikationsgüte (Efron, 1983; Hastie et al., 2009; Kohavi, 1995; Kohavi & John, 1997; Stone, 1974). Durch Kreuzvalidierung wird die Klassifikator-Induktion von einer Schleife über die Validierungsdurchgänge umschlossen. In Abbildung 9 wird dies noch einmal grafisch veranschaulicht.

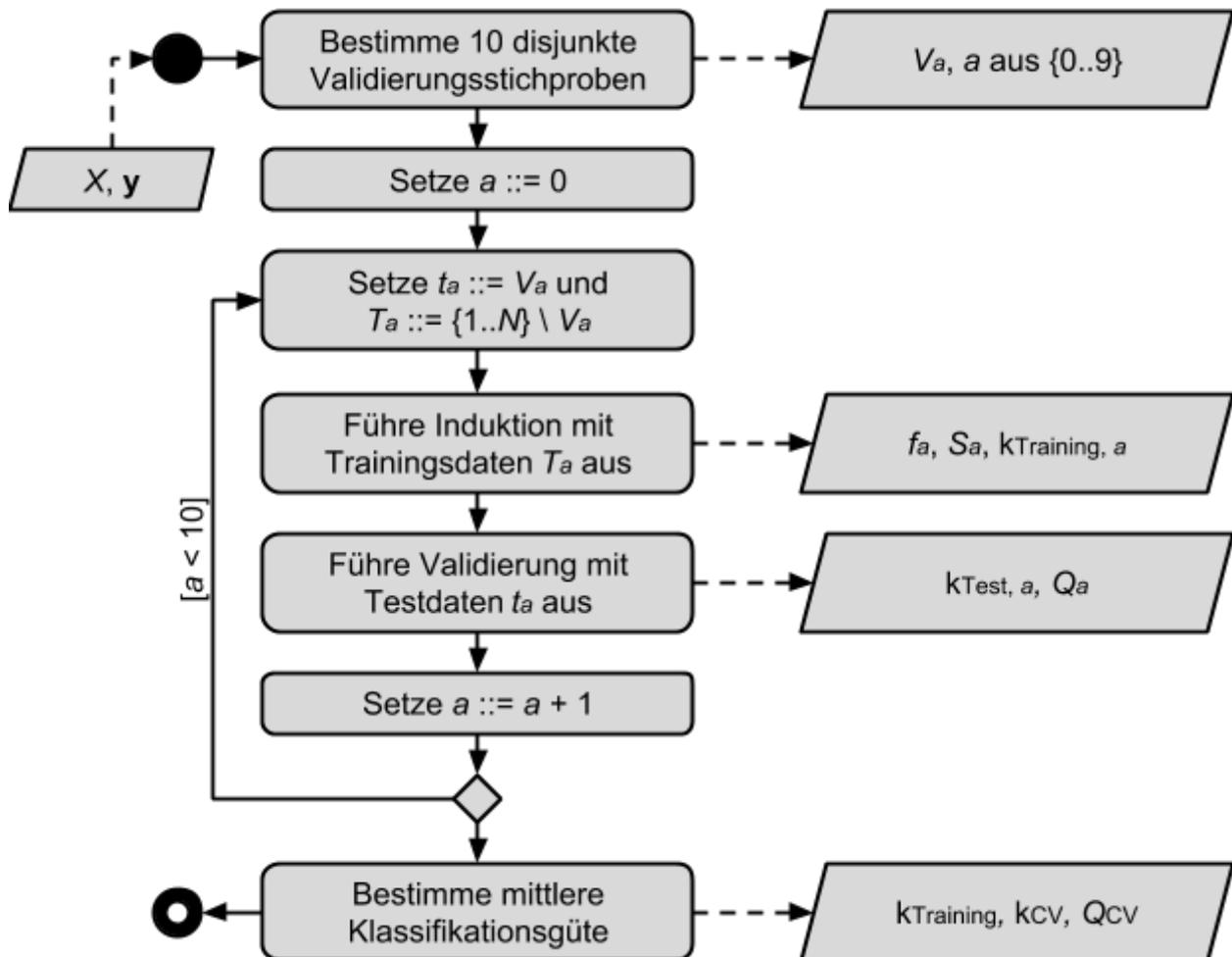


Abbildung 9. Grundlegender Aufbau eines Klassifikator-Induktors bei 10-fach Kreuzvalidierung (Flussdiagramm). Die verwendeten Formelzeichen entsprechen jenen im Text (eigene Darstellung).

5.7.3 Stabilität

Im vorigen Abschnitt wurde bereits beschrieben, dass es bei Zugrundelegung unterschiedlicher Stichproben bei Kreuzvalidierung zu unterschiedlich ausgewählten Prädiktor-Teilmengen kommen kann. Ein Maß, das die Einschätzung der Übereinstimmung der Prädiktor-Teilmengen über die zehn Validierungsdurchgänge hinweg erlaubt, ist die Stabilität. Als Maße für die Stabilität liegen zahlreiche Vorschläge vor (Kalousis, Prados & Hilario, 2005; Krizek, 2008; Kuncheva, 2007; Somol & Novovicova, 2010). In der vorliegenden Arbeit wird das Stabilitäts-Maß nach Kalousis, Prados und Hilario (2007) verwendet. Hierbei werden die Prädiktor-Teilmengen aus den zehn Validierungsdurchgängen paarweise mit einander verglichen, indem paarweise Übereinstimmungen bestimmt werden. Seien zwei Prädiktor-Teilmengen S_a und S_b gegeben, so schlagen Kalousis et al. (2007) als Maß der Übereinstimmung

$$s(a, b) := \frac{\text{card } S_a \cap S_b}{\text{card } S_a \cup S_b}$$

vor. Hierbei handelt es sich um das Verhältnis der Anzahl gemeinsamer Prädiktoren zur Anzahl aller in beiden Teilmengen enthaltenen Prädiktoren. Dieses Maß beträgt 1, wenn die Prädiktoren der beiden Prädiktor-Teilmengen komplett übereinstimmen. Es beträgt 0, wenn die beiden Prädiktor-Teilmengen kein gemeinsames Element enthalten. Als aggregiertes Stabilitätsmaß über alle Validierungsdurchgänge schlagen die Autoren den Mittelwert aller paarweisen Stabilitätsmaße, $s := \langle s(a, b) \rangle_{(a,b) \in \{0..9\}^2}$, vor. Dieses Stabilitätsmaß wird bei sämtlichen Experimentalreihen zur Klassifikation ermittelt und genannt.

5.8 ERSTE EXPERIMENTALREIHE: VOLLSTÄNDIGE SUCHE UND BAYES-KLASSIFIKATION

Bislang wurden grundsätzliche Spezifikationen und Validierungs-Kennwerte beschrieben, die für die vorliegende Arbeit bei der Klassifikator-Induktion eingesetzt werden. Offen sind noch das Vorgehen bei der Suche nach einer geeigneten Prädiktor-Teilmenge (Merkmals-Selektion) sowie die Art des zu induzierenden Klassifikators. In der ersten Experimentalreihe wird für die Merkmals-Selektion die vollständige Suche eingesetzt und es werden Bayes-Klassifikatoren induziert.

5.8.1 Vollständige Suche

Für die Suche nach geeigneten Prädiktor-Teilmengen (Merkmals-Selektion) stellt sich die Frage nach dem Vorgehen. Vollständige Suche entspricht vermutlich am nächsten einem intuitiven Vorgehen. Stehen M Prädiktoren zur Auswahl (in unserem Falle 25 EEB-Items), dann wären bei vollständiger Suche für jedes $l \in \{1..M\}$ jede mögliche Kombination von l Prädiktoren als Prädiktor-Teilmenge S zu verwenden und zu validieren. Schließlich würde jene Prädiktor-Teilmenge ausgewählt, die das beste Ergebnis der Validierung erzielt. Die Anzahl möglicher Prädiktor-Teilmengen der Länge l beträgt $M!/l!(M-l)!$. Bei Ziehung von einem Item aus den 25 EEB-Items erhält man 25 Möglichkeiten, bei zwei Items bereits 300 Möglichkeiten und bei drei Items 2 300 Möglichkeiten. Die Anzahl sämtlicher möglicher Prädiktor-Teilmengen erhält man durch die Summe über alle Längen l : $\sum_{l \in \{1..M\}} M!/l!(M-l)!$. Bei 25 EEB-Items sind dies 33 554 431 Möglichkeiten für Prädiktor-Teilmengen, die validiert werden können. Dies ist noch mit 18 zu multiplizieren, da 18 Kriterien separat vorhergesagt werden sollen. Da die Validierung zunächst einer Spezifizierung der Parameter für den Klassifikator bedarf (Lerner, s. o.), eine 10-fache Kreuzvalidierung durchgeführt werden soll und umfassende Analysen der Klassifikationsgüte folgen, ist bei vollständiger Suche mit einem erheblichen Zeitaufwand zu rechnen (Clarke et al., 2009). Andererseits ist durch den Beitrag von Hughes (1968) gesichert, dass die Messkomplexität nicht beliebig gesteigert werden muss. Wird die optimale Messkomplexität überschritten, kann keine

akzeptable Klassifikationsgüte bezüglich ungesehener Daten erreicht werden (a. a. O.). Daher wird in der ersten Experimentalreihe zwar vollständige Suche durchgeführt, diese beschränkt sich jedoch auf Prädiktor-Teilmengen der Längen 1 (bestes Einzelitem), 2 (bestes Paar) und 3 (bestes Tripel). In dieser Experimentalreihe werden die Rohdaten der EEB-Items verwendet. Wie weiter oben bereits ausgeführt wurde, ist die Kombination von mehr als drei nicht-dichotomisierten EEB-Items für die Vorhersage eines Kriteriums auf Grund der Erkenntnisse von Hughes (1968) nicht ratsam. Somit müssen nicht über 33 Millionen mögliche Prädiktor-Teilmengen, sondern nur 2 625 für jedes der 18 Kriterien validiert werden.

Für die Durchführung der vollständigen Suche wurde ein Algorithmus implementiert, mit dem die 2 625 möglichen Prädiktor-Teilmengen der Längen 1 bis 3 der Reihe nach validiert werden. Hiermit ist auch zugleich die Abbruchbedingung für den Klassifikator-Induktor definiert: Der Abbruch soll erfolgen, wenn alle 2 625 Kandidaten validiert wurden. Das beste Ergebnis der Validierung wird inklusive der zugehörigen Prädiktor-Teilmenge S und des zugehörigen Klassifikators f , separat für die Längen 1 bis 3, zwischengespeichert. Nach Durchlaufen des Algorithmus wird der Inhalt dieses Zwischenspeichers ausgegeben. Der Algorithmus wird für die 18 vorherzusagenden Kriterien separat jeweils mit 10-facher Kreuzvalidierung angewendet. Sämtliche Ergebnisse werden für die spätere Verschriftlichung in Text-Dateien auf dem Server gesichert.

5.8.2 Bayes-Klassifikation

Bayes-Klassifikation beruht auf bedingten Wahrscheinlichkeiten und setzt ein Klassensystem \mathcal{C} sowie ein diskretes System von Messwerten \mathcal{M} voraus. Der Lerner ℓ für einen Bayes-Klassifikator ermittelt, ausgehend von einer gegebenen Stichprobe $\{1..N\}$, die benötigten bedingten Wahrscheinlichkeiten. Voraussetzung ist, dass jeder Instanz $i \in \{1..N\}$ genau eine Klasse $c_i \in \mathcal{C}$ und ein Messwert $m_i \in \mathcal{M}$ zugeordnet ist. Im Folgenden bezeichnen wir mit $[c]_{\mathcal{C}} := \{i \in \{1..N\} | c_i = c\}$ die Menge aller Instanzen, die der Klasse $c \in \mathcal{C}$ zugeordnet sind und definieren analog die Mengen $[m]_{\mathcal{M}}$ bezüglich der Messwerte. Der Lerner ermittelt die bedingten Wahrscheinlichkeiten

$$p(m|c) = \frac{\text{card } [m]_{\mathcal{M}} \cap [c]_{\mathcal{C}}}{\text{card } [c]_{\mathcal{C}}}.$$

Sie geben den Anteil aller Instanzen der Klasse c an, der den Messwert m aufweist.

Die Anwendung eines Bayes-Klassifikators hat zum Ziel, eine Instanz auf Grund ihres Messwertes m einer Klasse c zuzuordnen. Voraussetzung ist die Kenntnis der bedingten

Wahrscheinlichkeiten $p(c|m)$, mit denen eine Instanz mit Messwert m den Klassen $c \in \mathcal{C}$ angehört. Diese sind gegeben durch

$$p(c|m) = \frac{\text{card } [m]_{\mathcal{M}} \cap [c]_{\mathcal{C}}}{\text{card } [m]_{\mathcal{M}}}.$$

Der Vergleich mit $p(m|c)$ liefert den verallgemeinerten Satz von Bayes (1764; Bronstein et al., 2006):

$$\frac{p(m|c) \text{card } [c]_{\mathcal{C}}}{N} = p(m|c) p(c) = p(c|m) p(m) = \frac{p(c|m) \text{card } [m]_{\mathcal{M}}}{N},$$

und somit

$$p(c|m) = \frac{p(m|c) p(c)}{p(m)}.$$

Der Bayes-Klassifikator B ordnet eine Instanz mit Messwert m jener Klasse c mit der höchsten bedingten Wahrscheinlichkeit $p(c|m)$ zu:

$$B(m) = \arg \max\{p(c|m) | c \in \mathcal{C}\} = \arg \max\{p(m|c) p(c) | c \in \mathcal{C}\}$$

(Duda, 1970; Fu, 1994; Hughes, 1968; Kohavi & John, 1997; Wald, 1949). Diese Vorschrift kann unter der Voraussetzung, dass die Klassen gleichverteilt sind, vereinfacht werden zu

$$B(m) = \arg \max\{p(m|c) | c \in \mathcal{C}\}.$$

Wir kommen nun zur Übertragung dieser Prinzipien auf das vorliegende Problem. Das Klassensystem ist gegeben als $\mathcal{C} = \{0, 1\}$ (0 = Kriterium niedrig ausgeprägt, 1 = Kriterium hoch ausgeprägt). Die Messwerte werden aus der Kombination von EEB-Items gebildet. Ein EEB-Item weist sieben Antwortkategorien $\{1..7\}$ auf, die als diskretes System von Messwerten betrachtet werden. Ist eine Stichprobe (Trainingsdaten) gegeben, so lassen sich die bedingten Wahrscheinlichkeiten $p(m|c)$, $(m, c) \in \{1..7\} \times \{0, 1\}$ bestimmen. Hat nun eine Person (Instanz aus den Testdaten) während einer Befragung bei diesem EEB-Item den Wert $m \in \{1..7\}$ angekreuzt, kann sie einer der Klassen 0 und 1 zugeordnet werden. Sie wird der Klasse 0 zugeordnet, wenn $p(m|0) p(0) > p(m|1) p(1)$, andernfalls der Klasse 1. Werden zwei EEB-Items kombiniert, so wird die 49-elementige Menge $\{1..7\}^2$ als diskretes System von Messwerten aufgefasst. D. h. der Messwert einer Instanz ist ein Paar $m = (u, v) \in \{1..7\}^2$, das die Antworten der entsprechenden Person bei diesen beiden EEB-Items repräsentiert. Die Vorgehensweise bei der Bayes-Klassifikation lässt sich auch auf diesen Fall uneingeschränkt anwenden. Allgemeiner kann von einer Prädiktor-Teilmenge S ausgegangen werden, die festlegt, welche EEB-Items kombiniert werden sollen. Das diskrete System von Messwerten ist

dann gegeben durch $\mathcal{M} = \{1..7\}^{\text{card } S}$. Der Messwert einer Instanz ist ein Tupel mit $\text{card } S$ Elementen, das die Antworten der entsprechenden Person bei diesen EEB-Items S repräsentiert. Die Vorgehensweise bei der Bayes-Klassifikation ist wiederum uneingeschränkt übertragbar. Sie ist darüber hinaus auf den Fall übertragbar, wenn nicht die Rohwerte der EEB-Items, sondern die dichotomisierten Werte verwendet werden. Dann ist das diskrete System von Messwerten gegeben durch $\mathcal{M} = \{0, 1\}^{\text{card } S}$. Letztere Wahl hat den Vorteil, dass auch bei der Kombination von vergleichsweise vielen EEB-Items noch Zellenbesetzungen erzielt werden, die zu verlässlichen bedingten Wahrscheinlichkeiten führen.

Die technische Umsetzung eines Bayes-Klassifikators ist vergleichsweise einfach. Sie erfordert im Wesentlichen die Programmierung von Schleifen, in denen Häufigkeiten von Tupeln gezählt werden. Mit diesen Häufigkeiten lassen sich anschließend die erforderlichen (bedingten) Wahrscheinlichkeiten berechnen, die für die folgende Anwendung des Klassifikators gespeichert werden. Die Wahrscheinlichkeiten $p(0)$ und $p(1)$ betragen bis auf minimale Rundungsfehler .500. Da der Rechenaufwand für die Berücksichtigung dieser Wahrscheinlichkeiten vernachlässigbar ist und mögliche Quellen für Schätzfehler gering gehalten werden sollen, werden sie dennoch bei der Klassifikation berücksichtigt.

Bayes-Klassifikatoren weisen eine Besonderheit auf, die sie für eine Überanpassung an Trainingsdaten besonders anfällig machen. Die Trainings-Korrektklassifikation nimmt mit steigender Messkomplexität nicht ab, vielmehr steigt sie in Abhängigkeit von der Messkomplexität monoton. Von Hughes (1968) wird nachgewiesen, dass dies für die Test-Korrektklassifikation keineswegs gilt. Vielmehr gibt es bei gegebenem Stichprobenumfang eine optimale Messkomplexität, bei der die Test-Korrektklassifikation maximal ist. Bei einer hohen Messkomplexität wird die Test-Korrektklassifikation von der Trainings-Korrektklassifikation stark überboten. Dieses Verhalten kann für Bayes-Klassifikatoren, aber auch für beliebige andere Klassifikatoren in CV-T-Diagrammen veranschaulicht werden. Hierin werden Cross-Validation-Korrektklassifikation und Test-Korrektklassifikation in Abhängigkeit von der Messkomplexität abgetragen. Im Ergebnisteil werden solche Diagramme abgebildet und für die Festlegung von Abbruchbedingungen bei heuristischer Suche (siehe unten) genutzt.

5.9 ZWEITE EXPERIMENTALREIHE: HEURISTISCHE SUCHE UND BINÄRLOGISTISCHE REGRESSION

Vollständige Suche führt immer zum besten Ergebnis und ist dann geeignet, wenn die Anzahl zur Verfügung stehender Prädiktoren und die Umfänge auszuwählender Prädiktor-Teilmenen vergleichsweise klein sind. Gehen wir von den Rohwerten der EEB-Items auf die

vollständig dichotomisierten Daten über, sind mehr als drei Prädiktoren auszuwählen. In einem solchen Falle greift man auf Heuristiken der Merkmals-Selektion zurück. In der zweiten Experimentalreihe werden zwei solcher Heuristiken angewendet. Darüber hinaus wird neben der Bayes-Klassifikation auch die binärlogistische Regression untersucht. Dies erfolgt zum einen, um einen Vergleich für die Bayes-Klassifikation zur Verfügung zu haben. Zum anderen soll damit ein Anschluss an die statistischen Standardverfahren der psychologischen Forschung hergestellt werden. Am Beispiel der binärlogistischen Regression wird gezeigt, wie solche Standardverfahren in den methodischen Rahmen der Klassifikator-Induktion eingebunden werden können.

5.9.1 Heuristische Suche

Bereits im Abschnitt zur vollständigen Suche weiter oben wurde beschrieben, dass der Aufwand für das Auffinden einer geeigneten (oder optimalen) Prädiktor-Teilmenge mit steigender Anzahl verfügbarer Prädiktoren schnell an die Grenzen des praktisch Machbaren stößt (Clarke et al., 2009). Daher weicht man bei der Merkmals-Selektion auf Heuristiken aus. Damit sind regelbasierte Strategien gemeint, die auf bestimmten Erfahrungen oder bestimmten Modellvorstellungen beruhen. Die Liste wissenschaftlich fundierter Heuristiken zur Merkmals-Selektion ist lang und wird beständig durch Verfeinerungen bereits etablierter Heuristiken oder neuartige Verfahren erweitert (Blum & Langley, 1997; Caruna & Freitag, 1994; Dash & Liu, 1997; Doumpos & Slappa, 2005; Guyon & Elisseeff, 2003; Jain & Zongker, 1997; Jain et al., 2000; Krizek, 2008; Kudo & Sklansky, 2000; Liu & Setiono, 1998; Liu & Yu, 2005; Molina et al., 2002; Pudil, Novovicova & Kittler, 1994; Reunanen, 2003; Siedlecki & Sklansky, 1989; Somol et al., 1999; Somol, Novovicova & Pudil, 2010; Yusta, 2009). Eine deutliche Trennlinie wird zwischen stochastischen und deterministischen Verfahren gezogen (Dash & Liu, 1997; Jain & Zongker, 1997; Krizek, 2008; Liu & Yu, 2005; Molina et al., 2002; Somol et al., 1999). Bei stochastischen Verfahren ist die Generierung von Prädiktor-Teilmengen auf die eine oder andere Art und Weise vom Zufall abhängig. Als Beispiele können die Las-Vegas-Methode oder genetische bzw. evolutive Algorithmen genannt werden (Liu & Setiono, 1998; Kudo & Sklansky, 2000; Siedlecki & Sklansky, 1989). Bei deterministischen Verfahren hängt die generierte Prädiktor-Teilmenge ausschließlich vom zu Grunde gelegten Datenmaterial (Trainingsdaten) ab. Beispielsweise fällt die vollständige Suche unter die deterministischen Verfahren. Solche Verfahren haben gegenüber den stochastischen Verfahren den Vorteil der Verifizierbarkeit. Wiederholt man ein Klassifikationsexperiment mit denselben Daten und denselben Spezifikationen des Klassifikator-Induktors (inklusive Merkmals-Selektion), dann wiederholt sich auch das Ergebnis.

In dieser zweiten Experimentalreihe werden die vollständig dichotomisierten Daten zur Klassifikator-Induktion verwendet. Anders, als bei der ersten Experimentalreihe ist auf Grund der Abhängigkeit der Korrektklassifikation von der Messkomplexität mit einem größeren Umfang ausgewählter Prädiktor-Teilmenge (ausgewählte EEB-Items) zu rechnen. Aus diesem Grunde werden hier statt vollständiger Suche heuristische Verfahren zur Merkmals-Selektion verwendet. Da die berichteten Forschungsbefunde möglichst nachvollziehbar sein sollen, fiel die Wahl auf zwei deterministische Verfahren. Das eine Verfahren nennt sich Sequential Forward Selection (SFS). Dies wird im deutschsprachigen Raum häufig unter dem Begriff der schrittweisen Vorwärtssuche eingesetzt. SFS findet Anschluss an die üblichen Methoden in der empirischen psychologischen Forschung (z. B. schrittweise binärlogistische Regression). Das andere Verfahren heißt Sequential Forward Floating Selection (SFFS). Es stellt eine Verfeinerung der SFS dar, bei der unter bestimmten Bedingungen bereits aufgenommene Merkmale wieder aus der Prädiktor-Teilmenge entfernt werden. SFFS kombiniert also die Vorwärtssuche mit der Rückwärtssuche (Pudil et al., 1994; Somol et al., 1999; Somol et al., 2010). Beide Verfahren werden für Klassifikationsprobleme empfohlen, die in ihrer Komplexität (Umfänge der Datensätze, Anzahl Klassen und möglicher Prädiktoren) dem hier untersuchten Problem entsprechen. Weitere auf diesen Verfahren aufbauende Heuristiken sind für komplexere Problemstellungen gut geeignet. Da sie jedoch mit einem erhöhten Rechenaufwand verbunden sind, bieten sie für das vorliegende Problem keinen Mehrwert.

Um SFS und SFFS zu implementieren, sind Definitionen von Unterprogrammen hilfreich. Gegeben sei eine Prädiktor-Teilmenge $S \subseteq \{1..M\}$. Mit

$$s^{ADD} := \arg \max\{\kappa(f(X, S \cup \{s\}), \mathbf{y}, T) | s \in \{1..M\} \setminus S\}$$

wird von den verbleibenden Prädiktoren $\{1..M\} \setminus S$ derjenige Prädiktor s^{ADD} gefunden, der die Trainings-Korrektklassifikation durch *Hinzufügen* zu S maximiert. Das Unterprogramm

$$ADD(S) := S \cup \{s^{ADD}\}$$

fügt der Prädiktor-Teilmenge S diesen Prädiktor s^{ADD} hinzu. SFS lässt sich damit bei noch näher zu definierender Abbruchbedingung θ (siehe weiter unten) wie folgt implementieren:

1. Setze $S ::= \emptyset$. Hierbei ist \emptyset die leere Menge, die kein Element enthält.
2. Lerne Klassifikator f .
3. Werte Abbruchbedingung $\theta(f, S)$ aus. Wenn sie erfüllt wird, gebe (f, S) aus, Ende.
4. Setze $S ::= ADD(S)$, gehe zu 2.

Für die Implementation von SFFS wird ein weiteres Unterprogramm benötigt. Mit

$$s^{RMV} := \arg \max\{\kappa(f(X, S \setminus \{s\}), \mathbf{y}, T) \mid s \in S\}$$

wird aus der gegebenen Prädiktor-Teilmenge S derjenige Prädiktor s^{RMV} gefunden, der die Trainings-Korrektklassifikation durch *Entfernen* aus S maximiert. Das Unterprogramm

$$RMV(S) := S \setminus \{s^{RMV}\}$$

entfernt diesen Prädiktor s^{RMV} aus der Prädiktor-Teilmenge S . Nun lässt sich auch die Implementation von SFFS angeben:

1. Setze $S ::= \emptyset$.
2. Setze $S ::= \text{ADD}(S)$.
3. Lerne Klassifikator f .
4. Werte Abbruchbedingung $\theta(f, S)$ aus. Wenn sie erfüllt wird, gebe (f, S) aus, Ende.
5. Setze $S^{RMV} ::= RMV(S)$ und lerne Klassifikator f^{RMV} .
6. Ist $\kappa(f^{RMV}(X, S^{RMV}), \mathbf{y}, T) \geq \kappa(f(X, S), \mathbf{y}, T)$ setze $S ::= S^{RMV}$, $f ::= f^{RMV}$, gehe zu 5.
7. Gehe zu 2.

Der Vorteil von SFFS gegenüber SFS besteht hauptsächlich in Zeile 6. Bei Verwendung von SFFS wird davon ausgegangen, dass sich die Trainings-Korrektklassifikation auch dann nicht verschlechtert, wenn geeignete Prädiktoren aus der Prädiktor-Teilmenge entfernt werden. Gegenüber SFS bietet dies die Möglichkeit, eine bessere Klassifikationsgüte bei gleichem oder geringerem Umfang ausgewählter Prädiktor-Teilmenen zu erzielen. Ob sich die Eignung dieser Heuristik für die vorliegende Problemstellung bestätigt, wird bei der Induktion von Bayes-Klassifikatoren geprüft.

5.9.2 Binärlogistische Regression

Gegeben sei wieder eine Prädiktor-Teilmenge $S \subseteq \{1..M\}$. Multiple binärlogistische Regression (BLR) ist geeignet, auf Grundlage der Daten (X, \mathbf{y}) eine Rechenvorschrift zu generieren, um in Abhängigkeit einer Instanz $\mathbf{x}_i := (x_{ij})_{j \in S}^T$ die Wahrscheinlichkeit $\pi(\mathbf{x}_i)$ vorherzusagen, mit der diese Instanz der Klasse 1 angehört. Der Ansatz der BLR besteht im binärlogistischen Regressionsmodell

$$\pi(\mathbf{x}_i) := \frac{\exp(\mathbf{x}_i^T \cdot \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \cdot \boldsymbol{\beta})}$$

Hierbei sei $x_0 = 1$ gesetzt, um einen konstanten Term β_0 berücksichtigen zu können. Dieses Modell ist durch Schätzung der Parameter $\boldsymbol{\beta}$ (reellwertiger Spaltenvektor mit $1 + \text{card } S$ Parametern) an die Daten (X, \mathbf{y}) anzupassen. Hierzu wird das binärlogistische Regressionsmodell zunächst durch Umformung in ein lineares Modell überführt. Es gilt

$$\exp(\mathbf{x}_i^T \cdot \boldsymbol{\beta}) = \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}$$

und somit die lineare Beziehung

$$\text{logit}(\pi(\mathbf{x}_i)) := \ln\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \mathbf{x}_i^T \cdot \boldsymbol{\beta}$$

(Agresti, 2002, 2007). Bei gegebenen Trainingsdaten $T \subseteq \{1..N\}$ erfolgt die Anpassung unter Verwendung der empirischen bedingten Wahrscheinlichkeiten

$$p(1|\mathbf{x}_i) = \frac{\text{card } [1]_{\{0,1\}} \cap [\mathbf{x}_i]_T}{\text{card } [\mathbf{x}_i]_T}$$

mit $[1]_{\{0,1\}} = \{j \in T | y_j = 1\}$ und $[\mathbf{x}_i]_T = \{j \in T | x_j = \mathbf{x}_i\}$. Es sind daher jene Parameter $\boldsymbol{\beta}$ zu schätzen, mit welchen

$$\text{logit}(p(1|\mathbf{x}_i)) = \mathbf{x}_i^T \cdot \boldsymbol{\beta} \text{ für alle } i \in T$$

bestmöglich erfüllt ist. Diese Anpassung erfolgt durch Maximum-Likelihood-Schätzung unter Verwendung der Newton-Raphson-Methode. Hierzu wird zunächst die Log-Likelihood-Funktion $L(\boldsymbol{\beta})$ bestimmt. Für das binärlogistische Regressionsmodell ist diese - ausgehend von der Binomialverteilung $B(k|p, n)$, $n := \text{card } T$, $p := \text{card } [1]_{\{0,1\}}/n$, $0 \leq k \leq n$ - gegeben durch

$$L(\boldsymbol{\beta}) = \left(\sum_{i \in T} y_i \mathbf{x}_i \right)^T \cdot \boldsymbol{\beta} - \sum_{i \in T} \text{card } [\mathbf{x}_i]_T \ln(1 + \exp(\mathbf{x}_i^T \cdot \boldsymbol{\beta}))$$

(Agresti, 2002, 2007). Die Newton-Raphson-Methode beginnt mit einer nahezu beliebigen ersten Festlegung $\boldsymbol{\beta}^{(0)}$ der Parameter in Iteration (0). Das Verfahren sei nun in Iteration (t) angelangt. Die Log-Likelihood-Funktion wird nach Taylor um die Schätzung $\boldsymbol{\beta}^{(t)}$ der Parameter bis zum quadratischen Term entwickelt (Bronstein et al., 2006):

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\beta}^{(t)}) + (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})^T \cdot \mathbf{u}(\boldsymbol{\beta}^{(t)}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})^T H(\boldsymbol{\beta}^{(t)}) (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}),$$

$$\text{mit } \mathbf{u}(\boldsymbol{\beta}) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_i} \right)_{i \in \{0\} \cup S} \text{ (Gradient) und } H(\boldsymbol{\beta}) = \left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j} \right)_{(i,j) \in (\{0\} \cup S)^2} \text{ (Hesse-Matrix).}$$

Am Maximum der Log-Likelihood-Funktion ist deren erste Ableitung gleich Null:

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \mathbf{u}(\boldsymbol{\beta}^{(t)}) + H(\boldsymbol{\beta}^{(t)}) (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) = \mathbf{0}.$$

Hiermit lassen sich die Parameter $\boldsymbol{\beta}^{(t+1)}$ der nächsten Iteration ($t + 1$) schätzen:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - H(\boldsymbol{\beta}^{(t)})^{-1} \mathbf{u}(\boldsymbol{\beta}^{(t)}).$$

Das Verfahren endet, wenn sich bei folgenden Iterationen keine nennenswerten Veränderungen der geschätzten Parameter mehr ergeben. Dass dieses Verfahren gegen ein Maximum konvergiert, ist durch die konkave Form von L garantiert (Agresti, 2002, 2007).

SageMathCloud™ bietet mit der Programmierschnittstelle statsmodels.api über die Programmiersprache Python einen Zugriff auf binärlogistische Regression (nähere Informationen, siehe <http://statsmodels.sourceforge.net/>). Die dortige Implementation verwendet das Newton-Raphson-Verfahren für die Parameter-Schätzung. Sie wird in einer Experimentalreihe in Verbindung mit Sequential Forward Selection (SFS) und 10-fach Kreuzvalidierung verwendet, um BLR-Klassifikatoren für die Vorhersage der 18 Kriterien auf der Grundlage von EEB-Items zu induzieren. Die Implementation liefert die geschätzten Parameter $\boldsymbol{\beta}$ zurück. Um einen BLR-Klassifikator zu testen, wird je Test-Instanz \mathbf{x}_i über das binärlogistische Regressionsmodell die Wahrscheinlichkeit $\pi(\mathbf{x}_i)$ geschätzt, mit der diese Instanz der Klasse 1 angehört. Ist $\pi(\mathbf{x}_i) > .500$, dann wird die Instanz der Klasse 1 zugewiesen, sonst der Klasse 0. Im Anschluss an diese Zuordnungen wird die Klassifikationsgüte eines BLR-Klassifikators (Korrektklassifikationen, Kontingenztafeln, Stabilität, s. o.) bestimmt. Mit Devroye (1988) könnte angenommen werden, dass es für einen BLR-Klassifikators als linearem Diskriminator keines Tests bedarf. Dieses Resultat gilt jedoch nur bei gleich verteilten und unabhängigen Prädiktoren. Mit Blick auf die deskriptiven Statistiken der EEB-Items ist diese Bedingung nicht erfüllt.

5.9.3 Festlegung von Abbruchbedingungen

Anders als bei der vollständigen Suche ist bei den Heuristiken SFS und SFFS die Abbruchbedingung θ nicht auf natürliche Weise gegeben. Die beiden Heuristiken würden ohne Abbruchbedingung stets den vollständigen Satz an Prädiktoren auswählen. Weiter oben wurde in Bezug auf die Bayes-Klassifikatoren erläutert, dass sich Trainings- und Test-Korrektklassifikation mit Zunahme der Messkomplexität bzw. der Anzahl ausgewählter Prädiktoren unterschiedlich verhalten: Die Trainings-Korrektklassifikation steigt monoton, während es für die Test-Korrektklassifikation ein Maximum gibt. Weiter wurde erläutert, dass

sich Trainings- und Test-Korrektklassifikation möglichst nicht signifikant unterscheiden sollten. Dies beugt einer Über- oder Unteranpassung des Klassifikators an die Trainingsdaten vor. Daher wird als Abbruchbedingung festgelegt:

$$\theta(f, S, X, \mathbf{y}, T) := \kappa(f(X, S), \mathbf{y}, T) \text{ ist signifikant größer als } \kappa(f(X, S), \mathbf{y}, t).$$

Auf Signifikanz wird mittels Wald-Test auf dem Niveau $p < .050$ entschieden. Während der Durchführung von SFS und SFFS wird derjenige induzierte Klassifikator $(f, S)^*$ gespeichert, der bislang die beste Test-Korrektklassifikation erzielt hat. Wird die Abbruchbedingung erfüllt, so wird dieser beste Klassifikator als Ergebnis des Klassifikator-Induktors ausgegeben. Er muss nicht zwingend mit dem Klassifikator (f, S) übereinstimmen, der zum Abbruch der Merkmals-Selektion geführt hat. Diese Modifikation der beiden Heuristiken ist damit begründet, dass die optimale Messkomplexität bei Erfüllung von θ schon überschritten worden sein kann.

SFS kann genutzt werden, um CV-T-Diagramme zu erstellen. In CV-T-Diagrammen wird das Verhalten von Trainings- und Test-Korrektklassifikation in Abhängigkeit von der Messkomplexität dargestellt. Zudem kann in einem CV-T-Diagramm der p -Wert des Wald-Tests in Abhängigkeit der Messkomplexität abgetragen werden. Solche Diagramme liegen sowohl für die Bayes- als auch die BLR-Klassifikation vor (siehe Ergebnisteil). Im Falle der Bayes-Klassifikation zeigt sich das von Hughes (1968) beschriebene Verhalten. Außerdem zeigt sich, dass mit θ Sätze um die vier bis fünf Prädiktoren ausgewählt werden. Dies entspricht der optimalen Messkomplexität nach Hughes (1968). Das Verhalten der BLR-Klassifikation unterscheidet sich insofern, dass sich Trainings- und Test-Korrektklassifikation auf einem gemeinsamen Niveau einpendeln. Um die Vergleichbarkeit der Experimentalreihen zu wahren, wird jedoch davon abgesehen, für die BLR-Klassifikation eine andere Abbruchbedingung zu definieren.

5.10 FINALE AUSWAHL UND KALIBRIERUNG VON KLASSIFIKATOREN: BOOTSTRAPPING

Ziel der vorliegenden Arbeit ist es, für beide Instrumente (vollständige EEB und WSIB) und jedes der 18 Kriterien je einen einzigen, möglichst optimalen Klassifikator zu bestimmen. Aus der Vielzahl jener Klassifikatoren, die in beiden Experimentalreihen induziert wurden, ist daher eine Auswahl zu treffen. Auf einer ersten Ebene werden die Cross-Validation-Resultate der Klassifikator-Induktoren verglichen. Die Beschränkung auf besonders geeignete Klassifikator-Induktoren kann die Suche nach geeigneten Klassifikatoren erleichtern. Auf einer zweiten Ebene sind die einzelnen Klassifikatoren untereinander zu vergleichen und der jeweils

beste Klassifikator auszuwählen. Dies erfolgt mittels Bootstrapping, also wiederholter Testung der Klassifikatoren mit unterschiedlichen Test-Stichproben. Ist ein Klassifikator ausgewählt, sind seine systematischen Fehler festzustellen, um sie bei späterer Verwendung des Klassifikators zu berücksichtigen. Dieser abschließende Schritt wird hier als Kalibrierung bezeichnet und erfolgt ebenfalls mittels Bootstrapping.

5.10.1 Vergleich von Klassifikator-Induktoren

In den beiden Experimentalreihen werden je Instrument (vollständige EEB oder WSIB) und je Kriterium sechs Spezifikationen von Klassifikator-Induktoren untersucht. Diesen sechs Induktoren sind die verwendeten Daten, die Zielfunktion bzw. die Beurteilung der Klassifikationsgüte sowie die 10-fach Kreuzvalidierung gemeinsam. Sie unterscheiden sich bei der Merkmals-Selektion (vollständige Suche, SFS, SFFS), bei der Art der induzierten Klassifikatoren (Bayes, binärlogistische Regression) sowie bei der Festlegung der Abbruchbedingung. Es stellt sich die Frage, ob einer dieser Induktoren den anderen überlegen ist, und falls ja, welcher dies ist. Da die Beurteilung der Klassifikationsgüte für alle Induktoren in gleicher Weise vorgenommen wird, können entsprechende Ergebnisse mit einander verglichen werden. Der Vergleich der Verfahren erfolgt durch Vergleich der Cross-Validation-Korrektklassifikationen mittels χ^2 -Test. D. h. für die Cross-Validation-Korrektklassifikationen κ_1 und κ_2 zweier Induktoren wird der entsprechende p -Wert der Größe $X^2 = \left(\frac{\kappa_1 - \kappa_2}{\sigma(\kappa_1 - \kappa_2)} \right)^2$ bestimmt (Agresti, 2002, 2007). Die Ergebnisse vorwegnehmend stellt sich heraus, dass diesbezüglich keiner der Induktoren vorzuziehen wäre. Dennoch weist binärlogistische Regression gegenüber der Bayes-Klassifikation einige Nachteile auf: Sie erfordert erheblich mehr Rechenaufwand, führt zu vergleichsweise geringeren Stabilitäten und neigt in Kombination mit SFS zu Extremen bei den Umfängen der ausgewählten Prädiktor-Teilmengen. Aus diesen Gründen beziehen sich die folgenden Analyse-Schritte ausschließlich auf Bayes-Klassifikatoren.

5.10.2 Bootstrapping

Bootstrapping ist ein Verfahren, mit dem trotz einer begrenzten Anzahl an Instanzen eine zuverlässige Schätzung statistischer Kennwerte erreicht werden kann. Hierzu wird aus dem Gesamtumfang $\{1..N\}$ aller Instanzen h Mal eine Zufallsstichprobe $t := (i_k)_{k \in \{1..n\}} \in \{1..N\}^n$ von n Instanzen (mit Zurücklegen) gezogen und die betreffenden Kennwerte z_j für die Zufallsstichproben $j \in \{1..h\}$ berechnet. Als Schätzung für den betreffenden Kennwert wird schließlich der Mittelwert über alle h gezogenen Zufallsstichproben gebildet: $\langle z \rangle := \langle z_j \rangle_{j \in \{1..h\}}$.

Darüber hinaus kann die Standardabweichung vermittels $\sigma^2(z) := \langle (z_j - \langle z \rangle)^2 \rangle_{j \in \{1..h\}}$, und somit auch ein Konfidenzintervall für den betreffenden Kennwert angegeben werden (Boos, 2003; Clarke et al., 2009; Davidson, Hinkley & Young, 2003; Efron, 1979, 2003; Hall, 2003; Hastie et al., 2009). Bootstrapping nutzt die Gesetze der großen Zahlen aus. Diesen Gesetzen nach konvergiert der durch unabhängige, gleichartige Zufallsexperimente ermittelte Erwartungswert eines Kennwertes $\langle z \rangle$ mit steigender Anzahl an Zufallsexperimenten immer stärker gegen den Erwartungswert der zu Grunde liegenden tatsächlichen Verteilung (Bronstein et al., 2006). Je höher also h gewählt wird, desto stärker kann dem durch Bootstrapping ermittelten Erwartungswert $\langle z \rangle$ vertraut werden. Die Wahl von h ist jedoch durch die Grenzen der praktischen Realisierbarkeit vieler Zufallsexperimente limitiert.

5.10.3 Finale Auswahl von Klassifikatoren

Über die verschiedenen Experimentalreihen zur Klassifikator-Induktion mit Kreuzvalidierung kommt es zu einer Vielzahl von Klassifikatoren, die als Kandidaten für die Vorhersage eines der 18 Kriterien in Frage kommen. Bei Zugrundelegung der 25 EEB-Items werden je Kriterium fünf Bayes-Experimentalreihen mit 10-facher Kreuzvalidierung durchgeführt. Das führt auf maximal 50 unterschiedliche Klassifikatoren. Bei Zugrundelegung des WSIB kommt man ebenfalls auf diese Anzahl. Es stellt sich daher die Frage, welcher dieser Kandidaten endgültig als geeignetster Klassifikator auszuwählen ist. Zur Lösung dieses Problems wird auf Bootstrapping zurückgegriffen. Zunächst werden je Instrument (vollständige EEB und WSIB) und je Kriterium alle ermittelten Prädiktor-Teilmenge inklusive Spezifikation der Merkmals-Selektion gesammelt. So erhält man eine Menge $\{S_i | i \in \{1..K\}\}$ von Kandidaten für Prädiktor-Teilmenge. Anschließend werden die Trainingsdaten mit dem Gesamtumfang aller Instanzen, $T = \{1..N\}$, gleichgesetzt. Der Lerner ℓ (s. o.) wird nun auf diese Trainingsdaten angesetzt und spezifiziert die Parameter des Bayes-Klassifikators: $f_i := \ell(X, \mathbf{y}, T, S_i)$. D. h. der Bayes-Klassifikator wird nicht mehr, wie zuvor bei Kreuzvalidierung, mit 90 % Instanzen der Induktionsstichproben, sondern mit 100 % trainiert. Nach diesen vorbereitenden Schritten setzt das Bootstrapping ein. Die Anzahl von Bootstrapping-Iterationen wird auf $h = 1\,000$ und der Umfang der einzelnen Zufallsstichproben auf $n = 100$ gesetzt. Als zu schätzende statistische Größe wird die Test-Korrektklassifikation gewählt. Somit erhält man 1 000 Zufallsstichproben $t_j, j \in \{1..1\,000\}$ sowie entsprechend 1 000 Test-Korrektklassifikationen $\kappa_{ij} := \kappa(f_i(X, S_i), \mathbf{y}, t_j)$ für den i -ten Kandidaten der K Klassifikatoren. Der Schätzwert für die Test-Korrektklassifikation des Klassifikators i ergibt sich schließlich durch Mittelwertbildung: $\langle \kappa_i \rangle = \langle \kappa_{ij} \rangle_{j \in \{1..1\,000\}}$ (Davidson et al., 2003; Efron, 1979, 2003). Sie wird

im Ergebnisteil mit $\kappa_{\text{Bootstrap}}$ bezeichnet. Mit Abbildung 10 wird das Vorgehen beim Bootstrapping noch einmal grafisch veranschaulicht.

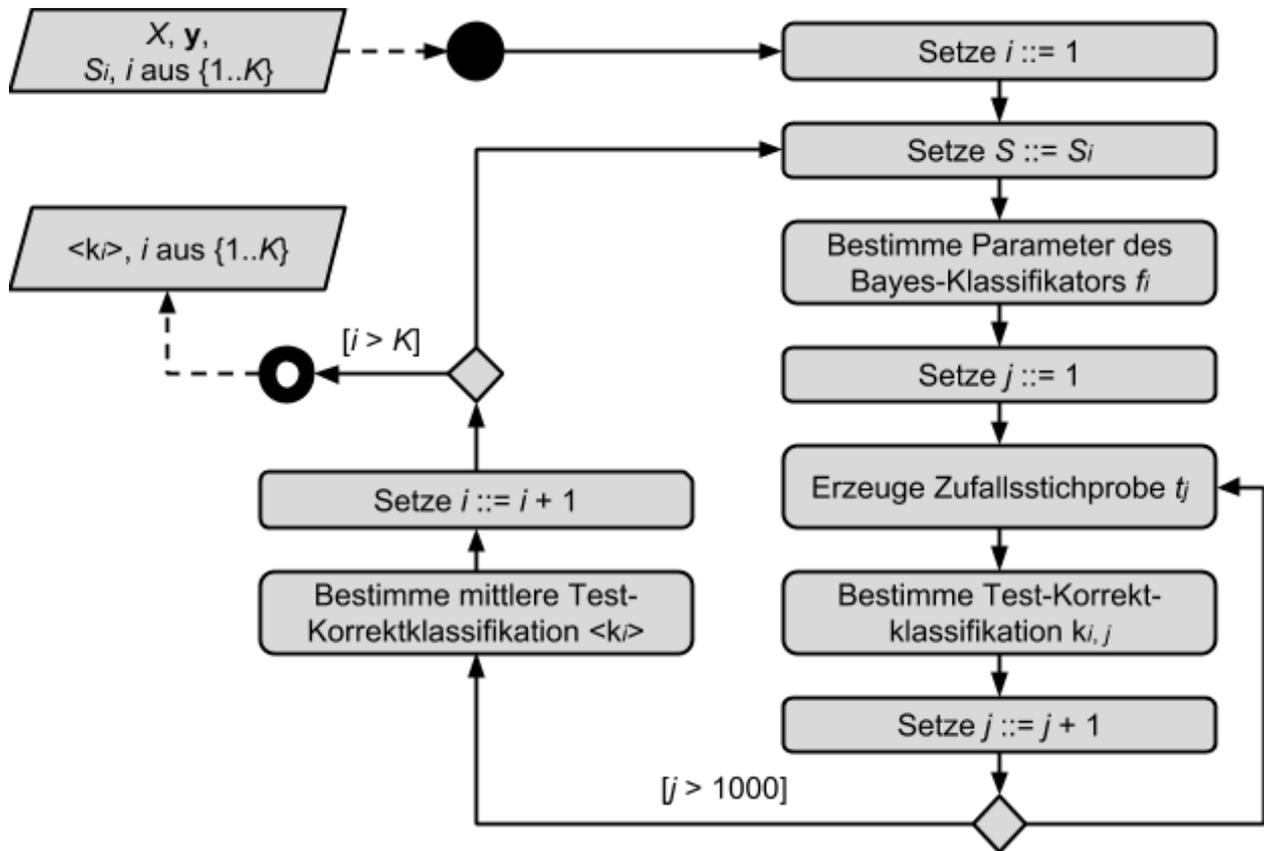


Abbildung 10. Schematische Darstellung des Vorgehens beim Bootstrapping zur Schätzung statistischer Kennwerte unter Verwendung eines Klassifikator-Induktors (eigene Darstellung).

Nach Durchführung des beschriebenen Algorithmus liegen für jeden Kandidaten je Instrument und je Kriterium die mittleren Test-Korrektklassifikationen vor. Diese sind die Grundlage für die endgültige Auswahl eines Klassifikators. Es wird jedoch nicht jener Klassifikator gewählt, der die höchste Test-Korrektklassifikation erzielt. Vielmehr wird auch hier den Ergebnissen von Hughes (1968) Rechnung getragen. Es wird jener Klassifikator gewählt, dessen mittlere Test-Korrektklassifikation am nächsten unterhalb der theoretisch maximal möglichen Test-Korrektklassifikation .750 liegt. Damit wird die Gefahr einer möglichen Unter- oder Überanpassung des Klassifikators an die Trainingsdaten abgemildert. Nach Durchführung der Prozedur liegt für jedes Instrument (vollständige EEB und WSIB) sowie für jedes Kriterium jeweils ein Klassifikator vor. Es kann vorkommen, dass ein Klassifikator bezüglich des WSIB bei einem Kriterium ein besseres Ergebnis erzielt, als der entsprechende Klassifikator bezüglich der vollständigen EEB. In diesem Falle wird nur der Klassifikator bezüglich des WSIB ausgewählt, da dieser in der vollständigen EEB enthalten ist. Somit liegt die Anzahl final ausgewählter Klassifikatoren zwischen 18 und $2 \times 18 = 36$ Stück.

5.10.4 Kalibrierung von Klassifikatoren

Bei der Entwicklung arbeitspsychologischer Analyseverfahren geht es weniger darum, personenbezogene Aussagen treffen zu können, sondern viel mehr um zuverlässige Aussagen über Personengruppen. Personengruppen sind häufig Beschäftigte in bestimmten Abteilungen, bestimmten Berufen oder bestimmten Funktionen innerhalb eines Unternehmens. Mit der Kalibrierung eines Klassifikators sind im Folgenden zwei Schritte gemeint: (1) Die Ermittlung und Dokumentation systematischer Fehler und (2) deren zukünftige Berücksichtigung beim Einsatz des Klassifikators zur Vorhersage von Kennwerten für Personengruppen. Für betrieblich relevante Personengruppen sollen mittels Einsatzes eines Klassifikators (a) zu erwartende relative Häufigkeiten der beiden Klassen 0 und 1 und (b) Erwartungswerte für alle 18 Kriterien vorhergesagt werden. Hierfür sollen die final ausgewählten Klassifikatoren (siehe voriger Abschnitt) eingesetzt werden. Bezüglich der relativen Häufigkeiten ist mit zwei Arten von Fehlern zu rechnen: Zum einen kann es bei der Vorhersage der beiden Klassen zu einer systematischen Verschiebung hin zu einer Klasse kommen. D. h. eine der beiden Klassen 0 und 1 kann bei der Vorhersage systematisch überrepräsentiert sein. Zum anderen kann es auch nach Berücksichtigung dieser Verschiebung noch zu einem mittleren Fehler bei der Vorhersage der beiden relativen Häufigkeiten kommen. Für die Vorhersage von Erwartungswerten ist ebenfalls mit einem mittleren Fehler zu rechnen. Die Ermittlung solcher systematischer Fehler erfolgt mittels Bootstrapping.

Die Ermittlung der Überrepräsentationen von Klassen bei der Vorhersage erfolgt bereits während der finalen Auswahl der Klassifikatoren. Dies ist möglich, da neben der Feststellung der mittleren Korrekturklassifikation auch jeweils eine Kontingenztafel über alle 1 000 Bootstrapping-Iterationen hinweg aufsummiert und ausgewertet wird. Diese Kontingenztafeln weisen Umfänge von $100 \times 1\,000 = 100\,000$ Klassifikationen einzelner Instanzen auf. Greifen wir auf die Variablenbezeichnungen für Kontingenztafeln aus Abbildung 8 zurück, so müssen die Randsummen $a + b$ sowie $c + d$ auf Grund der Dichotomisierung jeweils etwa 50 000 betragen. Ist bei der Vorhersage eine der beiden Klassen überrepräsentiert, dann unterscheiden sich die beiden Randsummen $a + c$ und $b + d$ wesentlich voneinander. Diese Überrepräsentationen können durch die Korrekturfaktoren $c_1 := \frac{a+b}{a+c}$ für Klasse 1 und $c_0 := \frac{c+d}{b+d}$ für Klasse 0 behoben werden. Sie werden im Ergebnisteil angegeben und bei der Schätzung von relativen Häufigkeiten für Personengruppen berücksichtigt.

Die mittleren Fehler bei der Schätzung von relativen Häufigkeiten und Erwartungswerten werden im Rahmen eines weiteren Bootstrapping-Verfahrens ermittelt. Auch hier werden die

Anzahl an Bootstrapping-Iterationen auf $h = 1\,000$ und die Umfänge der einzelnen Zufallsstichproben auf $n = 100$ gesetzt. In diesem Verfahren wird bei gegebenem Kriterium und gegebenem Instrument (vollständige EEB oder WSIB) für jede Zufallsstichprobe $t_j, j \in \{1..1\,000\}$ sowohl die relative Häufigkeit $\pi_{1,j}$ für Klasse 1 als auch der Erwartungswert μ_j durch Klassifikation geschätzt:

$$\pi_{1,j} := c_1 \frac{\text{card}\{i \in t_j \mid f(X, t_j)_i = 1\}}{100},$$

$$\mu_j := \pi_{1,j} \mu^{(1)} + (1 - \pi_{1,j}) \mu^{(0)}.$$

Hierbei sind $\mu^{(0)}$ und $\mu^{(1)}$ die Klassenmittelwerte des betreffenden Kriteriums (siehe deskriptiv-statistische Analyse weiter oben). Bei Schätzung der relativen Häufigkeit wird der Korrekturfaktor c_1 zur Beseitigung einer systematischen Überrepräsentation einer der beiden Klassen (s. o.) berücksichtigt. Da für jede Instanz aus den Induktionsstichproben außerdem die tatsächlichen Klassenzugehörigkeiten k_i sowie die tatsächlichen Skalenwerte $s_i, i \in \{1..N\}$ der Kriterien vorliegen, können auch die tatsächlichen relativen Häufigkeiten $p_{1,j}$ und arithmetischen Mittelwerte m_j für die Zufallsstichproben bestimmt werden:

$$p_{1,j} = \frac{\text{card}\{i \in t_j \mid k_i = 1\}}{100},$$

$$m_j = \langle s_i \rangle_{i \in t_j}.$$

Dies führt auf die Vorhersagefehler für die relativen Häufigkeiten und Erwartungswerte (Davidson et al., 2003; Efron, 1979, 2003):

$$\varepsilon(\pi_1)_j := p_{1,j} - \pi_{1,j}, \langle \varepsilon(\pi_1) \rangle = \langle \varepsilon(\pi_1)_j \rangle_{j \in \{1..1000\}},$$

$$\sigma^2(\varepsilon(\pi_1)) = \langle (\varepsilon(\pi_1)_j - \langle \varepsilon(\pi_1) \rangle)^2 \rangle_{j \in \{1..1000\}}.$$

$$\varepsilon(\mu)_j := m_j - \mu_j, \langle \varepsilon(\mu) \rangle = \langle \varepsilon(\mu)_j \rangle_{j \in \{1..1000\}},$$

$$\sigma^2(\varepsilon(\mu)) = \langle (\varepsilon(\mu)_j - \langle \varepsilon(\mu) \rangle)^2 \rangle_{j \in \{1..1000\}}.$$

Im Ergebnisteil werden arithmetische Mittelwerte und Standardabweichungen für diese Fehler je Kriterium und Instrument angegeben und später (s. u.) bei Schätzungen für Anwendungs- und Validierungszwecke berücksichtigt.

5.10.5 Vorhersagefehler von Erwartungswerten in Abhängigkeit des Stichprobenumfangs

In der Anwendung der Klassifikatoren ist die Vorhersage der Erwartungswerte von besonderer praktischer Relevanz, da sie anschaulicher ist, als die Vorhersage von relativen

Häufigkeiten. Hierbei ist zu erwarten, dass der Fehler bei der Vorhersage von Erwartungswerten vom Stichprobenumfang abhängig ist. Je geringer der Stichprobenumfang ist, desto größer wird der Vorhersagefehler sein. Aus diesem Grunde wird ein weiteres Bootstrapping-Experiment durchgeführt, bei dem der Stichprobenumfang von $n = 10$ bis $n = 100$ variiert und dabei der Vorhersagefehler $\langle \varepsilon(\mu) \rangle$ und die zugehörige Standardabweichung $\sigma_n(\varepsilon(\mu))$ bestimmt werden. Die Bootstrapping-Iterationen werden weiterhin mit $h = 1\,000$ festgesetzt. Ziel ist die empirische Ermittlung einer Funktion $f: \{10..100\} \rightarrow \mathbb{R}; n \mapsto \sigma_n(\varepsilon(\mu))$ für jedes der 18 Kriterien.¹² Mit dieser Funktion können Vertrauensintervalle im Anwendungsfalle an gegebene Stichprobenumfänge aus dem Zahlenbereich $\{10..100\}$ angepasst werden.

5.11 ANWENDUNG UND VALIDIERUNG DER ERMITTELTEN KLASSIFIKATOREN

Die Demonstration der praktischen Anwendungsmöglichkeiten hier präsentierter Methoden sowie die Validierung ermittelter Klassifikatoren beschränken sich zum einen auf Bayes-Klassifikatoren. Binärlogistische Regression wird nicht weiter untersucht (zur Begründung, siehe voriger Abschnitt). Zum anderen stehen hierfür Stichproben zur Verfügung, bei denen nicht die vollständige EEB sondern nur das Wuppertaler Screening Instrument psychische Beanspruchung (WSIB) erhoben wurde.

Die Anwendung erfolgt, wie bei der finalen Auswahl der Klassifikatoren (s. o.), in zwei Schritten: Zunächst wird die Wahrscheinlichkeit π_1 für Klasse 1 durch Klassifikation der zur Verfügung stehenden Instanzen auf Grundlage der Angaben im WSIB für alle 18 vorherzusagenden Kriterien geschätzt. Hierbei kommt je vorherzusagendem Kriterium der final ausgewählte Klassifikator f unter Verwendung der entsprechenden Prädiktor-Teilmenge S zum Einsatz. Erster Rechenschritt ist die Bestimmung der Schätzung π_1^* durch

$$\pi_1^* = c_1 \frac{\text{card}\{i \in \{1..N\} | f(X, S)_i = 1\}}{N}$$

Mit c_1 werden Verschiebungen auf Grund der Überrepräsentation einer der beiden Klassen bei der Vorhersage korrigiert (siehe Kalibrierung von Klassifikatoren weiter oben). Datenmatrix X und Stichprobenumfang N beziehen sich nun auf die Stichprobe, auf die der Klassifikator angewendet wird (Validierungs- bzw. Anwendungsstichprobe). Die korrigierte Schätzung π_1 erfolgt durch

¹² Mit \mathbb{R} wird die Menge der reellen Zahlen bezeichnet.

$$\pi_1 = \pi_1^* + \langle \varepsilon(\pi_1) \rangle.$$

$\langle \varepsilon(\pi_1) \rangle$ stellt den mittleren Fehler bei der Schätzung von π_1 dar, wie er bei der Kalibrierung ermittelt wurde. Wurde π_1^* ermittelt, dann kann vermittels der Formel

$$\mu = \pi_1^* \mu^{(1)} + (1 - \pi_1^*) \mu^{(0)} + \langle \varepsilon(\mu) \rangle$$

der Erwartungswert für ein Kriterium geschätzt werden. $\langle \varepsilon(\mu) \rangle$ ist der mittlere Fehler bei Schätzung des Erwartungswertes, wie er bei der Kalibrierung ermittelt wurde.

Da auch die Standardabweichungen der Fehler bei der Vorhersage von Wahrscheinlichkeiten und Erwartungswerten vorliegen, können die obigen Schätzungen um Konfidenzintervalle ergänzt werden. Für die Schätzungen der Wahrscheinlichkeiten für Klasse 1 sowie der Erwartungswerte werden die zweiseitigen 99,8 %-Konfidenzintervalle

$$99,8 \text{ \% -CI}[\blacksquare] = [\blacksquare - 3,174 \sigma(\varepsilon(\blacksquare)), \blacksquare + 3,174 \sigma(\varepsilon(\blacksquare))], \blacksquare \in \{\pi_1, \mu\},$$

ausgehend von den Umfängen $n = 100$ der beim Bootstrapping verwendeten Zufallsstichproben, angegeben (Davidson et al., 2003; Efron, 1979, 2003). Bei der Vorhersage von Erwartungswerten können die 99,8 %-Konfidenzintervalle auch für Stichprobenumfänge $n \in \{10..100\}$ angepasst werden (s. o.). Zur Validierung der Klassifikatoren wird geprüft, ob die aus den Messungen der Kriterien resultierenden arithmetischen Mittelwerte der Validierungsstichprobe tatsächlich innerhalb der geschätzten bzw. vorhergesagten 99,8 %-Konfidenzintervalle liegen. Die Breite dieser Konfidenzintervalle ist insofern angemessen, da die ermittelten Standardabweichungen vergleichsweise klein sind.

Sowohl die Ergebnisse der Validierung als auch die Schätzungen für Anwendungszwecke werden gesammelt und in Form von Profilen als Diagramme dargestellt. Die Schätzungen für Anwendungszwecke werden vor dem Hintergrund von deskriptiven Kennwerten (arithmetisches Mittel und Median) der Induktionsstichproben dargestellt. Dies ermöglicht den unmittelbaren visuellen Vergleich der Anwendungsstichprobe mit einer Vergleichsstichprobe (sogenanntes Benchmarking).

Neben der empirischen Validierung werden Schätzungen für Teilstichproben aus der Anwendungsstichprobe (Produktion und Führungskräfte) gegenübergestellt und mit dem verantwortlichen Gesundheitsmanager des betreffenden Unternehmens qualitativ diskutiert und interpretiert. Diese Form der Plausibilitätsprüfung soll weitergehende Hinweise auf die praktische Anwendbarkeit sowie die Validität der untersuchten Methoden liefern.

5.12 AUTOMATISIERTE DARSTELLUNG DER ERGEBNISSE

Die beschriebene methodische Vorgehensweise führt zu einem ungewöhnlich hohen Umfang an Einzelergebnissen. Dies stellt besondere Herausforderungen an die Qualität ihrer Verschriftlichung dar, sowohl im Hinblick auf die Korrektheit als auch im Hinblick auf die Lesbarkeit. Um die Korrektheit der dargestellten Ergebnisse zu sichern, wird ihre Verschriftlichung weitgehend automatisiert. Die Lesbarkeit wird zum einen durch eine hohe Strukturiertheit und zum anderen durch eine an den vorherzusagenden Kriterien orientierte Organisation der Darstellung trotz der vielen Einzelergebnisse hoffentlich ermöglicht.

Sämtliche Ergebnisse liegen in Form binärer Daten oder Textdateien auf einem Server von SageMathCloudTM vor. Es wurden Programme geschrieben, die diese Ergebnisse entweder in Form von Diagrammen oder in Form von HTML-Dokumenten formatieren. Diagramme konnten unmittelbar in die vorliegende Arbeit übernommen werden. Die HTML-Dokumente enthalten im Wesentlichen formatierte Tabellen mit Ergebnissen. Sie konnten mit Microsoft[®] Visual Basic[®] for Applications in Microsoft[®] Word-Dokumente umgewandelt und zusammengeführt werden. Die Formatierungen mussten nur noch geringfügig korrigiert und die Tabellen um beschreibende Texte ergänzt werden. Die weitgehende Automatisierung der Ergebnis-Darstellung erhöht die Effizienz bei der Text-Erstellung und verhindert zugleich Fehler bei der Übernahme der Vielzahl an berechneten Ergebnissen in den Text. Für die vorliegende Arbeit wurde somit ein Software-System entwickelt, das von der Aufbereitung der Daten bis hin zur Darstellung der Ergebnisse nur minimale manuelle Eingriffe erfordert und somit entsprechende Fehlerquellen weitgehend ausschaltet. Sind Fehler in den Ergebnis-Darstellungen (Tabellen und Diagramme) festzustellen, so sind sie zwingend auf Fehler in den implementierten und angewendeten Programmen zurückzuführen. Die Korrektheit der eingesetzten Algorithmen wurde jedoch durch intensive Prüfung (sogenanntes Debugging und Prüfung der Korrektheit durch Vergleich mit manuellen Beispielrechnungen bzw. überschaubaren Anwendungsfällen) nach bestem Wissen sichergestellt. Sämtliche Fließtexte im Ergebnis-Teil und sämtliche anderen Teile der vorliegenden Arbeit wurden manuell erstellt und sind somit von den genannten Vorteilen der Automatisierung ausgenommen.

6 ERGEBNISSE

Der Ergebnisteil der vorliegenden Arbeit beginnt mit der statistischen Beschreibung der EEB-Items. Hierauf folgt die Darstellung der CV-T-Diagramme, um das Verhalten von Trainings- und Test-Korrektklassifikation von Bayes-Klassifikatoren und binärlogistischer Regression in Abhängigkeit der Messkomplexität wiederzugeben. Den Kern des Ergebnisteils stellen die induzierten und ausgewählten Klassifikatoren dar. Für jedes der 18 Kriterien gibt es einen eigenen Bericht. Jeder Bericht beginnt mit einer Erläuterung und deskriptiv-statistischen Darstellung des entsprechenden Kriteriums sowie mit einer Tabelle von Korrelationen zwischen Kriterium und den einzelnen EEB-Items. Hieran schließen sich die Ergebnisse der beiden Experimentalreihen an. D. h. es werden die Cross-Validation-Ergebnisse aller sechs Spezifikationen von Klassifikator-Induktoren für das betreffende Kriterium dargestellt (in diesem Haupttext nur für das Kriterium „Arbeitsanforderungen“, sonst im Anhang A3). Es folgt die endgültige Auswahl von Klassifikatoren mittels Bootstrapping. Der Abschluss eines Berichtes wird durch die Ergebnisse zur Kalibrierung dieser Klassifikatoren gebildet. Nach diesen 18 Berichten werden die Vergleiche der Klassifikator-Induktoren dargestellt. Der Abschluss des Ergebnisteils enthält zum einen die Validierung der endgültigen Klassifikatoren anhand der Validierungsstichproben sowie die Darstellung vorhergesagter Belastungs- und Gesundheitsprofile auf Grundlage der Anwendungsstichprobe.

6.1 DESKRIPTIVE STATISTIK DER EEB-ITEMS

Im Zentrum der vorliegenden Arbeit stehen die Items aus der Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchung während der Arbeit (EEB). Kombinationen der Items werden als Prädiktoren für die Vorhersage der 18 dichotomisierten Kriterien ausgewählt. Daher erfolgt in diesem Abschnitt eine deskriptive Analyse auf Ebene einzelner EEB-Items. Neben Mittelwerten, Standardabweichungen und Stichprobenumfängen werden auch die relativen Häufigkeiten angegeben, mit denen je Item die sieben Antwortkategorien bei Befragungen gewählt wurden. Diese beschreibenden Kennwerte sind in Form von Diagrammen zusammengefasst. In Abbildung 11 finden sich die genannten Kennwerte, aufsteigend sortiert nach den Mittelwerten. Es zeigt sich, dass negativ konnotierte Eigenschaftswörter eher geringe Mittelwerte aufweisen, während die Mittelwerte für positiv konnotierte Eigenschaftswörter besonders hohe Mittelwerte aufweisen. Das Item „leistungsbereit“ weist mit 5,43 den höchsten Mittelwert auf, das Item „ängstlich“ mit 1,41 den niedrigsten. Da die Items in betrieblichen Kontexten beantwortet wurden, ist eine Tendenz zu sozial erwünschtem Antwortverhalten nicht auszuschließen.

Ergebnisse

Deskriptive Statistik der EEB-Items

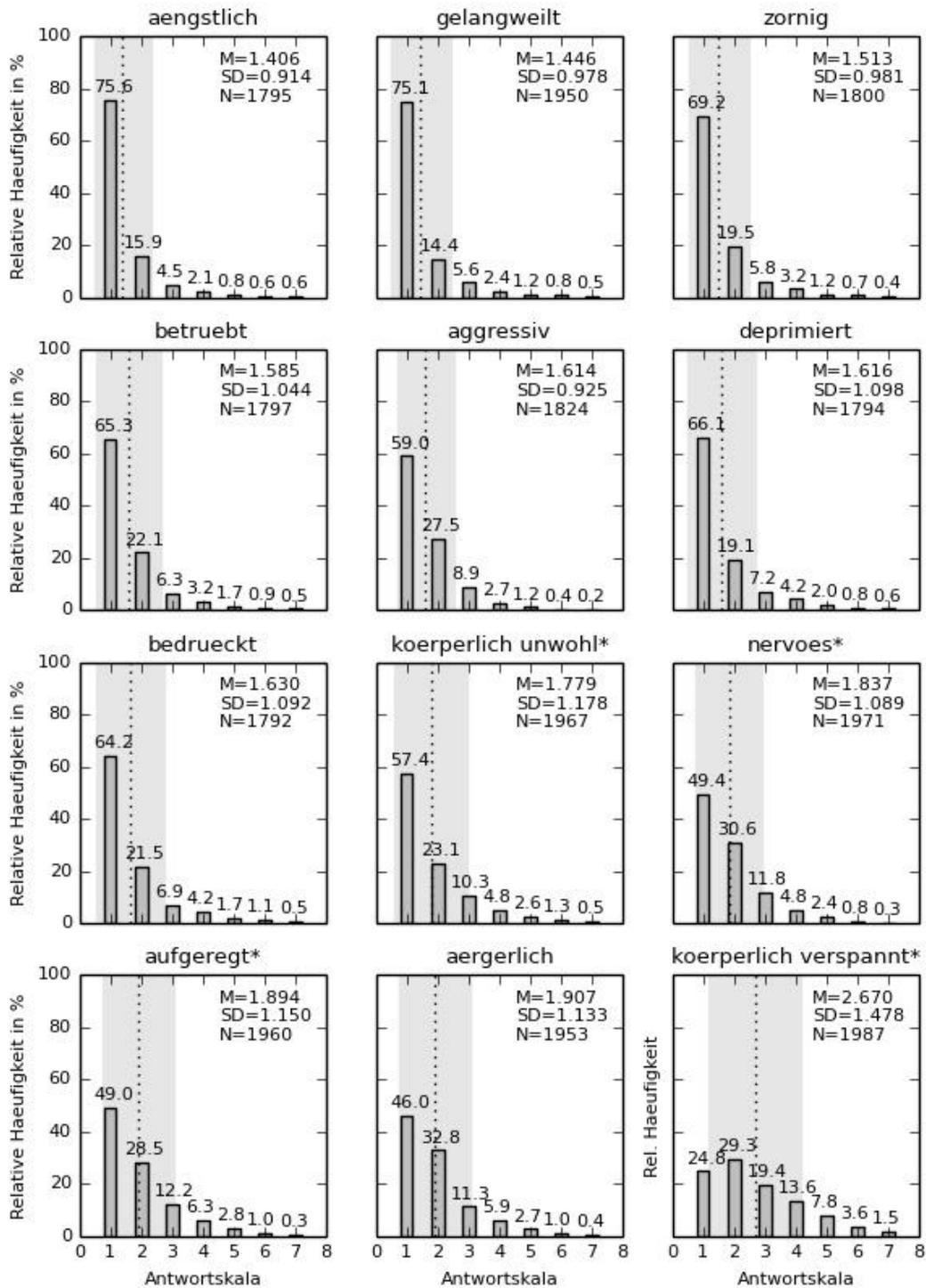


Abbildung 11. Deskriptive Statistik der einzelnen Items der Eigenschaftswörter zu Emotionen und Beanspruchungserleben während der Arbeit (EEB). Die Items sind nach Mittelwert aufsteigend sortiert. M = Mittelwert, SD = Standardabweichung, N = Stichprobenumfang. M ist als gepunktete Linie, der Bereich [M - SD, M + SD] als grauer Bereich eingezeichnet. *Dieses Item gehört auch zum Wuppertaler Screening Instrument psychische Beanspruchung (WSIB).

Ergebnisse

Deskriptive Statistik der EEB-Items

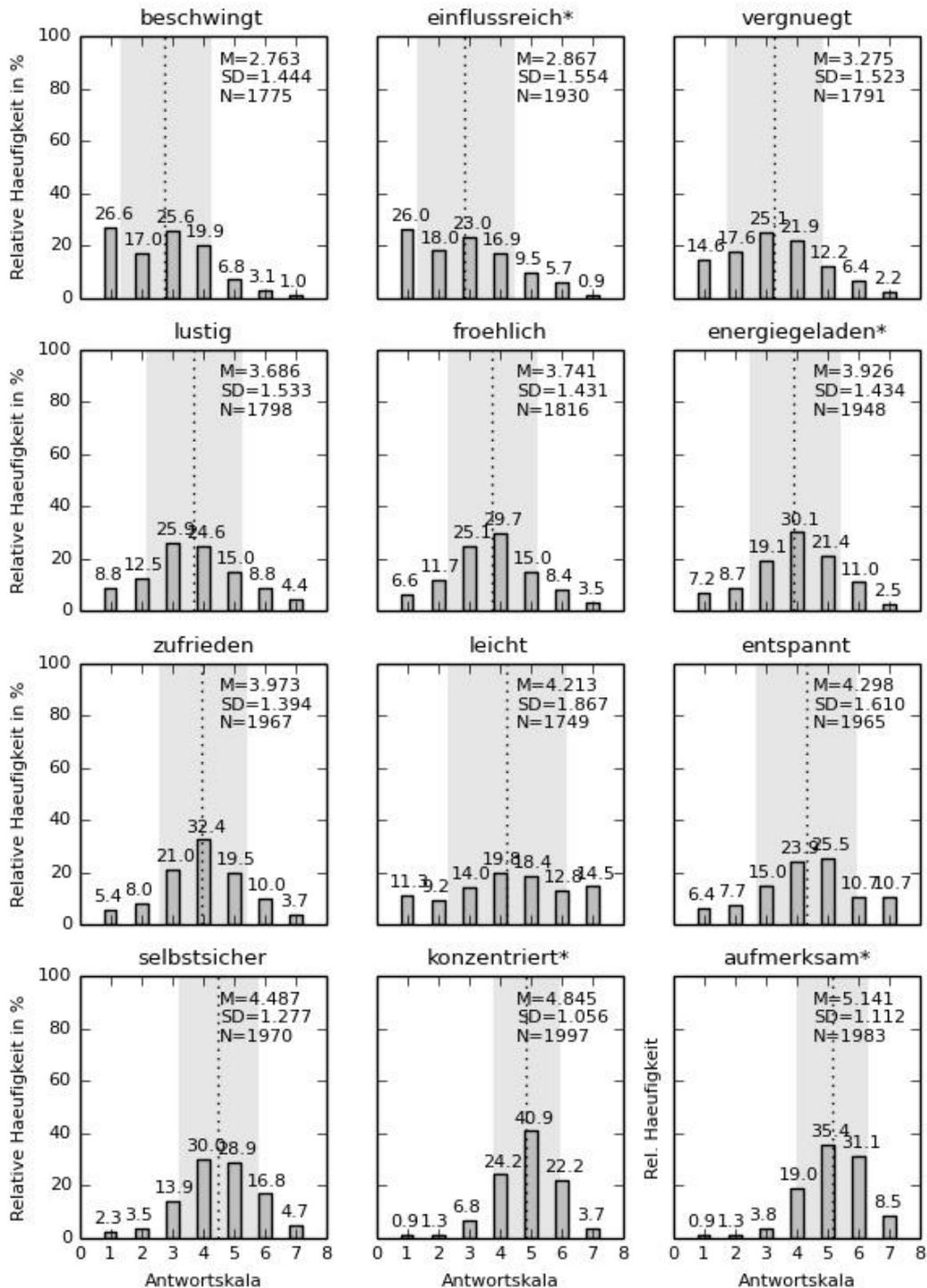


Abbildung 11 (Fortsetzung). Deskriptive Statistik der einzelnen Items der Eigenschaftswörter zu Emotionen und Beanspruchungserleben während der Arbeit (EEB). Die Items sind nach Mittelwert aufsteigend sortiert. M = Mittelwert, SD = Standardabweichung, N = Stichprobenumfang. M ist als gepunktete Linie, der Bereich [M - SD, M + SD] als grauer Bereich eingezeichnet. *Dieses Item gehört auch zum Wuppertaler Screening Instrument psychische Beanspruchung (WSIB).

Deskriptive Statistik der EEB-Items

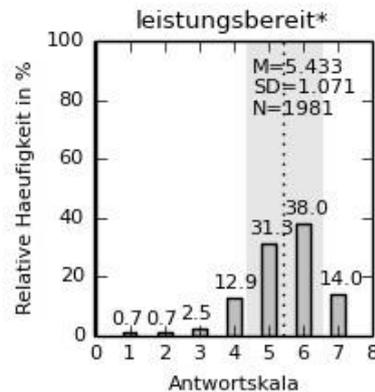


Abbildung 11 (Fortsetzung). Deskriptive Statistik der einzelnen Items der Eigenschaftswörter zu Emotionen und Beanspruchungserleben während der Arbeit (EEB). Die Items sind nach Mittelwert aufsteigend sortiert. M = Mittelwert, SD = Standardabweichung, N = Stichprobenumfang. M ist als gepunktete Linie, der Bereich [M - SD, M + SD] als grauer Bereich eingezeichnet. *Dieses Item gehört auch zum Wuppertaler Screening Instrument psychische Beanspruchung (WSIB).

Das Item „zufrieden“ markiert mit einer nahezu perfekten Normalverteilung den Übergang von negativ zu positiv konnotierten Befindenzuständen bei Beschäftigten in betrieblichen Kontexten. Dessen Normalverteilung kann anhand des Histogramms (Abbildung 11) ebenso deutlich festgemacht werden, wie anhand eines P-P-Plots (Abbildung 12).

P-P-Plot für das Item "zufrieden"

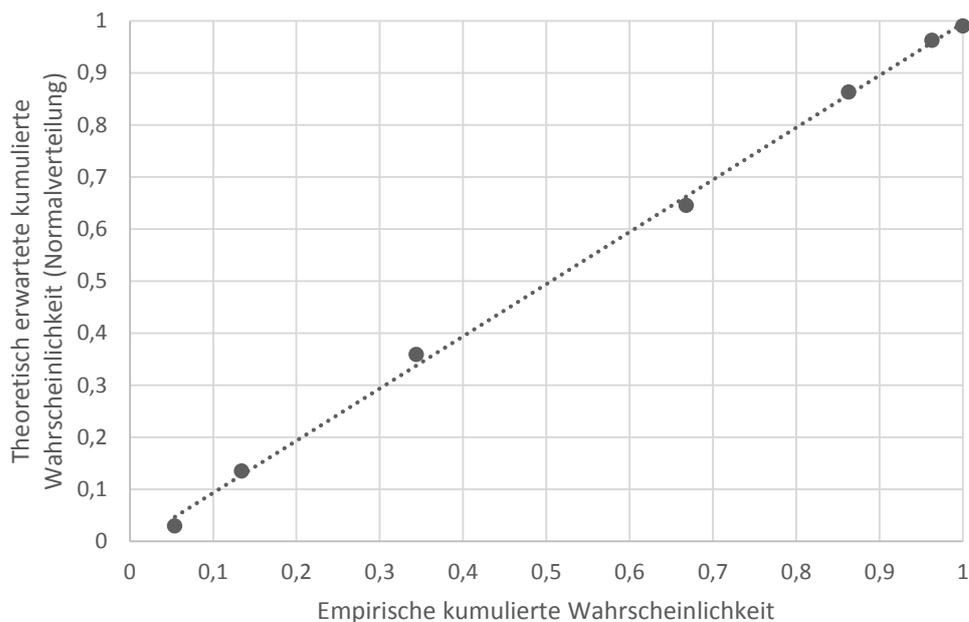


Abbildung 12. P-P-Plot für das Item „zufrieden“. Die gepunktete Linie stellt die Regressions-Gerade dar.

Der in Abbildung 12 dargestellte P-P-Plot basiert auf theoretisch erwarteten kumulierten Wahrscheinlichkeiten bei Annahme einer Normalverteilung mit den Parametern $M = 3,973$ und $SD = 1,394$. Diese wurden den empirischen kumulierten Wahrscheinlichkeiten gegenübergestellt. Die Regressionsgerade für das Item „zufrieden“ weist eine Steigung von 0,996 und einen Achsenabschnitt von 0,008 auf. Das Bestimmtheitsmaß R^2 beträgt für diese Regressionsgerade .999, der Standardfehler des Schätzwertes beträgt 0,015. Mit $F(1, 5) = 3\,968,57$; $p < .001$ kann diese starke Anpassung nicht als zufälliges Ergebnis interpretiert werden. Das Histogramm des Items „energiegeladen“ weist eine ähnlich gute Anpassung an eine Normalverteilung auf ($R^2 = .998$; $F(1, 5) = 2\,303,95$; $p < .001$).

6.2 CV-T-DIAGRAMME: KLASSIFIKATIONSGÜTE IN ABHÄNGIGKEIT DER MESSKOMPLEXITÄT

Für die Bestimmung von Abbruchkriterien der Klassifikator-Induktion ist das Verhalten der Klassifikationsgüte in Abhängigkeit der Messkomplexität eine hilfreiche vorbereitende Analyse. Darüber hinaus bietet diese Analyse die Verifikation der Korrektheit der implementierten Verfahren. Auf den Folgenden zwölf Seiten ist dieses Verhalten in Form von vier CV-T-Diagrammen dargestellt. Trainings- und Cross-Validation-Korrektklassifikation werden in Abhängigkeit der Messkomplexität dargestellt. Die Messkomplexität wird dabei mittels Sequential Forward Selection gesteigert. D. h. ausgehend von der leeren Menge (kein selektiertes Item) wird mit jedem Schritt dasjenige Item hinzugefügt, das zur größten Steigerung der Trainings-Korrektklassifikation führt. Trainings- und Cross-Validation-Korrektklassifikation werden je Schritt mittels Wald-Test auf signifikante Differenz geprüft. Zwei CV-T-Diagramme beziehen sich auf Bayes-Klassifikatoren auf Grundlage sämtlicher 25 EEB-Items (Abbildung 13) sowie auf Grundlage der neun WSIB-Items (Abbildung 14). Zwei weitere Diagramme beziehen sich auf binärlogistische Regression auf Grundlage sämtlicher 25 EEB-Items (Abbildung 15) sowie auf Grundlage der neun WSIB-Items (Abbildung 16). Die separate Darstellung von CV-T-Diagrammen je Instrument (vollständige EEB oder WSIB) ist angebracht, da jeweils mit unterschiedlichen Reihenfolgen der Aufnahme von WSIB-Items und daher mit jeweils unterschiedlichen CV-T-Diagrammen zu rechnen ist.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, EEB)

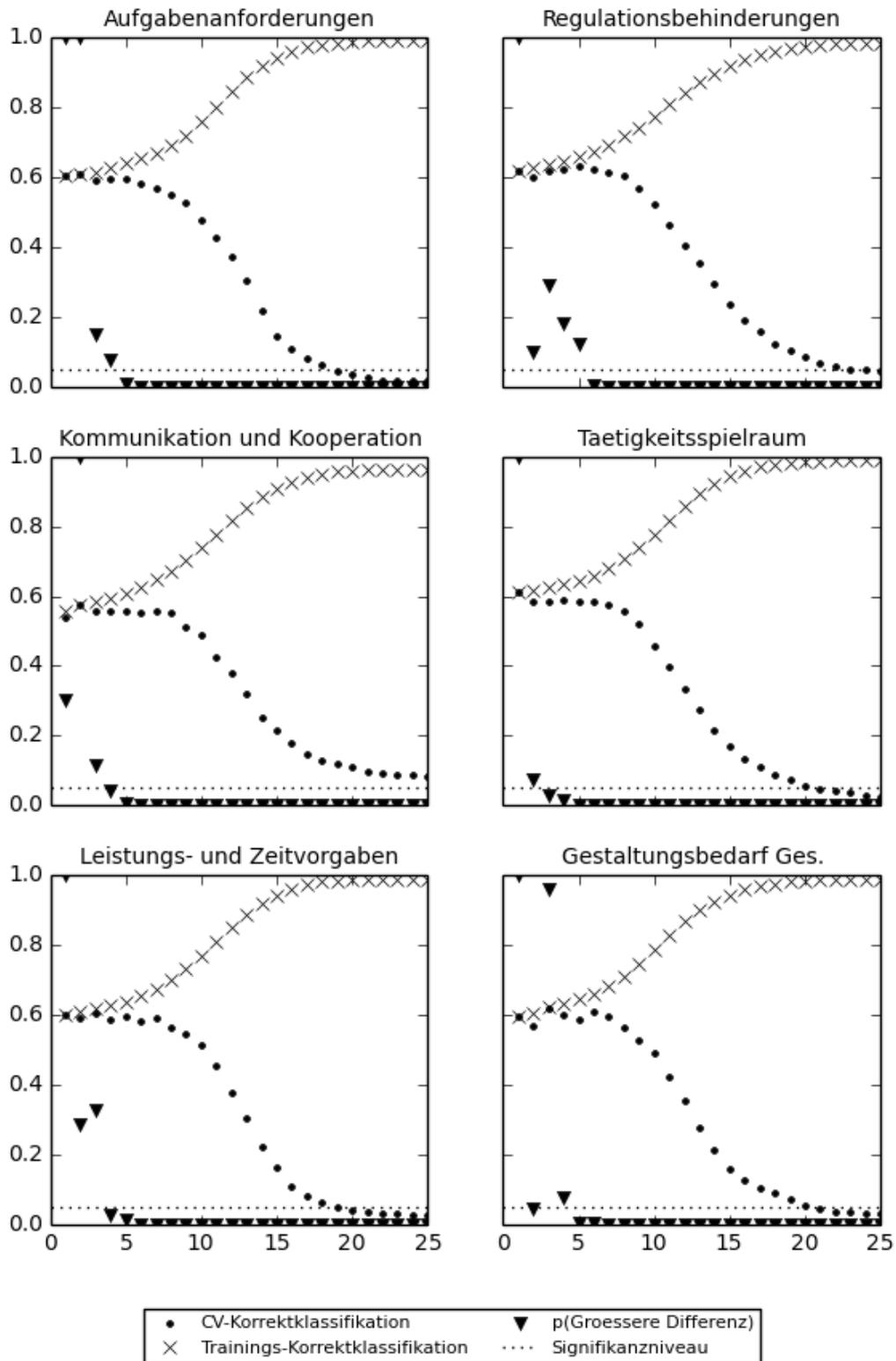


Abbildung 13. CV-T-Diagramme für Bayes-Klassifikatoren auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr mit einander vergleichbar.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, EEB)

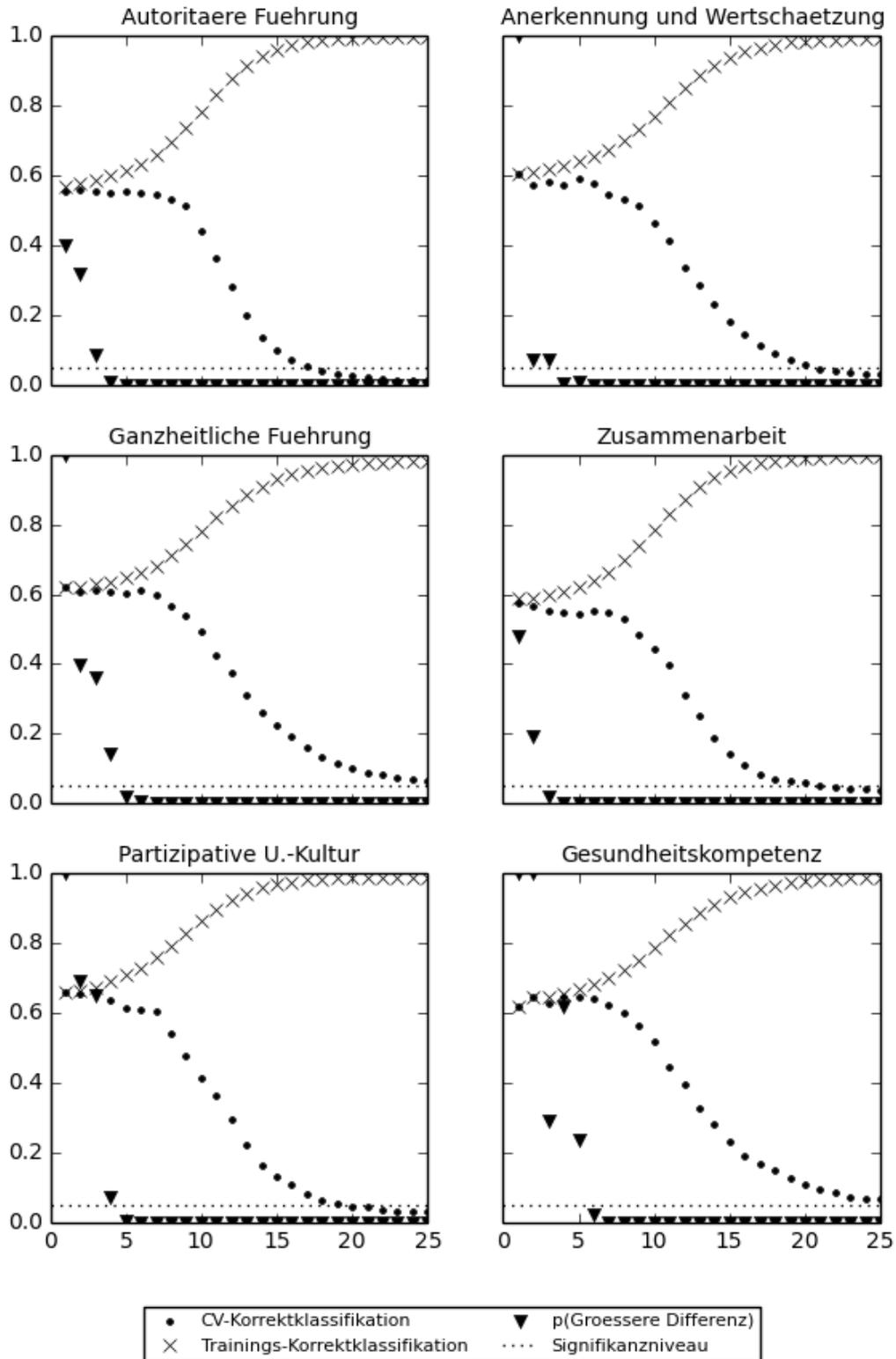


Abbildung 13 (Fortsetzung). CV-T-Diagramme für Bayes-Klassifikatoren auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr miteinander vergleichbar.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, EEB)

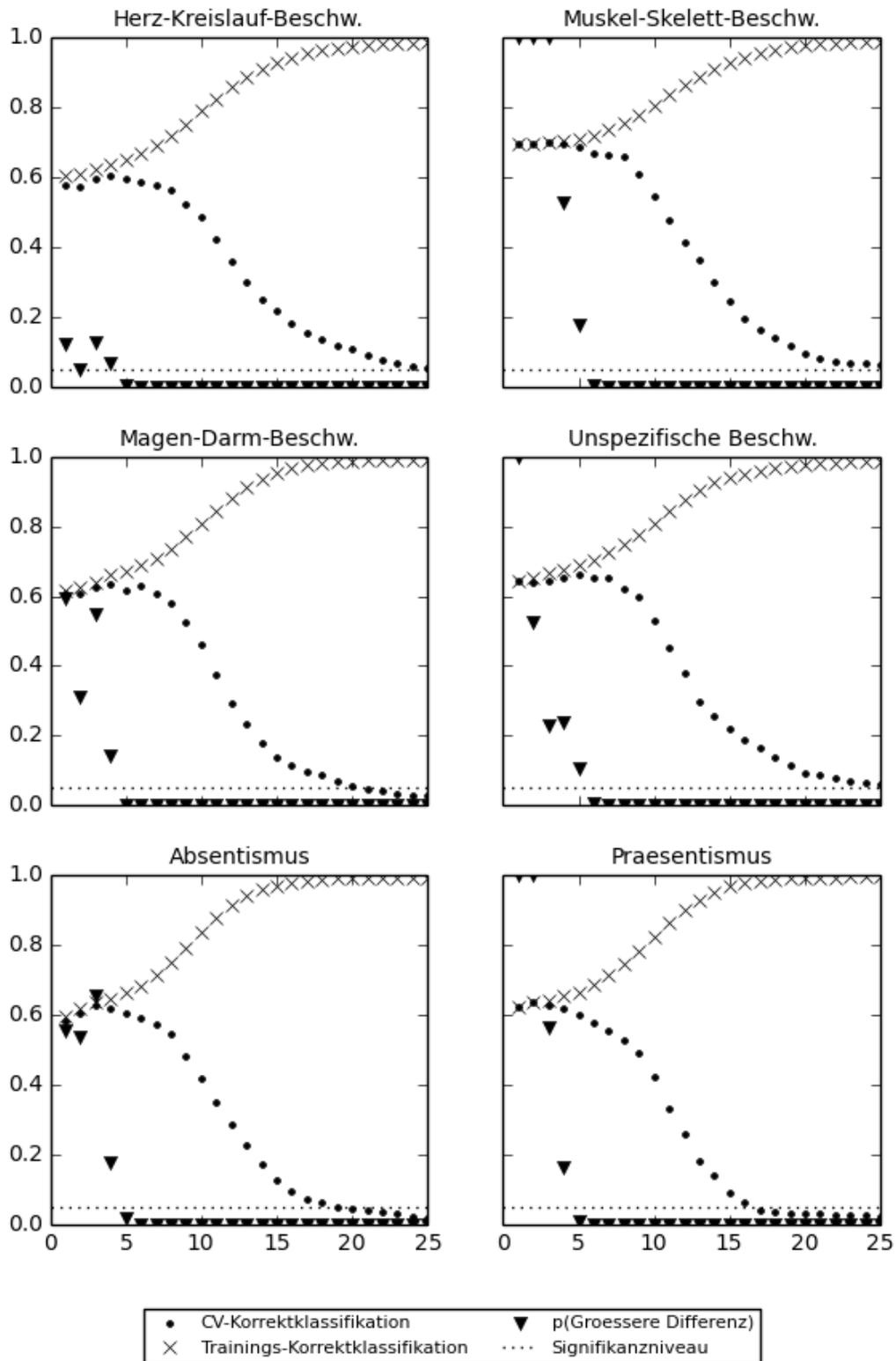


Abbildung 13 (Fortsetzung). CV-T-Diagramme für Bayes-Klassifikatoren auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr miteinander vergleichbar.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, WSIB)

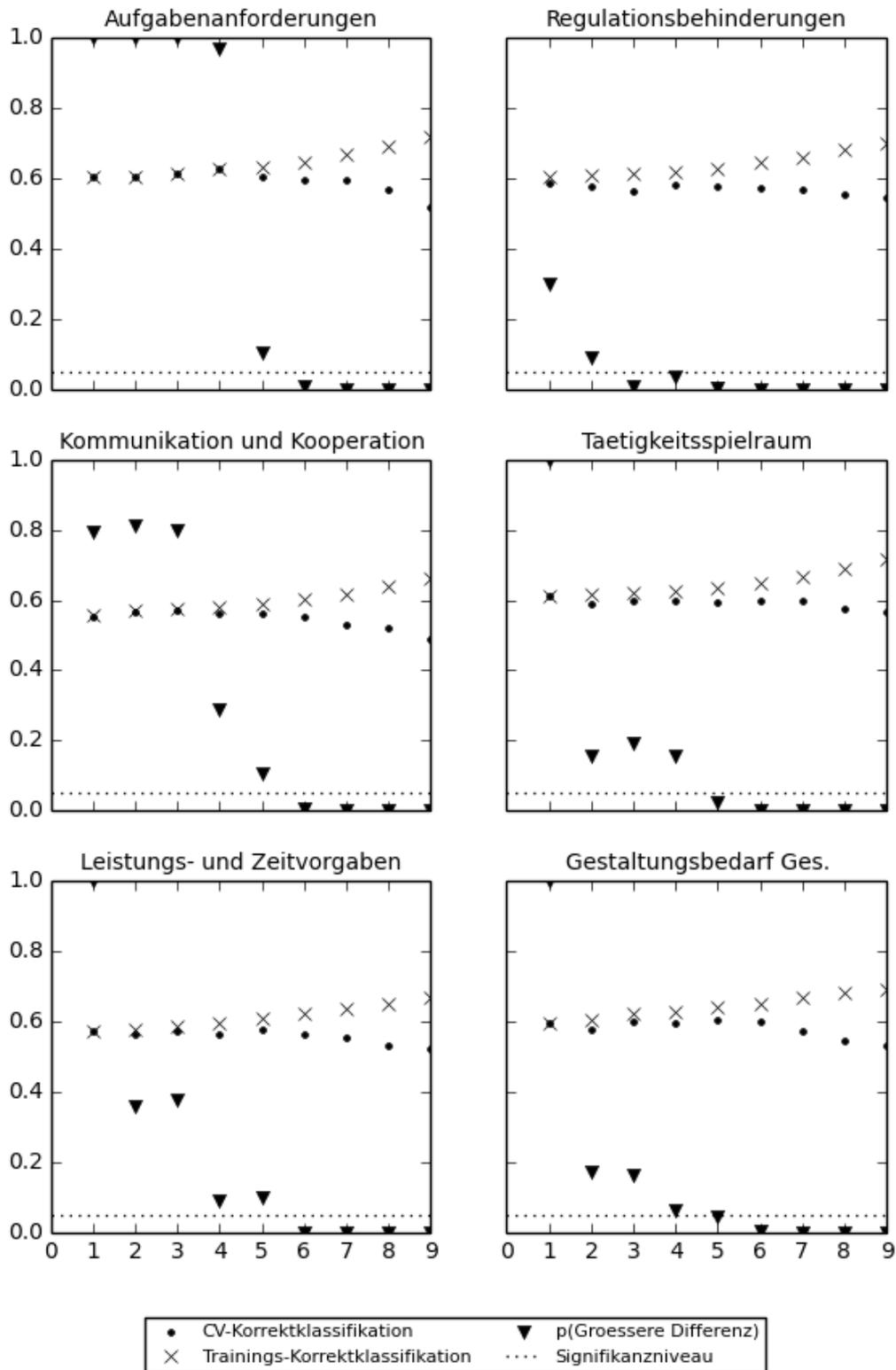


Abbildung 14. CV-T-Diagramme für Bayes-Klassifikatoren auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr mit einander vergleichbar.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, WSIB)

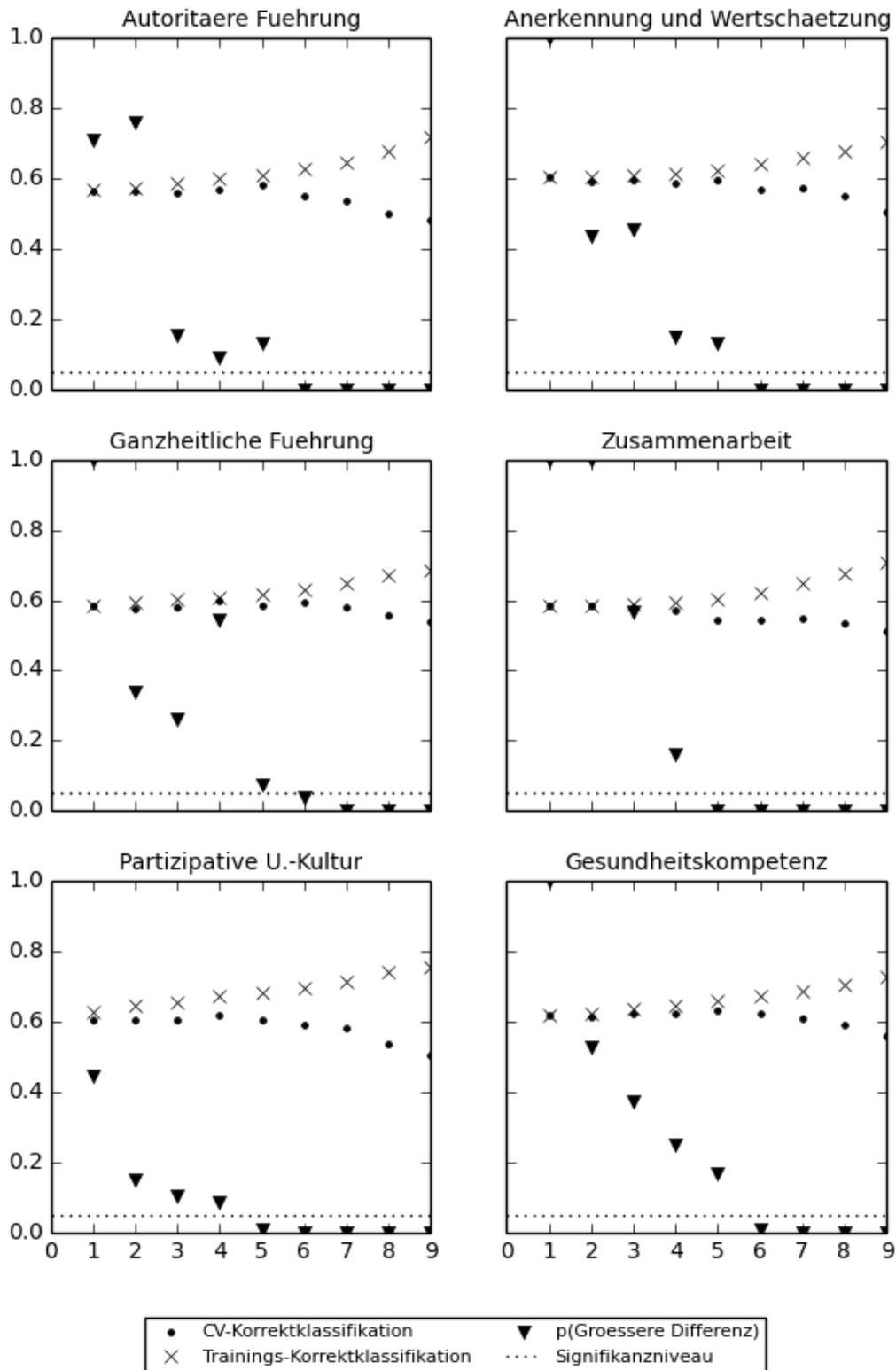


Abbildung 14 (Fortsetzung). CV-T-Diagramme für Bayes-Klassifikatoren auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr miteinander vergleichbar.

CV- und Trainings-Korrektklassifikation
(Bayes, Sequential Forward Selection, WSIB)

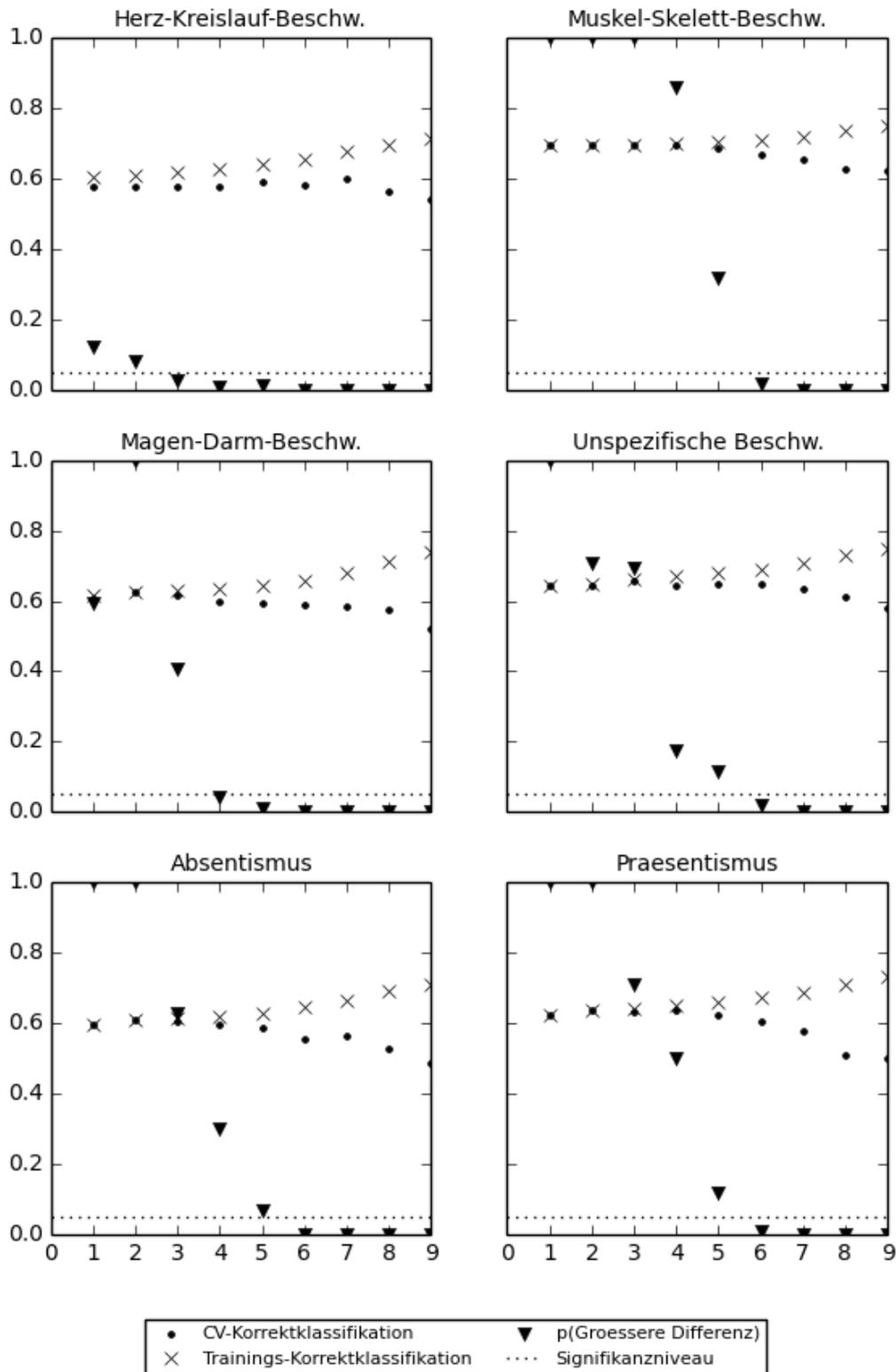


Abbildung 14 (Fortsetzung). CV-T-Diagramme für Bayes-Klassifikatoren auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr mit einander vergleichbar.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, EEB)

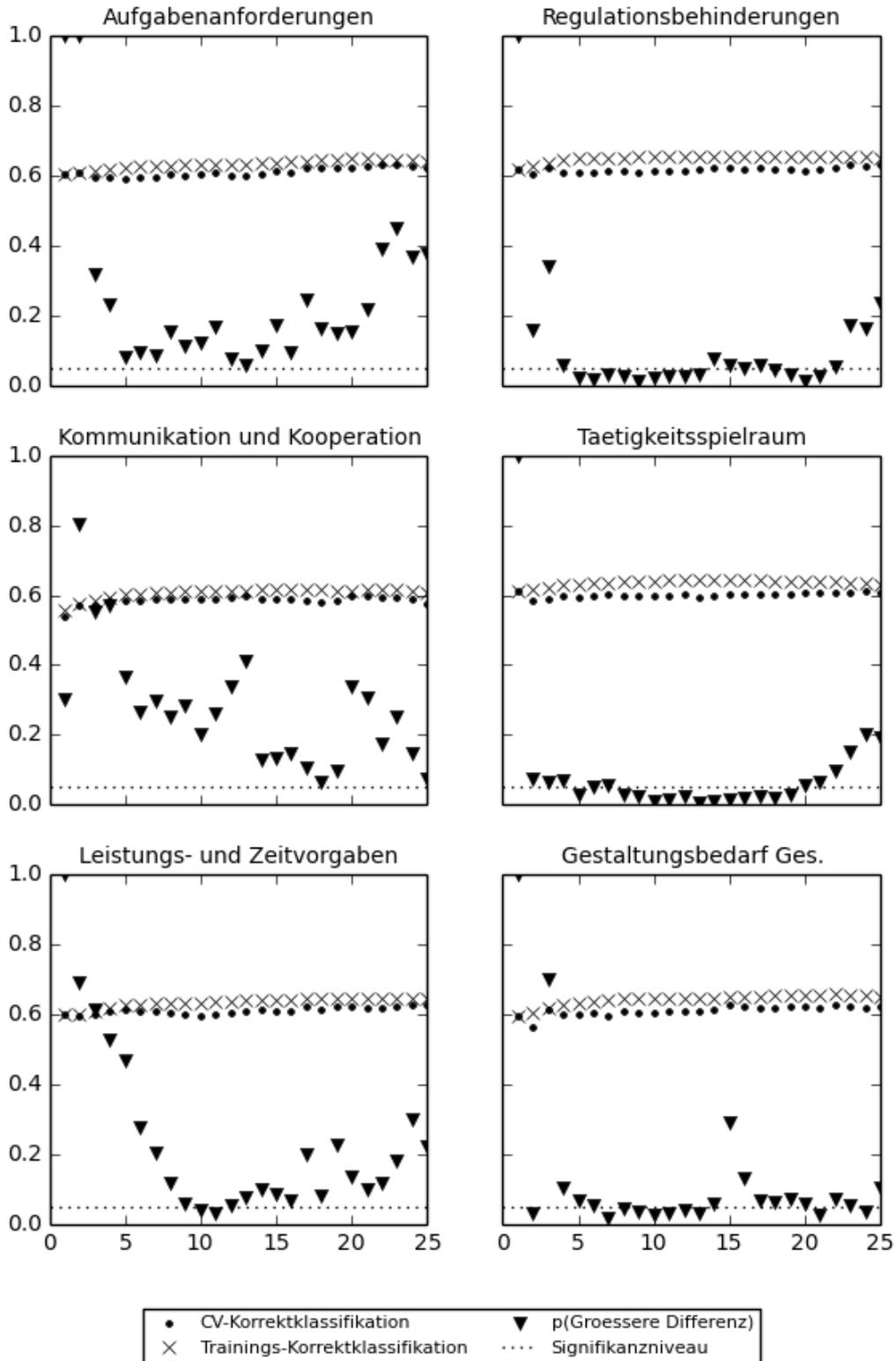


Abbildung 15. CV-T-Diagramme für binärologistische Regression auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind nicht mehr mit einander vergleichbar.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, EEB)

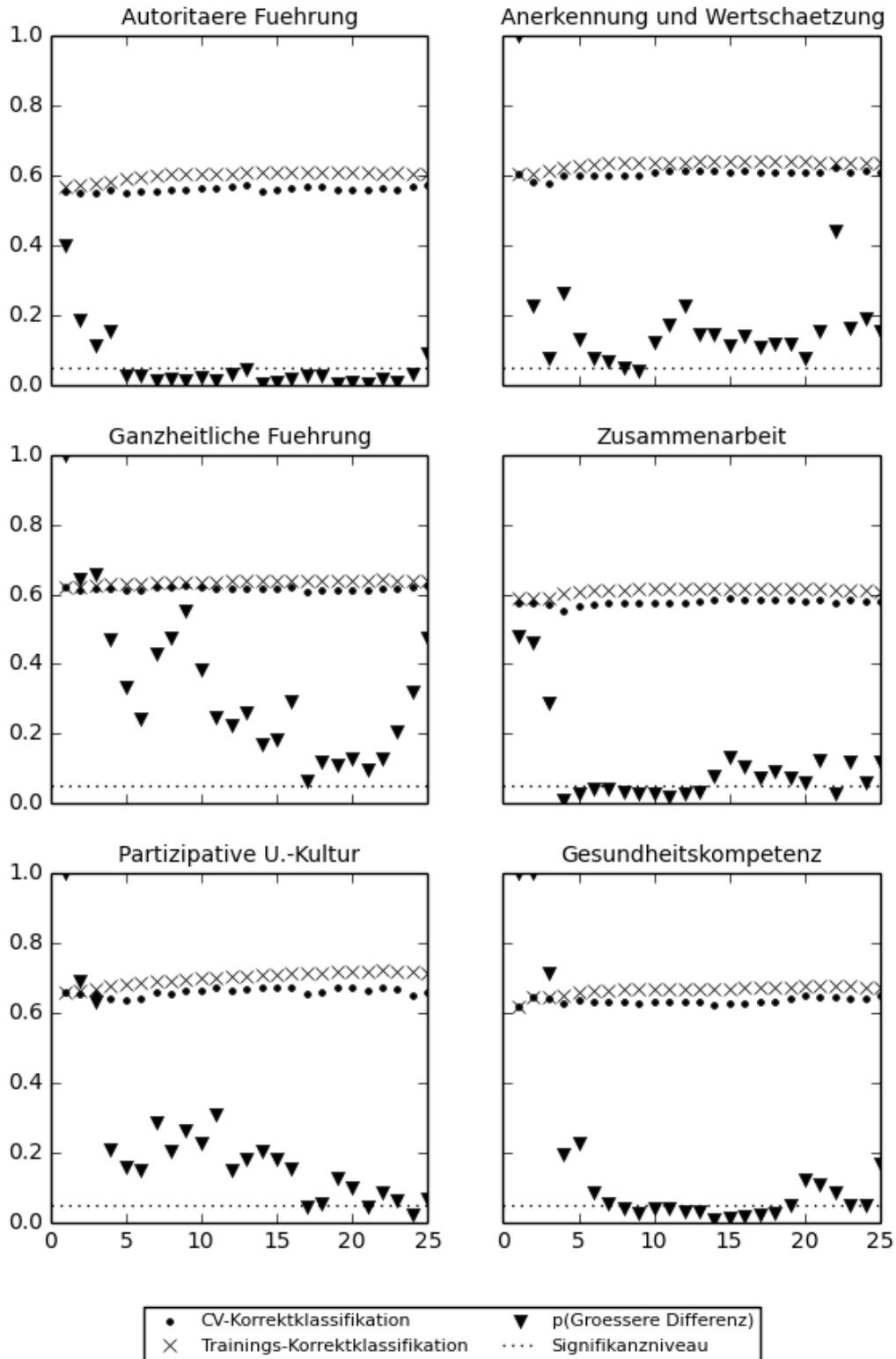


Abbildung 15 (Fortsetzung). CV-T-Diagramme für binärlogistische Regression auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind signifikant unterschiedlich.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, EEB)

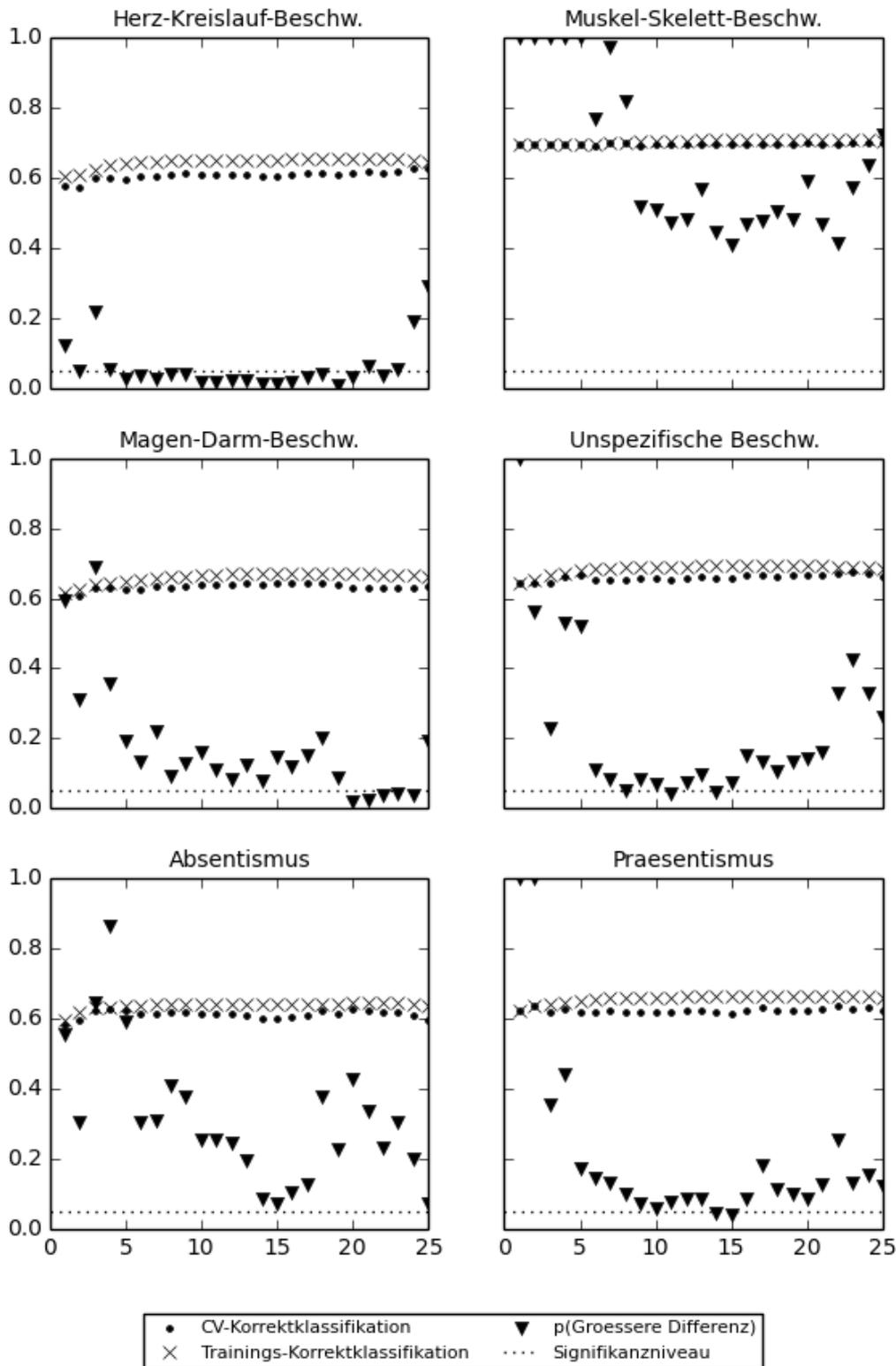


Abbildung 15 (Fortsetzung). CV-T-Diagramme für binärlogistische Regression auf Basis sämtlicher 25 EEB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind signifikant unterschiedlich.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, WSIB)

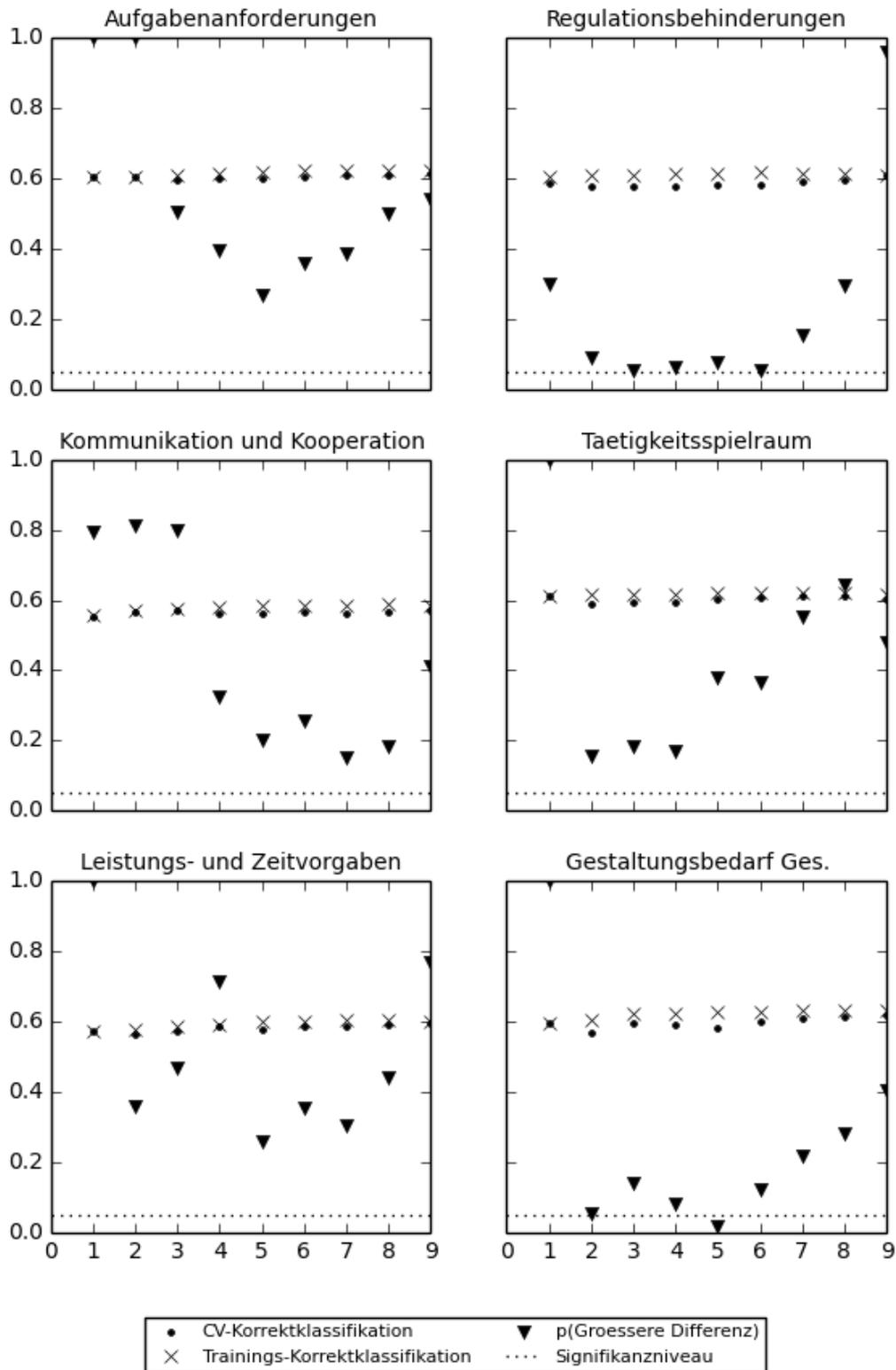


Abbildung 16. CV-T-Diagramme für binärlogistische Regression auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind signifikant unterschiedlich.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, WSIB)

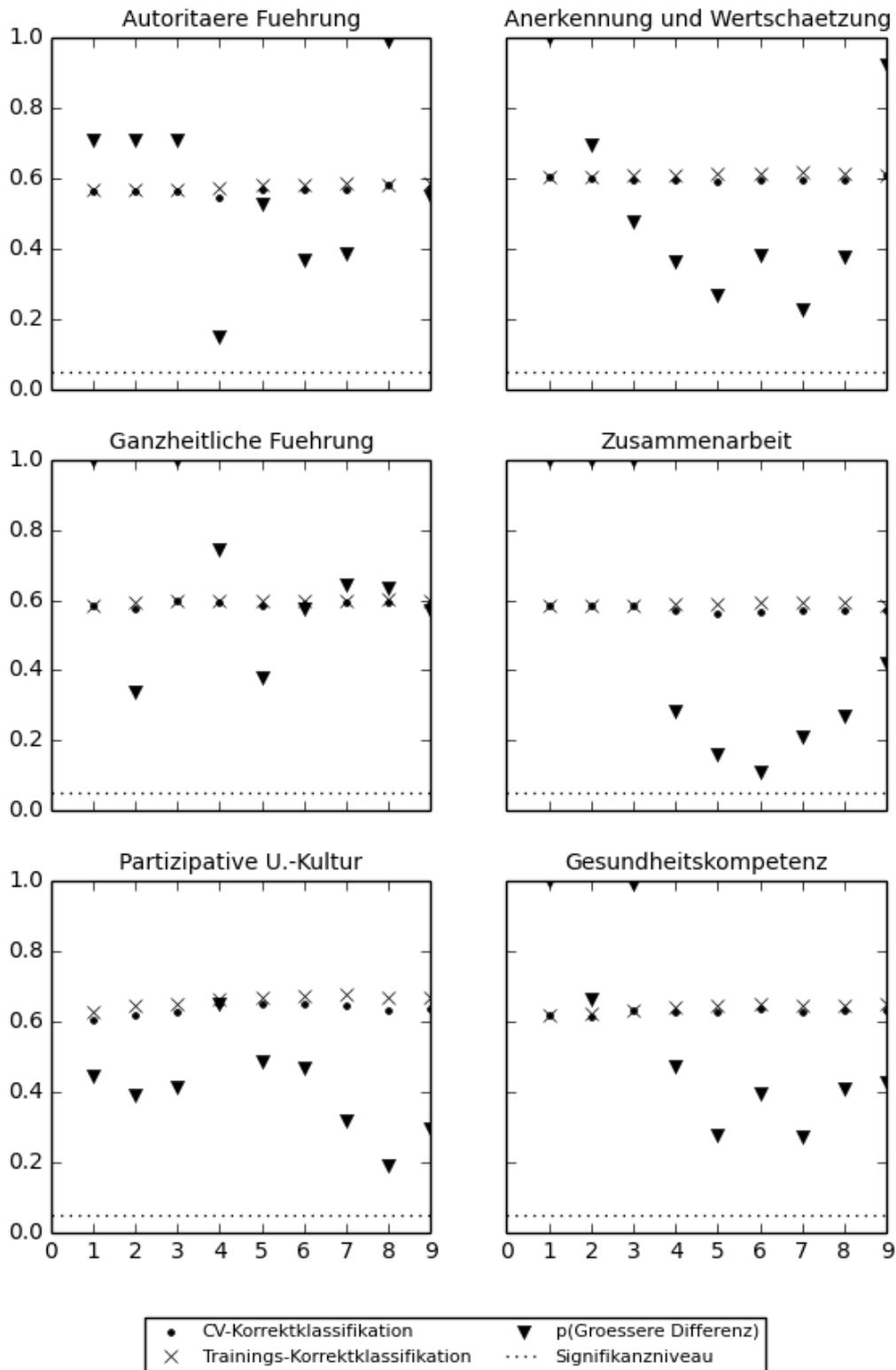


Abbildung 16 (Fortsetzung). CV-T-Diagramme für binärlogistische Regression auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind signifikant unterschiedlich.

CV- und Trainings-Korrektklassifikation
(BLR, Sequential Forward Selection, WSIB)

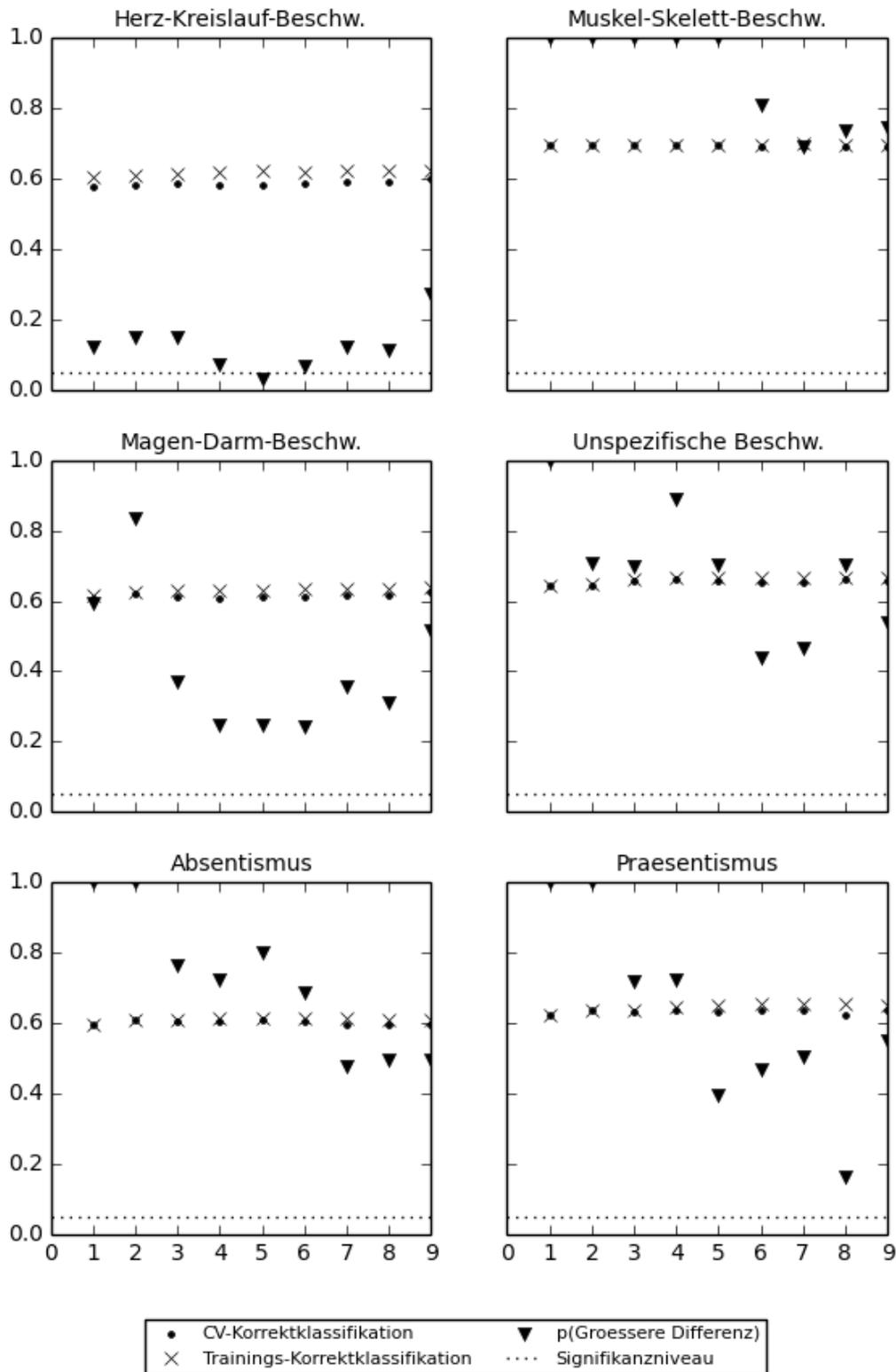


Abbildung 16 (Fortsetzung). CV-T-Diagramme für binärlogistische Regression auf Basis der neun WSIB-Items für jedes der 18 Merkmale. Dargestellt sind Trainings-Korrektklassifikation (Kreuze), Cross-Validation-Korrektklassifikation (Punkte) sowie p-Werte zur statistischen Einschätzung ihrer Differenz nach Wald-Test (Dreiecke) in Abhängigkeit der Messkomplexität. Letztere wurde mittels Sequential Forward Selection gesteigert. Die gepunktete Linie bei Abszissen-Wert .05 stellt das Signifikanzniveau dar. Liegt ein p-Wert darunter, dann ist die entsprechende Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation signifikant, d. h. CV- und Trainings-Korrektklassifikation sind signifikant unterschiedlich.

Die CV-T-Diagramme für Bayes-Klassifikatoren auf Grundlage sämtlicher 25 EEB-Items entsprechen dem Verhalten von Bayes-Klassifikatoren, das Hughes (1968) beschreibt: Die Trainings-Korrektklassifikation steigt mit der Messkomplexität in Richtung sicherer Klassifikation (optimale Anpassung an die Trainingsdaten), während die Cross-Validation-Korrektklassifikation nach einem Maximum stark abfällt. Die Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation wird ab einer Messkomplexität mit vier bis fünf EEB-Items signifikant. Auffällig ist, dass bereits mit dem ersten selegierten Item eine Korrektklassifikation erreicht wird, die mit dem jeweiligen Maximum vergleichbar ist (Abbildung 13). Betrachtet man Bayes-Klassifikatoren auf Grundlage der neun WSIB-Items, dann ist dasselbe Verhalten in der Tendenz erkennbar, jedoch wegen der geringeren Zahl zur Verfügung stehender Items nicht in derselben Deutlichkeit zu identifizieren (Abbildung 14). Die Analyse der CV-T-Diagramme für Bayes-Klassifikatoren ist Grundlage für die Festlegung von Abbruchbedingungen bei heuristischer Merkmals-Selektion. Fällt die Differenz zwischen Trainings- und Cross-Validation-Korrektklassifikation erstmalig signifikant aus, wird die Klassifikator-Induktion abgebrochen. Ergebnis der Klassifikator-Induktion ist derjenige Klassifikator mit der bis dahin größten Cross-Validation-Korrektklassifikation.

Das Verhalten binärlogistischer Regression unterscheidet sich qualitativ stark von jenem der Bayes-Klassifikation. Auch hier erreicht der Klassifikator mit dem ersten selegierten Item bereits eine Korrektklassifikation, die mit dem Maximum vergleichbar ist. Anschließend kommt es mit jeder weiteren Steigerung der Messkomplexität zu leichten Verbesserungen oder der Stagnation von Trainings- und Cross-Validation-Korrektklassifikation. Die Trainings-Korrektklassifikation ist in den meisten Fällen geringfügig größer, als die Cross-Validation-Korrektklassifikation. Signifikante Differenz zwischen den beiden Korrektklassifikationen wird nur in sechs Fällen bei niedriger Messkomplexität (nach maximal fünf Items) erreicht. Bei Muskel-Skelett-Beschwerden wird sie sogar gar nicht erreicht (Abbildung 15 und Abbildung 16). Ungeachtet dessen wird zum Zwecke der Vergleichbarkeit auch bei binärlogistischer Regression an oben beschriebener Abbruch-Bedingung festgehalten. Insbesondere, da das Maximum der Cross-Validation-Korrektklassifikation oftmals schon bei vergleichsweise geringer Messkomplexität erreicht wird.

6.3 ERGEBNISSE DER KLASSIFIKATOR-INDUKTION

Im vorliegenden Abschnitt werden die Ergebnisse der Klassifikator-Induktion sowie die endgültig ausgewählten und kalibrierten Klassifikatoren dargestellt. Dies erfolgt in identischer Weise für jedes der 18 vorherzusagenden Kriterien. In Tabelle 3 sind diese Kriterien, unterteilt

Ergebnisse

nach Arbeitsaufgabe und -situation, Führung und Zusammenarbeit sowie Gesundheitsindikatoren, aufgeführt. Die Darstellung der Ergebnisse zur Klassifikator-Induktion erfolgt in dieser Reihenfolge.

Tabelle 3. Vorherzusagende Kriterien, unterteilt nach Arbeitsaufgabe und -situation, Führung und Zusammenarbeit sowie Gesundheitsindikatoren.

Vorherzusagende Kriterien
<i>Arbeitsaufgabe und -situation</i>
Arbeitsanforderungen (ab S. 172)
Kommunikation und Kooperation (ab S. 184)
Leistungs- und Zeitvorgaben (ab S. 188)
Regulationsbehinderungen (ab S. 192)
Tätigkeitsspielraum (ab S. 196)
Gesamtbedarf für Arbeitsgestaltung (ab S. 200)
<i>Führung und Zusammenarbeit</i>
Autoritäre Führung (ab S. 204)
Bedeutung von Anerkennung und Wertschätzung (ab S. 208)
Ganzheitliche Führung (ab S. 212)
Partizipative Unternehmenskultur (ab S. 216)
Zusammenarbeit (ab S. 216)
<i>Gesundheitsindikatoren</i>
Gesundheitsbezogene Selbstwirksamkeit (Gesundheitskompetenz) (ab S. 224)
Herz-Kreislauf-Beschwerden (ab S. 228)
Magen-Darm-Beschwerden (ab S. 232)
Muskel-Skelett-Beschwerden (ab S. 236)
Unspezifische Beschwerden (ab S. 240)
Absentismustage (krankheitsbedingte Fehltage) (ab S. 244)
Präsentismustage (Anwesenheit bei der Arbeit trotz Krankheit) (ab S. 248)

6.3.1 Aufbau der Ergebnis-Berichte und Erläuterung verwendeter Kennwerte

Zu jedem der genannten 18 Kriterien liegt ein zwölf-seitiger Bericht im Anhang A3 vor, der die jeweiligen Ergebnisse zur Klassifikator-Induktion wiedergibt. In diesem Haupttext ist ein solcher Bericht vollständig nur beispielhaft für Aufgabenanforderungen angegeben. Für die weiteren Kriterien beschränken sich hier die Ergebnisberichte auf die wesentlichen Elemente: Deskriptiv-statistische Ergebnisse sowie die Darstellung endgültig ausgewählter

Klassifikatoren. Ein Bericht (sowohl in diesem Haupttext als auch im Anhang A3) beginnt mit Erläuterungen und deskriptiven Kennwerten des jeweiligen Kriteriums. Als Kennwerte werden angegeben: Stichprobenumfang (N), Modus, Median, arithmetisches Mittel (M), Standardabweichung (SD), Minimum (Min), Maximum (Max) sowie die beiden Klassenmittelwerte μ_0 und μ_1 . Als weiteres deskriptives Ergebnis werden die Spearman-Rang-Korrelationen (ρ) zwischen dem Kriterium und den einzelnen EEB-Items angegeben. Diese werden gemäß Cohen (1992) hinsichtlich ihrer praktischen Relevanz bewertet. D. h. es werden folgende Formulierungen angewandt: Praktisch nicht relevant für $|\rho| < .1$, geringer Effekt für $.1 \leq |\rho| < .3$, mittlerer Effekt für $.3 \leq |\rho| < .5$ und starker Effekt für $.5 \leq |\rho|$.

Auf die Beschreibung des Kriteriums folgen die Ergebnisse der ersten Reihe von Klassifikations-Experimenten (im Haupttext nur für Aufgabenanforderungen, für die weiteren Kriterien im Anhang A3). Je Cross-Validation-Iteration werden Trainings-Korrektklassifikation (κ_{Training}) sowie die ausgewählten Items in einer Tabelle zusammengefasst. In einer zweiten Tabelle sind jeweils die Cross-Validation-Ergebnisse zusammengetragen. Zentrales Ergebnis in dieser Tabelle ist die Cross-Validation-Korrektklassifikation (κ_{CV}). Diese wird zunächst mit der Korrektklassifikation bei konstanter Zuordnung ($\kappa_{\text{Konst.}}$) verglichen. Hierzu werden Erwartungswert ($\kappa_{\text{CV}} - \kappa_{\text{Konst.}}$), Standardabweichung ($SD[\kappa_{\text{CV}} - \kappa_{\text{Konst.}}]$), 95 %-Konfidenzintervall (95 %-CI $[\kappa_{\text{CV}} - \kappa_{\text{Konst.}}]$) sowie p -Wert nach einseitigem Wald-Test ($p[\text{Stichprobe } S: \kappa_{\text{Konst.}} > \kappa_{\text{CV}}]$) angegeben. Ein weiterer Vergleich erfolgt mit der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (κ_{Hughes}). Auch hier werden Erwartungswert, Standardabweichung und p -Wert angegeben. Die Stabilität der selektierten Prädiktor-Teilmengen wird als zusätzliche Einschätzung der Spezifikation des Klassifikator-Induktors angegeben. Der Vergleich zwischen mittlerer Trainings-Korrektklassifikation und Cross-Validation-Korrektklassifikation erfolgt mittels χ^2 -Test. Hierzu werden mittlere Trainings-Korrektklassifikation (κ_{Training}), Erwartungswert ($|\kappa_{\text{CV}} - \kappa_{\text{Training}}|$), Standardabweichung ($SD[|\kappa_{\text{CV}} - \kappa_{\text{Training}}|]$) sowie p -Wert nach zweiseitigem χ^2 -Test ($p[\text{größere Differenz}]$) angegeben. Den Abschluss der Cross-Validation-Ergebnisse stellt die Auswertung der Vier-Felder-Tafel dar. Neben der Vier-Felder-Tafel inklusive Randsummen werden ein Zusammenhangmaß (Cramers ϕ), geschätzter χ^2 -Wert ($X^2(1)$) sowie der entsprechende p -Wert ($p[\chi^2(1) > X^2(1)]$) angegeben. Die Spalten der beiden Tabellen beziehen sich auf das beste Einzelitem, das beste Paar sowie das beste Tripel nach vollständiger Suche. Die genannten beiden Tabellen werden einmal für die vollständige EEB und einmal für das WSIB dargestellt.

Die Ergebnisse der zweiten Experimental-Reihe zur Klassifikator-Induktion werden in gleicher Weise dargestellt. Lediglich auf den Vergleich mit der Korrektklassifikation nach Hughes (1968) wird verzichtet, da sich dieser Wert auf Grund verschieden umfangreicher ausgewählter Prädiktor-Teilmengen je Cross-Validation-Iteration nicht mehr eindeutig angeben lässt. Die Spalten der entsprechenden Tabellen beziehen sich nun auf binärlogistische Regression bei Sequential Forward Selection (BLR/SFS), Bayes-Klassifikation bei Sequential Forward Selection (Bayes/SFS) sowie Bayes-Klassifikation bei Sequential Forward Floating Selection (Bayes/SFFS).

Im Anschluss an die Darstellung der Experimental-Reihen zur Klassifikator-Induktion werden die Ergebnisse der Bootstrapping-Verfahren dargestellt (Haupttext und Anhang). Zu Beginn werden mittels Bootstrapping die Klassifikatoren für die beiden Instrumente (vollständige EEB und WSIB) endgültig ausgewählt. In einer Tabelle werden sämtliche Klassifikatoren durch die jeweils ausgewählten Items repräsentiert. Die mittlere Bootstrapping-Korrektklassifikation ($\kappa_{\text{Bootstrap}}$) sowie die zugehörige Standardabweichung ($SD[\kappa_{\text{Bootstrap}}]$) werden aufgeführt. Die Differenz dieser Korrektklassifikation zur Korrektklassifikation $\kappa_{\text{Hughes}} = .75$ bei unendlich vielen Messungen nach Hughes (1968) wird ebenfalls gelistet. Maßgabe für die endgültige Auswahl eines Klassifikators ist eine möglichst geringe Differenz, wobei $\kappa_{\text{Bootstrap}} \leq \kappa_{\text{Hughes}}$ gelten soll, um Über- oder Unteranpassungen an die Trainingsdaten zu vermeiden. Neben diesen Kennwerten sind je Instrument die Verfahren der Merkmals-Selektion aufgeführt, mit denen die jeweiligen Items ausgewählt wurden. Die endgültig ausgewählten Klassifikatoren sind in der Tabelle hervorgehoben. In der darauf folgenden Tabelle werden die Kennwerte der endgültig ausgewählten Klassifikatoren im Detail dargestellt. Zunächst werden die Items je Instrument genannt, anschließend wird die Bootstrapping-Korrektklassifikation, die zugehörige Standardabweichung sowie das 95 %-Konfidenzintervall ($95\% \text{-CI}[\kappa_{\text{Bootstrap}}]$) aufgeführt. Wurden für die beiden Instrumente (vollständige EEB und WSIB) unterschiedliche Klassifikatoren ausgewählt, dann erfolgt ein Vergleich der beiden Klassifikatoren. Die Differenz ihrer Korrektklassifikationen ($\Delta\kappa_{\text{Bootstrap}}$), die zugehörige Standardabweichung ($SD[\Delta\kappa_{\text{Bootstrap}}]$), das 95 %-Konfidenzintervall ($95\% \text{-CI}[\Delta\kappa_{\text{Bootstrap}}]$) sowie der p -Wert nach zweiseitigem Wald-Test ($p[\Delta\kappa_{\text{Bootstrap}} < 0]$) werden für diesen Vergleich angegeben. Inwiefern sich die Bootstrapping-Korrektklassifikation von der Korrektklassifikation nach Hughes (1968) unterscheidet, wird durch den Erwartungswert ($\kappa_{\text{Hughes}} - \kappa_{\text{Bootstrap}}$), die Standardabweichung ($SD[\kappa_{\text{Hughes}} - \kappa_{\text{Bootstrap}}]$), das 95 %-Konfidenzintervall ($95\% \text{-CI}[\kappa_{\text{Hughes}} - \kappa_{\text{Bootstrap}}]$) sowie den p -Wert

nach einseitigem Wald-Test ($p[\kappa_{\text{Hughes}} < \kappa_{\text{Bootstrap}}]$) angegeben. Wie bei der Darstellung der Experimental-Reihen zur Klassifikator-Induktion enthält auch diese Tabelle eine Auswertung der Vier-Felder-Tafel. Diese wird ergänzt um den Korrekturfaktor c zum Ausgleich der Überrepräsentation einer der beiden Klassen bei der Vorhersage (siehe die Ausführungen zur Kalibrierung der Klassifikatoren).

Den Abschluss eines Ergebnisberichts stellt eine Tabelle, die Ergebnisse der Kalibrierung der endgültig ausgewählten Klassifikatoren enthält (Haupttext und Anhang). Die ausgewählten Items werden nochmals genannt. Anschließend folgen mittlerer Fehler bei der Vorhersage der Wahrscheinlichkeit für Klasse 1 in einer Stichprobe ($\langle \varepsilon(\pi_1) \rangle$) sowie des vorhergesagten Erwartungswertes ($\langle \varepsilon(\mu) \rangle$) und die zugehörigen Varianzen ($\sigma^2(\varepsilon(\pi_1))$ und $\sigma^2(\varepsilon(\mu))$). Außerdem wird angegeben, mit welcher Korrektorklassifikation ($\kappa_{\text{Bootstrap}}$) vorhergesagt werden kann, ob eine Stichprobe über- oder unterdurchschnittlich viele Personen enthält, die Klasse 1 zugeordnet sind. Diese Bootstrap-Korrektklassifikation für Stichproben wird auch mit Hilfe einer entsprechenden Vier-Felder-Tafel ausgewertet. Hierbei kommen die bereits genannten Kennwerte zum Einsatz.

6.3.2 Aufgabenanforderungen

Aufgabenanforderungen umfassen Gestaltungsmerkmale der Arbeit, die Leistungen auf intellektueller, wissensbasierter und sensumotorischer Ebene beanspruchen. Sie geben Aufschluss, inwiefern eine Arbeitsaufgabe hierarchisch vollständig und vielfältig ist (Hacker & Sachse, 2014; Ulich, 2011). Die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999) berücksichtigt Aufgabenanforderungen mit vier Items (z. B. „Sie haben schwierige Aufgaben zu bearbeiten, die eine hohe Konzentration und Genauigkeit erfordern“). Diese sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 4 = "Trifft vollständig zu") zu bewerten. Nach Rekodierung auf eine dreistufige Bewertung (0 = geringer, 1 = mittlerer und 2 = hoher Gestaltungsbedarf) wird der Skalenwert durch Mittelwertbildung berechnet. In Tabelle 4 sind die deskriptiv-statistischen Kennwerte dieser Skala zusammengestellt.

Tabelle 4. Deskriptiv-statistische Kennwerte der Variablen Aufgabenanforderungen.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1544	1,00	1,00	1,07	0,51	0,00	2,00	0,66	1,48

Modus und Median bzw. Mittelwert (M) sind mit 1,00 bzw. 1,07 annähernd identisch. Die Stichprobe mit gültigem Skalenwert für Aufgabenanforderungen wurden für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 5 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Skalenwerten für Aufgabenanforderungen aufgelistet.

Tabelle 5. Korrelationen (Spearman ρ) der Variablen Aufgabenanforderungen mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.317*	2	aggressiv	-.025	3	nervös ^(a)	-.082
4	körperlich verspannt ^(a)	-.056	5	selbstsicher	-.156*	6	zornig	.021
7	aufmerksam ^(a)	-.273*	8	leistungsbereit ^(a)	-.243*	9	entspannt	-.101*
10	körperlich unwohl ^(a)	.072	11	energiegeladen ^(a)	-.133*	12	aufgeregt ^(a)	-.087
13	gelangweilt	.186*	14	beschwingt	-.069	15	fröhlich	-.005
16	ärgerlich	-.011	17	zufrieden	-.109*	18	ängstlich	-.047
19	lustig	.058	20	leicht	-.145*	21	deprimiert	-.001
22	bedrückt	-.012	23	einflussreich ^(a)	-.169*	24	vergnügt	.018
25	betrübt	-.002						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Das EEB-Item „konzentriert“ weist mit einer Korrelation von $-.317$ einen mittleren Effekt ($.3 < |\rho| \leq .5$) in Bezug auf Gestaltungsbedarf bei Aufgabenanforderungen auf, gefolgt von „aufmerksam“ mit $-.273$ (kleiner Effekt, $.1 < |\rho| \leq .3$). Die stärkste positive Korrelation mit

Ergebnisse

Gestaltungsbedarf bezüglich Aufgabenanforderungen besteht mit dem Item „gelangweilt“. Die Korrelation von .186 ist ebenfalls als kleiner Effekt zu interpretieren.

Tabelle 6. Selegierte Eigenschaftswörter der vollständigen EEB bei drei Bayes-Selektions-Experimenten (bestes Item, bestes Paar und bestes Tripel) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.

CV-Iteration	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter der vollständigen EEB		
	Bestes Item	Bestes Paar	Bestes Tripel
0	1	1, 23	1, 9, 11
	K _{Training} .606	.632	.683
1	1	1, 23	9, 11, 15
	K _{Training} .607	.631	.681
2	1	1, 12	9, 11, 14
	K _{Training} .600	.623	.680
3	1	1, 23	1, 9, 11
	K _{Training} .609	.626	.680
4	1	1, 12	11, 14, 20
	K _{Training} .601	.623	.677
5	1	1, 23	9, 11, 23
	K _{Training} .609	.627	.683
6	1	1, 23	9, 11, 23
	K _{Training} .607	.627	.677
7	1	1, 23	1, 9, 23
	K _{Training} .604	.634	.681
8	1	1, 23	9, 11, 15
	K _{Training} .599	.622	.677
9	1	1, 23	1, 9, 23
	K _{Training} .607	.632	.684

Stehen sämtliche EEB-Items zur Selektion zur Verfügung, so wird bei den kombinatorischen Experimenten (Bestes Einzelitem, Bestes Paar und Bestes Tripel) das Item 1 („konzentriert“) durchgehend als bestes Einzelitem gewählt. Für das beste Paar kommt bis auf zwei Ausnahmen das Item 23 („einflussreich“) hinzu. Bei der Auswahl des besten Tripels kommt es zu mehr Varianz. Hier stehen die Items 9 („entspannt“) und 11 („energiegeladen“) im Vordergrund (vgl. Tabelle 6). Die Stabilitäten dieser Kombinationen sind vergleichsweise hoch (vgl. auch Tabelle 7). Eine steigende Anzahl ausgewählte Items geht erwartungsgemäß mit einer steigenden Trainings-Korrektklassifikation einher.

Ergebnisse

Tabelle 7. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Bayes-Selektions-Experimente (bestes Item, bestes Paar, bestes Tripel). Geprüft wird die Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) nach Selektion von Eigenschaftswörtern der vollständigen EEB als Prädiktoren.

		Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter der vollständigen EEB								
		Bestes Item			Bestes Paar			Bestes Tripel		
		<i>Korrektklassifikation gegenüber Konstante</i>								
K _{CV}		.605			.610			.518		
K _{CV} - K _{Konst}		.105			.110			.018		
SD[K _{CV} - K _{Konst}]		.018			.018			.018		
95 %-CI[K _{CV} - K _{Konst}]		[.070, .140]			[.075, .145]			[-.018, .054]		
p[Stichprobe S: K _{Konst} > K _{CV}]		< .001			< .001			.162		
		<i>Korrektklassifikation gegenüber Hughes-Schätzung</i>								
K _{Hughes}		.730			.740			.704		
K _{Hughes} - K _{CV}		.125			.130			.186		
SD[K _{Hughes} - K _{CV}]		.017			.017			.017		
p[Klassifikator: K _{Hughes} < K _{CV}]		< .001			< .001			< .001		
		<i>Stabilität der Menge selezierter Eigenschaftswörter</i>								
Stabilität		1.000			.644			.089		
		<i>Trainings-Test-Drift der Korrektklassifikation</i>								
K _{Training}		.605			.628			.680		
K _{CV} - K _{Training}		.000			.018			.162		
SD[K _{CV} - K _{Training}]		.018			.018			.018		
p[größere Differenz]		.988			.317			< .001		
		<i>Auswertung der Vier-Felder-Tafel</i>								
Cramers ϕ		.223			.221			.036		
$\chi^2(1)$		77,02			74,98			1,90		
p[$\chi^2(1) > \chi^2(1)$]		< .001			< .001			.168		
Vierfeldertafel		1	0	Σ	1	0	Σ	1	0	Σ
(Zeilen: Messwerte,	1	333	439	772	447	321	768	381	363	744
Spalten: Vorhersage)	0	171	600	771	278	491	769	351	386	737
	Σ	504	1039	1543	725	812	1537	732	749	1481

Die beste Cross-Validation-Korrektklassifikation (Anwendung der Vorhersagemodelle auf ungesehene Daten) erzielt das beste Paar. Diese liegt mit .610 überzufällig über dem zu erwartenden Ergebnis von .500 bei konstanter Zuordnung. Das wird auch durch die Vierfeldertafel bestätigt. Die theoretisch bestmögliche Korrektklassifikation wird jedoch mit einer Differenz von .130 signifikant unterschritten. Ebenfalls unterschritten wird die

Ergebnisse

entsprechende Trainings-Korrektklassifikation. Der Unterschied ist jedoch nicht signifikant (vgl. Tabelle 7).

Tabelle 8. Selegierte Eigenschaftswörter des WSIB bei drei Bayes-Selektions-Experimenten (bestes Item, bestes Paar und bestes Tripel) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.

CV-Iteration	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter des WSIB		
	Bestes Item	Bestes Paar	Bestes Tripel
0	1	1, 23	4, 7, 11
	K _{Training} .606	.632	.679
1	1	1, 23	4, 11, 23
	K _{Training} .607	.631	.679
2	1	1, 12	4, 7, 11
	K _{Training} .600	.623	.671
3	1	1, 23	1, 11, 23
	K _{Training} .609	.626	.674
4	1	1, 12	4, 11, 23
	K _{Training} .601	.623	.669
5	1	1, 23	1, 11, 23
	K _{Training} .609	.627	.671
6	1	1, 23	4, 7, 11
	K _{Training} .607	.627	.670
7	1	1, 23	1, 11, 23
	K _{Training} .604	.634	.671
8	1	1, 23	1, 4, 11
	K _{Training} .599	.622	.670
9	1	1, 23	1, 11, 12
	K _{Training} .607	.632	.676

Stehen nur die Items des WSIB für die Selektion zur Verfügung, so ergeben sich für das beste Einzelitem sowie das beste Paar dieselben Kombinationen, wie bei Zugrundelegung der vollständigen EEB. Beim besten Tripel kommt es jedoch zu anderen Kombinationen, da nun z. B. das Item 9 („entspannt“) nicht mehr verfügbar ist (vgl. Tabelle 8). Die Rangfolge der Item-Kombinationen bezüglich der Trainings-Korrektklassifikation entspricht jener bei Verwendung der vollständigen EEB. Die zuvor (vollständige EEB) beobachtete hohe Stabilität zeigt sich entsprechend der ähnlichen Kombinationen auch bei Einschränkung auf das WSIB (Tabelle 9).

Ergebnisse

Tabelle 9. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Bayes-Selektions-Experimente (bestes Item, bestes Paar, bestes Tripel). Geprüft wird die Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) nach Selektion von Eigenschaftswörtern des WSIB als Prädiktoren.

		Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter des WSIB									
		Bestes Item			Bestes Paar			Bestes Tripel			
		<i>Korrektklassifikation gegenüber Konstante</i>									
K _{CV}		.605			.611			.548			
K _{CV} - K _{Konst}		.105			.111			.048			
SD[K _{CV} - K _{Konst}]		.018			.018			.018			
95 %-CI[K _{CV} - K _{Konst}]		[.070, .140]			[.076, .145]			[.013, .084]			
p[Stichprobe S: K _{Konst} > K _{CV}]		< .001			< .001			.004			
		<i>Korrektklassifikation gegenüber Hughes-Schätzung</i>									
K _{Hughes}		.730			.740			.704			
K _{Hughes} - K _{CV}		.125			.129			.156			
SD[K _{Hughes} - K _{CV}]		.017			.017			.017			
p[Klassifikator: K _{Hughes} < K _{CV}]		< .001			< .001			< .001			
		<i>Stabilität der Menge selezierter Eigenschaftswörter</i>									
Stabilität		1.000			.644			.156			
		<i>Trainings-Test-Drift der Korrektklassifikation</i>									
K _{Training}		.605			.628			.673			
K _{CV} - K _{Training}		.000			.017			.124			
SD[K _{CV} - K _{Training}]		.018			.018			.018			
p[größere Differenz]		1.00			.324			< .001			
		<i>Auswertung der Vier-Felder-Tafel</i>									
Cramers ϕ		.224			.221			.097			
$\chi^2(1)$		77,31			75,36			14,00			
p[$\chi^2(1) > \chi^2(1)$]		< .001			< .001			< .001			
Vierfeldertafel											
		1	0	Σ	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,		1	333	439	772	447	321	768	392	352	744
Spalten: Vorhersage)		0	171	601	772	278	492	770	324	429	753
		Σ	504	1040	1544	725	813	1538	716	781	1497

Wie bei der Auswahl aus der vollständigen EEB weist auch hier (WSIB) das beste Paar die beste Cross-Validation-Korrektklassifikation auf. Diese unterscheidet sich stark von der des besten Tripels. Sie unterscheidet sich nicht signifikant von der Trainings-Korrektklassifikation, liegt jedoch mit einer Differenz von .129 signifikant unterhalb der theoretisch bestmöglichen Korrektklassifikation (Tabelle 9).

Ergebnisse

Tabelle 10. Selegierte Eigenschaftswörter der vollständigen EEB bei drei Selektions-Experimenten (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.

CV-Iteration	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter der vollständigen EEB		
	BLR/SFS	Bayes/SFS	Bayes/SFFS
0	1	1, 6	1
κ_{Training}	.607	.609	.607
1	1, 2, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 24, 25	1	1
κ_{Training}	.654	.603	.603
2	1, 2, 3, 4, 5, 6, 7, 9, 12, 13, 16, 18, 19, 20, 21, 22, 23	1	7, 8, 11, 12, 13, 19, 23
κ_{Training}	.646	.600	.661
3	1, 6	1, 6, 7, 12, 16	6, 7, 8, 19
κ_{Training}	.612	.646	.618
4	1, 6	1, 6, 7, 12, 16, 25	6, 7, 8
κ_{Training}	.613	.655	.599
5	1, 2, 5, 6, 10, 12, 13, 15, 21, 22, 25	1, 4, 6, 8, 13, 20	7
κ_{Training}	.627	.651	.576
6	1	1	6, 7, 8
κ_{Training}	.602	.602	.602
7	1, 2, 3, 5, 6, 10, 12, 13, 14, 15, 17, 21, 25	1, 6, 8, 11, 13, 20, 23	1
κ_{Training}	.621	.661	.599
8	1, 6	1, 6, 7, 12, 16	1
κ_{Training}	.610	.643	.604
9	1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25	1	1
κ_{Training}	.660	.613	.613

Werden Experimente zur heuristischen Selektion von Items aus der vollständigen EEB durchgeführt (Sequential Forward Selection [SFS], Sequential Forward Floating Selection [SFFS]), so nimmt die Stabilität gegenüber vollständiger Suche stark ab (Tabelle 10 und Tabelle 11). Insbesondere unter Verwendung binärlogistischer Regression werden z. T. sehr umfangreiche Item-Kombinationen ausgewählt. Item 1 („konzentriert“) fällt als Item auf, das häufig als Einzelitem ausgewählt wird. Auf Grund der Trainings-Korrektklassifikationen lässt sich kein Algorithmus als überlegen identifizieren (Tabelle 10 und Tabelle 11).

Ergebnisse

Tabelle 11. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Selektions-Experimente (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]). Geprüft wird die Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) nach Selektion von Eigenschaftswörtern der vollständigen EEB als Prädiktoren.

	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter der vollständigen EEB									
	BLR/SFS			Bayes/SFS			Bayes/SFFS			
<i>Korrektklassifikation gegenüber Konstante</i>										
K_{CV}	.618			.624			.625			
$K_{CV} - K_{Konst}$.118			.124			.125			
$SD[K_{CV} - K_{Konst}]$.018			.018			.018			
95 %-CI[$K_{CV} - K_{Konst}$]	[.083, .153]			[.089, .159]			[.090, .160]			
$p[\text{Stichprobe S: } K_{Konst} > K_{CV}]$	< .001			< .001			< .001			
<i>Stabilität der Menge selezierter Eigenschaftswörter</i>										
Stabilität	.089			.156			.244			
<i>Trainings-Test-Drift der Korrektklassifikation</i>										
$K_{Training}$.625			.628			.608			
$ K_{CV} - K_{Training} $.007			.004			.017			
$SD[K_{CV} - K_{Training}]$.017			.017			.018			
$p[\text{größere Differenz}]$.683			.798			.333			
<i>Auswertung der Vier-Felder-Tafel</i>										
Cramers ϕ	.241			.257			.254			
$\chi^2(1)$	89,74			102,11			99,73			
$p[\chi^2(1) > \chi^2(1)]$	< .001			< .001			< .001			
Vierfeldertafel		1	0	Σ	1	0	Σ	1	0	Σ
(Zeilen: Messwerte,	1	397	374	771	380	390	770	414	357	771
Spalten: Vorhersage)	0	215	556	771	189	582	771	221	550	771
	Σ	612	930	1542	569	972	1541	635	907	1542

Bayes-Klassifikation führt bei der Cross-Validation-Korrektklassifikation zu besseren Ergebnissen, als binärlogistische Regression. Es finden sich keine signifikanten Unterschiede zu den Trainings-Korrektklassifikationen. Gegenüber vollständiger Suche sind die Cross-Validation-Korrektklassifikationen bei heuristischer Vorgehensweise durchgehend besser. Bei der Vorhersage ist eine Verschiebung hin zur Klasse 0 („Arbeitsanforderungen günstig ausgeprägt“) zu beobachten (Tabelle 11).

Ergebnisse

Tabelle 12. Selegierte Eigenschaftswörter des WSIB bei drei Selektions-Experimenten (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]) zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) und deren Korrektklassifikation (κ) auf Grundlage der Trainingsdaten.

CV-Iteration	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter des WSIB		
	BLR/SFS	Bayes/SFS	Bayes/SFFS
0	1 K _{Training} .607	1, 3, 7, 12 .631	1 .607
1	1 K _{Training} .603	1, 3, 7, 12 .629	1 .603
2	1 K _{Training} .600	1 .600	1 .600
3	1 K _{Training} .607	1, 3, 7 .613	1 .607
4	1 K _{Training} .607	1, 3, 7, 12 .632	4, 7, 10, 11, 12, 23 .635
5	1 K _{Training} .607	1, 3, 7, 12 .628	7 .576
6	1 K _{Training} .602	1, 3, 7, 12 .624	7, 8, 12 .600
7	1 K _{Training} .599	1, 3, 7, 12 .622	1 .599
8	1 K _{Training} .604	1, 3, 4, 7, 10, 11, 12 .661	7, 10, 11, 12 .606
9	1 K _{Training} .613	1 .613	1 .613

Werden die auswählbaren Items auf das WSIB eingeschränkt, dann sinkt entsprechend die Anzahl ausgewählter Items. Dennoch gibt es auch in diesem Falle bei Bayes-Klassifikation eine große Varianz in der Anzahl ausgewählter Items. Bei binärlogistischer Regression wird hingegen durchgehend Item 1 („konzentriert“) als Einzelitem gewählt. Die Trainings-Korrektklassifikationen sind etwas geringer als jene, die bei Zugrundelegung der vollständigen EEB erzielt werden konnten. Hinsichtlich der Trainings-Korrektklassifikationen hebt sich Bayes-Klassifikation (SFS) mit tendenziell höheren Werten hervor (Tabelle 10 und Tabelle 13).

Ergebnisse

Tabelle 13. Verschiedene Kennwerte zur Beurteilung der Klassifikations-Güte von Modellen auf Grundlage dreier Selektions-Experimente (Binärlogistische Regression, Sequential Forward Floating Selection [BLR/SFS], Bayes-Klassifikation, SFS [Bayes/SFS] und Bayes-Klassifikation, Sequential Forward Floating Selection [Bayes/SFFS]). Geprüft wird die Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) nach Selektion von Eigenschaftswörtern des WSIB als Prädiktoren.

	Experimente zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig), Eigenschaftswörter des WSIB									
	BLR/SFS			Bayes/SFS			Bayes/SFFS			
<i>Korrektklassifikation gegenüber Konstante</i>										
K _{CV}	.605			.636			.619			
K _{CV} - K _{Konst}	.105			.136			.119			
SD[K _{CV} - K _{Konst}]	.018			.018			.018			
95 %-CI[K _{CV} - K _{Konst}]	[.070, .140]			[.101, .170]			[.084, .153]			
p[Stichprobe S: K _{Konst} > K _{CV}]	< .001			< .001			< .001			
<i>Stabilität der Menge selezierter Eigenschaftswörter</i>										
Stabilität	1.000			.356			.333			
<i>Trainings-Test-Drift der Korrektklassifikation</i>										
K _{Training}	.605			.625			.605			
K _{CV} - K _{Training}	.000			.010			.014			
SD[K _{CV} - K _{Training}]	.018			.017			.018			
p[größere Differenz]	1.00			.551			.428			
<i>Auswertung der Vier-Felder-Tafel</i>										
Cramers ϕ	.224			.273			.241			
$\chi^2(1)$	77,35			114,90			89,83			
p[$\chi^2(1) > \chi^2(1)$]	< .001			< .001			< .001			
Vierfeldertafel		1	0	Σ	1	0	Σ	1	0	Σ
(Zeilen: Messwerte, Spalten: Vorhersage)	1	333	438	771	457	314	771	407	364	771
	0	171	600	771	247	523	770	224	547	771
	Σ	504	1038	1542	704	837	1541	631	911	1542

Bayes-Klassifikation mit Sequential Forward Selection (SFS) führt deutlich zur besten Cross-Validation-Korrektklassifikation, und ist somit den beiden anderen Algorithmen überlegen. Hier werden die Korrektklassifikation bei konstanter Zuordnung (signifikant) sowie die Trainings-Korrektklassifikation (nicht signifikant) überschritten. Dies wird auch durch die Vierfeldertafel bestätigt. Allerdings ist bei der Vorhersage eine Verschiebung hin zur Klasse 0 („Aufgabenanforderungen günstig ausgeprägt“) zu beobachten. Bayes-Klassifikation (SFFS) und binärlogistische Regression weisen eine vergleichbar hohe Cross-Validation-Korrektklassifikation auf (Tabelle 13).

Ergebnisse

Tabelle 14. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
1, 6	.616	.049	.134	SFS	
1	.606	.049	.144	SFS, SFFS, S	SFS, SFFS, S
1, 6, 7, 12, 16	.740	.044	.010	SFS	
1, 6, 7, 12, 16, 25	.777	.042	-.027	SFS	
1, 4, 6, 8, 13, 20	.857	.035	-.107	SFS	
1, 6, 8, 11, 13, 20, 23	.931	.025	-.181	SFS	
7, 8, 11, 12, 13, 19, 23	.942	.023	-.192	SFFS	
6, 7, 8, 19	.684	.044	.066	SFFS	
6, 7, 8	.624	.048	.126	SFFS	
7	.585	.049	.165	SFFS	SFFS
1, 23	.628	.047	.122	P	P
1, 12	.623	.046	.127	P	P
1, 9, 11	.674	.046	.076	T	
9, 11, 15	.673	.046	.077	T	
9, 11, 14	.664	.047	.086	T	
11, 14, 20	.668	.047	.082	T	
9, 11, 23	.669	.046	.081	T	
1, 9, 23	.674	.047	.076	T	
1, 3, 7, 12	.692	.048	.058		SFS
1, 3, 7	.646	.046	.104		SFS
1, 3, 4, 7, 10, 11, 12	.905	.030	-.155		SFS
4, 7, 10, 11, 12, 23	.895	.031	-.145		SFFS
7, 8, 12	.640	.048	.110		SFFS
7, 10, 11, 12	.723	.045	.027		SFFS
4, 7, 11	.664	.047	.086		T
4, 11, 23	.666	.047	.084		T
1, 11, 23	.666	.047	.084		T
1, 4, 11	.664	.048	.086		T
1, 11, 12	.663	.047	.087		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Bei Bootstrap-Experimenten zur Vorhersage von Aufgabenanforderungen werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (1: „konzentriert“, 6: „zornig“, 7: „aufmerksam“, 12: „aufgeregt“, 16: „ärgerlich“) und das WSIB (7: „aufmerksam“, 10: „körperlich unwohl“, 11: „energiegeladen“, 12: „aufgeregt“) ausgewählt. Beide Modelle

Ergebnisse

erzielen mit einer Bootstrap-Korrektklassifikation von .740 bzw. .723 vergleichbare Ergebnisse. Diese Bootstrap-Korrektklassifikationen sind auch vergleichbar mit der theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 14 und Tabelle 15).

Tabelle 15. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	konzentriert zornig aufmerksam aufgeregt ärgerlich	aufmerksam körperlich unwohl energiegeladen aufgeregt					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.740	.723					
SD[K _{Bootstrap}]	.044	.045					
95 %-CI[K _{Bootstrap}]	[.654, .827]	[.634, .811]					
<i>Vergleich EEB vs. WSIB</i>							
Δ K _{Bootstrap}		.018					
SD[Δ K _{Bootstrap}]		.063					
95 %-CI[Δ K _{Bootstrap}]		[-.106, .141]					
$p[\Delta$ K _{Bootstrap} < 0]		.390					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
K _{Hughes} - K _{Bootstrap}	.010	.027					
SD[K _{Hughes} - K _{Bootstrap}]	.044	.045					
95 %-CI[K _{Hughes} - K _{Bootstrap}]	[-.077, .096]	[-.061, .116]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.414	.274					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.483	.449					
$\chi^2(1)$	23315,35	20111,11					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	34922	15228	50150	33296	16665	49961
	0	10712	39075	49787	11028	38941	49969
	Σ	45634	54303	99937	44324	55606	99930
Korrekturfaktor	c	1,099	0,917		1,127	0,899	

Ergebnisse

Tabelle 16. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Arbeitsanforderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs N = 100 gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	konzentriert zornig aufmerksam aufgeregt ärgerlich	aufmerksam körperlich unwohl energiegeladen aufgeregt					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.002	-.003					
$\sigma(\varepsilon(\pi_1))$.054	.056					
$\langle \varepsilon(\mu) \rangle$	-0,002	-0,002					
$\sigma(\varepsilon(\mu))$	0,050	0,052					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.669	.698					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.338	.397					
$\chi^2(1)$	114,45	157,38					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	340	157	497	315	126	441
Spalten: Vorhersage)	0	174	329	503	176	383	559
	Σ	514	486	1000	491	509	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 35 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,3 Skalenpunkten. Dies entspricht 15 % der Gesamtskala, die von 0 bis 2 reicht. Werden Vorhersagen von Aufgabenanforderungen bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .669 (vollständige EEB) bzw. .698 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 16).

6.3.3 Kommunikation und Kooperation

Kommunikation und Kooperation umfasst Gestaltungsmerkmale der Arbeit, die Zusammenarbeit und Absprachen mit anderen Kolleginnen und Kollegen erfordern. Sie geben Aufschluss, inwiefern Möglichkeiten zur sozialen Interaktion gegeben sind (Ulich, 2011). Die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999) berücksichtigt Kommunikation und Kooperation mit zwei Items (z. B. „Sie arbeiten vorwiegend allein“). Diese sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 4 = "Trifft vollständig zu") zu bewerten. Nach Rekodierung auf eine dreistufige Bewertung (0 = geringer, 1 = mittlerer und 2 = hoher Gestaltungsbedarf) wird der Skalenwert durch Mittelwertbildung berechnet. In Tabelle 17 sind die deskriptiv-statistischen Kennwerte dieser Skala zusammengestellt.

Tabelle 17. Deskriptiv-statistische Kennwerte der Variablen Kommunikation und Kooperation.

<i>N</i>	Modus	Median	<i>M</i>	<i>SD</i>	Min	Max	μ_0	μ_1
1532	1,50	1,50	1,27	0,48	0,00	2,00	0,90	1,65

Modus und Median bzw. Mittelwert (*M*) sind mit 1,50 bzw. 1,27 annähernd identisch. Die Stichprobe mit gültigem Skalenwert für Kommunikation und Kooperation wurde für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 18 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Skalenwerten für Kommunikation und Kooperation aufgelistet.

Tabelle 18. Korrelationen (Spearman ρ) der Variablen Kommunikation und Kooperation mit den einzelnen Items der EEB.

Nr. EEB-Item	ρ	Nr. EEB-Item	ρ	Nr. EEB-Item	ρ
1 konzentriert ^(a)	-.173*	2 aggressiv	-.021	3 nervös ^(a)	-.070
4 körperlich verspannt ^(a)	-.060	5 selbstsicher	-.081	6 zornig	-.015
7 aufmerksam ^(a)	-.140*	8 leistungsbereit ^(a)	-.134*	9 entspannt	-.050
10 körperlich unwohl ^(a)	-.022	11 energiegeladen ^(a)	-.045	12 aufgeregt ^(a)	-.070
13 gelangweilt	.084	14 beschwingt	.026	15 fröhlich	.026
16 ärgerlich	-.034	17 zufrieden	-.039	18 ängstlich	-.050
19 lustig	.063	20 leicht	-.072	21 deprimiert	-.032
22 bedrückt	-.038	23 einflussreich ^(a)	-.009	24 vergnügt	.056
25 betrübt	-.033				

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Es finden sich drei signifikante Zusammenhänge zwischen EEB-Items und Kommunikation und Kooperation: konzentriert ($\rho = -.173$), aufmerksam ($\rho = -.140$) und leistungsbereit ($\rho = -.134$). Diese Korrelationen sind als kleine Effekte ($.1 < |\rho| \leq .3$) zu interpretieren.

Ergebnisse

Tabelle 19. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Kommunikation und Kooperation (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
1, 6, 11, 14, 23, 25	.851	.036	-.101	SFS	
1, 6, 14, 23, 25	.757	.043	-.007	SFS	
1, 6, 8, 11, 14, 23, 25	.913	.028	-.163	SFS	
1, 2, 8, 12, 14, 17, 20	.946	.023	-.196	SFS	
1, 14	.585	.051	.165	SFS, P	
1, 2, 11, 14, 20	.821	.036	-.071	SFS	
1, 8, 14, 20	.719	.045	.031	SFS	
1, 6, 8, 25	.661	.047	.089	SFFS	
6, 7, 10, 11, 15, 25	.819	.039	-.069	SFFS	
1, 8, 25	.611	.048	.139	SFFS	
7, 25	.574	.048	.176	SFFS	
8, 14	.571	.051	.179	SFFS	
1, 6, 8, 11, 15, 25	.830	.038	-.080	SFFS	
1	.556	.050	.194	SFFS, S	SFS, SFFS, S
1, 6, 7, 15, 25	.722	.045	.028	SFFS	
14, 25	.567	.049	.183	SFFS	
14	.558	.050	.192	S	
7	.557	.051	.193	S	S
9, 15	.575	.048	.175	P	
7, 17	.583	.050	.167	P	
14, 23	.581	.048	.169	P	
1, 9	.585	.048	.165	P	
1, 16	.585	.047	.165	P	
9, 19, 23	.669	.049	.081	T	
4, 11, 20	.666	.047	.084	T	
1, 8, 10	.611	.048	.139		SFS
1, 8	.577	.048	.173		SFS
1, 3, 7, 8, 10, 23	.819	.039	-.069		SFS
7, 10	.573	.047	.177		SFFS
3, 7, 8, 10, 11, 12, 23	.893	.031	-.143		SFFS
1, 3, 8, 10, 12, 23	.829	.038	-.079		SFFS
7, 23	.585	.049	.165		P
3, 7	.574	.051	.176		P
1, 4	.576	.051	.174		P
4, 7, 23	.641	.049	.109		T
4, 11, 23	.638	.047	.112		T
1, 4, 11	.635	.048	.115		T
1, 11, 23	.634	.048	.116		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 20. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Kommunikation und Kooperation (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	konzentriert zornig aufmerksam fröhlich betrübt	körperlich verspannt aufmerksam einflussreich					
<i>Bootstrap-Korrektklassifikation</i>							
$K_{\text{Bootstrap}}$.722	.641					
$SD[K_{\text{Bootstrap}}]$.045	.049					
95 %-CI $[K_{\text{Bootstrap}}]$	[.633, .811]	[.544, .738]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.082					
$SD[\Delta K_{\text{Bootstrap}}]$.067					
95 %-CI $[\Delta K_{\text{Bootstrap}}]$		[-.050, .213]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.112					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.028	.109					
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.045	.049					
95 %-CI $[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$	[-.061, .117]	[.012, .206]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.271	.014					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.445	.281					
$\chi^2(1)$	19814,80	7915,20					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	34475	15426	49901	32227	17588	49815
	0	12312	37666	49978	18319	31799	50118
	Σ	46787	53092	99879	50546	49387	99933
Korrekturfaktor	c	1,067	0,941		0,986	1,015	

Bei Bootstrap-Experimenten zur Vorhersage von Kommunikation und Kooperation werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (1: „konzentriert“, 6: „zornig“, 7: „aufmerksam“, 15: „fröhlich“, 25: „betrübt“) und das WSIB (4: „körperlich verspannt“, 7: „aufmerksam“, 23: „einflussreich“) ausgewählt. Ihre Bootstrap-Korrektklassifikationen von .722 bzw. .641 unterscheiden sich nicht signifikant. Die Bootstrap-Korrektklassifikation der Kombination für das WSIB unterschreitet die Korrektklassifikation nach Hughes (1968) jedoch signifikant auf dem Niveau $p < .05$ (Tabelle 19 und Tabelle 20).

Ergebnisse

Tabelle 21. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Kommunikation und Kooperation (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	konzentriert zornig aufmerksam fröhlich betrübt	körperlich verspannt aufmerksam einflussreich					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.002	.000					
$\sigma(\varepsilon(\pi_1))$.054	.061					
$\langle \varepsilon(\mu) \rangle$	-0,001	-0,002					
$\sigma(\varepsilon(\mu))$	0,050	0,053					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.645	.596					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.290	.193					
$\chi^2(1)$	83,97	37,09					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	311	183	494	306	215	521
Spalten: Vorhersage)	0	172	334	506	189	290	479
	Σ	483	517	1000	495	505	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für die vollständige EEB im Promillebereich, für das WSIB sind sie nach Rundung Null. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 34 % bis 39 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,3 Skalenpunkten. Dies entspricht 15 % der Gesamtskala, die von 0 bis 2 reicht. Werden Vorhersagen von Aufgabenanforderungen bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .645 (vollständige EEB) bzw. .596 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 21).

6.3.4 Leistungs- und Zeitvorgaben

Leistungs- und Zeitvorgaben umfasst Gestaltungsmerkmale der Arbeit, die von den Beschäftigten die Einhaltung bestimmter Vorgaben in Bezug auf Leistung und Zeit abverlangen und so zu einer äußeren Kontrolle führen. Die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999) berücksichtigt Leistungs- und Zeitvorgaben mit zwei Items (z. B. „Sie erhalten Leistungsvorgaben, Ihre Arbeit wird kontrolliert“). Diese sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 4 = "Trifft vollständig zu") zu bewerten. Nach Rekodierung auf eine dreistufige Bewertung (0 = geringer, 1 = mittlerer und 2 = hoher Gestaltungsbedarf) wird der Skalenwert durch Mittelwertbildung berechnet. In Tabelle 22 sind die deskriptiv-statistischen Kennwerte dieser Skala zusammengestellt.

Tabelle 22. Deskriptiv-statistische Kennwerte der Variablen Leistungs- und Zeitvorgaben.

<i>N</i>	Modus	Median	<i>M</i>	<i>SD</i>	Min	Max	μ_0	μ_1
1551	1,00	1,00	0,99	0,59	0,00	2,00	0,55	1,44

Modus und Median bzw. Mittelwert (*M*) sind mit 1,00 bzw. 0,99 identisch. Die Stichprobe mit gültigem Skalenwert für Leistungs- und Zeitvorgaben wurde für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 23 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Skalenwerten für Leistungs- und Zeitvorgaben aufgelistet.

Tabelle 23. Korrelationen (Spearman ρ) der Variablen Leistungs- und Zeitvorgaben mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	.101*	2	aggressiv	.012	3	nervös ^(a)	.003
4	körperlich verspannt ^(a)	.034	5	selbstsicher	.051	6	zornig	.063
7	aufmerksam ^(a)	.129*	8	leistungsbereit ^(a)	.093*	9	entspannt	.003
10	körperlich unwohl ^(a)	.077	11	energiegeladen ^(a)	.018	12	aufgeregt ^(a)	.018
13	gelangweilt	-.049	14	beschwingt	.009	15	fröhlich	.008
16	ärgerlich	.029	17	zufrieden	.013	18	ängstlich	.040
19	lustig	.025	20	leicht	-.014	21	deprimiert	.016
22	bedrückt	.028	23	einflussreich ^(a)	-.041	24	vergnügt	.028
25	betrübt	.024						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Es finden sich drei signifikante Zusammenhänge zwischen EEB-Items und Leistungs- und Zeitvorgaben: konzentriert ($\rho = -.101$), aufmerksam ($\rho = -.129$) und leistungsbereit ($\rho = -.093$). Die beiden ersten Korrelationen sind als kleine Effekte ($.1 < |\rho| \leq .3$) zu interpretieren, die letzte ist praktisch nicht relevant ($\rho \leq .1$).

Ergebnisse

Tabelle 24. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Leistungs- und Zeitvorgaben (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
7, 15, 20	.683	.049	.067	SFS	
7, 11, 13, 14, 17, 20	.909	.029	-.159	SFS	
20	.601	.050	.149	SFS, SFFS, S	
5, 7, 15, 20	.770	.044	-.020	SFS	
15, 20	.625	.048	.125	SFS	
7, 9	.616	.049	.134	SFFS	
1, 7, 9, 11, 19, 24	.912	.029	-.162	SFFS	
1, 7, 9, 11, 14	.846	.038	-.096	SFFS	
1, 7, 9, 11, 14, 24	.924	.027	-.174	SFFS	
20, 24	.633	.048	.117	P	
7, 20	.634	.047	.116	P	
20, 23, 24	.692	.046	.058	T	
14, 20, 23	.692	.047	.058	T	
17, 19, 20	.691	.046	.059	T	
11, 19, 20	.691	.046	.059	T	
4, 15, 20	.683	.048	.067	T	
14, 19, 20	.690	.046	.060	T	
1, 3, 4, 7, 10, 23	.872	.035	-.122		SFS
1, 3, 4, 7, 10, 12, 23	.917	.026	-.167		SFS
7	.570	.049	.180		SFS, SFFS, S
1, 4, 7, 10, 23	.791	.041	-.041		SFS
1, 7, 12	.625	.051	.125		SFS
1, 3, 7, 12, 23	.764	.042	-.014		SFS
1, 4, 7, 10, 12, 23	.865	.034	-.115		SFS
1, 4, 10, 12, 23	.781	.041	-.031		SFFS
8	.557	.049	.193		SFFS
1, 4, 8, 23	.714	.047	.036		SFFS
1, 3, 8, 11	.691	.046	.059		SFFS
1, 3, 8, 11, 23	.810	.039	-.060		SFFS
8, 23	.594	.050	.156		P
7, 23	.596	.051	.154		P
11, 23	.585	.049	.165		P
4, 7	.590	.050	.160		P
4, 8, 23	.646	.047	.104		T
4, 11, 23	.650	.047	.100		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 25. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Leistungs- und Zeitvorgaben (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

EEB und WSIB																	
Items	konzentriert körperlich verspannt leistungsbereit einflussreich																
<i>Bootstrap-Korrektklassifikation</i>																	
K _{Bootstrap}	.714																
SD[K _{Bootstrap}]	.047																
95 %-CI[K _{Bootstrap}]	[.622, .807]																
<i>Vergleich mit optimaler Korrektklassifikation K_{Hughes} = .75 nach Hughes</i>																	
K _{Hughes} - K _{Bootstrap}	.036																
SD[K _{Hughes} - K _{Bootstrap}]	.047																
95 %-CI[K _{Hughes} - K _{Bootstrap}]	[-.057, .128]																
p [K _{Hughes} < K _{Bootstrap}]	.226																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.434																
$\chi^2(1)$	18859,24																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;"></th> <th style="width: 15%; text-align: center;">1</th> <th style="width: 15%; text-align: center;">0</th> <th style="width: 65%; text-align: center;">Σ</th> </tr> </thead> <tbody> <tr> <td style="border-right: none;">(Zeilen: Messwerte,</td> <td style="text-align: center;">39955</td> <td style="text-align: center;">10241</td> <td style="text-align: center;">50196</td> </tr> <tr> <td style="border-right: none;">Spalten: Vorhersage)</td> <td style="text-align: center;">18312</td> <td style="text-align: center;">31492</td> <td style="text-align: center;">49804</td> </tr> <tr> <td style="border-right: none;">Σ</td> <td style="text-align: center;">58267</td> <td style="text-align: center;">41733</td> <td style="text-align: center;">100000</td> </tr> </tbody> </table>		1	0	Σ	(Zeilen: Messwerte,	39955	10241	50196	Spalten: Vorhersage)	18312	31492	49804	Σ	58267	41733	100000
	1	0	Σ														
(Zeilen: Messwerte,	39955	10241	50196														
Spalten: Vorhersage)	18312	31492	49804														
Σ	58267	41733	100000														
Korrekturfaktor	c	0,861	1,193														

Bei Bootstrap-Experimenten zur Vorhersage von Leistungs- und Zeitvorgaben wird sowohl für die vollständige EEB als auch das WSIB eine gemeinsame Item-Kombination gewählt: 1: „konzentriert“, 4: „körperlich verspannt“, 8: „leistungsbereit“ und 23: „einflussreich“. Ihre Bootstrap-Korrektklassifikation von .714 unterschreitet die Korrektklassifikation nach Hughes (1968) nicht signifikant (Tabelle 24 und Tabelle 25). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 32 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,36 Skalenpunkten. Dies entspricht 18 % der Gesamtskala, die von 0 bis 2 reicht. Werden Vorhersagen von Leistungs- und Zeitvorgaben bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .648 korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 26).

Ergebnisse

Tabelle 26. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Leistungs- und Zeitvorgaben (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs N = 100 gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

EEB und WSIB																	
Items	konzentriert körperlich verspannt leistungsbereit einflussreich																
<i>Mittlere Vorhersagefehler</i>																	
$\langle \varepsilon(\pi_1) \rangle$	-.002																
$\sigma(\varepsilon(\pi_1))$.051																
$\langle \varepsilon(\mu) \rangle$	-0,004																
$\sigma(\varepsilon(\mu))$	0,059																
<i>Bootstrap-Korrektklassifikation für Stichproben</i>																	
K _{Bootstrap}	.648																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.297																
$\chi^2(1)$	88,23																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 5%;"></th> <th style="width: 45%;">1</th> <th style="width: 45%;">0</th> <th style="width: 5%;">Σ</th> </tr> </thead> <tbody> <tr> <th style="width: 5%;">1</th> <td style="background-color: #d3d3d3;">331</td> <td>161</td> <td>492</td> </tr> <tr> <th style="width: 5%;">0</th> <td>191</td> <td style="background-color: #d3d3d3;">317</td> <td>508</td> </tr> <tr> <th style="width: 5%;">Σ</th> <td>522</td> <td>478</td> <td>1000</td> </tr> </tbody> </table>		1	0	Σ	1	331	161	492	0	191	317	508	Σ	522	478	1000
	1	0	Σ														
1	331	161	492														
0	191	317	508														
Σ	522	478	1000														
(Zeilen: Messwerte, Spalten: Vorhersage)																	

6.3.5 Regulationsbehinderungen

Regulationsbehinderungen umfassen Gestaltungsmerkmale der Arbeit, die die Regulation der Aufgabenbewältigung durch überflüssige und z. T. nicht zu bewältigende Zusatzerfordernisse stört oder behindert (Greiner et al., 1987; Leitner et al., 1987; vgl. Ulich, 2011). Die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999) berücksichtigt Regulationsbehinderungen mit fünf Items (z. B. „Sie erhalten ungenaue und schwer durchschaubare Arbeitsaufträge“). Diese sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 4 = "Trifft vollständig zu") zu bewerten. Nach Rekodierung auf eine dreistufige Bewertung (0 = geringer, 1 = mittlerer und 2 = hoher Gestaltungsbedarf) wird der Skalenwert durch Mittelwertbildung berechnet. In Tabelle 27 sind die deskriptiv-statistischen Kennwerte dieser Skala zusammengestellt.

Tabelle 27. Deskriptiv-statistische Kennwerte der Variablen Regulationsbehinderungen.

<i>N</i>	Modus	Median	<i>M</i>	<i>SD</i>	Min	Max	μ_0	μ_1
1534	1,40	1,20	1,21	0,47	0,00	2,00	0,83	1,59

Median und Mittelwert (*M*) sind mit 1,20 bzw. 1,21 annähernd identisch. Die Stichprobe mit gültigem Skalenwert für Regulationsbehinderungen wurde für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 28 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Skalenwerten für Regulationsbehinderungen aufgelistet.

Tabelle 28. Korrelationen (Spearman ρ) der Variablen Regulationsbehinderungen mit den einzelnen Items der EEB.

Nr. EEB-Item	ρ	Nr. EEB-Item	ρ	Nr. EEB-Item	ρ
1 konzentriert ^(a)	-.097*	2 aggressiv	.212*	3 nervös ^(a)	.171*
4 körperlich verspannt ^(a)	.231*	5 selbstsicher	-.098*	6 zornig	.211*
7 aufmerksam ^(a)	-.177*	8 leistungsbereit ^(a)	-.123*	9 entspannt	.032
10 körperlich unwohl ^(a)	.257*	11 energiegeladen ^(a)	-.160*	12 aufgeregt ^(a)	.105*
13 gelangweilt	.187*	14 beschwingt	-.102*	15 fröhlich	-.170*
16 ärgerlich	.260*	17 zufrieden	-.322*	18 ängstlich	.118*
19 lustig	-.068	20 leicht	.017	21 deprimiert	.275*
22 bedrückt	.280*	23 einflussreich ^(a)	-.161*	24 vergnügt	-.120*
25 betrübt	.249*				

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Nur die EEB-Items „lustig“, „leicht“ und „entspannt“ korrelieren *nicht* signifikant mit Regulationsbehinderungen. Alle anderen Items weisen mit Ausnahme von „zufrieden“ ($\rho = -.322$) einen geringen Effekt ($.1 < |\rho| \leq .3$) auf.

Ergebnisse

Tabelle 29. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Regulationsbehinderungen (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes - } K_{\text{Bootstrap}}}$	EEB	WSIB
16, 17, 22	.670	.047	.080	SFS	
22	.620	.049	.130	SFS, SFFS, S	
9, 10, 16, 17, 22	.796	.039	-.046	SFS	
10, 16, 17, 18, 21, 22	.774	.040	-.024	SFS	
7, 9, 10, 16, 17, 21, 22	.882	.032	-.132	SFS	
6, 10, 16, 17, 21, 22, 25	.809	.042	-.059	SFS	
10, 16, 17, 22	.712	.044	.038	SFS	
2, 4, 17, 21	.734	.044	.016	SFFS	
2, 10, 16, 17, 21	.758	.044	-.008	SFFS	
6, 10, 16, 17, 21, 25	.783	.041	-.033	SFFS	
2, 4, 10, 16, 17, 18, 21	.857	.036	-.107	SFFS	
4, 17, 21	.685	.045	.065	SFFS	
2, 7, 10, 16, 17, 21	.816	.039	-.066	SFFS	
10, 16, 17, 21	.714	.047	.036	SFFS	
4, 17	.643	.049	.107	P	
17, 22	.639	.048	.111	P	
4, 17, 19	.699	.046	.051	T	
4, 9, 17	.692	.046	.058	T	
10	.606	.049	.144		SFS, SFFS, S
3, 4, 10, 11	.731	.044	.019		SFS
7, 10	.614	.046	.136		SFS, SFFS
1, 4, 7, 10	.722	.045	.028		SFS
4	.599	.050	.151		SFS, SFFS, S
4, 7	.623	.048	.127		SFS, P
7, 10, 12, 23	.696	.045	.054		SFFS
1, 3, 4, 7	.702	.047	.048		SFFS
1, 3, 4, 7, 11	.811	.039	-.061		SFFS
3, 4, 7, 11, 12, 23	.896	.030	-.146		SFFS
3, 4, 11	.657	.047	.093		SFFS
10, 11	.620	.050	.130		P
4, 10	.622	.049	.128		P
4, 11	.617	.048	.133		P
4, 10, 11	.671	.047	.079		T
4, 11, 23	.668	.049	.082		T
4, 7, 11	.666	.046	.084		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Bei Bootstrap-Experimenten zur Vorhersage von Regulationsbehinderungen werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (2: „aggressiv“, 4: „körperlich verspannt“, 17: „zufrieden“, 21: „deprimiert“) und das WSIB (3: „nervös“, 4: „körperlich

Ergebnisse

verspannt“, 10: „körperlich unwohl“, 11: „energiegeladen“) ausgewählt. Beide Modelle erzielen mit einer Bootstrap-Korrektklassifikation von .734 bzw. .731 vergleichbare Ergebnisse. Diese Bootstrap-Korrektklassifikationen sind auch vergleichbar mit der theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 29 und Tabelle 30).

Tabelle 30. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Regulationsbehinderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	aggressiv körperlich verspannt zufrieden deprimiert	nervös körperlich verspannt körperlich unwohl energiegeladen					
<i>Bootstrap-Korrektklassifikation</i>							
$K_{\text{Bootstrap}}$.734	.731					
$SD[K_{\text{Bootstrap}}]$.044	.044					
95 %-CI [$K_{\text{Bootstrap}}$]	[.648, .820]	[.645, .818]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.003					
$SD[\Delta K_{\text{Bootstrap}}]$.062					
95 %-CI [$\Delta K_{\text{Bootstrap}}$]		[-.119, .125]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.479					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.016	.019					
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.044	.044					
95 %-CI [$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.070, .102]	[-.068, .105]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.361	.334					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.491	.472					
$\chi^2(1)$	24080,48	22299,56					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel		1	0	Σ	1	0	Σ
(Zeilen: Messwerte,	1	29375	20675	50050	31332	18600	49932
Spalten: Vorhersage)	0	5893	44057	49950	8289	41779	50068
	Σ	35268	64732	100000	39621	60379	100000
Korrekturfaktor	c	1,419	0,772		1,260	0,829	

Ergebnisse

Tabelle 31. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Regulationsbehinderungen (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs N = 100 gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	aggressiv körperlich verspannt zufrieden deprimiert	nervös körperlich verspannt körperlich unwohl energiegeladen					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.004	.002					
$\sigma(\varepsilon(\pi_1))$.060	.059					
$\langle \varepsilon(\mu) \rangle$	0,003	0,002					
$\sigma(\varepsilon(\mu))$	0,051	0,052					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.688	.673					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.377	.346					
$\chi^2(1)$	142,21	119,68					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	330	139	469	332	164	496
	0	173	358	531	163	341	504
	Σ	503	497	1000	495	505	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 38 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,3 Skalenpunkten. Dies entspricht 15 % der Gesamtskala, die von 0 bis 2 reicht. Werden Vorhersagen von Regulationsbehinderungen bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .688 (vollständige EEB) bzw. .673 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 31).

6.3.6 Tätigkeitsspielraum

Tätigkeitsspielraum umfasst Gestaltungsmerkmale der Arbeit, die von der Planung bis zur Durchführung einer Arbeitshandlung Entscheidungs- und Gestaltungsmöglichkeiten für die Beschäftigten einräumen. Sie geben Aufschluss, inwiefern eine Arbeitsaufgabe sequenziell vollständig und verantwortungsvoll ist (Hacker & Sachse, 2014; Ulich, 2011). Die Synthetische Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999) berücksichtigt Tätigkeitsspielraum mit drei Items (z. B. „Sie müssen oft Entscheidungen treffen und tragen die Verantwortung dafür“). Diese sind auf einer fünfstufigen Skala (0 = "Trifft gar nicht zu" bis 4 = "Trifft vollständig zu") zu bewerten. Nach Rekodierung auf eine dreistufige Bewertung (0 = geringer, 1 = mittlerer und 2 = hoher Gestaltungsbedarf) wird der Skalenwert durch Mittelwertbildung berechnet. In Tabelle 32 sind die deskriptiv-statistischen Kennwerte dieser Skala zusammengestellt.

Tabelle 32. Deskriptiv-statistische Kennwerte der Variablen Tätigkeitsspielraum.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1541	1,33	1,33	1,18	0,52	0,00	2,00	0,76	1,60

Modus und Median bzw. Mittelwert (M) sind mit 1,33 bzw. 1,18 vergleichbar ausgeprägt. Die Stichprobe mit gültigem Skalenwert für Tätigkeitsspielraum wurde für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 33 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Skalenwerten für Tätigkeitsspielraum aufgelistet.

Tabelle 33. Korrelationen (Spearman ρ) der Variablen Tätigkeitsspielraum mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.228*	2	aggressiv	-.003	3	nervös ^(a)	.011
4	körperlich verspannt ^(a)	.013	5	selbstsicher	-.231*	6	zornig	.038
7	aufmerksam ^(a)	-.251*	8	leistungsbereit ^(a)	-.243*	9	entspannt	-.056
10	körperlich unwohl ^(a)	.100*	11	energiegeladen ^(a)	-.160*	12	aufgeregt ^(a)	-.006
13	gelangweilt	.180*	14	beschwingt	-.104*	15	fröhlich	-.103*
16	ärgerlich	.033	17	zufrieden	-.196*	18	ängstlich	.047
19	lustig	-.026	20	leicht	-.086	21	deprimiert	.050
22	bedrückt	.039	23	einflussreich ^(a)	-.320*	24	vergnügt	-.077
25	betrübt	.067						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Je geringer der Gestaltungsbedarf für Tätigkeitsspielraum ausgeprägt ist, desto einflussreicher ($\rho = -.320$), aufmerksamer ($\rho = -.251$) und leistungsbereiter ($\rho = -.243$) fühlen sich die Beschäftigten während der Arbeit. Gleichzeitig fühlen sie sich umso weniger

Ergebnisse

gelangweilt ($\rho = .180$). Während die Korrelation mit „einflussreich“ als mittlerer Effekt ($.3 < |\rho| \leq .5$) zu interpretieren ist, sind alle weiteren signifikanten Korrelationen als kleine Effekte ($.1 < |\rho| \leq .3$) zu betrachten.

Tabelle 34. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Tätigkeitsspielraum (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
23	.611	.049	.139	SFS, SFFS, S	SFS, SFFS, S
1, 5, 19, 23	.766	.042	-.016	SFS	
1, 5, 6, 8, 12, 23	.846	.036	-.096	SFS	
5, 11, 13	.647	.047	.103	SFFS	
5, 7, 8	.644	.047	.106	SFFS	
2, 5, 11, 13, 16, 25	.785	.040	-.035	SFFS	
2, 5, 7, 8, 21	.735	.045	.015	SFFS	
5, 23	.638	.048	.112	P	
14, 23	.639	.047	.111	P	
8, 23	.636	.049	.114	P	P
13, 23	.636	.049	.114	P	
5, 23, 24	.695	.046	.055	T	
5, 19, 23	.694	.045	.056	T	
7, 20, 23	.689	.047	.061	T	
5, 15, 23	.691	.045	.059	T	
20, 23, 24	.690	.046	.060	T	
5, 17, 19	.680	.046	.070	T	
1, 3, 8, 23	.725	.046	.025		SFS
1, 3, 4, 23	.734	.044	.016		SFS
3, 4, 7, 10, 23	.811	.040	-.061		SFS
1, 4, 7, 10, 11, 12, 23	.945	.022	-.195		SFS
1, 3, 23	.653	.047	.097		SFS
7, 8	.598	.052	.152		SFFS
1, 7, 8	.639	.049	.111		SFFS
4, 23	.623	.048	.127		P
7, 23	.633	.050	.117		P
1, 8, 23	.673	.046	.077		T
7, 11, 23	.675	.044	.075		T
4, 10, 23	.670	.047	.080		T
7, 8, 23	.673	.046	.077		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 35. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Tätigkeitsspielraum (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	aggressiv selbstsicher aufmerksam leistungsbereit deprimiert	konzentriert nervös körperlich verspannt einflussreich					
<i>Bootstrap-Korrektklassifikation</i>							
$K_{\text{Bootstrap}}$.735	.734					
$SD[K_{\text{Bootstrap}}]$.045	.044					
95 %-CI $[K_{\text{Bootstrap}}]$	[.646, .823]	[.648, .820]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.001					
$SD[\Delta K_{\text{Bootstrap}}]$.063					
95 %-CI $[\Delta K_{\text{Bootstrap}}]$		[-.122, .124]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.494					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.015	.016					
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.045	.044					
95 %-CI $[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$	[-.073, .104]	[-.070, .102]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.369	.357					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.470	.472					
$\chi^2(1)$	22093,58	22280,70					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	35654	14347	50001	40110	9940	50050
	0	12153	37790	49943	16670	33280	49950
	Σ	47807	52137	99944	56780	43220	100000
Korrekturfaktor	c	1,046	0,958		0,881	1,156	

Bei Bootstrap-Experimenten zur Vorhersage von Tätigkeitsspielraum werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (2: „aggressiv“, 5: „selbstsicher“, 7: „aufmerksam“, 8: „leistungsbereit“, 21: „deprimiert“) und das WSIB (1: „konzentriert“, 3: „nervös“, 4: „körperlich verspannt“, 23: „einflussreich“) ausgewählt. Beide Modelle erzielen mit einer Bootstrap-Korrektklassifikation von .735 bzw. .734 vergleichbare Ergebnisse. Diese Bootstrap-Korrektklassifikationen sind auch vergleichbar mit der theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 34 und Tabelle 35).

Ergebnisse

Tabelle 36. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Tätigkeitsspielraum (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs N = 100 gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	aggressiv selbstsicher aufmerksam leistungsbereit deprimiert	konzentriert nervös körperlich verspannt einflussreich					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.002	.000					
$\sigma(\varepsilon(\pi_1))$.052	.051					
$\langle \varepsilon(\mu) \rangle$	0,000	-0,002					
$\sigma(\varepsilon(\mu))$	0,052	0,051					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.686	.687					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.373	.373					
$\chi^2(1)$	138,94	139,37					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	347	169	516	373	173	546
	0	145	339	484	140	314	454
	Σ	492	508	1000	513	487	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für die vollständige EEB im Promillebereich, für das WSIB sind sie nach Rundung Null. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 33 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind identisch bzw. nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,3 Skalenpunkten. Dies entspricht 15 % der Gesamtskala, die von 0 bis 2 reicht. Werden Vorhersagen von Tätigkeitsspielraum bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .686 (vollständige EEB) bzw. .687 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 36).

6.3.7 Gesamtbedarf für Arbeitsgestaltung

Gesamtbedarf für Arbeitsgestaltung ist ein aggregiertes Merkmal der Synthetischen Beanspruchungs- und Arbeitsanalyse (SynBA; Wieland-Eckelmann et al., 1999). Der Rohwert dieses Merkmals ergibt sich durch Summation der Gestaltungsbedarfe für die fünf Merkmalsbereiche der SynBA: Aufgabenanforderungen, Kommunikation und Kooperation, Leistungs- und Zeitvorgaben, Regulationsbehinderungen und Tätigkeitsspielraum. Er kann daher Werte zwischen 0 (kein Gesamtbedarf für Arbeitsgestaltung) und 10 (besonders hoher Gesamtbedarf für Arbeitsgestaltung) annehmen. In Tabelle 37 sind die deskriptiv-statistischen Kennwerte dieses Kennwertes zusammengestellt.

Tabelle 37. Deskriptiv-statistische Kennwerte der Variablen Gesamtbedarf für Arbeitsgestaltung.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1466	6,43	5,70	5,71	1,32	1,70	10,00	4,67	6,75

Während Median und Mittelwert (M) mit 5,7 identisch sind, fällt der Modus mit 6,43 etwas höher aus. Die Stichprobe mit gültigem Kennwert für Gesamtgestaltungsbedarf wurde für die folgenden Analysen am Median in die Kategorien 0 = „günstig“ und 1 = „ungünstig“ unterteilt. In Tabelle 38 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Kennwerten für Gesamtgestaltungsbedarf aufgelistet.

Tabelle 38. Korrelationen (Spearman ρ) der Variablen Gesamtbedarf für Arbeitsgestaltung mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.268*	2	aggressiv	.052	3	nervös ^(a)	.009
4	körperlich verspannt ^(a)	.048	5	selbtsicher	-.181*	6	zornig	.111*
7	aufmerksam ^(a)	-.261*	8	leistungsbereit ^(a)	-.239*	9	entspannt	-.051
10	körperlich unwohl ^(a)	.184*	11	energiegeladen ^(a)	-.178*	12	aufgeregt ^(a)	-.019
13	gelangweilt	.209*	14	beschwingt	-.099*	15	fröhlich	-.096*
16	ärgerlich	.095*	17	zufrieden	-.239*	18	ängstlich	.038
19	lustig	.020	20	leicht	-.105*	21	deprimiert	.108*
22	bedrückt	.105*	23	einflussreich ^(a)	-.268*	24	vergnügt	-.038
25	betrübt	.112*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Die EEB-Items „konzentriert“ und „einflussreich“ weisen die höchste negative Korrelation von -.268 auf, gefolgt von „aufmerksam“ mit einer Korrelation von -.261. Die höchste positive Korrelation mit Gesamtgestaltungsbedarf besteht mit dem Item „körperlich unwohl“ ($\rho = .184$). Keine der Korrelationen kommt über einen kleinen Effekt ($.1 < |\rho| \leq .3$) hinaus. Einige signifikante Korrelationen sind praktisch nicht relevant ($|\rho| \leq .1$).

Ergebnisse

Tabelle 39. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Arbeitsgestaltung gesamt (0 = günstig, 1 = ungünstig). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	KBootstrap	SD[KBootstrap]	KHughes - KBootstrap	EEB	WSIB
1, 10, 13	.649	.049	.101	SFS	
1	.592	.049	.158	SFS, SFFS, S	SFS, SFFS, S
1, 9, 21, 23	.735	.044	.015	SFS	
1, 7, 10	.657	.050	.093	SFS	SFS
1, 7, 9, 17, 19, 23	.939	.024	-.189	SFS	
1, 10, 16	.654	.048	.096	SFS	
1, 7, 10, 11, 14, 17	.884	.031	-.134	SFS	
7, 9, 10, 23	.760	.043	-.010	SFFS	
7, 10, 17	.663	.049	.087	SFFS	
7, 9, 13, 23	.716	.044	.034	SFFS	
2, 6, 7, 9, 10, 17, 25	.871	.034	-.121	SFFS	
7	.581	.049	.169	S	SFFS, S
1, 10	.627	.047	.123	P	SFS, P
11, 20, 23	.683	.048	.067	T	
9, 11, 19	.678	.046	.072	T	
19, 20, 23	.687	.047	.063	T	
11, 19, 23	.683	.047	.067	T	
9, 19, 23	.681	.047	.069	T	
11, 17, 19	.676	.048	.074	T	
1, 4, 7, 10, 11, 23	.888	.032	-.138		SFS
1, 3, 8, 10, 23	.785	.041	-.035		SFS
1, 4, 7, 10, 11	.793	.040	-.043		SFS
3, 4, 7, 8, 10, 11, 12, 23	.956	.020	-.206		SFFS
3, 7, 10, 23	.711	.045	.039		SFFS
3, 4, 7, 10, 23	.798	.040	-.048		SFFS
7, 10, 23	.652	.046	.098		SFFS
3, 4, 7, 10, 11, 23	.893	.032	-.143		SFFS
4, 11, 23	.665	.048	.085		T
1, 4, 10	.658	.048	.092		T
1, 4, 23	.656	.047	.094		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Bei Bootstrap-Experimenten zur Vorhersage des Gesamtgestaltungsbedarfs werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (1: „konzentriert“, 9: „entspannt“, 21: „deprimiert“, 23: „einflussreich“) und das WSIB (3: „nervös“, 7:

Ergebnisse

„aufmerksam“, 10: „körperlich unwohl“, 23: „einflussreich“) ausgewählt. Beide Modelle erzielen mit einer Bootstrap-Korrektklassifikation von .735 bzw. .711 vergleichbare Ergebnisse. Diese Bootstrap-Korrektklassifikationen sind auch vergleichbar mit der theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 39 und Tabelle 40).

Tabelle 40. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Arbeitsgestaltung gesamt (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB						
Items	konzentriert entspannt deprimiert einflussreich	nervös aufmerksam körperlich unwohl einflussreich						
<i>Bootstrap-Korrektklassifikation</i>								
K _{Bootstrap}	.735	.711						
SD[K _{Bootstrap}]	.044	.045						
95 %-CI[K _{Bootstrap}]	[.649, .820]	[.623, .799]						
<i>Vergleich EEB vs. WSIB</i>								
$\Delta K_{\text{Bootstrap}}$.024						
SD[$\Delta K_{\text{Bootstrap}}$]		.063						
95 %-CI[$\Delta K_{\text{Bootstrap}}$]		[-.099, .147]						
$p[\Delta K_{\text{Bootstrap}} < 0]$.352						
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>								
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.015	.039						
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.044	.045						
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.070, .101]	[-.049, .127]						
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.364	.193						
<i>Auswertung der Vier-Felder-Tafel</i>								
Cramers ϕ	.479	.422						
$\chi^2(1)$	22941,07	17804,93						
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001						
Vierfeldertafel								
(Zeilen: Messwerte, Spalten: Vorhersage)	1	31985	18109	50094	1	35703	14378	50081
	0	8407	41499	49906	0	14495	35351	49846
	Σ	40392	59608	100000	Σ	50198	49729	99927
Korrekturfaktor	c	1,240	0,837		c	0,998	1,002	

Ergebnisse

Tabelle 41. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Arbeitsgestaltung gesamt (0 = günstig, 1 = ungünstig) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs N = 100 gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	konzentriert entspannt deprimiert einflussreich	nervös aufmerksam körperlich unwohl einflussreich					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.002	.000					
$\sigma(\varepsilon(\pi_1))$.059	.054					
$\langle \varepsilon(\mu) \rangle$	-0,005	-0,004					
$\sigma(\varepsilon(\mu))$	0,140	0,130					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.672	.621					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.344	.241					
$\chi^2(1)$	118,41	57,92					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	333	159	492	288	183	471
Spalten: Vorhersage)	0	169	339	508	196	333	529
	Σ	502	498	1000	484	516	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für die vollständige EEB im Promillebereich, für das WSIB sind sie nach Rundung Null. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 34 % bis 37 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,84 Skalenpunkten. Dies entspricht 8,4 % der Gesamtskala, die von 0 bis 10 reicht. Werden Vorhersagen des Gesamtgestaltungsbedarfs bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .672 (vollständige EEB) bzw. .621 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe vorkommen (Tabelle 41).

6.3.8 Autoritäre Führung

Autoritäre Führung ist eine Teilskala des Fragebogens zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Sie umfasst Verhaltensweisen der Führungskräfte, die bei den Geführten als Stressoren wirksam werden können (z. B. „Persönlichen Ärger oder Ärger mit der Geschäftsleitung lässt er/sie [die Führungskraft] an uns aus“). Die insgesamt vier Items sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Der Rohwert dieser Variablen ergibt sich durch Mittelwertbildung. In Tabelle 42 sind die deskriptiv-statistischen Kennwerte dieser Variablen zusammengefasst.

Tabelle 42. Deskriptiv-statistische Kennwerte der Variablen Autoritäre Führung.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1342	1,00	1,25	1,33	0,79	0,00	4,00	0,70	1,96

Modus und Median fallen etwas niedriger aus, als der Mittelwert (*M*). Insgesamt fallen die zentralen Maße für autoritäre Führung auf der fünfstufigen Skala vergleichsweise niedrig aus. Die Stichprobe mit gültigem Rohwert für autoritäre Führung wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 43 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für autoritäre Führung aufgelistet.

Tabelle 43. Korrelationen (Spearman ρ) der Variablen Autoritäre Führung mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.067	2	aggressiv	.182*	3	nervös ^(a)	.111*
4	körperlich verspannt ^(a)	.138*	5	selbstsicher	-.014	6	zornig	.202*
7	aufmerksam ^(a)	-.084	8	leistungsbereit ^(a)	-.087	9	entspannt	-.012
10	körperlich unwohl ^(a)	.220*	11	energiegeladen ^(a)	-.066	12	aufgeregt ^(a)	.093
13	gelangweilt	.144*	14	beschwingt	-.078	15	fröhlich	-.076
16	ärgerlich	.191*	17	zufrieden	-.190*	18	ängstlich	.097*
19	lustig	-.018	20	leicht	-.063	21	deprimiert	.177*
22	bedrückt	.152*	23	einflussreich ^(a)	-.066	24	vergnügt	-.038
25	betrübt	.176*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Signifikante Korrelationen sind praktisch nicht relevant ($|\rho| \leq .1$) (z. B. „ängstlich“) oder sie sind als geringe Effekte ($.1 < |\rho| \leq .3$) zu interpretieren. Die stärkste positive Beziehung von autoritärer Führung besteht mit „körperlich unwohl“ ($\rho = .220$), die stärkste negative mit „zufrieden“ ($\rho = -.190$).

Ergebnisse

Tabelle 44. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von autoritärer Führung (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
10	.592	.048	.158	SFS, SFFS, S	SFS, SFFS, S
1, 10, 17, 25	.688	.048	.062	SFS	
7, 9, 10, 13, 17, 20, 21	.918	.027	-.168	SFS	
10, 13	.608	.049	.142	SFS, P	
17	.578	.050	.172	SFFS	
12, 16, 17, 20	.743	.044	.007	SFFS	
9, 10	.604	.047	.146	P	
5, 17	.605	.049	.145	P	
5, 20	.599	.050	.151	P	
10, 16	.604	.048	.146	P	
9, 17, 19	.681	.046	.069	T	
9, 11, 24	.673	.048	.077	T	
4, 17, 20	.670	.047	.080	T	
4, 20, 23	.672	.046	.078	T	
4, 9, 15	.676	.046	.074	T	
1, 3, 7, 8, 10	.742	.043	.008		SFS
1, 7, 8, 10, 23	.769	.042	-.019		SFS
1, 10, 23	.639	.048	.111		SFS
3, 4, 8, 23	.710	.044	.040		SFFS
4	.557	.050	.193		SFFS
1, 3, 4, 8	.690	.046	.060		SFFS
3, 4, 7, 8	.691	.046	.059		SFFS
10, 23	.600	.048	.150		P
7, 10	.602	.049	.148		P
4, 10	.599	.048	.151		P
10, 11	.597	.048	.153		P
7, 10, 11	.651	.050	.099		T
1, 4, 23	.646	.047	.104		T
1, 4, 10	.652	.049	.098		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 45. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von autoritärer Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	aufgeregt ärgerlich zufrieden leicht	konzentriert nervös aufmerksam leistungsbereit körperlich unwohl					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.743	.742					
SD[K _{Bootstrap}]	.044	.043					
95 %-CI[K _{Bootstrap}]	[.655, .830]	[.657, .826]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.001					
SD[$\Delta K_{\text{Bootstrap}}$]		.062					
95 %-CI[$\Delta K_{\text{Bootstrap}}$]		[-.120, .122]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.494					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.007	.008					
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.044	.043					
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.080, .095]	[-.076, .093]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.434	.423					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.499	.484					
$\chi^2(1)$	24886,74	23453,39					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	31569	18602	50171	35470	14578	50048
	0	7138	42691	49829	11237	38642	49879
	Σ	38707	61293	100000	46707	53220	99927
Korrekturfaktor	c	1,296	0,813		1,072	0,937	

Bei Bootstrap-Experimenten zur Vorhersage von autoritärer Führung wird für die vollständige EEB die Item-Kombination 12: „aufgeregt“, 16: „ärgerlich“, 17: „zufrieden“, 20: „leicht“ endgültig ausgewählt. Dieses Modell weist eine Bootstrap-Korrektklassifikation von .743 auf. Für das WSIB ist es die Item-Kombination 1: „konzentriert“, 3: „nervös“, 7: „aufmerksam“, 8: „leistungsbereit“ und 10: „körperlich unwohl“ mit einer Bootstrap-Korrektklassifikation von .742. Beide Bootstrap-Korrektklassifikationen unterscheiden sich nicht signifikant untereinander und von der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (Tabelle 44 und Tabelle 45). Die mittleren Fehler bei der Vorhersage der

Ergebnisse

Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 33 % bis 37 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,5 Skalenpunkten. Dies entspricht 12,5 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von autoritärer Führung bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .682 (vollständige EEB) bzw. .680 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ausgeprägt ist (Tabelle 46).

Tabelle 46. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von autoritärer Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	aufgeregt ärgerlich zufrieden leicht	konzentriert nervös aufmerksam leistungsbereit körperlich unwohl					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.002	.005					
$\sigma(\varepsilon(\pi_1))$.059	.052					
$\langle \varepsilon(\mu) \rangle$	0,003	0,003					
$\sigma(\varepsilon(\mu))$	0,085	0,080					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.682	.680					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.364	.359					
$\chi^2(1)$	132,62	129,08					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	347	152	499	321	154	475
Spalten: Vorhersage)	0	166	335	501	166	359	525
	Σ	513	487	1000	487	513	1000

6.3.9 Bedeutung von Anerkennung und Wertschätzung

Bedeutung von Anerkennung und Wertschätzung ist eine Teilskala des Fragebogens zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Mit den zugehörigen beiden Items wird gefragt, wieviel Bedeutung Anerkennung und Wertschätzung durch die Führungskraft für die Beschäftigten haben (z. B. „Anerkennung und Wertschätzung durch unsere Führungskraft ist mir wichtig“). Die Items sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Der Rohwert dieser Variablen ergibt sich durch Mittelwertbildung. In Tabelle 47 sind die deskriptiv-statistischen Kennwerte dieser Variablen zusammengefasst.

Tabelle 47. Deskriptiv-statistische Kennwerte der Variablen Bedeutung von Anerkennung und Wertschätzung.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1342	3,00	3,00	2,93	0,94	0,00	4,00	2,24	3,63

Die zentralen Maße (Modus, Median und Mittelwert [M]) weisen vergleichbare Werte auf und sind vergleichsweise hoch auf der fünfstufigen Skala angesiedelt. Die Stichprobe mit gültigem Rohwert für Bedeutung von Anerkennung und Wertschätzung wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 48 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Bedeutung von Anerkennung und Wertschätzung aufgelistet.

Tabelle 48. Korrelationen (Spearman ρ) der Variablen Bedeutung von Anerkennung und Wertschätzung mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	.165*	2	aggressiv	-.047	3	nervös ^(a)	.066
4	körperlich verspannt ^(a)	.037	5	selbstsicher	.082	6	zornig	-.015
7	aufmerksam ^(a)	.218*	8	leistungsbereit ^(a)	.246*	9	entspannt	.133*
10	körperlich unwohl ^(a)	-.038	11	energiegeladen ^(a)	.095*	12	aufgeregt ^(a)	.068
13	gelangweilt	-.075	14	beschwingt	.074	15	fröhlich	.124*
16	ärgerlich	-.035	17	zufrieden	.139*	18	ängstlich	.022
19	lustig	.077	20	leicht	.146*	21	deprimiert	-.007
22	bedrückt	.002	23	einflussreich ^(a)	.097*	24	vergnügt	.060
25	betrübt	-.003						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Die signifikanten Korrelationen sind durchgehend positiv und sind praktisch nicht bedeutsam ($|\rho| \leq .1$) („energiegeladen“, „einflussreich“) oder als kleine Effekte ($.1 < |\rho| \leq .3$) zu interpretieren. Die höchste Korrelation besteht mit „leistungsbereit“ ($\rho = .246$), gefolgt von „aufmerksam“ ($\rho = .218$).

Ergebnisse

Tabelle 49. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Bedeutung von Anerkennung und Wertschätzung (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	K _{Bootstrap}	SD[K _{Bootstrap}]	K _{Hughes} - K _{Bootstrap}	EEB	WSIB
5, 8, 9, 20	.790	.041	-.040	SFS	
8, 9	.643	.047	.107	SFS, P	
8	.611	.050	.139	SFS, SFFS, S	SFS, SFFS, S
5, 8, 9, 18, 20	.836	.037	-.086	SFS	
5, 8, 9, 14, 18, 20	.911	.029	-.161	SFS	
5, 8, 9, 18	.719	.046	.031	SFS	
1, 5, 9, 12, 20	.859	.036	-.109	SFFS	
1, 5, 9, 20	.773	.043	-.023	SFFS	
20	.590	.049	.160	SFFS	
1, 20	.613	.049	.137	SFFS	
5, 6, 9, 20	.761	.042	-.011	SFFS	
8, 20	.641	.048	.109	P	
11, 15, 20	.695	.047	.055	T	
5, 9, 20	.699	.047	.051	T	
5, 9, 17	.695	.045	.055	T	
5, 9, 24	.691	.046	.059	T	
8, 9, 20	.694	.047	.056	T	
5, 20, 24	.692	.044	.058	T	
1, 3, 8	.648	.046	.102		SFS
1, 3, 7, 8, 11	.784	.042	-.034		SFS
1, 3, 8, 10, 12	.755	.044	-.005		SFS
1, 3, 4, 7, 8, 11	.870	.034	-.120		SFS
1, 3, 7, 11, 12, 23	.876	.033	-.126		SFFS
1, 3, 4	.639	.048	.111		SFFS
1, 3, 7, 11, 12	.792	.040	-.042		SFFS
8, 11	.622	.049	.128		P
8, 10	.623	.050	.127		P
3, 8	.621	.048	.129		P
4, 8	.616	.048	.134		P
8, 11, 23	.661	.048	.089		T
3, 8, 23	.659	.048	.091		T
4, 8, 23	.658	.049	.092		T
4, 7, 11	.656	.048	.094		T
1, 3, 11	.654	.048	.096		T
4, 8, 11	.663	.046	.087		T

Anmerkungen: K_{Bootstrap} = Bootstrap-Korrektklassifikation, K_{Hughes} = .750 = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 50. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Bedeutung von Anerkennung und Wertschätzung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	selbstsicher leistungsbereit entspannt ängstlich	körperlich verspannt leistungsbereit energiegeladen					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.719	.663					
SD[K _{Bootstrap}]	.046	.046					
95 %-CI[K _{Bootstrap}]	[.629, .809]	[.574, .753]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.056					
SD[$\Delta K_{\text{Bootstrap}}$]		.065					
95 %-CI[$\Delta K_{\text{Bootstrap}}$]		[-.071, .182]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.194					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.031	.087					
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.046	.046					
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.059, .121]	[-.003, .176]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.250	.028					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.443	.328					
$\chi^2(1)$	19629,36	10743,91					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	32165	17763	49928	31348	18807	50155
	0	10325	39747	50072	14863	34982	49845
	Σ	42490	57510	100000	46211	53789	100000
Korrekturfaktor	c	1,175	0,871		1,085	0,927	

Bei Bootstrap-Experimenten zur Vorhersage der Bedeutung von Anerkennung und Wertschätzung werden zwei unterschiedliche Item-Kombinationen gewählt. Für die vollständige EEB sind es die Items 5 („selbstsicher“), 8 („leistungsbereit“), 9 („entspannt“) und 18 („ängstlich“). Für das WSIB sind es die Items 4 („körperlich verspannt“), 8 („leistungsbereit“) und 11 („energiegeladen“). Das Modell der vollständigen EEB weist mit .719 eine bessere Korrektklassifikation auf, als das des WSIB mit .663. Dieser Unterschied ist jedoch nicht statistisch bedeutsam. Die Korrektklassifikationen fallen vergleichsweise niedriger aus, als die theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 49 und Tabelle 104). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 sind

Ergebnisse

Null (EEB) bzw. liegen im Promillebereich (WSIB). Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 37 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,6 Skaleneinheiten. Dies entspricht 16 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von der Bedeutung von Anerkennung und Wertschätzung bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .633 (vollständige EEB) bzw. .599 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ausgeprägt ist (Tabelle 51).

Tabelle 51. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Bedeutung von Anerkennung und Wertschätzung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	selbstsicher leistungsbereit entspannt ängstlich	körperlich verspannt leistungsbereit energiegeladen					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.000	-.003					
$\sigma(\varepsilon(\pi_1))$.058	.060					
$\langle \varepsilon(\mu) \rangle$	-0,001	-0,006					
$\sigma(\varepsilon(\mu))$	0,103	0,106					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.633	.599					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.266	.198					
$\chi^2(1)$	70,53	39,02					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	303	188	491	258	168	426
	0	179	330	509	233	341	574
	Σ	482	518	1000	491	509	1000

6.3.10 Ganzheitliche Führung

Ganzheitliche Führung ist eine Teilskala des Fragebogens zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Sie fasst aufgabenorientierte (z. B. „Meine Führungskraft nimmt sich ausreichend Zeit, Aufgaben mit mir abzusprechen“) und mitarbeiterorientierte Führung (z. B. „Wir MitarbeiterInnen haben leichten Zugang zu unserem/unserer Vorgesetzten“) zusammen. Die insgesamt sechs Items sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Der Rohwert dieser Variablen ergibt sich durch Mittelwertbildung. In Tabelle 52 sind die deskriptiv-statistischen Kennwerte dieser Variablen zusammengefasst.

Tabelle 52. Deskriptiv-statistische Kennwerte der Variablen Ganzheitliche Führung.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1347	2,67	2,33	2,25	0,80	0,00	4,00	1,61	2,90

Die Werte für Mittelwert (M), Median und Modus steigen in dieser Reihenfolge. Die Verteilung der Werte ist daher linksschief. Die Stichprobe mit gültigem Rohwert für ganzheitliche Führung wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 53 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für ganzheitliche Führung aufgelistet.

Tabelle 53. Korrelationen (Spearman ρ) der Variablen Ganzheitliche Führung mit den einzelnen Items der EEB.

Nr. EEB-Item	ρ	Nr. EEB-Item	ρ	Nr. EEB-Item	ρ
1 konzentriert ^(a)	.173*	2 aggressiv	-.175*	3 nervös ^(a)	-.077
4 körperlich verspannt ^(a)	-.138*	5 selbstsicher	.084	6 zornig	-.189*
7 aufmerksam ^(a)	.185*	8 leistungsbereit ^(a)	.188*	9 entspannt	.037
10 körperlich unwohl ^(a)	-.201*	11 energiegeladen ^(a)	.169*	12 aufgeregt ^(a)	-.026
13 gelangweilt	-.196*	14 beschwingt	.171*	15 fröhlich	.235*
16 ärgerlich	-.191*	17 zufrieden	.378*	18 ängstlich	-.068
19 lustig	.151*	20 leicht	.058	21 deprimiert	-.187*
22 bedrückt	-.191*	23 einflussreich ^(a)	.234*	24 vergnügt	.184*
25 betrübt	-.187*				

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Signifikante Korrelationen sind mit Ausnahme von „zufrieden“ ($\rho = .378$) gering ($.1 < |\rho| \leq .3$). Die stärkste negative Korrelation mit ganzheitlicher Führung weist das EEB-Item „körperlich unwohl“ mit $\rho = -.201$ auf.

Ergebnisse

Tabelle 54. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von ganzheitlicher Führung (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
13, 17	.656	.049	.094	SFS, P	
9, 13, 17	.681	.047	.069	SFS	
9, 11, 13, 17	.757	.043	-.007	SFS	
17	.643	.047	.107	SFS, SFFS, S	
3, 5, 6, 13, 17, 18	.804	.040	-.054	SFS	
1, 6, 13, 17, 18, 22, 23	.838	.038	-.088	SFS	
13, 14, 15	.644	.049	.106	SFFS	
9, 17	.653	.049	.097	P	
17, 21	.655	.046	.095	P	
6, 17	.659	.047	.091	P	
17, 20	.651	.048	.099	P	
11, 17, 20	.715	.045	.035	T	
14, 17, 20	.704	.047	.046	T	
8	.568	.048	.182		SFS, SFFS
1, 3, 4, 10, 11, 12, 23	.929	.026	-.179		SFS
3, 4, 8, 10, 11, 23	.897	.031	-.147		SFS
23	.583	.048	.167		SFS, SFFS, S
3, 4, 11, 23	.743	.043	.007		SFS
3, 4, 23	.652	.048	.098		SFS
3, 4, 7, 8, 10, 11, 23	.944	.023	-.194		SFS
1, 10, 23	.643	.049	.107		SFS
10	.573	.050	.177		SFFS, S
1, 3, 4, 7, 8, 12	.820	.039	-.070		SFFS
1, 3, 4, 7, 8, 10, 12	.867	.034	-.117		SFFS
4, 7, 8, 10	.688	.049	.062		SFFS
3, 4, 7, 8	.705	.046	.045		SFFS
4, 23	.608	.048	.142		P
10, 23	.608	.049	.142		P
4, 7, 23	.653	.047	.097		T
4, 11, 23	.667	.048	.083		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 55. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von ganzheitlicher Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

EEB und WSIB																				
Items	nervös körperlich verspannt energiegeladen einflussreich																			
<i>Bootstrap-Korrektklassifikation</i>																				
$K_{\text{Bootstrap}}$.743																			
$SD[K_{\text{Bootstrap}}]$.043																			
95 %-CI [$K_{\text{Bootstrap}}$]	[.657, .828]																			
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>																				
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.007																			
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.043																			
95 %-CI [$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.078, .093]																			
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.432																			
<i>Auswertung der Vier-Felder-Tafel</i>																				
Cramers ϕ	.495																			
$\chi^2(1)$	24469,77																			
$p[\chi^2(1) > \chi^2(1)]$	< .001																			
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;">1</th> <th style="text-align: center;">0</th> <th style="text-align: center;">Σ</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">(Zeilen: Messwerte, Spalten: Vorhersage)</td> <td style="text-align: center;">1</td> <td style="text-align: center;">41755</td> <td style="text-align: center;">8065</td> <td style="text-align: center;">49820</td> </tr> <tr> <td></td> <td style="text-align: center;">0</td> <td style="text-align: center;">17680</td> <td style="text-align: center;">32500</td> <td style="text-align: center;">50180</td> </tr> <tr> <td></td> <td style="text-align: center;">Σ</td> <td style="text-align: center;">59435</td> <td style="text-align: center;">40565</td> <td style="text-align: center;">100000</td> </tr> </tbody> </table>		1	0	Σ	(Zeilen: Messwerte, Spalten: Vorhersage)	1	41755	8065	49820		0	17680	32500	50180		Σ	59435	40565	100000
	1	0	Σ																	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	41755	8065	49820																
	0	17680	32500	50180																
	Σ	59435	40565	100000																
Korrekturfaktor	c	0,838	1,237																	

Bei Bootstrap-Experimenten zur Vorhersage von ganzheitlicher Führung wird eine für die vollständige EEB sowie das WSIB gemeinsame Item-Kombination (3: „nervös“, 4: „körperlich verspannt“, 11: „energiegeladen“, 23: „einflussreich“) ausgewählt. Dieses Modell weist eine Bootstrap-Korrektklassifikation von .743 auf. Sie ist nicht signifikant kleiner, als die theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 54 und Tabelle 55). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 29 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,45 Skaleneinheiten. Dies entspricht 12 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von ganzheitlicher Führung bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .659

Ergebnisse

korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ausgeprägt ist (Tabelle 56).

Tabelle 56. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von ganzheitlicher Führung (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

EEB und WSIB																	
Items	nervös körperlich verspannt energiegeladen einflussreich																
<i>Mittlere Vorhersagefehler</i>																	
$\langle \varepsilon(\pi_1) \rangle$.002																
$\sigma(\varepsilon(\pi_1))$.046																
$\langle \varepsilon(\mu) \rangle$	0,004																
$\sigma(\varepsilon(\mu))$	0,078																
<i>Bootstrap-Korrektklassifikation für Stichproben</i>																	
$K_{\text{Bootstrap}}$.659																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.318																
$\chi^2(1)$	101,19																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel (Zeilen: Messwerte, Spalten: Vorhersage)	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>1</th> <th>0</th> <th>Σ</th> </tr> </thead> <tbody> <tr> <th>1</th> <td style="background-color: #cccccc;">330</td> <td>175</td> <td>505</td> </tr> <tr> <th>0</th> <td>166</td> <td style="background-color: #cccccc;">329</td> <td>495</td> </tr> <tr> <th>Σ</th> <td>496</td> <td>504</td> <td>1000</td> </tr> </tbody> </table>		1	0	Σ	1	330	175	505	0	166	329	495	Σ	496	504	1000
	1	0	Σ														
1	330	175	505														
0	166	329	495														
Σ	496	504	1000														

6.3.11 Partizipative Unternehmenskultur

Partizipative Unternehmenskultur ist eine Teilskala des Fragebogens zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Sie umfasst Möglichkeiten der Mitbestimmung der Beschäftigten im Unternehmen (z. B. „Wenn man eine gute Idee hat, kann man sie in unserer Firma auch verwirklichen“). Die insgesamt drei Items sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Der Rohwert dieser Variablen ergibt sich durch Mittelwertbildung. In Tabelle 57 sind die deskriptiv-statistischen Kennwerte dieser Variablen zusammengefasst.

Tabelle 57. Deskriptiv-statistische Kennwerte der Variablen Partizipative Unternehmenskultur.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1344	1,75	1,75	1,75	0,82	0,00	4,00	1,10	2,40

Die zentralen Maße (Modus, Median und Mittelwert [M]) weisen identische Werte auf und sind mit 1,75 vergleichsweise niedrig auf der fünfstufigen Skala angesiedelt. Die Stichprobe mit gültigem Rohwert für partizipative Unternehmenskultur wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 58 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für partizipative Unternehmenskultur aufgelistet.

Tabelle 58. Korrelationen (Spearman ρ) der Variablen Partizipative Unternehmenskultur mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	.153*	2	aggressiv	-.173*	3	nervös ^(a)	-.058
4	körperlich verspannt ^(a)	-.159*	5	selbstsicher	.069	6	zornig	-.172*
7	aufmerksam ^(a)	.116*	8	leistungsbereit ^(a)	.139*	9	entspannt	.028
10	körperlich unwohl ^(a)	-.202*	11	energiegeladen ^(a)	.147*	12	aufgeregt ^(a)	-.022
13	gelangweilt	-.221*	14	beschwingt	.190*	15	fröhlich	.213*
16	ärgerlich	-.233*	17	zufrieden	.369*	18	ängstlich	-.059
19	lustig	.123*	20	leicht	.045	21	deprimiert	-.179*
22	bedrückt	-.179*	23	einflussreich ^(a)	.345*	24	vergnügt	.163*
25	betrübt	-.186*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Signifikante Korrelationen mittleren Effekts ($.3 < |\rho| \leq .5$) bestehen mit den Items „zufrieden“ ($\rho = .369$) und „einflussreich“ ($\rho = .345$). Alle weiteren signifikanten Korrelationen sind als kleine Effekte ($.1 < |\rho| \leq .3$) zu interpretieren. Der höchste negative Zusammenhang besteht mit dem Item „ärgerlich“ ($\rho = -.233$).

Ergebnisse

Tabelle 59. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von partizipativer Unternehmenskultur (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
3, 17, 23	.697	.045	.053	SFS	
2, 10, 13, 17, 19, 20, 23	.935	.025	-.185	SFS	
2, 3, 8, 13, 17, 23	.851	.035	-.101	SFS	
13, 17	.646	.048	.104	SFS, SFFS	
17	.636	.046	.114	SFS, SFFS, S	
3, 8, 13, 17, 23, 25	.840	.036	-.090	SFS	
17, 23	.657	.045	.093	SFS, P	
3, 10, 17, 19, 23	.837	.037	-.087	SFS	
3, 13, 14, 16, 17, 23	.874	.033	-.124	SFS	
2, 13, 17, 20	.715	.045	.035	SFFS	
3, 16, 23	.680	.046	.070	SFFS	
4, 13, 14, 17	.734	.045	.016	SFFS	
13, 14, 19, 22, 23	.811	.039	-.061	SFFS	
3, 14, 16, 23	.743	.044	.007	SFFS	
23	.628	.047	.122	SFFS, S	SFS, SFFS, S
3, 23	.648	.049	.102	SFFS	SFS, P
10, 23	.652	.047	.098	P	P
9, 17, 23	.710	.045	.040	T	
9, 17, 20	.714	.045	.036	T	
9, 20, 23	.705	.046	.045	T	
1, 3, 4, 8, 11, 23	.916	.028	-.166		SFS
1, 3, 10, 23	.736	.045	.014		SFS
1, 3, 10, 12, 23	.786	.041	-.036		SFS
1, 4, 7, 8, 10, 12	.835	.037	-.085		SFFS
1, 4	.595	.048	.155		SFFS
1, 4, 7, 8, 10, 11, 12	.912	.029	-.162		SFFS
4, 10, 23	.688	.046	.062		T
4, 11, 23	.687	.046	.063		T
3, 4, 23	.691	.048	.059		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 60. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von partizipativer Unternehmenskultur (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	nervös beschwingt ärgerlich einflussreich	konzentriert nervös körperlich unwohl einflussreich					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.743	.736					
SD[K _{Bootstrap}]	.044	.045					
95 %-CI[K _{Bootstrap}]	[.656, .830]	[.649, .823]					
<i>Vergleich EEB vs. WSIB</i>							
ΔK _{Bootstrap}	.007						
SD[ΔK _{Bootstrap}]	.063						
95 %-CI[ΔK _{Bootstrap}]	[-.116, .130]						
p[ΔK _{Bootstrap} < 0]	.456						
<i>Vergleich mit optimaler Korrektklassifikation K_{Hughes} = .75 nach Hughes</i>							
K _{Hughes} - K _{Bootstrap}	.007	.014					
SD[K _{Hughes} - K _{Bootstrap}]	.044	.045					
95 %-CI[K _{Hughes} - K _{Bootstrap}]	[-.080, .094]	[-.073, .101]					
p[K _{Hughes} < K _{Bootstrap}]	.436	.376					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers φ	.491	.473					
χ ² (1)	24129,25	22330,18					
p[χ ² (1) > χ ² (1)]	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	33572	16558	50130	35360	14646	50006
	0	9161	40709	49870	11766	38228	49994
	Σ	42733	57267	100000	47126	52874	100000
Korrekturfaktor	c	1,173	0,871		1,061	0,946	

Bei Bootstrap-Experimenten zur Vorhersage von partizipativer Unternehmenskultur werden zwei unterschiedliche Item-Kombinationen gewählt. Für die vollständige EEB sind es die Items 3 („nervös“), 14 („beschwingt“), 16 („ärgerlich“) und 23 („einflussreich“). Die Bootstrap-Korrektklassifikation für dieses Modell beträgt .743. Für das WSIB sind es die Items 1 („konzentriert“), 3 („nervös“), 10 („körperlich unwohl“) und 23 („einflussreich“) mit einer Bootstrap-Korrektklassifikation von .736. Die Korrektklassifikationen unterscheiden sich nicht signifikant untereinander und von der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (Tabelle 59 und Tabelle 60). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen im Promillebereich. Das 99,8 %-Konfidenzintervall für

Ergebnisse

diese Vorhersage umfasst einen Bereich von etwa 36 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,5 Skalenpunkten. Dies entspricht 12,5 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von partizipativer Unternehmenskultur bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .647 (EEB) bzw. .650 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ausgeprägt ist (Tabelle 61).

Tabelle 61. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von partizipativer Unternehmenskultur (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	nervös beschwingt ärgert einflussreich	konzentriert nervös körperlich unwohl einflussreich					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.004	.002					
$\sigma(\varepsilon(\pi_1))$.056	.056					
$\langle \varepsilon(\mu) \rangle$	-0,005	0,001					
$\sigma(\varepsilon(\mu))$	0,085	0,084					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.647	.650					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.295	.300					
$\chi^2(1)$	87,10	89,93					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	330	192	522	318	177	495
Spalten: Vorhersage)	0	161	317	478	173	332	505
	Σ	491	509	1000	491	509	1000

6.3.12 Zusammenarbeit

Zusammenarbeit ist eine Teilskala des Fragebogens zu Führung und Zusammenarbeit (FFZ; Scherrer & Wieland, 2006; Wieland & Scherrer, 2007; Wieland, Winizuk & Hammes, 2009). Mit den zugehörigen drei Items wird gefragt, inwiefern sich die Beschäftigten innerhalb einer Abteilung gegenseitig unterstützen und anerkennen (z. B. „Neue Ideen und Problemlösungen geben wir Kolleginnen und Kollegen untereinander weiter“). Die Items sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Der Rohwert dieser Variablen ergibt sich durch Mittelwertbildung. In Tabelle 62 sind die deskriptiv-statistischen Kennwerte dieser Variablen zusammengefasst.

Tabelle 62. Deskriptiv-statistische Kennwerte der Variablen Zusammenarbeit.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1342	3,00	2,50	2,53	0,89	0,00	4,00	1,83	3,24

Modus, Median und Mittelwert (*M*) nehmen in dieser Reihenfolge steigende Werte an. Diese Werte fallen im Vergleich zur fünfstufigen Skala vergleichsweise hoch aus. Die Stichprobe mit gültigem Rohwert für Zusammenarbeit wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 63 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Zusammenarbeit aufgelistet.

Tabelle 63. Korrelationen (Spearman ρ) der Variablen Zusammenarbeit mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	.138*	2	aggressiv	-.141*	3	nervös ^(a)	-.050
4	körperlich verspannt ^(a)	-.058	5	selbstsicher	.104*	6	zornig	-.135*
7	aufmerksam ^(a)	.184*	8	leistungsbereit ^(a)	.206*	9	entspannt	.124*
10	körperlich unwohl ^(a)	-.129*	11	energiegeladen ^(a)	.104*	12	aufgeregt ^(a)	-.060
13	gelangweilt	-.152*	14	beschwingt	.163*	15	fröhlich	.175*
16	ärgerlich	-.133*	17	zufrieden	.240*	18	ängstlich	-.054
19	lustig	.100*	20	leicht	.154*	21	deprimiert	-.092
22	bedrückt	-.131*	23	einflussreich ^(a)	.146*	24	vergnügt	.144*
25	betrübt	-.136*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Signifikante Korrelationen sind praktisch nicht relevant ($|\rho| \leq .1$) („lustig“) oder sie sind als geringe Effekte ($.1 < |\rho| \leq .3$) zu interpretieren. Die stärkste positive Beziehung von Zusammenarbeit besteht mit „zufrieden“ ($\rho = .240$), die stärkste negative mit „gelangweilt“ ($\rho = -.152$).

Ergebnisse

Tabelle 64. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Zusammenarbeit (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
17	.577	.049	.173	SFS, SFFS, S	
13, 14, 20	.665	.048	.085	SFS	
13, 15, 17, 20	.746	.044	.004	SFS	
3, 8, 17, 20	.752	.042	-.002	SFS	
7, 8, 13, 20	.707	.044	.043	SFS	
13, 20	.614	.047	.136	SFS	
7, 11, 13, 15, 17, 20, 22	.933	.024	-.183	SFS	
7, 17, 22	.646	.045	.104	SFS	
7, 13, 14, 16, 20	.806	.039	-.056	SFFS	
7, 11, 13, 14, 16, 19, 20, 21	.965	.019	-.215	SFFS	
11, 13, 14, 15, 20	.829	.038	-.079	SFFS	
3, 13, 17	.634	.049	.116	SFFS	
7, 8, 13, 16, 20, 22	.822	.037	-.072	SFFS	
7, 13, 14, 15, 20, 22	.876	.033	-.126	SFFS	
20	.578	.050	.172	S	
5, 20	.624	.048	.126	P	
17, 20	.623	.049	.127	P	
9, 15, 20	.696	.047	.054	T	
5, 20, 24	.691	.047	.059	T	
5, 17, 20	.689	.047	.061	T	
19, 20, 23	.689	.046	.061	T	
8	.571	.049	.179		SFS, SFFS, S
1	.564	.049	.186		SFS
3, 4, 7, 8, 10, 23	.864	.034	-.114		SFS
3, 8, 10	.623	.048	.127		SFS
3, 7, 8, 10, 23	.767	.042	-.017		SFS
1, 8, 10, 23	.701	.047	.049		SFS
1, 7, 8, 10	.685	.047	.065		SFS
1, 4, 7, 8, 10, 12, 23	.915	.028	-.165		SFS
1, 7, 10, 23	.699	.045	.051		SFFS
1, 7, 10	.622	.048	.128		SFFS
1, 10	.590	.049	.160		SFFS
1, 3, 7, 10, 23	.777	.041	-.027		SFFS
7	.568	.049	.182		S
1, 8	.594	.049	.156		P
8, 10	.590	.048	.160		P
1, 7	.591	.048	.159		P
4, 11, 23	.650	.047	.100		T
1, 4, 7	.642	.047	.108		T
4, 7, 11	.644	.048	.106		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 65. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Zusammenarbeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	gelangweilt fröhlich zufrieden leicht	konzentriert leistungsbereit körperlich unwohl einflussreich					
<i>Bootstrap-Korrektklassifikation</i>							
$K_{\text{Bootstrap}}$.746	.701					
$SD[K_{\text{Bootstrap}}]$.044	.047					
95 %-CI [$K_{\text{Bootstrap}}$]	[.660, .832]	[.609, .793]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.045					
$SD[\Delta K_{\text{Bootstrap}}]$.064					
95 %-CI [$\Delta K_{\text{Bootstrap}}$]		[-.081, .171]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.243					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.004	.049					
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.044	.047					
95 %-CI [$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.082, .090]	[-.043, .141]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.465	.150					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.493	.404					
$\chi^2(1)$	24257,05	16323,70					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	37865	12061	49926	36786	13041	49827
	0	13322	36752	50074	16823	33350	50173
	Σ	51187	48813	100000	53609	46391	100000
Korrekturfaktor	c	0,975	1,026		0,929	1,082	

Bei Bootstrap-Experimenten zur Vorhersage von Zusammenarbeit werden zwei unterschiedliche Item-Kombinationen gewählt. Für die vollständige EEB sind es die Items 13 („gelangweilt“), 15 („fröhlich“), 17 („zufrieden“) und 20 („leicht“). Die Bootstrap-Korrektklassifikation für dieses Modell beträgt .746. Für das WSIB sind es die Items 1 („konzentriert“), 8 („leistungsbereit“), 10 („körperlich unwohl“) und 23 („einflussreich“) mit einer Bootstrap-Korrektklassifikation von .701. Die Korrektklassifikationen unterscheiden sich nicht signifikant untereinander und von der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (Tabelle 64 und Tabelle 65). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen im Promillebereich. Das 99,8 %-Konfidenzintervall für

Ergebnisse

diese Vorhersage umfasst einen Bereich von etwa 33 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,55 Skalenpunkten. Dies entspricht 13,5 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von Zusammenarbeit bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .695 (EEB) bzw. .669 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ausgeprägt ist (Tabelle 66).

Tabelle 66. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Zusammenarbeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	gelangweilt fröhlich zufrieden leicht	konzentriert leistungsbereit körperlich unwohl einflussreich					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.003	.006					
$\sigma(\varepsilon(\pi_1))$.050	.052					
$\langle \varepsilon(\mu) \rangle$	-0,006	0,007					
$\sigma(\varepsilon(\mu))$	0,087	0,091					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
KBootstrap	.695	.669					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.390	.338					
$\chi^2(1)$	151,97	114,12					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	339	151	490	345	161	506
Spalten: Vorhersage)	0	154	356	510	170	324	494
	Σ	493	507	1000	515	485	1000

6.3.13 Gesundheitsbezogene Selbstwirksamkeit

Gesundheitsbezogene Selbstwirksamkeit bzw. Gesundheitskompetenz orientiert sich am Konzept der Selbstwirksamkeit nach Bandura (1977). Die zehn Items des Gesundheitskompetenz-Fragebogens (GKF; Wieland, 2010a; Wieland & Hammes, 2009, 2010) orientieren sich am Fragebogen zur allgemeinen Selbstwirksamkeit von Schwarzer und Jerusalem (1999). Sie wurden für die Domäne gesundheitsbezogenen Verhaltens angepasst (z. B. "Wenn ich Pläne bezüglich meiner Gesundheit mache, dann kann ich diese auch umsetzen.") und sind auf einer fünfstufigen Skala zu beantworten (0 = "Trifft gar nicht zu" bis 4 = "Trifft immer zu"). Die Rohwerte der Skala ergeben sich nach Umpolung zweier Items durch Mittelwertbildung über die Items. Die deskriptiv-statistischen Kennwerte dieser Rohwerte sind in Tabelle 67 zusammengefasst.

Tabelle 67. Deskriptiv-statistische Kennwerte der Variablen Gesundheitsbezogene Selbstwirksamkeit.

<i>N</i>	Modus	Median	<i>M</i>	<i>SD</i>	Min	Max	μ_0	μ_1
1421	2,50	2,50	2,50	0,67	0,00	4,00	1,97	3,03

Modus, Median und Mittelwert (*M*) erhalten durchgehend den Wert 2,50. Die Stichprobe mit gültigem Rohwert für Gesundheitskompetenz wurde für die folgenden Analysen am Median in die Kategorien 0 = „gering“ und 1 = „hoch“ unterteilt. In Tabelle 68 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Gesundheitskompetenz aufgelistet.

Tabelle 68. Korrelationen (Spearman ρ) der Variablen Gesundheitsbezogene Selbstwirksamkeit mit den einzelnen Items der EEB.

Nr. EEB-Item	ρ	Nr. EEB-Item	ρ	Nr. EEB-Item	ρ
1 konzentriert ^(a)	.157*	2 aggressiv	-.133*	3 nervös ^(a)	-.189*
4 körperlich verspannt ^(a)	-.244*	5 selbstsicher	.233*	6 zornig	-.122*
7 aufmerksam ^(a)	.214*	8 leistungsbereit ^(a)	.220*	9 entspannt	-.140*
10 körperlich unwohl ^(a)	-.284*	11 energiegeladen ^(a)	.215*	12 aufgeregt ^(a)	-.147*
13 gelangweilt	-.077	14 beschwingt	.173*	15 fröhlich	.204*
16 ärgerlich	-.155*	17 zufrieden	.248*	18 ängstlich	-.170*
19 lustig	.089	20 leicht	-.095*	21 deprimiert	-.204*
22 bedrückt	-.233*	23 einflussreich ^(a)	.219*	24 vergnügt	.165*
25 betrübt	-.228*				

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Nur die EEB-Items „gelangweilt“ und „lustig“ korrelieren nicht signifikant mit Gesundheitskompetenz. Alle anderen Items weisen einen geringen Effekt ($.1 < |\rho| \leq .3$) auf.

Ergebnisse

Tabelle 69. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von gesundheitsbezogener Selbstwirksamkeit (0 = gering, 1 = hoch). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
5, 10	.657	.046	.093	SFS, P	
10	.620	.049	.130	SFS, SFFS, S	SFS, SFFS, S
1, 4, 5, 10	.742	.045	.008	SFS	
4, 5, 8, 10, 18	.782	.041	-.032	SFS	
1, 4, 5, 8, 10, 18	.831	.037	-.081	SFS	
4, 5, 10, 13, 25	.787	.042	-.037	SFS	
4, 5, 25	.680	.047	.070	SFFS	
4, 5, 18	.665	.047	.085	SFFS	
5, 25	.643	.047	.107	SFFS	
4, 5, 16, 18	.718	.045	.032	SFFS	
5, 9, 20	.689	.047	.061	T	
5, 9, 19	.692	.046	.058	T	
5, 9, 10	.699	.046	.051	T	
4, 5, 20	.691	.048	.059	T	
5, 10, 20	.695	.046	.055	T	
4, 7, 10, 11	.739	.045	.011		SFS
8, 10, 11, 12, 23	.813	.040	-.063		SFS
4, 7, 10, 11, 23	.831	.038	-.081		SFS
3, 4, 7, 8, 10, 11, 23	.942	.024	-.192		SFS
10, 11, 23	.682	.046	.068		SFS, T
1, 7, 10, 11	.716	.045	.034		SFS
4, 7, 8	.657	.047	.093		SFFS
4, 7, 8, 11, 23	.834	.036	-.084		SFFS
4, 11, 23	.683	.045	.067		SFFS, T
4, 7	.626	.050	.124		SFFS
4, 7, 11, 12, 23	.845	.035	-.095		SFFS
4	.589	.051	.161		SFFS
1, 10	.639	.049	.111		P
10, 11	.636	.049	.114		P
8, 10	.636	.048	.114		P

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Bei Bootstrap-Experimenten zur Vorhersage von Gesundheitskompetenz werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (1: „konzentriert“, 4: „körperlich verspannt“, 5: „selbstsicher“, 10: „körperlich unwohl“) und das WSIB (4:

Ergebnisse

„körperlich verspannt“, 7: „aufmerksam“, 10: „körperlich unwohl“, „energiegeladen“) ausgewählt. Das erste Modell ist mit einer Bootstrap-Korrektklassifikation von .742 zwar etwas besser, als das zweite Modell mit .739. Dieser Unterschied ist jedoch nicht signifikant. Diese Bootstrap-Korrektklassifikationen sind nicht signifikant kleiner, als die theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 69 und Tabelle 70).

Tabelle 70. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von gesundheitsbezogener Selbstwirksamkeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	konzentriert körperlich verspannt selbstsicher körperlich unwohl	körperlich verspannt aufmerksam körperlich unwohl energiegeladen					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.742	.739					
SD[K _{Bootstrap}]	.045	.045					
95 %-CI[K _{Bootstrap}]	[.654, .830]	[.651, .826]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.003						
SD[$\Delta K_{\text{Bootstrap}}$]	.063						
95 %-CI[$\Delta K_{\text{Bootstrap}}$]	[-.121, .127]						
$p[\Delta K_{\text{Bootstrap}} < 0]$.479						
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.008	.011					
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.045	.045					
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.080, .096]	[-.076, .099]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.432	.402					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.499	.499					
$\chi^2(1)$	24878,09	24856,40					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	43354	6852	50206	43872	5984	49856
	0	18922	30872	49794	20125	30019	50144
	Σ	62276	37724	100000	63997	36003	100000
Korrekturfaktor	c	0,806	1,320		0,779	1,393	

Ergebnisse

Tabelle 71. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von gesundheitsbezogener Selbstwirksamkeit (0 = gering, 1 = hoch) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	konzentriert körperlich verspannt selbstsicher körperlich unwohl	körperlich verspannt aufmerksam körperlich unwohl energiegeladen					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.001	.004					
$\sigma(\varepsilon(\pi_1))$.044	.064					
$\langle \varepsilon(\mu) \rangle$	0,002	0,005					
$\sigma(\varepsilon(\mu))$	0,060	0,061					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.658	.664					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.318	.334					
$\chi^2(1)$	101,42	111,40					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	318	148	466	328	135	463
Spalten: Vorhersage)	0	194	340	534	201	336	537
	Σ	512	488	1000	529	471	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 28 % bis 41 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,35 Skalenpunkten. Dies entspricht etwa 9 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von Gesundheitskompetenz bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .658 (vollständige EEB) bzw. .664 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich hoch für eine Stichprobe ist (Tabelle 71).

6.3.14 Herz-Kreislauf-Beschwerden

Herz-Kreislauf-Beschwerden wurden mit dem Fragebogen zur Häufigkeit körperlicher Beschwerden (HkB; Wieland, Metz & Richter, 2002) erhoben. Dieses Merkmal umfasst sechs Items (z. B. „Wie oft haben Sie Schmerzen in der Herzgegend?“). Die Items werden auf einer fünfstufigen Ordinalskala (0 = "praktisch nie", 1 = "etwa 2 x im Jahr", 2 = "etwa 2 x im Monat", 3 = "etwa 3 x die Woche" und 4 = "fast täglich") von den Befragten eingeschätzt. Der Rohwert dieses Merkmals ist das Maximum, das bei diesen Items von einer Person angekreuzt wurde. Die deskriptiven Kennwerte dieses Merkmals sind in Tabelle 72 zusammengefasst.

Tabelle 72. Deskriptiv-statistische Kennwerte der Variablen Herz-Kreislauf-Beschwerden.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1306	2,00	2,00	1,78	1,13	0,00	4,00	0,91	2,66

Modus, Median und Mittelwert (M) liegen bei 2 („etwa 2 x im Monat“). Die Stichprobe mit gültigem Rohwert für Herz-Kreislauf-Beschwerden wurde für die folgenden Analysen am Median in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 73 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Herz-Kreislauf-Beschwerden aufgelistet.

Tabelle 73. Korrelationen (Spearman ρ) der Variablen Herz-Kreislauf-Beschwerden mit den einzelnen Items der EEB.

Nr. EEB-Item	ρ	Nr. EEB-Item	ρ	Nr. EEB-Item	ρ
1 konzentriert ^(a)	-.072	2 aggressiv	.149*	3 nervös ^(a)	.210*
4 körperlich verspannt ^(a)	.294*	5 selbstsicher	-.088	6 zornig	.083
7 aufmerksam ^(a)	-.070	8 leistungsbereit ^(a)	-.085	9 entspannt	.075
10 körperlich unwohl ^(a)	.309*	11 energiegeladen ^(a)	-.090	12 aufgeregt ^(a)	.126*
13 gelangweilt	.046	14 beschwingt	-.122*	15 fröhlich	-.148*
16 ärgerlich	.113*	17 zufrieden	-.176*	18 ängstlich	.185*
19 lustig	-.105*	20 leicht	.050	21 deprimiert	.220*
22 bedrückt	.225*	23 einflussreich ^(a)	-.135*	24 vergnügt	-.099*
25 betrübt	.191*				

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit den Herz-Kreislaufbeschwerden. Das Item „körperlich verspannt“ kommt mit einem Korrelationskoeffizienten $|\rho| > .3$ als einziges über einen schwachen Effekt hinaus. Diese beiden genannten Items werden gefolgt von „bedrückt“ und „betrübt“, welche als Facetten von Depressivität interpretiert werden können. Die stärkste negative Korrelation besteht zwischen Herz-Kreislauf-Beschwerden und dem EEB-Item „zufrieden“.

Ergebnisse

Tabelle 74. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Herz-Kreislauf-Beschwerden (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
4, 18, 20, 22	.724	.045	.026	SFS, SFFS	
4, 10, 18, 20	.736	.044	.014	SFS	
1, 7, 9, 10	.720	.046	.030	SFS	
4, 10, 20	.700	.045	.050	SFS	
1, 3, 10, 16, 17, 20	.885	.032	-.135	SFS	
4, 18, 20	.673	.046	.077	SFS	
10	.605	.048	.145	SFS, SFFS, S	SFS, SFFS, S
4, 18	.618	.049	.132	SFS, SFFS	
4, 18, 20, 25	.718	.046	.032	SFFS	
4	.600	.050	.150	SFFS, S	SFFS, S
7, 10, 18	.642	.048	.108	SFFS	
4, 20	.637	.049	.113	P	
4, 10	.632	.048	.118	P	SFS, P
4, 9	.627	.048	.123	P	
4, 21	.627	.049	.123	P	
4, 11, 20	.701	.044	.049	T	
4, 5, 20	.703	.045	.047	T	
4, 20, 24	.704	.048	.046	T	
4, 9, 15	.697	.045	.053	T	
4, 19, 20	.702	.046	.048	T	
3, 4, 10, 11, 12	.800	.041	-.050		SFS
3, 4, 7, 10, 11, 12, 23	.937	.023	-.187		SFS
1, 3, 4, 10	.717	.047	.033		SFS
3, 4, 10	.662	.048	.088		SFS
3, 4, 7, 10, 11	.816	.037	-.066		SFS
1, 3, 10	.639	.047	.111		SFFS
3, 4, 7, 11, 12, 23	.909	.028	-.159		SFFS
3, 7, 10, 12, 23	.778	.042	-.028		SFFS
1, 3, 4	.656	.047	.094		SFFS
3, 4	.626	.049	.124		P
4, 10, 11	.688	.046	.062		T
4, 7, 10	.675	.046	.075		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 75. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Herz-Kreislauf-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	körperlich verspannt körperlich unwohl ängstlich leicht	konzentriert nervös körperlich verspannt körperlich unwohl					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.736	.717					
SD[K _{Bootstrap}]	.044	.047					
95 %-CI[K _{Bootstrap}]	[.649, .823]	[.625, .808]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.019					
SD[$\Delta K_{\text{Bootstrap}}$]		.064					
95 %-CI[$\Delta K_{\text{Bootstrap}}$]		[-.107, .146]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.382					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.014	.033					
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.044	.047					
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.073, .101]	[-.058, .125]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.377	.237					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.490	.452					
$\chi^2(1)$	23988,23	20406,27					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	30528	19740	50268	28621	21277	49898
Spalten: Vorhersage)	0	6658	43074	49732	7054	43048	50102
	Σ	37186	62814	100000	35675	64325	100000
Korrekturfaktor	c	1,352	0,792		1,399	0,779	

Bei Bootstrap-Experimenten zur Vorhersage von Herz-Kreislauf-Beschwerden werden zwei sehr unterschiedliche Item-Kombinationen für die vollständige EEB (4: „körperlich verspannt“, 10: „körperlich unwohl“, 18: „ängstlich“, 20: „leicht“) und das WSIB (1: „konzentriert“, 3: „nervös“, 4: „körperlich verspannt“, 10: „körperlich unwohl“) ausgewählt. Das erste Modell ist mit einer Bootstrap-Korrektklassifikation von .736 zwar etwas besser, als das zweite Modell mit .717. Dieser Unterschied ist jedoch nicht signifikant. Außerdem sind diese Bootstrap-Korrektklassifikationen nicht signifikant kleiner, als die theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 74 und Tabelle 75).

Ergebnisse

Tabelle 76. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Herz-Kreislauf-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	körperlich verspannt körperlich unwohl ängstlich leicht	konzentriert nervös körperlich verspannt körperlich unwohl					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.003	.004					
$\sigma(\varepsilon(\pi_1))$.061	.064					
$\langle \varepsilon(\mu) \rangle$	0,006	0,005					
$\sigma(\varepsilon(\mu))$	0,122	0,122					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.677	.644					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.348	.288					
$\chi^2(1)$	120,96	82,97					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte, Spalten: Vorhersage)	1	390	173	563	322	181	503
	0	150	287	437	175	322	497
	Σ	540	460	1000	497	503	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 40 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,7 Skaleneinheiten. Dies entspricht etwa 18 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von Herz-Kreislauf-Beschwerden bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .677 (vollständige EEB) bzw. .644 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 76).

6.3.15 Magen-Darm-Beschwerden

Magen-Darm-Beschwerden wurden mit dem Fragebogen zur Häufigkeit körperlicher Beschwerden (HkB; Wieland, Metz & Richter, 2002) erhoben. Dieses Merkmal umfasst zwei Items (z. B. „Wie oft reagiert Ihr Magen empfindlich (Durchfall, Verdauungsbeschwerden oder Ähnliches)?“). Die Items werden auf einer fünfstufigen Ordinalskala (0 = "praktisch nie", 1 = "etwa 2 x im Jahr", 2 = "etwa 2 x im Monat", 3 = "etwa 3 x die Woche" und 4 = "fast täglich") von den Befragten eingeschätzt. Der Rohwert dieses Merkmals ist das Maximum, das bei diesen Items von einer Person angekreuzt wurde. Die deskriptiven Kennwerte dieses Merkmals sind in Tabelle 77 zusammengefasst.

Tabelle 77. Deskriptiv-statistische Kennwerte der Variablen Magen-Darm-Beschwerden.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1328	1,00	1,00	1,24	1,07	0,00	4,00	0,44	2,05

Modus und Median sind mit einem Wert von 1,00 niedriger als der Mittelwert (M). Es muss also von einer Rechtsschiefe ausgegangen werden. Die Stichprobe mit gültigem Rohwert für Magen-Darm-Beschwerden wurde für die folgenden Analysen am Median in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 78 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Magen-Darm-Beschwerden aufgelistet.

Tabelle 78. Korrelationen (Spearman ρ) der Variablen Magen-Darm-Beschwerden mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.073	2	aggressiv	.161*	3	nervös ^(a)	.225*
4	körperlich verspannt ^(a)	.294*	5	selbstsicher	-.136*	6	zornig	.152*
7	aufmerksam ^(a)	-.038	8	leistungsbereit ^(a)	-.027	9	entspannt	.102*
10	körperlich unwohl ^(a)	.285*	11	energiegeladen ^(a)	-.093	12	aufgeregt ^(a)	.207*
13	gelangweilt	.050	14	beschwingt	-.099*	15	fröhlich	-.154*
16	ärgerlich	.194*	17	zufrieden	-.142*	18	ängstlich	.194*
19	lustig	-.117*	20	leicht	.126*	21	deprimiert	.218*
22	bedrückt	.250*	23	einflussreich ^(a)	-.125*	24	vergnügt	-.099*
25	betrübt	.228*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit Magen-Darm-Beschwerden. Diese Items kommen mit Korrelationskoeffizienten $|\rho| < .3$ nicht über einen schwachen Effekt hinaus. Die stärkste negative Korrelation besteht mit dem EEB-Item „fröhlich“.

Ergebnisse

Tabelle 79. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Magen-Darm-Beschwerden (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
4, 10, 20	.690	.047	.060	SFS	
4, 10	.638	.048	.112	SFS	SFS, P
4	.614	.050	.136	SFS, SFFS, S	SFS, SFFS, S
3, 4, 18, 20	.741	.043	.009	SFS	
3, 4, 10, 20	.767	.045	-.017	SFS	
3, 4, 10, 11, 20	.853	.034	-.103	SFS	
2, 3, 4, 10, 11, 20	.897	.030	-.147	SFS	
3, 4, 13, 20, 22	.790	.041	-.040	SFFS	
10	.610	.048	.140	SFFS	SFFS
10, 18	.615	.048	.135	SFFS	
5, 10, 20, 22	.728	.043	.022	SFFS	
3, 5, 6, 10, 20	.796	.039	-.046	SFFS	
5, 6, 10, 20	.727	.044	.023	SFFS	
4, 20	.650	.047	.100	P	
4, 17, 20	.721	.043	.029	T	
4, 11, 20	.719	.045	.031	T	
4, 19, 20	.712	.046	.038	T	
3, 4, 8, 10, 11, 12	.866	.034	-.116		SFS
1, 4, 8, 10	.728	.045	.022		SFS
3, 4, 7, 10, 23	.818	.039	-.068		SFS
1, 4, 11, 12, 23	.854	.035	-.104		SFFS
1, 3, 10, 11	.718	.045	.032		SFFS
3, 8, 10, 11, 12, 23	.869	.035	-.119		SFFS
1, 10, 11, 12, 23	.805	.038	-.055		SFFS
1, 3, 8, 10, 11, 12, 23	.929	.026	-.179		SFFS
1, 10, 11	.649	.050	.101		SFFS
3, 10, 11, 12	.709	.046	.041		SFFS
4, 11, 23	.680	.047	.070		T
4, 7, 11	.680	.047	.070		T
4, 7, 10	.680	.048	.070		T
4, 7, 12	.668	.049	.082		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 80. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Magen-Darm-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	nervös körperlich verspannt ängstlich leicht	konzentriert körperlich verspannt leistungsbereit körperlich unwohl					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.741	.728					
SD[K _{Bootstrap}]	.043	.045					
95 %-CI[K _{Bootstrap}]	[.657, .825]	[.640, .817]					
<i>Vergleich EEB vs. WSIB</i>							
ΔK _{Bootstrap}		.012					
SD[ΔK _{Bootstrap}]		.062					
95 %-CI[ΔK _{Bootstrap}]		[-.109, .134]					
p[ΔK _{Bootstrap} < 0]		.420					
<i>Vergleich mit optimaler Korrektklassifikation K_{Hughes} = .75 nach Hughes</i>							
K _{Hughes} - K _{Bootstrap}	.009	.022					
SD[K _{Hughes} - K _{Bootstrap}]	.043	.045					
95 %-CI[K _{Hughes} - K _{Bootstrap}]	[-.075, .093]	[-.067, .110]					
p[K _{Hughes} < K _{Bootstrap}]	.416	.316					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers φ	.487	.470					
χ ² (1)	23745,84	22124,87					
p[χ ² (1) > χ ² (1)]	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	33148	16733	49881	30658	19454	50112
Spalten: Vorhersage)	0	9173	40946	50119	7700	42188	49888
	Σ	42321	57679	100000	38358	61642	100000
Korrekturfaktor	c	1,179	0,869		1,306	0,809	

Bei Bootstrap-Experimenten zur Vorhersage von Magen-Darm-Beschwerden werden zwei unterschiedliche Item-Kombinationen für die vollständige EEB (3: „nervös“, 4: „körperlich verspannt“, 18: „ängstlich“, 20: „leicht“) und das WSIB (1: „konzentriert“, 4: „körperlich verspannt“, 8: „leistungsbereit“, 10: „körperlich unwohl“) ausgewählt. Das erste Modell ist mit einer Bootstrap-Korrektklassifikation von .741 etwas besser, als das zweite Modell mit .728. Dieser Unterschied ist jedoch nicht signifikant. Die Bootstrap-Korrektklassifikationen sind nicht signifikant kleiner, als die theoretisch bestmögliche Korrektklassifikation nach Hughes (1968) (Tabelle 79 und Tabelle 80). Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegen für beide Instrumente im Promillebereich. Das 99,8 %-

Ergebnisse

Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 36 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,7 Skalenpunkten. Dies entspricht etwa 18 % der Gesamtskala, die von 0 bis 4 reicht. Werden Bootstrap-Experimente für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .701 (vollständige EEB) bzw. .670 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 81).

Tabelle 81. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Magen-Darm-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	nervös körperlich verspannt ängstlich leicht	konzentriert körperlich verspannt leistungsbereit körperlich unwohl					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$	-.001	-.003					
$\sigma(\varepsilon(\pi_1))$.055	.058					
$\langle \varepsilon(\mu) \rangle$	-0,002	-0,008					
$\sigma(\varepsilon(\mu))$	0,108	0,108					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
K _{Bootstrap}	.701	.670					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.403	.340					
$\chi^2(1)$	162,52	115,64					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	355	135	490	347	156	503
Spalten: Vorhersage)	0	164	346	510	174	323	497
	Σ	519	481	1000	521	479	1000

6.3.16 Muskel-Skelett-Beschwerden

Muskel-Skelett-Beschwerden wurden mit dem Fragebogen zur Häufigkeit körperlicher Beschwerden (HkB; Wieland, Metz & Richter, 2002) erhoben. Dieses Merkmal umfasst vier Items (z. B. „Wie oft haben Sie Rückenschmerzen?“). Die Items werden auf einer fünfstufigen Ordinalskala (0 = "praktisch nie", 1 = "etwa 2 x im Jahr", 2 = "etwa 2 x im Monat", 3 = "etwa 3 x die Woche" und 4 = "fast täglich") von den Befragten eingeschätzt. Der Rohwert dieses Merkmals ist das Maximum, das bei diesen Items von einer Person angekreuzt wurde. Die deskriptiven Kennwerte dieses Merkmals sind in Tabelle 82 zusammengefasst.

Tabelle 82. Deskriptiv-statistische Kennwerte der Variablen Muskel-Skelett-Beschwerden.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1321	2,00	2,00	2,48	1,19	0,00	4,00	1,47	3,48

Modus und Median sind mit einem Wert von 2,00 niedriger als der Mittelwert (M). Es muss also von einer Rechtsschiefe ausgegangen werden. Die Stichprobe mit gültigem Rohwert für Muskel-Skelett-Beschwerden wurde für die folgenden Analysen am Median in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 83 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Muskel-Skelett-Beschwerden aufgelistet.

Tabelle 83. Korrelationen (Spearman ρ) der Variablen Muskel-Skelett-Beschwerden mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.083	2	aggressiv	.105*	3	nervös ^(a)	.145*
4	körperlich verspannt ^(a)	.505*	5	selbstsicher	-.069	6	zornig	.087
7	aufmerksam ^(a)	-.037	8	leistungsbereit ^(a)	-.062	9	entspannt	.148*
10	körperlich unwohl ^(a)	.312*	11	energiegeladen ^(a)	-.117*	12	aufgeregt ^(a)	.103*
13	gelangweilt	.004	14	beschwingt	-.180*	15	fröhlich	-.179*
16	ärgerlich	.130*	17	zufrieden	-.232*	18	ängstlich	.139*
19	lustig	-.124*	20	leicht	.126*	21	deprimiert	.193*
22	bedrückt	.199*	23	einflussreich ^(a)	-.217*	24	vergnügt	-.157*
25	betrübt	.164*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit den Muskel-Skelett-Beschwerden. Diese Items kommen mit einem Korrelationskoeffizienten $|\rho| > .3$ über einen schwachen Effekt hinaus. Die stärkste negative Korrelation besteht zwischen Muskel-Skelett-Beschwerden und dem EEB-Item „zufrieden“.

Ergebnisse

Tabelle 84. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
1, 4, 18	.717	.045	.033	SFS	
4	.697	.045	.053	SFS, SFFS, S	SFS, SFFS, S
1, 4, 8, 18	.754	.043	-.004	SFS	
1, 4, 8, 9, 18	.833	.037	-.083	SFS	
10	.619	.049	.131	SFFS	SFFS
4, 6	.714	.045	.036	P	
4, 9	.708	.045	.042	P	
4, 14	.715	.045	.035	P	
4, 17, 20	.758	.042	-.008	T	
4, 11, 20	.758	.044	-.008	T	
4, 19, 20	.754	.043	-.004	T	
4, 9, 11	.754	.044	-.004	T	
4, 20, 24	.754	.044	-.004	T	
4, 14, 20	.756	.043	-.006	T	
1, 3, 4, 8, 10	.824	.038	-.074		SFS
1, 3, 4, 10	.771	.040	-.021		SFS
4, 23	.712	.045	.038		P
4, 8	.708	.044	.042		P
4, 8, 23	.737	.044	.013		T
4, 11, 23	.744	.044	.006		T
4, 7, 23	.739	.042	.011		T
4, 7, 11	.741	.044	.009		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Bei Bootstrap-Experimenten zur Vorhersage von Muskel-Skelett-Beschwerden wird sowohl für die vollständige EEB als auch das WSIB dieselbe Item-Kombination (4: „körperlich verspannt“, 11: „energiegeladen“, 23: „einflussreich“) ausgewählt. Diese Item-Kombination erzielt eine Bootstrap-Korrektklassifikation von .744. Sie liegt damit knapp unterhalb der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (Tabelle 84 und Tabelle 85).

Ergebnisse

Tabelle 85. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

EEB und WSIB	
Items	körperlich verspannt energiegeladen einflussreich
<i>Bootstrap-Korrektklassifikation</i>	
$K_{\text{Bootstrap}}$.744
$SD[K_{\text{Bootstrap}}]$.044
95 %-CI[$K_{\text{Bootstrap}}$]	[.659, .829]
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>	
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.006
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.044
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.079, .091]
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.446
<i>Auswertung der Vier-Felder-Tafel</i>	
Cramers ϕ	.488
$\chi^2(1)$	23832,04
$p[\chi^2(1) > \chi^2(1)]$	< .001
Vierfeldertafel	
(Zeilen: Messwerte, Spalten: Vorhersage)	
1	1
0	0
Σ	Σ
Korrekturfaktor	c
	1
	0
	Σ
	1,022
	0,979
	100000

Der mittlere Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegt im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 31 %. Der mittlere Fehler bei der Vorhersage von Erwartungswerten ist nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,7 Skalenpunkten. Dies entspricht etwa 18 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von Muskel-Skelett-Beschwerden bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .672 korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 86).

Ergebnisse

Tabelle 86. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

EEB und WSIB																	
Items	körperlich verspannt energiegeladen einflussreich																
<i>Mittlere Vorhersagefehler</i>																	
$\langle \varepsilon(\pi_1) \rangle$.003																
$\sigma(\varepsilon(\pi_1))$.049																
$\langle \varepsilon(\mu) \rangle$	0,005																
$\sigma(\varepsilon(\mu))$	0,109																
<i>Bootstrap-Korrektklassifikation für Stichproben</i>																	
$K_{\text{Bootstrap}}$.672																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.345																
$\chi^2(1)$	119,24																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel (Zeilen: Messwerte, Spalten: Vorhersage)	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>1</th> <th>0</th> <th>Σ</th> </tr> </thead> <tbody> <tr> <th>1</th> <td style="background-color: #cccccc;">341</td> <td>180</td> <td>521</td> </tr> <tr> <th>0</th> <td>148</td> <td style="background-color: #cccccc;">331</td> <td>479</td> </tr> <tr> <th>Σ</th> <td>489</td> <td>511</td> <td>1000</td> </tr> </tbody> </table>		1	0	Σ	1	341	180	521	0	148	331	479	Σ	489	511	1000
	1	0	Σ														
1	341	180	521														
0	148	331	479														
Σ	489	511	1000														

6.3.17 Unspezifische Beschwerden

Unspezifische Beschwerden wurden mit dem Fragebogen zur Häufigkeit körperlicher Beschwerden (HkB; Wieland, Metz & Richter, 2002) erhoben. Dieses Merkmal umfasst sechs Items (z. B. „Wie oft fühlen Sie sich schnell müde?“). Die Items werden auf einer fünfstufigen Ordinalskala (0 = "praktisch nie", 1 = "etwa 2 x im Jahr", 2 = "etwa 2 x im Monat", 3 = "etwa 3 x die Woche" und 4 = "fast täglich") von den Befragten eingeschätzt. Der Rohwert dieses Merkmals ist das Maximum, das bei diesen Items von einer Person angekreuzt wurde. Die deskriptiven Kennwerte dieses Merkmals sind in Tabelle 87 zusammengefasst.

Tabelle 87. Deskriptiv-statistische Kennwerte der Variablen Unspezifische Beschwerden.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
1304	2,00	3,00	2,65	1,06	0,00	4,00	1,78	3,52

Modus, Median und Mittelwert (M) unterscheiden sich paarweise. Eine klare Tendenz, z. B. der Schiefe, ist auf Grund dieser zentralen Maße nicht abzuleiten. Die Stichprobe mit gültigem Rohwert für unspezifische Beschwerden wurde für die folgenden Analysen am Median in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 88 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für unspezifische Beschwerden aufgelistet.

Tabelle 88. Korrelationen (Spearman ρ) der Variablen Unspezifische Beschwerden mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.093	2	aggressiv	.185*	3	nervös ^(a)	.296*
4	körperlich verspannt ^(a)	.387*	5	selbstsicher	-.139*	6	zornig	.160*
7	aufmerksam ^(a)	-.063	8	leistungsbereit ^(a)	-.089	9	entspannt	.186*
10	körperlich unwohl ^(a)	.353*	11	energiegeladen ^(a)	-.142*	12	aufgeregt ^(a)	.232*
13	gelangweilt	.042	14	beschwingt	-.183*	15	fröhlich	-.203*
16	ärgerlich	.205*	17	zufrieden	-.255*	18	ängstlich	.243*
19	lustig	-.115*	20	leicht	.154*	21	deprimiert	.272*
22	bedrückt	.316*	23	einflussreich ^(a)	-.225*	24	vergnügt	-.150*
25	betrübt	.291*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit den unspezifischen Beschwerden. Diese Items kommen mit einem Korrelationskoeffizienten $|\rho| > .3$ über einen schwachen Effekt hinaus. Dies gilt auch noch für das Item „bedrückt“. Die stärkste negative Korrelation besteht zwischen unspezifischen Beschwerden und dem EEB-Item „zufrieden“.

Ergebnisse

Tabelle 89. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
3, 4, 9, 10, 18, 22, 23	.902	.030	-.152	SFS	
4, 9, 10, 18, 22, 23	.869	.033	-.119	SFS	
4, 10, 18, 22, 23	.787	.041	-.037	SFS	
4	.645	.049	.105	SFS, SFFS, S	SFS, SFFS, S
4, 10, 18, 22	.731	.045	.019	SFS	
3, 4, 10, 22	.745	.044	.005	SFS	
10	.639	.047	.111	SFFS, S	SFFS, S
10, 18, 22	.676	.047	.074	SFFS	
9, 10, 18, 22	.740	.044	.010	SFFS	
10, 18	.651	.048	.099	SFFS	
9, 10, 11, 14, 18, 22	.862	.034	-.112	SFFS	
4, 23	.665	.047	.085	P	P
4, 22	.662	.046	.088	P	
4, 17	.665	.049	.085	P	
4, 9	.662	.047	.088	P	
4, 9, 11	.728	.046	.022	T	
4, 17, 20	.727	.046	.023	T	
4, 9, 15	.723	.044	.027	T	
4, 14, 20	.728	.045	.022	T	
4, 9, 19	.726	.046	.024	T	
4, 10, 11, 23	.779	.041	-.029		SFS
1, 4, 10, 11, 23	.868	.034	-.118		SFS
4, 10, 23	.700	.047	.050		SFS
4, 10	.657	.047	.093		SFS
3, 4, 10, 11, 23	.849	.035	-.099		SFS
1, 3, 4, 7, 10, 11, 23	.945	.023	-.195		SFS
3, 10, 11, 23	.748	.044	.002		SFFS
3, 7, 8, 10, 11, 12, 23	.916	.027	-.166		SFFS
1, 3, 10, 23	.740	.045	.010		SFFS
4, 11	.662	.046	.088		P
4, 11, 23	.721	.043	.029		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

Ergebnisse

Tabelle 90. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

EEB und WSIB																	
Items	nervös körperlich unwohl energiegeladen einflussreich																
<i>Bootstrap-Korrektklassifikation</i>																	
K _{Bootstrap}	.748																
SD[K _{Bootstrap}]	.044																
95 %-CI[K _{Bootstrap}]	[.662, .835]																
<i>Vergleich mit optimaler Korrektklassifikation K_{Hughes} = .75 nach Hughes</i>																	
K _{Hughes} - K _{Bootstrap}	.002																
SD[K _{Hughes} - K _{Bootstrap}]	.044																
95 %-CI[K _{Hughes} - K _{Bootstrap}]	[-.085, .088]																
p [K _{Hughes} < K _{Bootstrap}]	.485																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.509																
$\chi^2(1)$	25910,88																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="border-right: none;"></th> <th style="border-left: none;">1</th> <th style="border-left: none;">0</th> <th style="border-left: none;">Σ</th> </tr> </thead> <tbody> <tr> <td style="border-right: none;">(Zeilen: Messwerte,</td> <td style="background-color: #e0e0e0;">31850</td> <td style="background-color: #e0e0e0;">18089</td> <td style="background-color: #e0e0e0;">49939</td> </tr> <tr> <td style="border-right: none;">Spalten: Vorhersage)</td> <td style="background-color: #e0e0e0;">7078</td> <td style="background-color: #e0e0e0;">42983</td> <td style="background-color: #e0e0e0;">50061</td> </tr> <tr> <td style="border-right: none;">Σ</td> <td style="background-color: #e0e0e0;">38928</td> <td style="background-color: #e0e0e0;">61072</td> <td style="background-color: #e0e0e0;">100000</td> </tr> </tbody> </table>		1	0	Σ	(Zeilen: Messwerte,	31850	18089	49939	Spalten: Vorhersage)	7078	42983	50061	Σ	38928	61072	100000
	1	0	Σ														
(Zeilen: Messwerte,	31850	18089	49939														
Spalten: Vorhersage)	7078	42983	50061														
Σ	38928	61072	100000														
Korrekturfaktor	c	1,283	0,820														

Bei Bootstrap-Experimenten zur Vorhersage von unspezifischen Beschwerden wird sowohl für die vollständige EEB als auch das WSIB dieselbe Item-Kombination (3: „nervös“, 10: „körperlich unwohl“, 11: „energiegeladen“, 23: „einflussreich“) ausgewählt. Diese Item-Kombination erzielt eine Bootstrap-Korrektklassifikation von .748. Sie liegt damit knapp unterhalb der theoretisch bestmöglichen Korrektklassifikation nach Hughes (1968) (Tabelle 89 und Tabelle 90). Der mittlere Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 liegt im Promillebereich. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 35 %. Der mittlere Fehler bei der Vorhersage von Erwartungswerten ist nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 0,7 Skaleneinheiten. Dies entspricht etwa 18 % der Gesamtskala, die von 0 bis 4 reicht. Werden Vorhersagen von unspezifischen Beschwerden bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von

Ergebnisse

.692 korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 91).

Tabelle 91. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Muskel-Skelett-Beschwerden (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

EEB und WSIB																	
Items	nervös körperlich unwohl energiegeladen einflussreich																
<i>Mittlere Vorhersagefehler</i>																	
$\langle \varepsilon(\pi_1) \rangle$	-.001																
$\sigma(\varepsilon(\pi_1))$.055																
$\langle \varepsilon(\mu) \rangle$	-0,002																
$\sigma(\varepsilon(\mu))$	0,108																
<i>Bootstrap-Korrektklassifikation für Stichproben</i>																	
$K_{\text{Bootstrap}}$.692																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.384																
$\chi^2(1)$	147,51																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel (Zeilen: Messwerte, Spalten: Vorhersage)	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 5%;"></th> <th style="width: 20%;">1</th> <th style="width: 20%;">0</th> <th style="width: 15%;">Σ</th> </tr> </thead> <tbody> <tr> <th style="width: 5%;">1</th> <td style="background-color: #d3d3d3;">374</td> <td>173</td> <td>547</td> </tr> <tr> <th style="width: 5%;">0</th> <td>135</td> <td style="background-color: #d3d3d3;">318</td> <td>453</td> </tr> <tr> <th style="width: 5%;">Σ</th> <td>509</td> <td>491</td> <td>1000</td> </tr> </tbody> </table>		1	0	Σ	1	374	173	547	0	135	318	453	Σ	509	491	1000
	1	0	Σ														
1	374	173	547														
0	135	318	453														
Σ	509	491	1000														

6.3.18 Absentismustage

Absentismustage wurden mit dem folgenden Item erhoben: „Wie viele Tage haben Sie in den letzten 12 Monaten auf Grund von Krankheit bei der Arbeit gefehlt?“ Die Befragten konnten die Anzahl der Tage in einem freien Textfeld angeben. In Tabelle 92 sind die deskriptiven Kennwerte dieses Items zusammengefasst.

Tabelle 92. Deskriptiv-statistische Kennwerte der Variablen Absentismustage.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
884	0,00	3,00	7,49	16,74	0,00	215,00	0,83	14,19

Während die meisten Befragten angegeben haben, nicht gefehlt zu haben, gab es auch solche, die recht hohen Absentismus, bis hin zu sämtlichen Werktagen im Jahr, aufwiesen. Die Stichprobe mit gültigen Angaben bei diesem Item wurde für die folgenden Analysen am Median mit drei Absentismustagen in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 93 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Absentismustage aufgelistet.

Tabelle 93. Korrelationen (Spearman ρ) der Variablen Absentismustage mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	-.046	2	aggressiv	.001	3	nervös ^(a)	.051
4	körperlich verspannt ^(a)	.229*	5	selbstsicher	.009	6	zornig	.065
7	aufmerksam ^(a)	-.011	8	leistungsbereit ^(a)	-.067	9	entspannt	.111
10	körperlich unwohl ^(a)	.212*	11	energiegeladen ^(a)	-.099	12	aufgeregt ^(a)	.038
13	gelangweilt	.037	14	beschwingt	-.053	15	fröhlich	-.082
16	ärgerlich	.093	17	zufrieden	-.119*	18	ängstlich	.117*
19	lustig	-.006	20	leicht	.060	21	deprimiert	.171*
22	bedrückt	.207*	23	einflussreich ^(a)	-.080	24	vergnügt	-.053
25	betrückt	.177*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit den angegebenen Absentismustagen. Diese beiden Items werden gefolgt von den Items „bedrückt“ und „betrückt“, welche als Facetten von Depressivität interpretiert werden können. Die stärkste negative Korrelation besteht zwischen Absentismustagen und dem EEB-Item „zufrieden“. Sämtliche Korrelationen kommen jedoch nicht über einen schwachen Effekt ($.1 \leq |\rho| < .3$) hinaus.

Ergebnisse

Tabelle 94. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Absentismustagen (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
4, 11, 17	.697	.046	.053	SFS	
2, 3, 4, 11, 17, 19, 22	.953	.021	-.203	SFS	
4	.601	.049	.149	SFS, SFFS, S	SFS, SFFS, S
4, 11, 22	.689	.047	.061	SFS	
4, 17	.636	.049	.114	SFS, SFFS	
4, 10, 11, 16, 17, 23	.916	.028	-.166	SFS	
3, 4, 11, 17, 22	.855	.035	-.105	SFS	
11, 22	.619	.048	.131	SFFS	
3, 9, 10, 11, 13, 22	.821	.037	-.071	SFFS	
4, 10, 11, 17	.768	.043	-.018	SFFS	
22	.592	.049	.158	SFFS, S	
1, 5, 9, 11, 14, 22	.912	.028	-.162	SFFS	
4, 11	.642	.048	.108	P	P
4, 5, 17	.709	.045	.041	T	
4, 11, 19	.709	.045	.041	T	
4, 5, 19	.711	.044	.039	T	
4, 11, 15	.718	.046	.032	T	
4, 8	.611	.049	.139		SFS
4, 7, 8, 10	.716	.046	.034		SFS
3, 4, 7, 8, 10, 11, 23	.956	.020	-.206		SFS
10	.571	.051	.179		SFFS
10, 11	.598	.048	.152		SFFS
4, 11, 23	.706	.046	.044		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

In Tabelle 94 sind die Ergebnisse von Bootstrap-Experimenten für alle Vorhersage-Modelle (Bayes-Klassifikation) zur Vorhersage von Absentismustagen zusammengefasst. Erwartungsgemäß erzielen Modelle mit steigender Anzahl ausgewählter Items tendenziell eine höhere Bootstrap-Korrektklassifikation. Das Modell, bei dem die Korrektklassifikation am nächsten unterhalb der optimalen Korrektklassifikation liegt, enthält als Auswahl aus der vollständigen EEB die Items 4 („körperlich verspannt“), 11 („energiegeladen“) und 15 („fröhlich“). Bei Auswahl der Items aus dem WSIB lauten die Items 4 („körperlich verspannt“), 7 („aufmerksam“), 8 („leistungsbereit“) und 10 („körperlich unwohl“).

Ergebnisse

Tabelle 95. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Absentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

	EEB	WSIB					
Items	körperlich verspannt energiegeladen fröhlich	körperlich verspannt aufmerksam leistungsbereit körperlich unwohl					
<i>Bootstrap-Korrektklassifikation</i>							
K _{Bootstrap}	.718	.716					
SD[K _{Bootstrap}]	.046	.046					
95 %-CI[K _{Bootstrap}]	[.628, .808]	[.625, .806]					
<i>Vergleich EEB vs. WSIB</i>							
$\Delta K_{\text{Bootstrap}}$.002					
SD[$\Delta K_{\text{Bootstrap}}$]		.065					
95 %-CI[$\Delta K_{\text{Bootstrap}}$]		[-.125, .130]					
$p[\Delta K_{\text{Bootstrap}} < 0]$.486					
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>							
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.032	.034					
SD[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	.046	.046					
95 %-CI[$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.058, .122]	[-.056, .125]					
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.242	.228					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.453	.459					
$\chi^2(1)$	20540,16	21084,61					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel							
(Zeilen: Messwerte, Spalten: Vorhersage)		1	0	Σ	1	0	Σ
	1	29176	20942	50118	27318	22780	50098
	0	7277	42605	49882	5665	44237	49902
	Σ	36453	63547	100000	32983	67017	100000
Korrekturfaktor	c	1,375	0,785		1,519	0,745	

Die beiden ausgewählten Modelle weisen mit .718 (vollständige EEB) und .716 (WSIB) fast identisch hohe Korrektklassifikationen auf und sind nicht signifikant kleiner als die optimal mögliche Korrektklassifikation (Tabelle 95). Die Korrekturfaktoren heben die durchgehend beobachtbare Verschiebung hin zur Klasse 0 („wenige“) bei der Vorhersage von Absentismustagen auf.

Ergebnisse

Tabelle 96. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Absentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

	EEB	WSIB					
Items	körperlich verspannt energiegeladen fröhlich	körperlich verspannt aufmerksam leistungsbereit körperlich unwohl					
<i>Mittlere Vorhersagefehler</i>							
$\langle \varepsilon(\pi_1) \rangle$.000	.001					
$\sigma(\varepsilon(\pi_1))$.064	.069					
$\langle \varepsilon(\mu) \rangle$	0,031	-0,067					
$\sigma(\varepsilon(\mu))$	1,795	1,680					
<i>Bootstrap-Korrektklassifikation für Stichproben</i>							
$K_{\text{Bootstrap}}$.791	.773					
<i>Auswertung der Vier-Felder-Tafel</i>							
Cramers ϕ	.583	.546					
$\chi^2(1)$	339,38	298,32					
$p[\chi^2(1) > \chi^2(1)]$	< .001	< .001					
Vierfeldertafel	1	0	Σ	1	0	Σ	
(Zeilen: Messwerte,	1	374	90	464	377	106	483
Spalten: Vorhersage)	0	119	417	536	121	396	517
	Σ	493	507	1000	498	502	1000

Die mittleren Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 sind identisch Null (EEB) bzw. liegen im Promillebereich (WSIB). Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 41 bis 44 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten sind nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 11 Tagen. Werden Vorhersagen der Absentismustage bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .791 (vollständige EEB) bzw. .773 (WSIB) korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 96).

6.3.19 Präsentismustage

Präsentismustage wurden mit dem folgenden Item erhoben: „Wie viele Tage waren Sie in den letzten 12 Monaten trotz Krankheit arbeiten?“ Die Befragten konnten die Anzahl der Tage in einem freien Textfeld angeben. In Tabelle 97 sind die deskriptiven Kennwerte dieses Items zusammengefasst.

Tabelle 97. Deskriptiv-statistische Kennwerte der Variablen Präsentismustage.

N	Modus	Median	M	SD	Min	Max	μ_0	μ_1
794	0,00	5,00	6,27	8,07	0,00	110,00	1,69	10,86

Während die meisten Befragten angegeben haben, nicht krank gearbeitet zu haben, gab es auch solche, die recht hohen Präsentismus, bis über 100 Werktagen im Jahr, aufwiesen. Die Stichprobe mit gültigen Angaben bei diesem Item wurde für die folgenden Analysen am Median mit fünf Präsentismustagen in die Kategorien 0 = „wenige“ und 1 = „viele“ unterteilt. In Tabelle 98 sind jedoch zunächst die Korrelationen (Spearman ρ) zwischen den Items der EEB und den Rohwerten für Präsentismustage aufgelistet.

Tabelle 98. Korrelationen (Spearman ρ) der Variablen Präsentismustage mit den einzelnen Items der EEB.

Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ	Nr.	EEB-Item	ρ
1	konzentriert ^(a)	.049	2	aggressiv	.068	3	nervös ^(a)	.112
4	körperlich verspannt ^(a)	.280*	5	selbstsicher	.040	6	zornig	.045
7	aufmerksam ^(a)	.079	8	leistungsbereit ^(a)	.071	9	entspannt	.150*
10	körperlich unwohl ^(a)	.288*	11	energiegeladen ^(a)	-.029	12	aufgeregt ^(a)	.110
13	gelangweilt	.004	14	beschwingt	-.056	15	fröhlich	-.064
16	ärgerlich	.104	17	zufrieden	-.140*	18	ängstlich	.109
19	lustig	-.009	20	leicht	.094	21	deprimiert	.171*
22	bedrückt	.181*	23	einflussreich ^(a)	-.078	24	vergnügt	-.068
25	betrübt	.197*						

^(a)Dieses Item gehört auch zum WSIB; * $p < .001$

Erwartungsgemäß korrelieren die beiden Items „körperlich verspannt“ und „körperlich unwohl“ am stärksten positiv mit den angegebenen Präsentismustagen. Diese beiden Items werden gefolgt von den Items „bedrückt“ und „betrübt“, welche als Facetten von Depressivität interpretiert werden können. Die stärkste negative Korrelation besteht zwischen Präsentismustagen und dem EEB-Item „zufrieden“. Sämtliche Korrelationen kommen jedoch nicht über einen schwachen Effekt ($.1 \leq |\rho| < .3$) hinaus.

Ergebnisse

Tabelle 99. Bootstrap-Experimente zur Auswahl geeigneter Bayes-Klassifikatoren. Es wurden je Item-Set 1000 Experimente mit Zufallsstichproben vom Umfang $N = 100$ durchgeführt. Die Modelle dienen zur Vorhersage von Präsentismustagen (0 = wenige, 1 = viele). Fett markiert ist das geeignetste Item-Set aus der vollständigen EEB sowie aus dem Screening (WSIB).

Items	$K_{\text{Bootstrap}}$	$SD[K_{\text{Bootstrap}}]$	$K_{\text{Hughes}} - K_{\text{Bootstrap}}$	EEB	WSIB
8, 9, 10	.673	.048	.077	SFS	
8, 10	.639	.048	.111	SFS	SFS, P
5, 8, 9, 10	.756	.044	-.006	SFS	
6, 8, 10, 24	.714	.044	.036	SFS	
10	.623	.048	.127	SFS, SFFS, S	SFS, SFFS, S
4, 6, 7, 8, 21	.767	.043	-.017	SFFS	
4, 8	.625	.049	.125	SFFS	SFFS
7, 10	.642	.049	.108	P	P
2, 4	.637	.048	.113	P	
4, 21	.629	.049	.121	P	
4, 5, 19	.715	.044	.035	T	
4, 11, 14	.709	.044	.041	T	
4, 11, 19	.707	.045	.043	T	
4, 7, 8, 10	.733	.046	.017		SFS
3, 4, 7, 8, 10, 12	.851	.035	-.101		SFS
4, 7, 8, 11, 23	.867	.034	-.117		SFFS
4, 10, 11	.699	.049	.051		T
4, 11, 23	.692	.047	.058		T

Anmerkungen: $K_{\text{Bootstrap}}$ = Bootstrap-Korrektklassifikation, $K_{\text{Hughes}} = .750$ = Hughes-Schätzung für einen optimalen Klassifikator, S = bestes Einzel-Item, P = bestes Item-Paar, T = bestes Item-Tripel, SFS = Sequential Forward Selection, SFFS = Sequential Forward Floating Selection

In Tabelle 99 sind die Ergebnisse von Bootstrap-Experimenten für alle Vorhersage-Modelle (Bayes-Klassifikation) zur Vorhersage von Präsentismustagen zusammengefasst. Erwartungsgemäß erzielen Modelle mit steigender Anzahl ausgewählter Items tendenziell eine höhere Bootstrap-Korrektklassifikation. Das Modell, bei dem die Korrektklassifikation am nächsten unterhalb der optimalen Korrektklassifikation liegt, enthält als Auswahl aus der vollständigen EEB sowie des WSIB die Items 4 („körperlich verspannt“), 7 („aufmerksam“), 8 („leistungsbereit“) und 10 („körperlich unwohl“). Dieses Modell erzielt eine Bootstrap-Korrektklassifikation von .733, welche nicht signifikant von der theoretisch bestmöglichen Korrektklassifikation abweicht (Tabelle 100).

Ergebnisse

Tabelle 100. Evaluations-Kennwerte des besten Item-Sets zur personenbezogenen Vorhersage von Präsentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB.

EEB und WSIB																									
Items	körperlich verspannt aufmerksam leistungsbereit körperlich unwohl																								
<i>Bootstrap-Korrektklassifikation</i>																									
$K_{\text{Bootstrap}}$.733																								
$SD[K_{\text{Bootstrap}}]$.046																								
95 %-CI [$K_{\text{Bootstrap}}$]	[.643, .823]																								
<i>Vergleich mit optimaler Korrektklassifikation $K_{\text{Hughes}} = .75$ nach Hughes</i>																									
$K_{\text{Hughes}} - K_{\text{Bootstrap}}$.017																								
$SD[K_{\text{Hughes}} - K_{\text{Bootstrap}}]$.046																								
95 %-CI [$K_{\text{Hughes}} - K_{\text{Bootstrap}}$]	[-.073, .107]																								
$p[K_{\text{Hughes}} < K_{\text{Bootstrap}}]$.355																								
<i>Auswertung der Vier-Felder-Tafel</i>																									
Cramers ϕ	.481																								
$\chi^2(1)$	23161,08																								
$p[\chi^2(1) > \chi^2(1)]$	< .001																								
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;">1</th> <th style="text-align: center;">0</th> <th style="text-align: center;">Σ</th> </tr> </thead> <tbody> <tr> <td style="text-align: left;">(Zeilen: Messwerte, Spalten: Vorhersage)</td> <td style="text-align: center;">1</td> <td style="text-align: center;">30448</td> <td style="text-align: center;">19629</td> <td style="text-align: center;">50077</td> </tr> <tr> <td></td> <td style="text-align: center;">0</td> <td style="text-align: center;">7087</td> <td style="text-align: center;">42836</td> <td style="text-align: center;">49923</td> </tr> <tr> <td></td> <td style="text-align: center;">Σ</td> <td style="text-align: center;">37535</td> <td style="text-align: center;">62465</td> <td style="text-align: center;">100000</td> </tr> <tr> <td style="text-align: left;">Korrekturfaktor</td> <td style="text-align: center;">c</td> <td style="text-align: center;">1,334</td> <td style="text-align: center;">0,799</td> <td></td> </tr> </tbody> </table>		1	0	Σ	(Zeilen: Messwerte, Spalten: Vorhersage)	1	30448	19629	50077		0	7087	42836	49923		Σ	37535	62465	100000	Korrekturfaktor	c	1,334	0,799	
	1	0	Σ																						
(Zeilen: Messwerte, Spalten: Vorhersage)	1	30448	19629	50077																					
	0	7087	42836	49923																					
	Σ	37535	62465	100000																					
Korrekturfaktor	c	1,334	0,799																						

Der mittlere Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 ist identisch Null. Das 99,8 %-Konfidenzintervall für diese Vorhersage umfasst einen Bereich von etwa 41 %. Der mittlere Fehler bei der Vorhersage von Erwartungswerten ist nahe Null. Hier umfasst das 99,8 %-Konfidenzintervall einen Bereich von etwa 11 Tagen. Werden Vorhersagen der Präsentismustage bei Bootstrap-Experimenten für 1000 zufällige Stichproben mit einem Umfang von je 100 Instanzen durchgeführt, so kann mit einer Wahrscheinlichkeit von .662 korrekt vorhergesagt werden, ob diese über- oder unterdurchschnittlich häufig für eine Stichprobe auftreten (Tabelle 101).

Ergebnisse

Tabelle 101. Evaluative Kennwerte des besten Item-Sets zur stichprobenbezogenen Vorhersage von Präsentismustagen (0 = wenige, 1 = viele) mittels Items aus der vollständigen EEB sowie des WSIB. Mittels Bootstrapping wurden 1000 Stichproben des Umfangs $N = 100$ gezogen. Dargestellt sind die mittleren Vorhersagefehler und ihre Standardabweichungen für die Wahrscheinlichkeit der Klasse 1 sowie für Erwartungswerte. Ausgewertet wurde außerdem die Wahrscheinlichkeit, mit der der Schluss von vorhergesagter Über- oder Unterschreitung der Basisrate (50 %) für die Klasse 1 auf die entsprechende gemessene Über- oder Unterschreitung zulässig ist.

EEB und WSIB																	
Items	körperlich verspannt aufmerksam leistungsbereit körperlich unwohl																
<i>Mittlere Vorhersagefehler</i>																	
$\langle \varepsilon(\pi_1) \rangle$.000																
$\sigma(\varepsilon(\pi_1))$.064																
$\langle \varepsilon(\mu) \rangle$	0,031																
$\sigma(\varepsilon(\mu))$	1,795																
<i>Bootstrap-Korrektklassifikation für Stichproben</i>																	
$K_{\text{Bootstrap}}$.662																
<i>Auswertung der Vier-Felder-Tafel</i>																	
Cramers ϕ	.325																
$\chi^2(1)$	105,68																
$p[\chi^2(1) > \chi^2(1)]$	< .001																
Vierfeldertafel	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;"></th> <th style="width: 20%; text-align: center;">1</th> <th style="width: 20%; text-align: center;">0</th> <th style="width: 55%; text-align: center;">Σ</th> </tr> </thead> <tbody> <tr> <td style="width: 5%; text-align: center;">1</td> <td style="text-align: center;">321</td> <td style="text-align: center;">153</td> <td style="text-align: center;">474</td> </tr> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">185</td> <td style="text-align: center;">341</td> <td style="text-align: center;">526</td> </tr> <tr> <td style="text-align: center;">Σ</td> <td style="text-align: center;">506</td> <td style="text-align: center;">494</td> <td style="text-align: center;">1000</td> </tr> </tbody> </table>		1	0	Σ	1	321	153	474	0	185	341	526	Σ	506	494	1000
	1	0	Σ														
1	321	153	474														
0	185	341	526														
Σ	506	494	1000														
(Zeilen: Messwerte, Spalten: Vorhersage)																	

6.4 VORHERSAGEFEHLER IN ABHÄNGIGKEIT DES STICHPROBENUMFANGES

In den Ergebnisberichten oben wurden mittlere Fehler $\langle \varepsilon(\mu) \rangle$ bei der Vorhersage von Erwartungswerten sowie die zugehörigen Standardabweichungen $\sigma(\varepsilon(\mu))$ angegeben. Insbesondere für die Standardabweichungen ist zu erwarten, dass diese mit steigendem Stichprobenumfang n sinken. Mit Abbildung 17 wird diese Erwartung am Beispiel von Aufgabenanforderungen empirisch bestätigt. Dieser Abbildung liegen Bootstrapping-Experimente mit $h = 1\,000$ Iterationen für Stichprobenumfänge zwischen $n = 10$ und $n = 100$ Instanzen zugrunde.

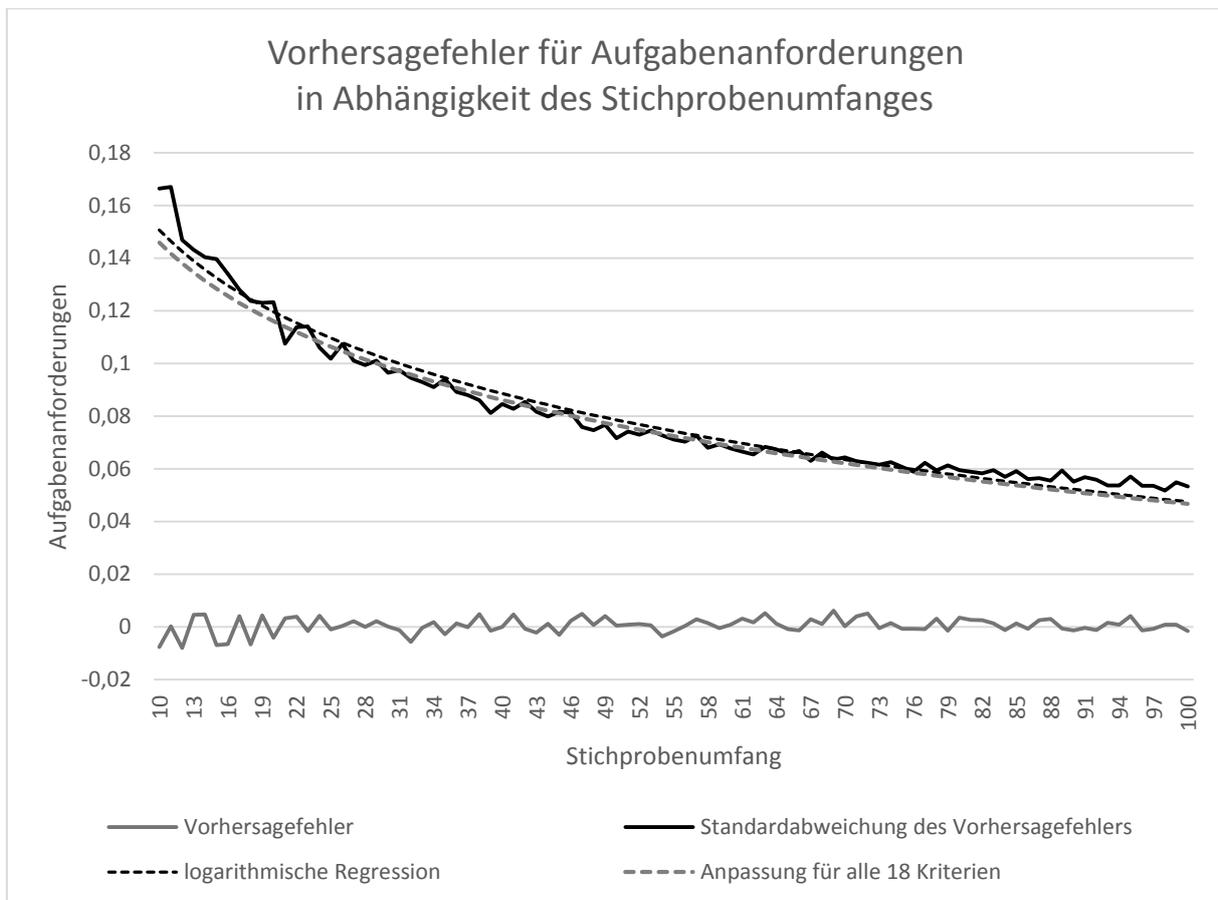


Abbildung 17. Vorhersagefehler und die Standardabweichung dieses Fehlers in Abhängigkeit des Stichprobenumfanges am Beispiel von Aufgabenanforderungen. Dargestellt sind auch die Anpassung an die empirischen Daten durch logarithmische Regression sowie die Anpassung, die sich aus der Betrachtung aller 18 Kriterien ergibt.

Die Standardabweichung des Vorhersagefehlers in Abhängigkeit des Stichprobenumfanges lässt sich für alle 18 Kriterien logarithmisch anpassen (logarithmische Regression). D. h. es gilt für alle 18 Kriterien ein Zusammenhang der Form

$$\sigma_n(\varepsilon(\mu)) = \sigma(X)(a_X - b_X \ln n)$$

mit den Konstanten a_X, b_X sowie dem Stichprobenumfang $n \in \{10..100\}$ bzw. dessen natürlichem Logarithmus $\ln n$. $\sigma(X)$ ist die Standardabweichung des Kriteriums X , wie sie sich

Ergebnisse

für die Induktionsstichproben ergibt (siehe Kennwert SD bei den deskriptiv-statistischen Angaben in den Ergebnisberichten oben). In Tabelle 102 sind die Parameter und Fit-Indizes dieser logarithmischen Regressionen für alle 18 Kriterien zusammen gestellt.

Tabelle 102. Parameter und Fit-Indizes der logarithmischen Regressionen der Standardabweichungen von Vorhersagefehlern in Abhängigkeit des Stichprobenumfangs $n \in \{10..100\}$ für alle 18 Kriterien.

Kriterium (X)	$\sigma(X)$	a_X	b_X	R^2	F^*
Aufgabenanforderungen	0,51	0,498	0,0878	.985	2930,01
Kommunikation und Kooperation	0,48	0,531	0,0940	.985	2854,05
Leistungs- und Zeitvorgaben	0,59	0,469	0,0831	.987	3237,24
Regulationsbehinderungen	0,47	0,526	0,0926	.987	3277,02
Tätigkeitsspielraum	0,52	0,451	0,0798	.986	3091,67
Gesamtbedarf für Arbeitsgestaltung	1,32	0,479	0,0849	.984	2762,72
Autoritäre Führung	0,79	0,489	0,0863	.985	2846,64
Bedeutung von Anerkennung und Wertschätzung	0,94	0,534	0,0946	.985	2975,24
Ganzheitliche Führung	0,80	0,459	0,0815	.983	2546,17
Partizipative Unternehmenskultur	0,82	0,475	0,0839	.985	2959,22
Zusammenarbeit	0,89	0,481	0,0846	.985	2950,28
Gesundheitsbezogene Selbstwirksamkeit	0,67	0,434	0,0767	.986	3107,91
Herz-Kreislauf-Beschwerden	1,13	0,516	0,0909	.987	3341,94
Magen-Darm-Beschwerden	1,07	0,490	0,0866	.987	3197,58
Muskel-Skelett-Beschwerden	1,19	0,465	0,0823	.982	2461,62
Unspezifische Beschwerden	1,06	0,504	0,0893	.986	3049,17
Absentismustage	16,74	0,482	0,0849	.984	2684,80
Präsentismustage	8,07	0,472	0,0829	.985	2953,68

* $df_1 = 1, df_2 = 88, p < .001$

Tabelle 102 ist zu entnehmen, dass die Parameter a_X und b_X für alle 18 Kriterien eine vergleichbare Höhe aufweisen. D. h. es liegt eine gesetzmäßige Abhängigkeit zwischen allen beteiligten Parametern vor. Um gemeinsame Parameter a und b für alle 18 Kriterien zu bestimmen, wurden lineare Regressionen mit $\sigma(X)$ als abhängiger und $\sigma(X)/a_X$ bzw. $\sigma(X)/b_X$ als unabhängiger Variable durchgeführt. Der gemeinsame Parameter a ergibt sich zu $a = 0,480$ mit einer Standardabweichung von $\sigma(a) = 0,002$ ($1 - R^2 = 1 \cdot 10^{-4}, F(1, 16) = 76134,40, p < .001$). Der gemeinsame Parameter b ergibt sich zu $b = 0,0844$ mit einer Standardabweichung von $\sigma(b) = 0,0003$ ($1 - R^2 = 1 \cdot 10^{-4}, F(1, 16) = 71821,71, p < .001$). Die konstanten Terme dieser Regressionen sowie die Fehler der ermittelten gemeinsamen Konstanten sind vernachlässigbar. Somit lassen sich die Standardabweichungen

der Vorhersagefehler für Erwartungswerte für alle 18 Kriterien wie folgt in Abhängigkeit des Stichprobenumfanges $n \in \{10..100\}$ bestimmen:

$$\sigma_n(\varepsilon(\mu)) = \sigma(X)(0,48 - 0,0844 \ln n) = 0,1688 \sigma(X) \ln\left(\frac{17,18}{\sqrt{n}}\right).$$

6.5 VERGLEICH DER KLASSIFIKATOR-INDUKTOREN

Für die Klassifikator-Induktion kamen vollständige Suche in Verbindung mit Bayes-Klassifikation, Sequential Forward Selection in Verbindung mit binärlogistischer Regression und Bayes-Klassifikation und Sequential Forward Floating Selection in Verbindung mit Bayes-Klassifikation zum Einsatz. Es stellt sich die Frage, ob eines dieser Verfahren den anderen in Bezug auf die Klassifikationsgüte überlegen ist. Diese Frage kann durch paarweisen Vergleich der entsprechenden Korrektklassifikationen (χ^2 -Test) je Kriterium beantwortet werden. Hierbei beschränke ich mich im Folgenden auf den Vergleich der Klassifikation in Bezug auf die vollständige EEB. Vollständige Suche wird ausschließlich mit Sequential Forward Selection (SFS) verglichen, da dieses Verfahren sowohl in Verbindung mit binärlogistischer Regression als auch mit Bayes-Klassifikation zum Einsatz kam. Wird vollständige Suche mit SFS verglichen, so wird jeweils die höchste erzielte Cross-Validation-Korrektklassifikation (bestes Einzelitem, bestes Paar oder bestes Tripel) herangezogen. Binärlogistische Regression wird ausschließlich mit Bayes-Klassifikation in Verbindung mit SFS verglichen, da Sequential Forward Floating Selection (SFFS) nicht in Verbindung mit binärlogistischer Regression verwendet wurde. Bei dem Vergleich von SFS mit SFFS handelt es sich dementsprechend immer um den Vergleich von Bayes-Klassifikatoren.

Der Vergleich zweier Verfahren erfolgt durch einen χ^2 -Test je Kriterium. Hierzu werden die Stichprobenumfänge je Kriterium, die beiden jeweiligen Cross-Validation-Korrektklassifikationen, ihre Differenzen mit der zugehörigen Standardabweichungen und der Schätzwert für χ^2 genannt. Das Signifikanzniveau wird mit $\alpha = .05$ festgelegt. Die Schätzwerte für χ^2 werden entsprechend markiert.

In den folgenden Tabellen sind die Ergebnisse der Vergleiche zusammengefasst: (1) Vergleich von Bayes-Klassifikation in Verbindung mit vollständiger Suche und mit SFS (Tabelle 103), (2) Vergleich von Bayes-Klassifikation in Verbindung mit vollständiger Suche und binärlogistischer Regression in Verbindung mit SFS (Tabelle 104), (3) Vergleich von Bayes-Klassifikation mit binärlogistischer Regression in Verbindung mit SFS (Tabelle 105) und (4) Vergleich von SFS mit SFFS in Verbindung mit Bayes-Klassifikation (Tabelle 106).

Ergebnisse

Tabelle 103. Vergleich der Cross-Validation-Korrektklassifikationen bei Bayes-Klassifikation in Kombination mit vollständiger Suche und mit Sequential Forward Selection (SFS).

Kriterium	N	CV-Korrektklassifikation		Differenz	SD	X ² (1)
		vollständige Suche	SFS			
Aufgabenanforderungen	1542	.610	.624	-.014	.018	0,64
Kommunikation und Kooperation	1530	.533	.534	-.001	.018	0,00
Leistungs- und Zeitvorgaben	1549	.599	.562	.037	.018	4,36*
Regulationsbehinderungen	1532	.620	.587	.033	.018	3,49
Tätigkeitsspielraum	1539	.604	.624	-.020	.018	1,30
Gesamtbedarf für Arbeitsgestaltung	1464	.614	.624	-.010	.018	0,31
Autoritäre Führung	1471	.588	.597	-.009	.018	0,25
Bedeutung von Anerkennung und Wertschätzung	1330	.578	.633	-.055	.019	8,45*
Ganzheitliche Führung	1457	.641	.658	-.017	.018	0,93
Partizipative Unternehmenskultur	523	.628	.642	-.014	.030	0,22
Zusammenarbeit	1456	.584	.594	-.010	.018	0,30
Gesundheitsbezogene Selbstwirksamkeit	1420	.649	.606	.043	.018	5,63*
Herz-Kreislauf-Beschwerden	1305	.580	.618	-.038	.019	3,93*
Magen-Darm-Beschwerden	1327	.629	.598	.031	.019	2,69
Muskel-Skelett-Beschwerden	1320	.696	.645	.051	.018	7,79*
Unspezifische Beschwerden	1303	.636	.684	-.048	.019	6,71*
Absentismustage	884	.604	.643	-.039	.023	2,87
Präsentismustage	794	.616	.651	-.035	.024	2,10

* $p < .05$

Bayes-Klassifikation in Kombination mit SFS führt in 13 Fällen zu einer höheren Cross-Validation-Korrektklassifikation, als in Kombination mit vollständiger Suche. Bei drei dieser Fälle ist dieser Unterschied statistisch bedeutsam. Allerdings führt auch vollständige Suche in drei Fällen zu einer signifikant höheren Cross-Validation-Korrektklassifikation, als SFS (Tabelle 103).

Ergebnisse

Tabelle 104. Vergleich der CV-Korrektklassifikationen bei binärlogistischer Regression in Kombination mit Sequential Forward Selection (BLR/SFS) und Bayes-Klassifikation mit vollständiger Suche.

Kriterium	N	CV-Korrektklassifikation			SD	X ² (1)
		vollständige Suche	BLR/SFS	Differenz		
Aufgabenanforderungen	1542	.610	.605	.005	.018	0,08
Kommunikation und Kooperation	1530	.533	.582	-.049	.018	7,46*
Leistungs- und Zeitvorgaben	1549	.599	.602	-.003	.018	0,03
Regulationsbehinderungen	1532	.620	.603	.017	.018	0,93
Tätigkeitsspielraum	1539	.604	.613	-.009	.018	0,26
Gesamtbedarf für Arbeitsgestaltung	1464	.614	.617	-.003	.018	0,03
Autoritäre Führung	1471	.588	.591	-.003	.018	0,03
Bedeutung von Anerkennung und Wertschätzung	1330	.578	.656	-.078	.019	17,23*
Ganzheitliche Führung	1457	.641	.650	-.009	.018	0,26
Partizipative Unternehmenskultur	523	.628	.667	-.039	.030	1,75
Zusammenarbeit	1456	.584	.598	-.014	.018	0,59
Gesundheitsbezogene Selbstwirksamkeit	1420	.649	.632	.017	.018	0,89
Herz-Kreislauf-Beschwerden	1305	.580	.581	-.001	.019	0,00
Magen-Darm-Beschwerden	1327	.629	.631	-.002	.019	0,01
Muskel-Skelett-Beschwerden	1320	.696	.696	-.000	.018	0,00
Unspezifische Beschwerden	1303	.636	.664	-.028	.019	2,25
Absentismustage	884	.604	.617	-.013	.023	0,31
Präsentismustage	794	.616	.647	-.031	.024	1,64

* $p < .05$

Binärlogistische Regression in Verbindung mit Sequential Forward Selection führt in 15 Fällen zu einer höheren Cross-Validation-Korrektklassifikation als Bayes-Klassifikation in Verbindung mit vollständiger Suche. Diese Unterschiede werden jedoch nur in zwei Fällen signifikant (Tabelle 104).

Ergebnisse

Tabelle 105. Vergleich der Cross-Validation-Korrektklassifikationen bei binärlogistischer Regression (BLR) und Bayes-Klassifikation in Verbindung mit Sequential Forward Selection.

Kriterium	N	CV-Korrektklassifikation			SD	X ² (1)
		Bayes	BLR	Differenz		
Aufgabenanforderungen	1542	.624	.605	.019	.018	1,18
Kommunikation und Kooperation	1530	.534	.582	-.048	.018	7,16*
Leistungs- und Zeitvorgaben	1549	.562	.602	-.040	.018	5,10*
Regulationsbehinderungen	1532	.587	.603	-.016	.018	0,81
Tätigkeitsspielraum	1539	.624	.613	.011	.018	0,39
Gesamtbedarf für Arbeitsgestaltung	1464	.624	.617	.007	.018	0,15
Autoritäre Führung	1471	.597	.591	.006	.018	0,11
Bedeutung von Anerkennung und Wertschätzung	1330	.633	.656	-.023	.019	1,54
Ganzheitliche Führung	1457	.658	.650	.008	.018	0,21
Partizipative Unternehmenskultur	523	.642	.667	-.025	.029	0,72
Zusammenarbeit	1456	.594	.598	-.004	.018	0,05
Gesundheitsbezogene Selbstwirksamkeit	1420	.606	.632	-.026	.018	2,04
Herz-Kreislauf-Beschwerden	1305	.618	.581	.037	.019	3,73
Magen-Darm-Beschwerden	1327	.598	.631	-.033	.019	3,05
Muskel-Skelett-Beschwerden	1320	.645	.696	-.051	.018	7,79*
Unspezifische Beschwerden	1303	.684	.664	.020	.018	1,19
Absentismustage	884	.643	.617	.026	.023	1,28
Präsentismustage	794	.651	.647	.004	.024	0,03

* $p < .05$

Binärlogistische Regression und Bayes-Klassifikationen führen in Verbindung mit Sequential Forward Selection in jeweils neun Fällen zu einer höheren Cross-Validation-Korrektklassifikation. Signifikante Unterschiede finden sich in drei Fällen für eine höhere Cross-Validation-Korrektklassifikation bei binärlogistischer Regression (Tabelle 105).

Ergebnisse

Tabelle 106. Vergleich der Cross-Validation-Korrektklassifikationen bei Bayes-Klassifikation in Verbindung mit Sequential Forward Selection (SFS) und Sequential Forward Floating Selection (SFFS).

Kriterium	N	CV-Korrektklassifikation		Differenz	SD	X ² (1)
		SFS	SFFS			
Aufgabenanforderungen	1542	.624	.625	-.001	.017	0,00
Kommunikation und Kooperation	1530	.534	.590	-.056	.018	9,78*
Leistungs- und Zeitvorgaben	1549	.562	.623	-.061	.018	11,98*
Regulationsbehinderungen	1532	.587	.652	-.065	.018	13,79*
Tätigkeitsspielraum	1539	.624	.633	-.009	.017	0,27
Gesamtbedarf für Arbeitsgestaltung	1464	.624	.616	.008	.018	0,20
Autoritäre Führung	1471	.597	.602	-.005	.018	0,08
Bedeutung von Anerkennung und Wertschätzung	1330	.633	.636	-.003	.019	0,03
Ganzheitliche Führung	1457	.658	.648	.010	.018	0,32
Partizipative Unternehmenskultur	523	.642	.676	-.034	.029	1,35
Zusammenarbeit	1456	.594	.614	-.020	.018	1,22
Gesundheitsbezogene Selbstwirksamkeit	1420	.606	.669	-.063	.018	12,25*
Herz-Kreislauf-Beschwerden	1305	.618	.629	-.011	.019	0,34
Magen-Darm-Beschwerden	1327	.598	.658	-.060	.019	10,26*
Muskel-Skelett-Beschwerden	1320	.645	.707	-.062	.018	11,63*
Unspezifische Beschwerden	1303	.684	.675	.009	.018	0,24
Absentismustage	884	.643	.649	-.006	.023	0,07
Präsentismustage	794	.651	.646	.005	.024	0,04

* $p < .05$

Bayes-Klassifikation in Verbindung mit Sequential Forward Floating Selection führt in 14 Fällen zu einer besseren Cross-Validation-Korrektklassifikation als in Verbindung mit Sequential Forward Selection. In sechs dieser Fälle ist der Unterschied signifikant (Tabelle 106).

Zusammenfassend ist festzustellen, dass das heuristische Selektionsverfahren Sequential Forward Selection unabhängig vom Klassifikator (Bayes oder binärlogistische Regression) geringfügig bessere Cross-Validation-Korrektklassifikationen erwarten lässt als vollständige Suche (Tabelle 103 und Tabelle 104). Auch, wenn auf binärlogistische Regression drei Fälle signifikant höherer Cross-Validation-Korrektklassifikationen im Vergleich mit Bayes-Klassifikation zurückzuführen sind, wird dieser Vergleich als unentschieden bewertet. Beide Verfahren erzielen gleich häufig eine etwas höhere Cross-Validation-Korrektklassifikation (Tabelle 105). Der Vergleich der beiden heuristischen Selektionsverfahren, Sequential Forward Selection (SFS) und Sequential Forward Floating Selection (SFFS), fällt hingegen deutlicher aus: SFFS ist sowohl hinsichtlich der Häufigkeit, mit der es höhere Cross-Validation-Korrektklassifikationen erzielt, als auch hinsichtlich signifikant besserer Ergebnisse SFS überlegen (Tabelle 106).

6.6 VALIDIERUNG ENDGÜLTIG AUSGEWÄHLTER KLASSIFIKATOREN

Um die endgültig ausgewählten Klassifikatoren zu validieren, werden diese auf die Validierungsstichproben angewendet. Die Validierungsstichproben sind in keines der oben beschriebenen Experimentalreihen zur Klassifikator-Induktion eingeflossen. Sie stellen somit für die Klassifikatoren völlig unbekannte Datensätze dar. Sie enthalten außerdem die meisten (Hochtechnologieunternehmen) bzw. alle (Wasserversorger) Kriterien. Da jedoch nur die Items aus dem WSIB enthalten sind, können nur die Klassifikatoren bezüglich des WSIB validiert werden. Nach Aufbereitung und Bereinigung enthält die Stichprobe des Wasserversorgers zwischen $N = 117$ (Präsentismustage) und $N = 150$ (ganzheitliche Führung) Instanzen. Diese Umfänge und alle weiteren deskriptiv-statistischen Kennwerte der Kriterien für diese Stichprobe sind in Tabelle 107 zusammengefasst. Außerdem findet sich dort je Kriterium ein Vergleich mit den Induktionsstichproben (t -Test für zwei unabhängige Stichproben).

Die Ergebnisse der Validierung werden in Form zweier Profile dargestellt. Eines bezieht sich auf die Vorhersage der Wahrscheinlichkeit für Klasse 1 (Abbildung 18), die andere bezieht sich auf die Vorhersage von Erwartungswerten (Abbildung 19). In beiden Profilen sind die gemessenen Werte (obere Markierung), die vorhergesagten Werte (untere, starke Markierung) und die 99,8 %-Konfidenzintervalle (seitliche, schwache Markierungen) für jedes Kriterium enthalten. Die Messungen ergeben sich direkt aus den Daten. Die vorhergesagten Werte und Konfidenzintervalle werden berechnet, wie im Teil zur Methodik beschrieben. Aus den Profilen lässt sich ablesen, ob die gemessenen Werte innerhalb der vorhergesagten 99,8 %-Konfidenzintervalle liegen.

Ergebnisse

Tabelle 107. Deskriptiv-statistische Kennwerte der Validierungsstudie und Vergleich mit den Induktionsstichproben (t-Test für zwei unabhängige Stichproben). Die genannten statistischen Kennwerte sind wie für die Induktionsstichproben definiert.

Kriterium	Validierungsstichprobe							Induktionsstichproben			Vergleich
	N	Min	Max	Modus	Median	M	SD	N	M	SD	T
Aufgabenanforderungen	146	0,00	2,00	1,00	1,00	1,06	0,49	1544	1,07	0,51	-0,19
Kommunikation und Kooperation	146	0,00	2,00	1,00	1,00	1,17	0,42	1532	1,27	0,48	-2,40*
Leistungs- und Zeitvorgaben	147	0,00	2,00	0,50	0,50	0,78	0,56	1551	0,99	0,59	-4,23*
Regulationsbehinderungen	143	0,00	2,00	1,00	1,20	1,16	0,44	1534	1,21	0,47	-1,20
Tätigkeitsspielraum	143	0,00	2,00	1,33	1,00	1,11	0,50	1541	1,18	0,52	-1,50
Gesamtbedarf für Arbeitsgestaltung	133	2,25	8,40	4,52	5,28	5,25	1,30	1466	5,71	1,32	-3,85*
Autoritäre Führung	149	0,00	3,25	1,25	1,00	1,12	0,74	1342	1,33	0,79	-3,11*
Bedeutung von Anerkennung und Wertschätzung	148	0,00	4,00	3,00	3,00	3,04	0,82	1342	2,93	0,94	1,33
Ganzheitliche Führung	150	0,60	4,00	2,67	2,67	2,58	0,69	1347	2,25	0,8	4,91*
Partizipative Unternehmenskultur	149	0,25	4,00	2,00	2,50	2,35	0,77	1344	1,75	0,82	8,59*
Zusammenarbeit	149	0,00	4,00	3,00	3,00	2,79	0,74	1342	2,53	0,89	3,42*
Gesundheitsbezogene Selbstwirksamkeit	140	0,70	3,70	2,70	2,60	2,53	0,58	1421	2,50	0,67	0,46
Herz-Kreislauf-Beschwerden	146	0,00	4,00	2,00	2,00	1,68	0,98	1306	1,78	1,13	-0,98
Magen-Darm-Beschwerden	143	0,00	4,00	1,00	1,00	1,06	0,87	1328	1,24	1,07	-1,91
Muskel-Skelett-Beschwerden	144	0,00	4,00	2,00	2,00	2,43	1,10	1321	2,48	1,19	-0,48
Unspezifische Beschwerden	144	0,00	4,00	2,00	2,00	2,44	1,01	1304	2,65	1,06	-2,29*
Absentismustage	128	0,00	42,00	0,00	3,00	5,29	7,72	884	7,49	16,74	-1,46
Präsentismustage	117	0,00	30,00	0,00	3,00	5,17	6,61	794	6,27	8,07	-1,41

* $p < .05$

Tabelle 107 ist zu entnehmen, dass sich die Validierungsstichprobe in acht Merkmalen signifikant von den Induktionsstichproben unterscheidet. Drei Merkmale davon fallen auf Arbeitsaufgaben und -situation, vier auf Führung und Zusammenarbeit und nur eines auf Gesundheitsindikatoren. Sämtliche Unterschiede weisen auf signifikant oder tendenziell günstigere Werte in der Validierungsstichprobe hin.

Ergebnisse

Profil-Vorhersage fuer einen Wasserwirtschaftsverband Grundlage: WSIB, div. Taetigkeiten

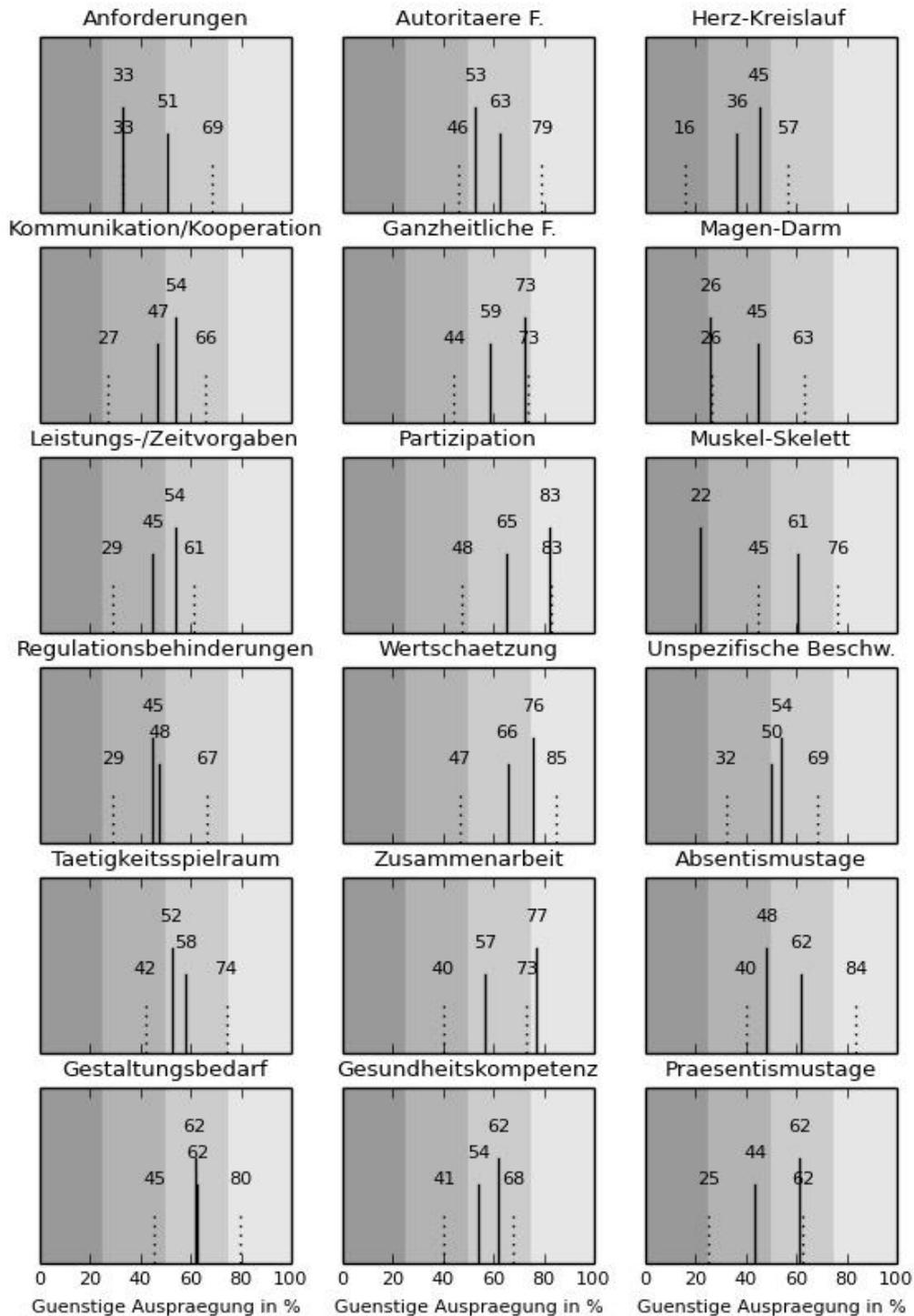


Abbildung 18. Validierung der Klassifikatoren bezüglich des WSIB zur Vorhersage der Wahrscheinlichkeiten fuer Klasse 1 je Kriterium. Dargestellt sind gemessene Wahrscheinlichkeit aus den Daten (laengere durchgezogene Linie), vorhergesagte Wahrscheinlichkeit (kuerzere durchgezogene Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden kurzen, gestrichelten Linien). Die einzelnen Werte werden so dargestellt, dass das jeweilige Kriterium umso unguenstiger ausgepraegt ist, je kleiner die Wahrscheinlichkeit ist. Dies wird auch durch die Graustufen des Hintergrundes symbolisiert.

Ergebnisse

Kennwert-Vorhersage fuer einen Wasserwirtschaftsverband Grundlage: WSIB, div. Taetigkeiten

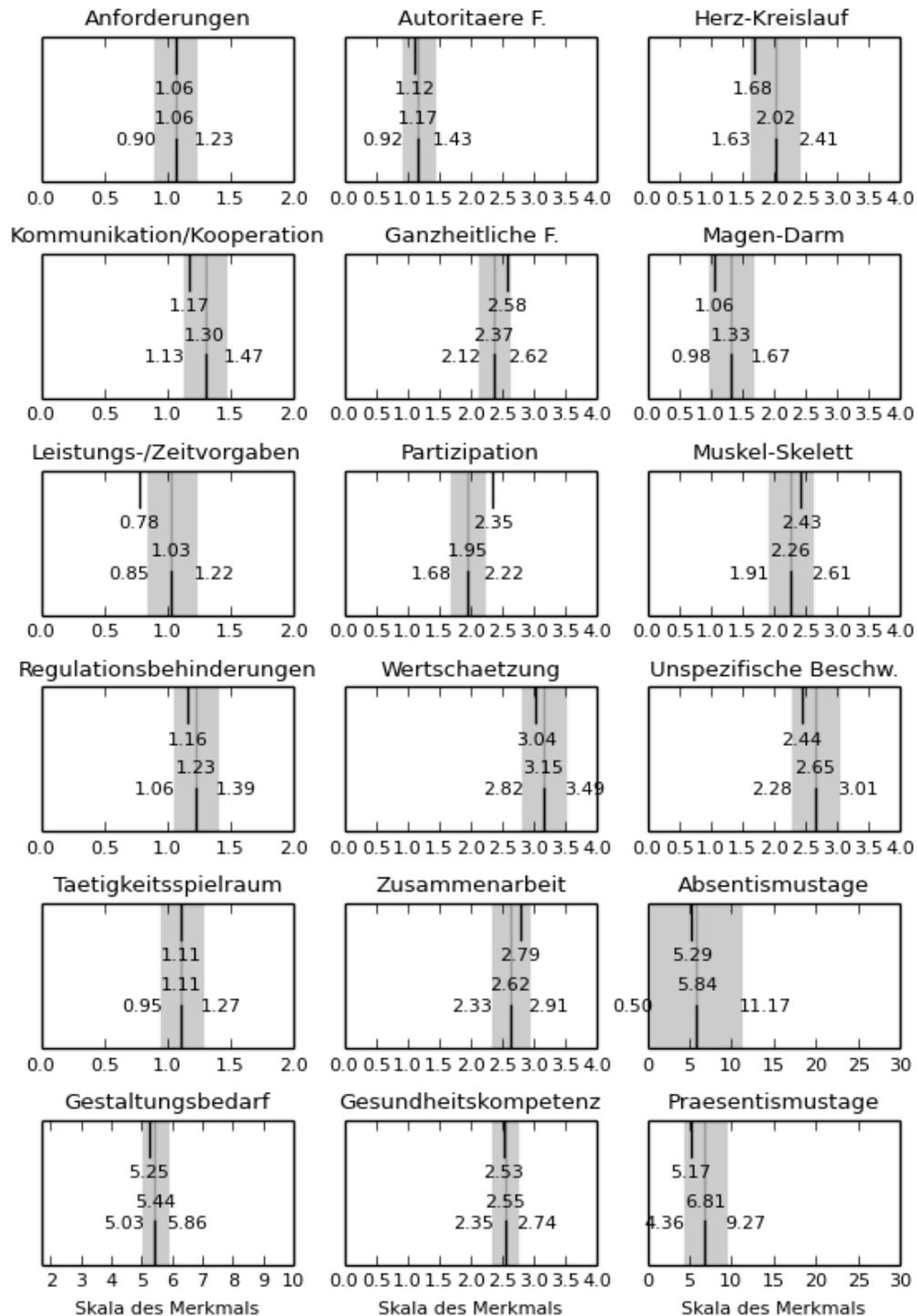


Abbildung 19. Validierung der Klassifikatoren bezüglich des WSIB zur Vorhersage der Erwartungswerte je Kriterium. Dargestellt sind die gemessenen arithmetischen Mittel aus den Daten (obere Markierung), die vorhergesagten Erwartungswerte (untere schwarze Markierung) sowie die vorhergesagten 99,8 %-Konfidenzintervalle (grau hinterlegter Bereich um den vorhergesagten Erwartungswert).

Beiden Profilen ist zu entnehmen, dass der gemessene Wert trotz signifikanter Unterschiede zwischen Validierungs- und Induktionsstichprobe (Tabelle 107) in 16 von 18 Fällen innerhalb der vorhergesagten Konfidenzintervalle liegt. Für Muskel-Skelett-Beschwerden liegt die gemessene Wahrscheinlichkeit für günstige Werte weit unterhalb der Vorhersage. D. h. die Beschäftigten der Validierungsstichprobe berichten sehr viel häufiger von überdurchschnittlich vielen Muskel-Skelett-Beschwerden, als vorhergesagt (Abbildung 18). Dementsprechend ist auch das gemessene arithmetische Mittel für Muskel-Skelett-Beschwerden etwas höher als der vorhergesagte Erwartungswert. Hier fällt der gemessene Wert jedoch noch in das 99,8 %-Konfidenzintervall (Abbildung 19). Zusammenarbeit wird von den Beschäftigten der Validierungsstichprobe wesentlich häufiger als überdurchschnittlich gut beschrieben, als vorhergesagt (Abbildung 18). Dies zeigt sich auch in einem höheren gemessenen arithmetischen Mittel im Vergleich zum vorhergesagten Erwartungswert. Wiederum liegt der Messwert hier innerhalb des vorhergesagten 99,8 %-Konfidenzintervalls (Abbildung 19). Der Gestaltungsbedarf für Leistungs- und Zeitvorgaben wird von den Beschäftigten der Validierungsstichprobe im Mittel wesentlich geringer eingeschätzt, als vorhergesagt (Abbildung 19). Dementsprechend ist die gemessene Wahrscheinlichkeit für günstige Leistungs- und Zeitvorgaben auch etwas höher, als die vorhergesagte. Sie liegt jedoch innerhalb des vorhergesagten 99,8 %-Konfidenzintervalls (Abbildung 18). Partizipative Unternehmenskultur wird von den Beschäftigten der Validierungsstichprobe im Mittel wesentlich günstiger bewertet, als vorhergesagt (Abbildung 19). Auch dies spiegelt sich in einer höheren gemessenen Wahrscheinlichkeit für überdurchschnittliche partizipative Unternehmenskultur im Vergleich zur Vorhersage wider. Auch hier liegt der gemessene Wert gerade noch im 99,8 %-Konfidenzintervall (Abbildung 18). Die Differenz zwischen gemessenem arithmetischem Mittel und vorhergesagtem Erwartungswert beträgt für Leistungs- und Zeitvorgaben 0,25 Skaleneinheiten (Abbildung 19). Dieser Wert liegt noch weit unterhalb der Standardabweichung von Leistungs- und Zeitvorgaben, die 0,59 Skaleneinheiten misst (Tabelle 22). Genauso liegt die Differenz zwischen gemessenem arithmetischem Mittel und vorhergesagtem Erwartungswert für partizipative Unternehmenskultur mit 0,40 Skaleneinheiten (Abbildung 19) noch weit unterhalb der Standardabweichung von partizipativer Unternehmenskultur. Diese umfasst 0,82 Skaleneinheiten (Tabelle 57). Für die verbleibenden Kriterien gelten diese Verhältnisse in stärkerer Weise. Es kann daher - unter angemessener Berücksichtigung der 99,8 %-Konfidenzintervalle - von einer guten Vorhersagekraft der validierten Klassifikatoren ausgegangen werden.

Ergebnisse

Die Validierungsstichprobe aus dem Hochtechnologieunternehmen wird dazu verwendet, die funktionale Abhängigkeit des Vorhersagefehlers $\sigma(\varepsilon(\mu))$ von Erwartungswerten vom Stichprobenumfang n zu prüfen. Weiter oben wurde festgestellt, dass sich dieser Fehler mittels der Formel

$$\sigma_n(\varepsilon(\mu)) = 0,1688 \sigma(X) \ln\left(\frac{17,18}{\sqrt{n}}\right)$$

bestimmen lässt. Da die Stichprobe aus dem Hochtechnologieunternehmen 11 Teilstichproben mit n zwischen 7 und 22 Teilnehmern enthält, lässt sich diese Formel leicht empirisch prüfen. In Tabelle 108 sind die Ergebnisse dieser Prüfungen zusammengefasst. In Anhang A5 sind die Einzelergebnisse (empirische Mittelwerte, vorhergesagte Erwartungswerte und zugehörige 99,8 %-Konfidenzintervalle) für alle 11 Teilstichproben sowie die Gesamtstichprobe und die jeweils 16 Kriterien tabellarisch aufgeführt.

Tabelle 108. Anzahl korrekt vorhergesagter Erwartungswerte für 16 Kriterien unter Anwendung des funktionalen Zusammenhanges zwischen dem entsprechenden Vorhersagefehler und dem Stichprobenumfang.

Bereich	Tätigkeit	Stichprobenumfang	Anzahl korrekt vorhergesagter Erwartungswerte für 16 Kriterien
A	Fertigung (MA)	22	15
B	Fertigung (FK)	12	16
C	Einkauf (MA)	16	15
D	Einkauf (MA)	19	15
E	Einkauf (FK)	8	16
F	Technik (MA)	18	14
G	Technik (MA)	22	15
H	Technik (MA)	11	16
I	Technik (FK)	8	16
J	Entwicklung	9	16
K	Entwicklung	7	15
Gesamt	Diverse	152*	16

MA = Mitarbeiter ohne Führungsverantwortung, FK = Führungskräfte

*Umfang der Gesamtstichprobe nach Aufbereitung und Bereinigung der Daten; für die Gesamtstichprobe wurde der empirisch ermittelte Vorhersagefehler für $n = 100$ verwendet

Durch diese Ergebnisse wird die Vorhersagekraft der validierten Klassifikatoren weiter unterstützt. Unter Einbeziehung der Ergebnisse für den Wasserversorger ist zusammenfassend festzustellen, dass die Vorhersage von Erwartungswerten unter Berücksichtigung des Stichprobenumfangs und entsprechender 99,8 %-Konfidenzintervalle in 201 von 210 Fällen gelingt. Dies entspricht einer Quote von 95,7 % korrekter Vorhersagen.

6.7 PROFIL-BESTIMMUNG: DEMONSTRATION DER ANWENDUNGSMÖGLICHKEITEN

Um die Anwendungsmöglichkeiten der generierten Klassifikatoren zu demonstrieren, werden ähnliche Profile wie oben bei der Validierung für die Anwendungsstichprobe bestimmt. Auch die Anwendungsstichprobe ist in keine der durchgeführten Experimental-Reihen zur Klassifikator-Induktion eingeflossen. D. h. auch diese Daten sind den Klassifikatoren völlig unbekannt. Hinzu kommt die Besonderheit, dass für diese Stichprobe ausschließlich die Daten aus dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB) vorliegen. Es wird demonstriert, dass allein auf Basis dieser neun Items Profile bestimmt werden können, die sämtliche 18 Kriterien (Arbeitsaufgabe und -situation, Führung und Zusammenarbeit, Gesundheitsindikatoren) umfassen. Die Anwendungsstichprobe umfasst $N = 551$ Instanzen. Da für alle Beschäftigte aus dieser Stichprobe auch die jeweilige Abteilung festgehalten wurde, können funktionsbezogene Teilstichproben gezogen werden. Eine Teilstichprobe umfasst 36 Führungskräfte, eine weitere 151 Beschäftigte der Produktion.

Auf den folgenden Seiten finden sich jeweils zwei vorhergesagte Profile für diese drei Stichproben. Eines der Profile sagt die Wahrscheinlichkeit für eine günstige Ausprägung sämtlicher 18 Kriterien voraus. Ein weitere sagt die Erwartungswerte für diese Kriterien voraus. Alle vorhergesagten Kennwerte werden mit dem 99,8 %-Konfidenzintervall versehen, wie weiter oben im Teil zur Methodik beschrieben.

Das jeweilige Profil zur Vorhersage der Wahrscheinlichkeit günstiger Ausprägungen symbolisiert durch Graustufen im Hintergrund, ob der angegebene Wert günstig oder ungünstig ist. Je dunkler der Hintergrund, desto ungünstiger der vorhergesagte Wert. Das jeweilige Profil zur Vorhersage der Erwartungswerte ermöglicht ein Benchmarking auf Grundlage der Induktionsstichproben. Dies ist ebenfalls durch eine entsprechende Gestaltung des Hintergrundes sowie die bekannten deskriptiv-statistischen Kennwerte der Kriterien möglich. Eine weiße Linie markiert das arithmetische Mittel des jeweiligen Kriteriums. Der Median markiert zwei Bereiche: Einen dunkelgrauen Bereich, in dem ungünstige Werte liegen und einen hellgrauen Bereich, in dem günstige Werte liegen. Beiden Profilen ist also direkt zu entnehmen, ob die vorhergesagten Werte günstig oder ungünstig zu bewerten sind. Zugleich ermöglichen die Profile einen unternehmensinternen Vergleich. Dies wird im Folgenden anhand der gesamten Anwendungsstichprobe (Abbildung 20 und Abbildung 21), der Führungskräfte (Abbildung 22 und Abbildung 23) sowie der Beschäftigten aus der Produktion verdeutlicht (Abbildung 24 und Abbildung 25).

Ergebnisse

Profil-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 551, div. Taetigkeiten

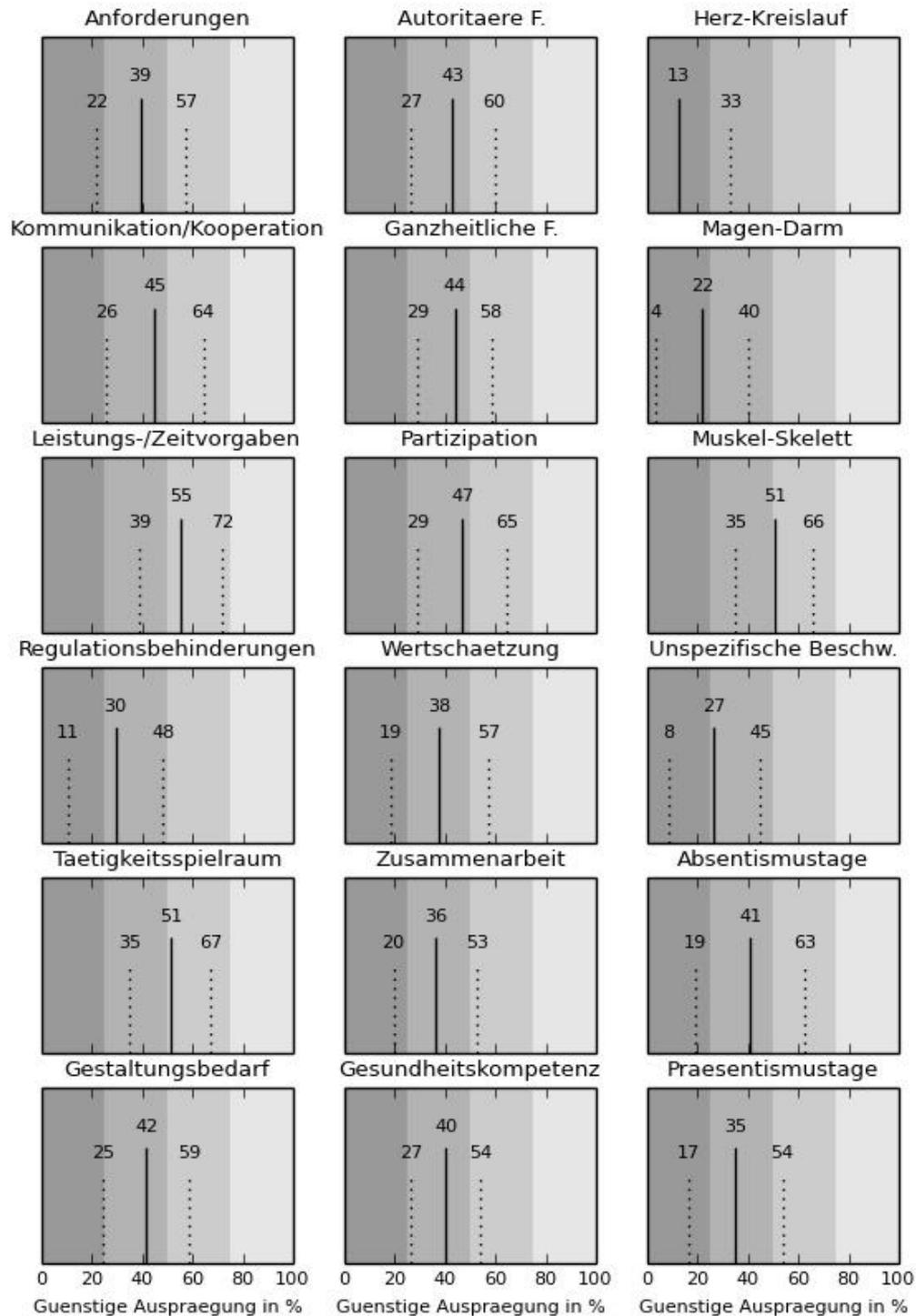


Abbildung 20. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die gesamte Anwendungstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB. Dargestellt sind die vorhergesagte Wahrscheinlichkeit (durchgezogene Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden kurzen, gestrichelten Linien). Die einzelnen Werte werden so dargestellt, dass das jeweilige Kriterium umso ungünstiger ausgeprägt ist, je kleiner die Wahrscheinlichkeit ist. Dies wird auch durch die Graustufen des Hintergrundes symbolisiert.

Ergebnisse

Kennwert-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 551, div. Taetigkeiten

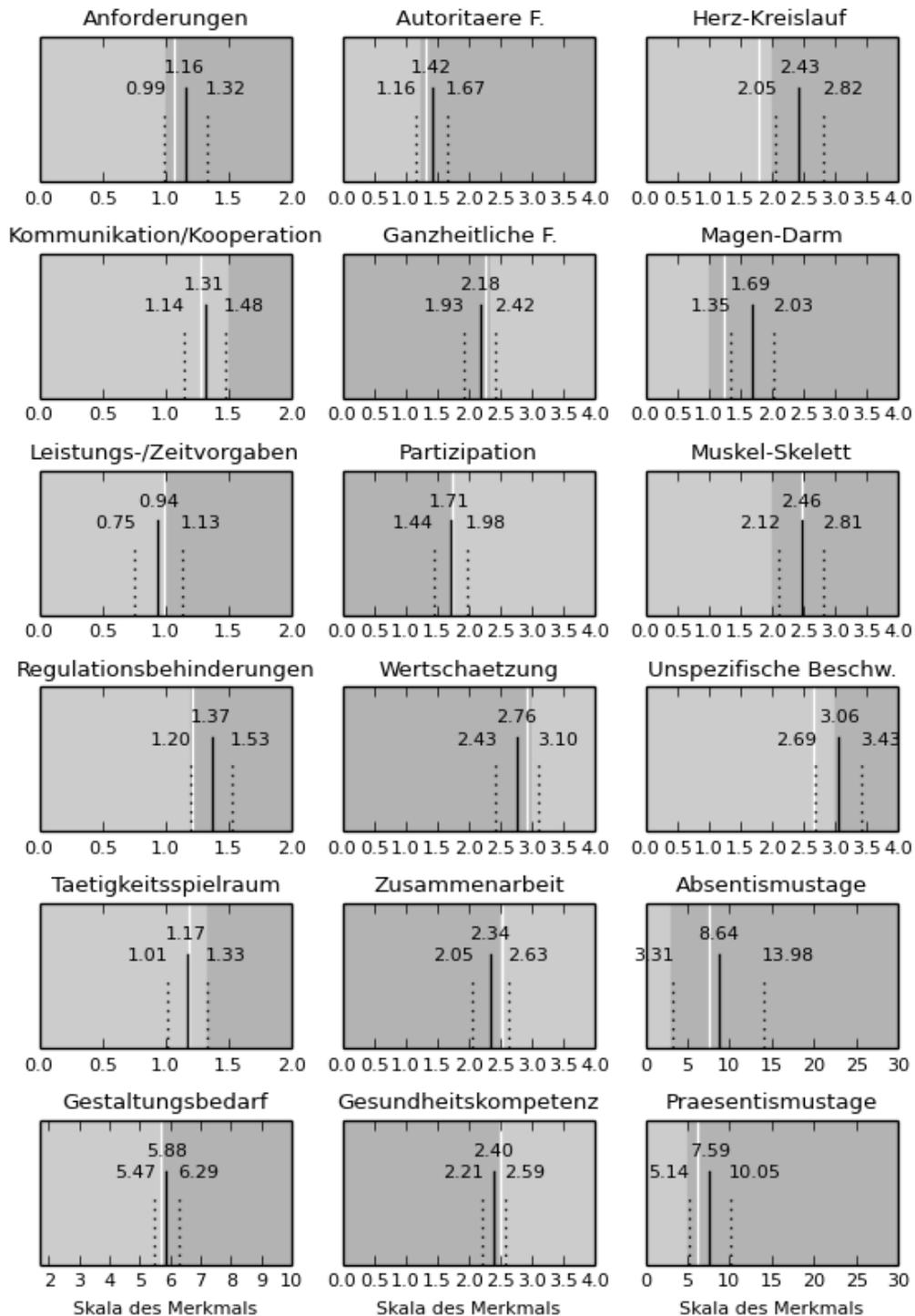


Abbildung 21. Vorhersage der Erwartungswerte je Kriterium fuer die gesamte Anwendungstichprobe mit Hilfe der Klassifikatoren bezueglich des WSIB. Dargestellt sind der vorhergesagte Erwartungswerte (durchgezogene schwarze Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden gepunkteten Linien). Weiße, durchgezogene Linie: Arithmetisches Mittel des Kennwertes bei der Induktionsstichprobe. Dunkelgrauer Hintergrund: Ungueltige Werte nach Mediansplit der Induktionsstichprobe. Hellgrauer Hintergrund: Gueltige Werte nach Mediansplit der Induktionsstichprobe.

Ergebnisse

Profil-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 36, Fuehrungskraefte

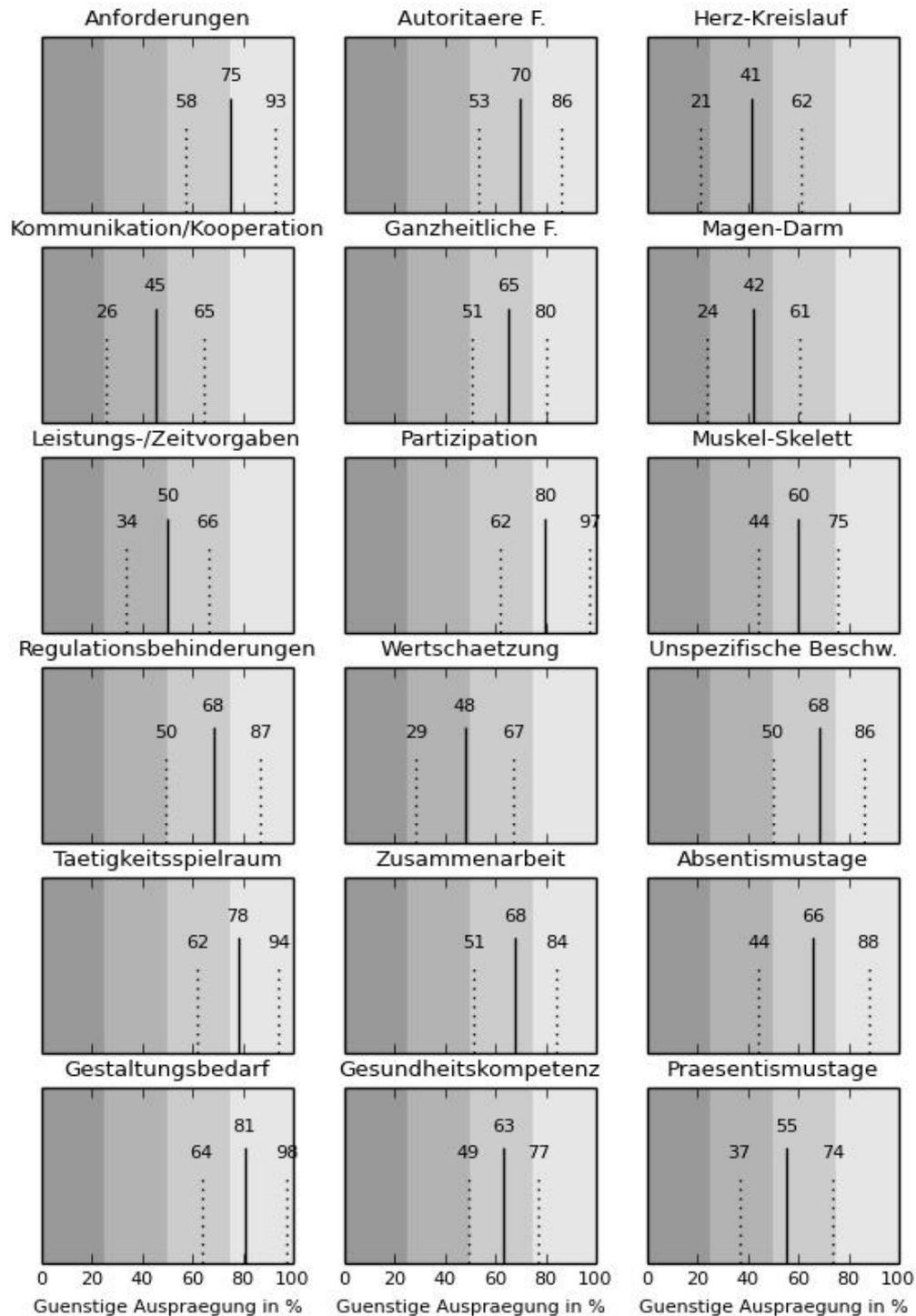


Abbildung 22. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die Teilstichprobe „Führungskräfte“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB. Dargestellt sind die vorhergesagte Wahrscheinlichkeit (durchgezogene Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden kurzen, gestrichelten Linien). Die einzelnen Werte werden so dargestellt, dass das jeweilige Kriterium umso ungünstiger ausgeprägt ist, je kleiner die Wahrscheinlichkeit ist. Dies wird auch durch die Graustufen des Hintergrundes symbolisiert.

Ergebnisse

Kennwert-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 36, Fuehrungskraefte

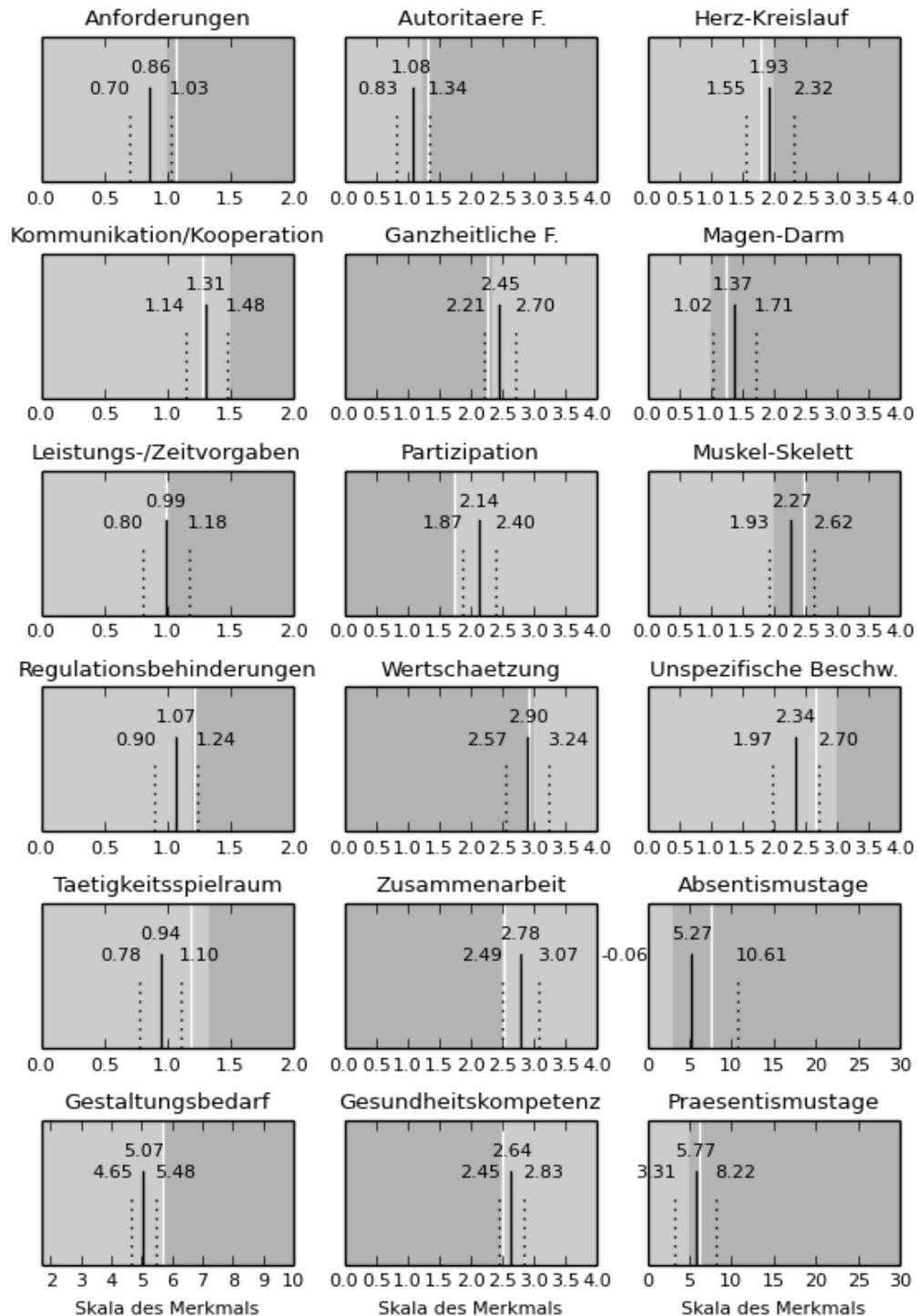


Abbildung 23. Vorhersage der Erwartungswerte je Kriterium für die Teilstichprobe „Führungskräfte“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB. Dargestellt sind der vorhergesagte Erwartungswerte (durchgezogene schwarze Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden gepunkteten Linien). Weiße, durchgezogene Linie: Arithmetisches Mittel des Kennwertes bei der Induktionsstichprobe. Dunkelgrauer Hintergrund: Ungünstige Werte nach Mediansplit der Induktionsstichprobe. Hellgrauer Hintergrund: Günstige Werte nach Mediansplit der Induktionsstichprobe.

Ergebnisse

Profil-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 151, Produktion

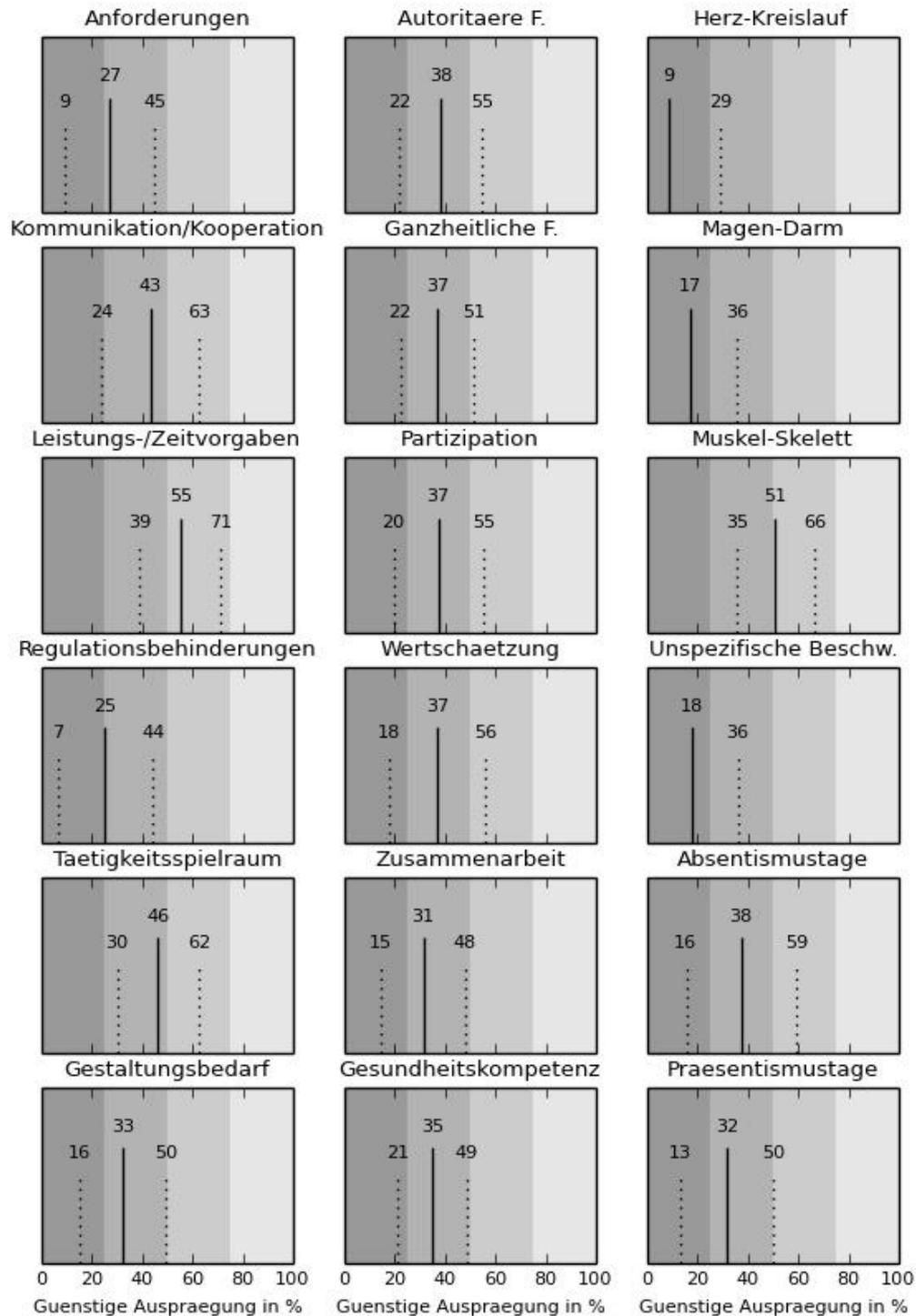


Abbildung 24. Vorhersage der Wahrscheinlichkeiten für Klasse 1 je Kriterium für die Teilstichprobe „Produktion“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB. Dargestellt sind die vorhergesagte Wahrscheinlichkeit (durchgezogene Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden kurzen, gestrichelten Linien). Die einzelnen Werte werden so dargestellt, dass das jeweilige Kriterium umso ungünstiger ausgeprägt ist, je kleiner die Wahrscheinlichkeit ist. Dies wird auch durch die Graustufen des Hintergrundes symbolisiert.

Ergebnisse

Kennwert-Vorhersage fuer ein Technologieunternehmen Grundlage: WSIB, N = 151, Produktion

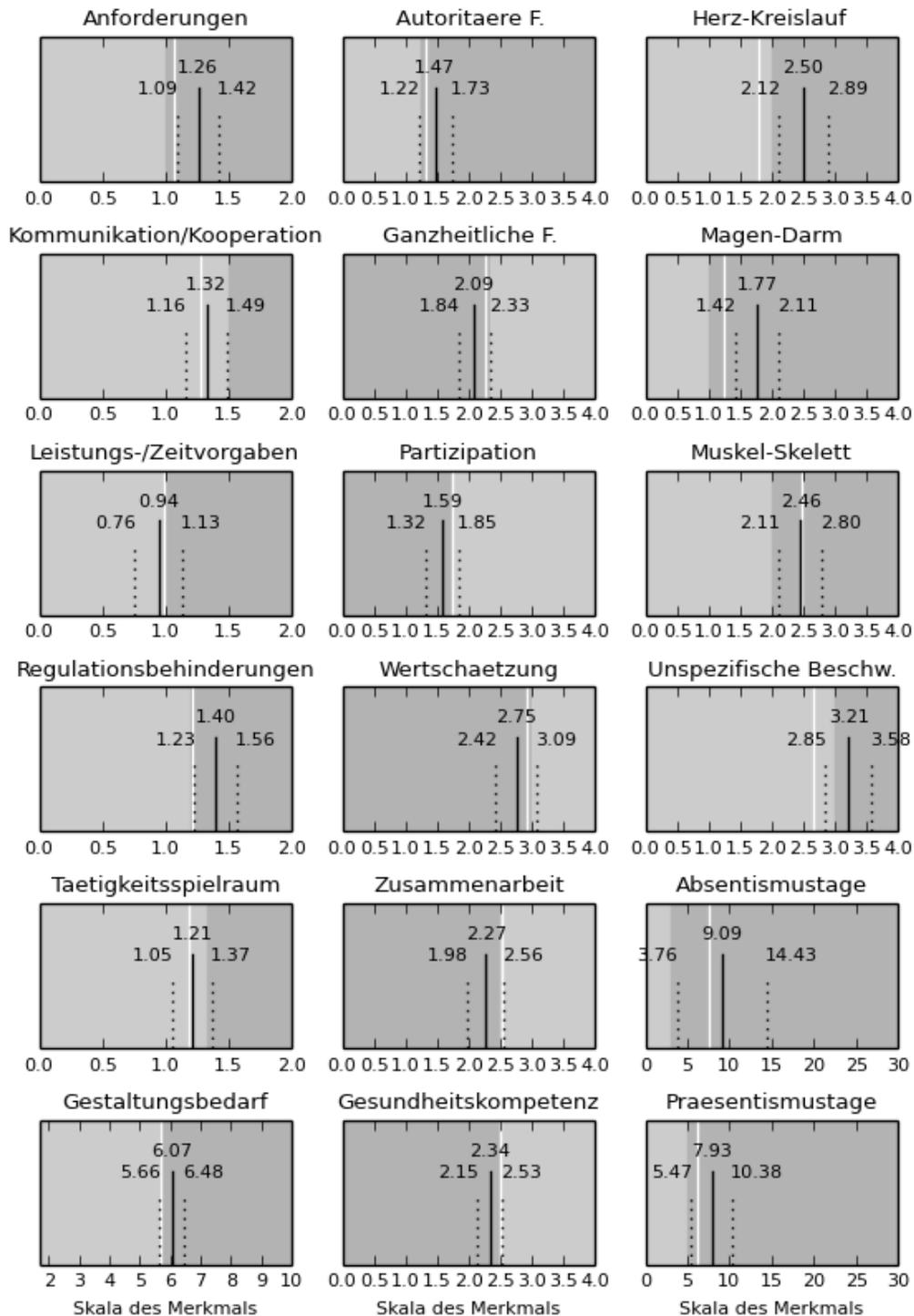


Abbildung 25. Vorhersage der Erwartungswerte je Kriterium für die Teilstichprobe „Produktion“ der Anwendungsstichprobe mit Hilfe der Klassifikatoren bezüglich des WSIB. Dargestellt sind der vorhergesagte Erwartungswerte (durchgezogene schwarze Linie) und vorhergesagtes 99,8 %-Konfidenzintervall (die beiden gepunkteten Linien). Weiße, durchgezogene Linie: Arithmetisches Mittel des Kennwertes bei der Induktionsstichprobe. Dunkelgrauer Hintergrund: Ungünstige Werte nach Mediansplit der Induktionsstichprobe. Hellgrauer Hintergrund: Günstige Werte nach Mediansplit der Induktionsstichprobe.

Interpretieren wir zunächst die Profile für die gesamte Anwendungsstichprobe, so fällt zweierlei auf (Abbildung 20 und Abbildung 21): Erstens finden sich die meisten vorhergesagten Kennwerte tendenziell in der Nähe der arithmetischen Mittelwerte der Induktionsstichproben. Dies kann auf die Breite der berücksichtigten Tätigkeiten und Abteilungen bzw. Funktionen zurückzuführen sein. Zweitens ist ein Großteil der vorhergesagten Kennwerte als eher ungünstig zu interpretieren. Dies ist auf eine selektive Teilnahme der Abteilungen in dem Unternehmen zurückzuführen: Es nehmen überwiegend solche Abteilungen freiwillig an dem entsprechenden Programm des Unternehmens teil, die für sich bereits einen gewissen Bedarf für gesundheitsfördernde betriebliche Maßnahmen festgestellt haben (mündliche Mitteilung des Gesundheitsmanagers in dem entsprechenden Unternehmen). Diese beiden Profile erscheinen also vor dem Hintergrund der betrieblichen Praxis plausibel.

Die Kennwerte in den Profilen für Führungskräfte sind überwiegend als günstig zu interpretieren (Abbildung 22 und Abbildung 23). Dies betrifft vor allem – mit Ausnahme von Leistungs- und Zeitvorgaben – die Merkmale der Arbeitsgestaltung sowie – mit Ausnahme der Bedeutung von Anerkennung und Wertschätzung – die Merkmale von Führung und Zusammenarbeit. Auch die gesundheitsbezogene Selbstwirksamkeit wird als eher günstig vorhergesagt. Bei den Gesundheitsindikatoren fallen jedoch nur die unspezifischen Beschwerden mit einer eher günstigen Prognose auf. Ungeachtet dessen bestätigt sich hier ein aus der Literatur bekanntes Phänomen: Führungskräfte verfügen tendenziell eher über vollständige Tätigkeiten und einen gut gestalteten Arbeitsplatz sowie über eine gute Gesundheit (z. B. Wieland & Scherrer, 2007). Kontrastiert man die Vorhersagen für Führungskräfte mit jenen für die Beschäftigten aus der Produktion (Abbildung 24 und Abbildung 25), so wird diese Plausibilitätsbetrachtung erhärtet. Für Beschäftigte aus der Produktion werden überwiegend wesentlich schlechtere Kennwerte vorhergesagt, als für Führungskräfte. Dies betrifft bei den Merkmalen der Arbeitsgestaltung vor allem die Regulationsbehinderungen und den Gesamtbedarf an Arbeitsgestaltung. Die Kennwerte für Führung und Zusammenarbeit werden – mit Ausnahme der Bedeutung von Anerkennung und Wertschätzung – wesentlich ungünstiger vorhergesagt. Im Vergleich zu Führungskräften wird für die Beschäftigten aus der Produktion auch eine wesentlich geringere gesundheitsbezogene Selbstwirksamkeit vorhergesagt (vgl. Wieland & Hammes, 2010; Wieland & Scherrer, 2007). Im direkten Vergleich der Profile von Führungskräften und Beschäftigten in der Produktion bestätigt sich darüber hinaus, dass erstere – hier via Vorhersage – über einen besseren Gesundheitszustand verfügen. Die Produktionsbereiche wurden in das Projekt des Unternehmens aufgenommen, da sie vor allem mit Gesundheitsproblemen aufgefallen sind (mündliche Mitteilung des Gesundheitsmanagers).

Ergebnisse

Zusammenfassend lässt sich festhalten, dass die vorhergesagten Profile sowohl im Hinblick auf bekannte empirische Befunde als auch im Hinblick auf die Umstände in den befragten Bereichen bzw. dem Unternehmen plausibel erscheinen. Unter Beachtung der Ergebnisse zur Validierung der Klassifikatoren ist jedoch anzumerken: Die angegebenen Konfidenzintervalle sind unbedingt als solche bei der Interpretation zu beachten. D. h. es können nur tendenzielle Interpretationen vorgenommen werden. Jeder einzelne der vorhergesagten Kennwerte steht bei der Interpretation zur Diskussion. Diese Diskussion sollte mit den Betroffenen im befragten Unternehmen durchgeführt werden (z. B. im Rahmen von Workshops). Um die tatsächlichen Ausprägungen der vorhergesagten Kriterien zu ermitteln und darauf basierende Maßnahmen der betrieblichen Gesundheitsförderung abzuleiten, bedarf es einer entsprechenden Messung. Das Unternehmen, in dem die Anwendungsstichprobe erhoben wurde, führt sowohl Workshops als auch zukünftig detaillierte Messungen arbeitspsychologischer Merkmale durch.

7 DISKUSSION, ANWENDUNGSBEZUG UND AUSBLICK

Zum Abschluss der vorliegenden Arbeit sollen die Untersuchungsbefunde eingehend diskutiert und interpretiert sowie ihre praktische Relevanz dargelegt werden. Zu Beginn werden die Gültigkeitsbereiche der vorliegenden Befunde eingegrenzt. Dies erfolgt unter einer kritischen Betrachtung der methodischen Herangehensweise. Anschließend werden die Befunde unter dem Aspekt der Klassifikation genauer betrachtet. Hierbei steht im Vordergrund, inwiefern sich die theoretischen Überlegungen zur Klassifikation in den Befunden widerspiegeln. Die inhaltliche Interpretation der Befunde vollzieht sich vor dem Hintergrund der Struktur der EEB sowie den theoretischen Überlegungen über die Zusammenhänge zwischen psychischer Belastung, psychischer Beanspruchung und langfristigen Folgen. D. h. es wird der Versuch unternommen, erhaltene Klassifikatoren mit theoretisch zu erwartenden Beanspruchungsmustern zu vergleichen. Im Ausblick wird letztlich auf die Anwendungsmöglichkeiten der erhaltenen Klassifikatoren im Speziellen sowie die eingesetzte Methodik im Allgemeinen eingegangen.

7.1 EINGRENZUNG DER GÜLTIGKEITSBEREICHE VORLIEGENDER UNTERSUCHUNGSBEFUNDE

Den in dieser Arbeit präsentierten Vorgehensweisen, Ergebnissen und Klassifikatoren liegen Befragungsdaten aus Umsetzungsprojekten zu betrieblicher Gesundheitsförderung zu Grunde. Hierbei wurden verschiedene Branchen, Tätigkeiten und Personengruppen berücksichtigt (Baumgart et al., 2002; Hammes et al., 2009, 2010; Scherrer, 2002; Wieland & Hammes, 2009, 2014b; Wieland et al., 2001, 2006, 2009). Innerhalb solcher Kontexte sind die beschriebenen Untersuchungen relevant. Eine Übertragung auf andere als arbeitspsychologische und betriebliche Kontexte setzt die Erhebung entsprechender Daten und deren Verwendung bei der Klassifikator-Induktion voraus. Da die gewählten Methoden unabhängig kontextueller Inhalte auf ähnlich strukturierte Daten (Beschreibung von Objekten mittels diskret kodierter Messwerte) übertragbar sind, ist die Methodik weit über den betrieblichen bzw. arbeitspsychologischen Kontext hinaus einsetzbar. Über psychologische Anwendungsdisziplinen hinaus ist die Anwendung der Methodik beispielsweise überall dort denkbar, wo umfangreiche standardisierte Befragungen vorgenommen werden (z. B. human-, sozial- und wirtschaftswissenschaftliche Domänen). Im Folgenden steht jedoch die Eingrenzung des Gültigkeitsbereichs unter Berücksichtigung der verwendeten Datenbasis im Vordergrund.

Es wurden ausschließlich Daten aus Querschnittsuntersuchungen verwendet. Diese Einschränkung ist der Herkunft der verwendeten Daten geschuldet. Die Erhebungen fanden im Rahmen von Umsetzungsprojekten zu betrieblicher Gesundheitsförderung in betrieblichen

Settings statt. Aus betrieblicher Sicht bedeutet eine Längsschnittuntersuchung sowohl ökonomisch als auch politisch einen erheblichen Aufwand. Aus diesem Grunde finden in Betrieben nur äußerst selten Längsschnittuntersuchungen (z. B. zur Evaluation) statt (Busch & Werner, 2011; Elkeles & Beck, 2010). Die hier präsentierten Ergebnisse sind daher nicht kausal zu interpretieren. Sie stellen viel mehr korrelative Ergebnisse dar (Bortz & Döring, 2009). Dies war auch Intention der Untersuchungen: Ausgehend vom WITAL (Wuppertaler Inventar zur ressourcenorientierten Arbeits- und Organisationsgestaltung; Wieland, 2014b) wurde untersucht, inwieweit Verteilungen auf den Antwortkategorien der EEB-Items den Schluss auf Verteilungen und Erwartungswerte der 18 Kriterien als feste Bestandteile des WITAL erlauben. Es ging um die Frage, in welchem Maße das aufwändige WITAL durch die viel ökonomischer anzuwendende EEB (Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchung während der Arbeit) im Sinne eines Screenings ersetzt werden kann. Diese Fragestellung hat demnach einen festen Messzeitpunkt zur Prämisse. Ungeachtet dessen wurden die EEB-Items als Prädiktoren gewählt, da aus theoretischer Sicht ein kausaler Zusammenhang zwischen psychischer Belastung, psychischer Beanspruchung und langfristigen Folgen anzunehmen ist und die EEB psychische Beanspruchung misst. Es bleibt festzuhalten: Die Ergebnisse sind höchstens in einem theoretischen Sinne kausal zu interpretieren, sind jedoch aus methodischer Sicht korrelativ und beziehen sich bereits aus der Fragestellung heraus auf einen festen Messzeitpunkt. Wird mit Hilfe der hier induzierten Klassifikatoren ein bestimmtes Profil aus einer betrieblichen Erhebung mit dem WSIB berechnet, so ist es wie folgt zu interpretieren: Hätte man zum Erhebungszeitpunkt statt des WSIB das WITAL eingesetzt, so würden die gemessenen Verteilungen und Erwartungswerte mit hoher Wahrscheinlichkeit (zu ca. 95 %) innerhalb der vorhergesagten bzw. berechneten Konfidenzintervalle liegen.

Diese Interpretation ist jedoch noch weiter einzuschränken. Bei den zur Klassifikator-Induktion und Validierung berücksichtigten Erhebungen wurde die EEB bzw. das WSIB als fester Bestandteil des WITAL eingesetzt. In der Regel wurden die EEB-Items beantwortet, nachdem zuvor umfangreiche Fragen zum Gesundheitszustand, zur Arbeitsaufgabe und zur physischen und psychosozialen Arbeitssituation gestellt wurden. Insbesondere da die Instruktion der EEB verlangt, dass man bei der Bewertung der EEB-Items Bezug auf seine Arbeit im Allgemeinen nimmt, kann ein Übertragungseffekt (Primacy-Recency-Effekt; Bortz & Döring, 2009) nicht ausgeschlossen werden. Ein solcher Übertragungseffekt würde bei der alleinigen Verwendung der EEB bzw. des WSIB als Screening-Instrument ausbleiben. Dies könnte zu Verschiebungen in den vorhergesagten Profilen führen. Die geführte Argumentation

lässt sich analog auf das WSIB als Bestandteil der vollständigen EEB übertragen. Die WSIB-Items wurden in den Induktionsstichproben als Bestandteile der vollständigen EEB erhoben, so dass auch hier Übertragungseffekte nicht ausgeschlossen werden können. In den Validierungsstichproben wurden nur die WSIB-Items als Bestandteile des WITAL erhoben. Die nachgewiesene Vorhersagekraft des WSIB im Rahmen der Validierung kann als Indiz dafür interpretiert werden, dass etwaige Übertragungseffekte anderer EEB-Items auf die WSIB-Items für die vorgenommenen Klassifikationen wenig bedeutsam sind. Dies insbesondere deshalb, weil die Validierungsstichprobe aus der Wasserwirtschaft insgesamt ein günstigeres Profil als die Induktionsstichproben mit acht signifikant günstigeren Ausprägungen der Kriterien aufweist. In der Anwendungsstichprobe wurde das WSIB als Screening eingesetzt. Hier deuten die plausiblen Ergebnisse für die unterschiedlichen Gruppen (Gesamtstichprobe, Führungskräfte und Beschäftigte in der Produktion) darauf hin, dass etwaige Verschiebungen der Profile systematisch sind. D. h. die Profile bilden ungeachtet etwaiger Übertragungseffekte die Verhältnisse zwischen Gruppen plausibel ab. Eine Fortführung der hier vorgestellten Untersuchungen könnte ein zweistufiges Erhebungsdesign mit drei Gruppen zu Grunde legen. In der ersten Stufe würde ein Screening-Instrument (entweder EEB oder WSIB) in zwei Gruppen (Gruppen 1 und 2) eingesetzt bzw. nichts erhoben werden (Gruppe 3 bzw. Kontrollgruppe). In der zweiten Stufe würde das gesamte WITAL mit (Gruppen 1 und 3) und ohne (Gruppe 2) eingesetztem Screening-Instrument zum Einsatz kommen. Eine solche Erhebung ließe die Quantifizierung solcher Übertragungseffekte zu. In dem Unternehmen, in dem die Anwendungsstichprobe erhoben wurde, zeichnet sich die Möglichkeit der Untersuchung von Übertragungseffekten ab. Dort ist der baldige Einsatz des WITAL parallel zum WSIB geplant.

Die berichteten Korrekturklassifikationen (Vorhersagegenauigkeiten) resultieren aus der 10-fach Kreuzvalidierung sowie den Bootstrapping-Experimenten. D. h. ihnen liegen ausschließlich die Daten der Induktionsstichproben zu Grunde. Vorhersagegenauigkeiten liegen sowohl für die vollständige EEB als auch das WSIB vor. Die abschließende externe Validierung mit Hilfe einer separaten Validierungsstichprobe konnte nur für das WSIB vorgenommen werden. Auch die Demonstration der Anwendungsmöglichkeiten bezieht sich ausschließlich auf das WSIB. Damit sind zwar die Vorhersagegenauigkeiten für alle Klassifikatoren und beide Instrumente angegeben. Die Ergebnisse der abschließenden externen Validierung gelten jedoch nur für Klassifikatoren, die auf dem WSIB beruhen. Es zeigt sich jedoch, dass sich die Vorhersagegenauigkeiten zweier endgültig ausgewählter Klassifikatoren (bezüglich EEB oder WSIB) für keines der Kriterien signifikant voneinander unterscheiden. Ein Klassifikator bezüglich des WSIB weist für alle Kriterien eine Vorhersagegenauigkeit auf,

die mit der Vorhersagegenauigkeit eines Klassifikators bezüglich der vollständigen EEB vergleichbar oder mit ihr identisch ist. Dies spricht dafür, dass es praktisch kaum einen Unterschied macht, ob man einen Klassifikator bezüglich des WSIB oder der vollständigen EEB einsetzt. Und dies bedeutet, dass der neun Item umfassende WSIB ohne besondere Verluste die vollständige, 25 Items umfassende EEB im Zusammenspiel mit den Klassifikatoren als Screening ersetzen kann. Zusammenfassend liegt die abschließende externe Validierung nur für Klassifikatoren bezüglich des WSIB vor, was jedoch aus der Perspektive der praktischen Anwendung keine Einschränkung bedeutet.

Im Fokus der vorgestellten Untersuchungen stand die Kriteriums-Validität bzw. die Vorhersagegenauigkeit. Die Objektivität der verwendeten Instrumente kann durch die hohe Standardisierung als gesichert angesehen werden. Reliabilitäten der verwendeten Instrumente wurden in der vorliegenden Arbeit weder geprüft noch berichtet. Dies geschah aus mehreren Gründen. Da in der vorliegenden Arbeit statt der individuellen Messwerte die *Verteilungen der Instanzen auf die diskreten Messwertspektren* die empirische Grundlage bilden, spielen die Reliabilitäten der individuellen Messwerte eine höchstens sekundäre Rolle. Zu fragen wäre vielmehr, ob die vorgefundenen *Verteilungen* reliabel sind. Mit den zur Verfügung stehenden Daten lässt sich diese Frage nicht beantworten, da je Stichprobe ausschließlich Verteilungen zu einem Messzeitpunkt vorliegen und sich eine theoretisch anzunehmende Grundgesamtheit als Vergleichs-Standard nicht unanfechtbar definieren lässt. Werden beispielsweise alle Erwerbstätigen oder alle sozialversicherungspflichtig Erwerbstätigen oder Beschäftigte einer bestimmten Branche, eines bestimmten Berufs oder einer bestimmten Funktion einbezogen? Muss man zwischen verschiedenen Betriebsgrößenklassen unterscheiden? Sollte man global, national oder regional definieren? Eine sinnvolle Definition von Retest-Reliabilitäten für Verteilungen könnte darüber hinaus an (z. T. gewünschten) betrieblichen Dynamiken scheitern. Während interne und Retest-Reliabilität für Verteilungen in betrieblichen Stichproben kaum sinnvoll definierbar sind, verbleibt die Split- (Half-) Reliabilität.¹³ Bootstrapping kann als Analogon für diese Form der Reliabilität betrachtet werden: Bei einer gegebenen Stichprobe sollte jede zufällig gezogene Teilstichprobe in ihren Merkmalen eine ähnliche Verteilung aufweisen, wie die Ausgangsstichprobe. Die berichteten engen Konfidenzintervalle spiegeln die Ähnlichkeit der Verteilungen in Teilstichproben bei Bootstrapping-Experimenten wider.

¹³ Dies heißt nicht, dass derartige Reliabilitäten für Verteilungen prinzipiell nicht definierbar sind. Der kinetischen Gastheorie liegen Grundannahmen über die Verteilung kinetischer Energie für die Bestandteile eines Gases zu Grunde. Reliabilität und Validität einer Messung von Stoffmenge, Temperatur, Druck oder Volumen kann direkt an den theoretischen Vorhersagen unter Berücksichtigung der Randbedingungen geprüft werden. Volumeneinheiten werden als Messgrößen über die entsprechenden Zusammenhänge normiert.

Über diesen Fokus hinaus werden arbeitspsychologische Instrumente zur Analyse der Arbeitssituation häufig nicht einem testtheoretischen Paradigma folgend entwickelt. Viele Merkmale einer Arbeitssituation lassen sich nach dem Schema „vorhanden/mehr oder weniger vorhanden/nicht vorhanden“ bewerten, so dass oft wenige Items, manchmal gar nur eines zur Bewertung ausreicht. Interne Reliabilitäten lassen sich bei solchen Instrumenten selten sinnvoll angeben. Andererseits tragen derart ökonomische Verfahren den Anforderungen der praktischen Anwendung in Betrieben Rechnung (Amelang & Schmidt-Atzert, 2006). Gegenstand der Arbeitspsychologie ist das menschliche Erleben und Verhalten in der Arbeit. Daher sind Zusammenhänge zwischen individuellen psychologischen Merkmalen und Merkmalen der Arbeitssituation in der Arbeitspsychologie von besonderer Bedeutung. Mit der Konzentration auf Vorhersagegenauigkeiten wurde in den vorliegenden Untersuchungen dieser Betrachtung der Vorzug vor testtheoretischen Analysen der verwendeten Instrumente gegeben. Dies geschah selbstverständlich, ohne testtheoretischen Konzepten in ihrem Anwendungsbereich (z. B. Messung von Persönlichkeitseigenschaften, klinische Diagnostik, Personalauswahl) ihre Bedeutung und Verdienste absprechen zu wollen (Amelang & Schmidt-Atzert, 2006). Im Gegenteil wurde mit der gesundheitsbezogenen Selbstwirksamkeit (Gesundheitskompetenz) ein Kriterium einbezogen, das durchaus akzeptable psychometrische Kennwerte aufweist (Wieland & Hammes, 2009). Es bleibt jedoch festzuhalten, dass die vorliegenden Untersuchungen ausschließlich Aufschluss über Vorhersagegenauigkeiten der induzierten Klassifikatoren und keinerlei Anhaltspunkte über die psychometrischen Eigenschaften der eingesetzten Instrumente liefern. Letztere sind anderen Veröffentlichungen zu entnehmen (Baumgart et al., 2002; Hammes et al., 2009, 2010; Hammes & Wieland, 2009, 2014b; Scherrer, 2002; Wieland et al., 2001, 2006, 2009; Wieland & Hammes, 2009; Wieland-Eckelmann et al., 1999).

Vergleiche von Klassifikator-Induktoren unterliegen den folgenden Einschränkungen: (a) Vollständige Suche, heuristische Suche und die Bootstrapping-Experimente wurden mit unterschiedlich kodierten Daten durchgeführt. (b) Binärlogistische Regression wurde unter vollständiger Suche nicht eingesetzt. (c) Binärlogistische Regression wurde bei der endgültigen Auswahl, Kalibrierung und Validierung von Klassifikatoren (Bootstrapping-Experimente) ausgeschlossen. (d) Aussagen über andere Selektionsverfahren oder Klassifikatoren lassen sich nicht treffen. Gültigkeitsbereiche entsprechender Ergebnisse und Aussagen werden im Folgenden genauer eingegrenzt.

Bei der vollständigen Suche wurden die Rohwerte der EEB-Items verwendet, während bei der heuristischen Suche die dichotomisierten Daten zum Einsatz kamen. Dieser Entscheidung

lag die Frage zu Grunde, ob bereits mit einem oder wenigen EEB-Items hinreichend gute Vorhersagegenauigkeiten erzielt werden können. Da mit sieben Antwortkategorien bereits nach zwei bis drei Items die optimale Messkomplexität erreicht bzw. überschritten ist (Hughes, 1968), lag die Verwendung der Rohdaten bei vollständiger Suche nahe. Bei heuristischer Suche wurden die dichotomisierten Daten verwendet, um die Überlegenheit eines der beiden Selektions-Verfahren (Sequential Forward Selection [SFS] oder Sequential Forward Floating Selection [SFFS]) gegenüber dem anderen festzustellen. Es wurde davon ausgegangen, dass diese Frage nur bei einem hinreichend großen geeigneten Suchraum beantwortet werden kann. Durch die Dichotomisierung der Daten wurde sichergestellt, dass etwa vier bis fünf EEB-Items für eine hinreichend gute Vorhersagegenauigkeit ausgewählt werden müssen und die resultierenden Zellen hinreichend stark besetzt sind. Bei den Bootstrapping-Experimenten zur endgültigen Auswahl von Klassifikatoren wurde schließlich auf die dichotomisierten Daten zurückgegriffen, um auch bei kleinen Stichproben (z. B. Abteilungen mit zehn bis 20 Beschäftigten) noch brauchbare Zellbesetzungen sicherzustellen. Resultat dieser Entscheidungen ist, dass die Item-Kombinationen aus der vollständigen Suche nur in vier Fällen unter den endgültig ausgewählten Klassifikatoren zu finden sind (s. u.). Dies bedeutet jedoch nicht, dass vollständige Suche weniger geeignet ist, als heuristische Suche. Im Gegenteil führt vollständige Suche aus theoretischer Perspektive immer zu einem optimalen Ergebnis. Voraussetzung ist jedoch, dass die Kodierung der Daten erhalten bleibt. Ein fairer Vergleich von vollständiger und heuristischer Suche sollte in folgenden Untersuchungen unter gleichen Bedingungen (u. A. gleiche Kodierung der Daten) erfolgen.

Mit binärlogistischer Regression ist im Vergleich zur Bayes-Klassifikation ein bedeutend höherer Rechenaufwand verbunden (Doumpos & Slappa, 2005). Dieser erhöhte Rechenaufwand macht sich bei vollständiger Suche und Bootstrapping-Experimenten besonders bemerkbar, da hier sehr häufig Klassifikatoren induziert werden müssen. Die eingesetzte Analyse- und Entwicklungsumgebung SageMathCloud[®] kam hier an ihre technischen Grenzen. Es kam zu Abbrüchen (Time-Out) bzw. zum Einfrieren des Systems. Dieser Umstand hat zur Folge, dass keine binärlogistischen Regressions-Modelle nach vollständiger Suche zum Vergleich vorliegen. Aus demselben Grunde liegen solche Modelle auch nicht als endgültig ausgewählte Klassifikatoren vor. Über ihre externe Validität liegen keine Befunde vor. Faire Vergleiche zwischen Bayes-Klassifikatoren und binärlogistischen Regressions-Modellen sind daher in Bezug auf die Vorhersagegenauigkeit nur unter heuristischer Suche möglich. Andererseits weist dieser Umstand auf die entscheidende Stärke der Bayes-Klassifikatoren hin: Sie sind bei vergleichbarer Vorhersagegenauigkeit wesentlich effizienter als binärlogistische

Regression. Zugleich wird dadurch bestätigt, dass heuristische Suche gegenüber vollständiger Suche einen starken Effizienz-Vorteil aufweist (a. a. O.).

Wie bereits einleitend im Kapitel über das methodische Vorgehen erläutert wurde, ist die Auswahl an Klassifikatoren und Selektions-Algorithmen nicht zuletzt durch ihre Anpassungsfähigkeit an unterschiedlichste Problemstellungen nahezu unbegrenzt (Clarke et al., 2009; Hastie et al., 2009; Jain et al., 2000; Kotsiantis et al., 2006; Krizek, 2008; Maimon & Rokach, 2005; Somol et al., 1999; Yusta, 2009). Die hier getroffene Auswahl erfolgte unter bestimmten Vorannahmen und theoretischen Überlegungen, die weiter oben dargelegt wurden. Es ist nicht auszuschließen, dass andere Wissenschaftlerinnen und Wissenschaftler eine gut begründete andere Auswahl treffen würden. Somit sind die hier berichteten Befunde auf die verwendeten Verfahren und Algorithmen beschränkt. Ob und inwiefern andere Verfahren und Algorithmen unter Verwendung der gleichen Datenbasis zu besseren Ergebnissen führen, kann nur durch entsprechende (theoretische oder experimentelle) Untersuchungen entschieden werden.

7.2 METHODISCHE DISKUSSION VORLIEGENDER UNTERSUCHUNGSBEFUNDE

Insgesamt ist – unter Berücksichtigung weiter oben dargestellter Einschränkungen – festzustellen, dass die eingesetzten Klassifikator-Induktoren in ihrem Verhalten in allen Aspekten den theoretischen Erwartungen entsprechen. Dies soll im Folgenden verdeutlicht werden an (a) den CV-T-Diagrammen, (b) den Vorhersagegenauigkeiten in Abhängigkeit der verwendeten Selektions-Verfahren und Klassifikatoren, (c) den ausgewählten Item-Kombinationen, (d) dem Vergleich der Klassifikatoren untereinander sowie (e) den Ergebnissen der Kalibrierung und Validierung der Klassifikatoren.

7.2.1 Diskussion und Interpretation der CV-T-Diagramme

Bereits im Ergebnisteil wurde dargelegt, dass die CV-T-Diagramme für Bayes-Klassifikatoren das theoretisch zu erwartende Verhalten bei steigender Messkomplexität widerspiegeln. Es wurde festgestellt: Die Trainings-Korrektklassifikation steigt monoton mit zunehmender Messkomplexität. Bei maximaler Messkomplexität (Einbezug aller 25 EEB-Items) wird fast perfekte Klassifikation auf den Trainingsdaten erzielt. Die Cross-Validation-Korrektklassifikation steigt bis zu einem Maximum an und fällt anschließend monoton. Bei maximaler Messkomplexität gelingt die Klassifikation auf den Testdaten nicht mehr. Bei hinreichend großem Umfang an Instanzen wäre hier eine Stagnation bei $\kappa_{CV} = .5$ zu erwarten (Hughes, 1968; Kanal & Chandrasekaran, 1971; Narendra & Fukunaga, 1977; Raudys, 2006; Reunanen, 2003). Dass die CV-Korrektklassifikation auf Null sinkt, liegt an zu geringen

Zellbesetzungen: Bei 25 dichotomisierten EEB-Items gibt es 2^{25} bzw. gut 33 Millionen mögliche Zellen. Bei einem Umfang von etwa 100 Instanzen in den Trainingsdaten sind nach dem Training für maximal 100 Zellen die benötigten bedingten Wahrscheinlichkeiten definiert. D. h. für fast alle möglichen Zellen sind die bedingten Wahrscheinlichkeiten nicht definiert. Daher können die Instanzen in den Testdaten nicht mehr sinnvoll den beiden Klassen zugeordnet werden. Umgekehrt ist es sehr unwahrscheinlich, dass zwei Instanzen der Trainingsdaten bei allen 25 EEB-Items denselben Wert aufweisen. D. h. die meisten Instanzen der Trainingsdaten besetzen einzeln eine eigene Zelle. Die für eine solche Zelle ermittelte bedingte Wahrscheinlichkeit sorgt für eine sichere Zuordnung der entsprechenden Instanz zur korrekten Klasse. Dies erklärt die nahezu perfekte Trainings-Korrektklassifikation bei maximaler Messkomplexität.

Die Betrachtung der maximalen Messkomplexität veranschaulicht eine Überanpassung eines Bayes-Klassifikators an die Trainingsdaten. Um diese zu vermeiden, wurde festgelegt, dass eine heuristische Suche bei signifikantem Unterschied zwischen Trainings- und Cross-Validation-Korrektklassifikation abbrechen soll (Kanal & Chandrasekaran, 1971; Kohavi & Sommerfield, 1995; Raudys, 2006; Reunanen, 2003). Dieses Abbruchkriterium führt bei der Induktion von Bayes-Klassifikatoren zu einem Umfang von zwei bis fünf ausgewählten EEB-Items als Prädiktoren. Dieser Umfang entspricht weitgehend einer optimalen Messkomplexität (Hughes, 1968). Zugleich ist den CV-T-Diagrammen zu entnehmen, dass bereits mit dem ersten ausgewählten EEB-Item eine CV-Korrektklassifikation erzielt wird, die mit der maximalen CV-Korrektklassifikation vergleichbar ist (vgl. Hua et al., 2005). Bei Auswahl des ersten Items sind Trainings- und CV-Korrektklassifikation weitgehend identisch. Dies ist darauf zurückzuführen, dass Trainings- und Testdaten zufällige Teilstichproben aus den Induktionsstichproben sind. Bei einem dichotomisierten EEB-Item stehen nur zwei Zellen zur Besetzung zur Verfügung. Auf Grund der Zufälligkeit der beiden Teilstichproben ist davon auszugehen, dass sich die Instanzen in Abhängigkeit der Ausprägungen der Kriterien jeweils in gleichen Verhältnissen auf diese beiden Zellen verteilen. Die mit Hilfe der Trainingsdaten ermittelten bedingten Wahrscheinlichkeiten sollten demnach die Verhältnisse in den Testdaten weitgehend korrekt wiedergeben, so dass es zu weitgehend identischen Korrektklassifikationen kommt. Solange die durch Item-Kombination erzeugte Anzahl an Zellen gering genug bleibt, um weiterhin gleiche Verhältnisse in den Trainings- und Testdaten abbilden zu können, bleiben Trainings- und CV-Korrektklassifikation vergleichbar. Im vorliegenden Falle gilt dies für bis zu fünf EEB-Items. Erst bei höherer Messkomplexität kommt es zu den beschriebenen Differenzen zwischen Trainings- und CV-Korrektklassifikation.

Auf Grund der Zufälligkeit der Trainings- und Test-Teilstichproben sollte in beiden Teilstichproben das Quotenverhältnis $\frac{p_1}{1-p_1}$ zwischen den Wahrscheinlichkeiten für Klasse 0 und Klasse 1 annähernd identisch sein. Dies gilt unabhängig von ausgewählten Prädiktoren. Auch die Verteilung der Instanzen auf die beiden Zellen jedes einzelnen dichotomisierten EEB-Items sollte in beiden Teilstichproben annähernd identisch sein. Diese beiden Umstände erklären die Charakteristika der CV-T-Diagramme für binärlogistische Regression. Das binärlogistische Regressionsmodell wird durch Schätzung der Parameter β derart angepasst, dass das Quotenverhältnis unabhängig von der Anzahl verwendeter Prädiktoren bestmöglich vorhergesagt werden kann (Agresti, 2002, 2007). Da sich die Verhältnisse in beiden Teilstichproben sowohl bezüglich des Quotenverhältnisses als auch bezüglich der einzelnen EEB-Items weitgehend gleichen, ist ein mit den Trainingsdaten induziertes binärlogistisches Regressionsmodell auch ein geeignetes Modell für die Testdaten. Damit ist klar, warum Trainings- und CV-Korrektklassifikation unabhängig von der Messkomplexität nahezu identisch sind und, beginnend mit dem ersten ausgewählten Item, auf etwa gleichem Niveau bleiben. Dennoch weisen die CV-T-Diagramme signifikante Unterschiede zwischen Trainings- und CV-Korrektklassifikation nach, wobei in der Regel die Trainings-Korrektklassifikation etwas besser ausgeprägt ist, als die CV-Korrektklassifikation. Dies ist insofern plausibel, da die Regressions-Modelle unter Verwendung der Trainingsdaten induziert werden und somit besser an diese als an die Testdaten angepasst sind.

Das charakteristische Verhalten von Trainings- und CV-Korrektklassifikation in Abhängigkeit der Messkomplexität ist bei Induktion von sowohl Bayes-Klassifikatoren als auch von binärlogistischen Regressions-Modellen für alle 18 Kriterien identisch. Es ist darüber hinaus, wie nun ausführlich dargelegt wurde, theoretisch erklärbar. Dies weist zum einen auf die Korrektheit der implementierten Klassifikator-Induktoren hin. Zum anderen werden theoretische Erwartungen bezüglich dieser Induktoren bestätigt.

7.2.2 Methodische Diskussion und Interpretation von Vorhersagegenauigkeiten

Für alle 18 Kriterien wird mindestens ein Klassifikator induziert, dessen CV-Korrektklassifikation auf dem Niveau $p < .001$ signifikant besser ist, als die Korrektklassifikation $\kappa_{\text{Konst.}} = .500$ bei konstanter Zuordnung der Instanzen zu einer Klasse. Bei heuristischer Suche wird dieses Signifikanzniveau von jedem induzierten Klassifikator erreicht. Bei Regulationsbehinderungen, Tätigkeitsspielraum, partizipativer Unternehmenskultur, Gesundheitskompetenz sowie Rücken- und unspezifischen Beschwerden führt vollständige Suche ebenfalls durchgehend zu Klassifikatoren, die dieses Signifikanzniveau

erreichen. Für weitere zehn Kriterien führt vollständige Suche sowohl bezüglich der vollständigen EEB als auch des WSIB zu mindestens einem Klassifikator, der dieses Signifikanzniveau erreicht. Für Zusammenarbeit kann mit Hilfe der vollständigen Suche kein Klassifikator bezüglich des WSIB gefunden werden, der dieses Signifikanzniveau erreicht. Die für Kommunikation und Kooperation durch vollständige Suche ermittelten Klassifikatoren erreichen durchgehend nicht dieses Signifikanzniveau.

Bleiben wir bei der vollständigen Suche und betrachten den Vergleich mit der theoretisch zu erwartenden mittleren Korrektklassifikation nach Hughes (1968). Hier ist festzustellen, dass die Korrektklassifikation fast aller ermittelten Klassifikatoren auf dem Niveau $p < .001$ signifikant unterhalb dieser Marke liegt. Ausnahmen finden sich nur bei der Vorhersage von Muskel-Skelett-Beschwerden. Das beste Einzelitem („körperlich verspannt“) erzielt hier eine besonders gute CV-Korrektklassifikation von .696. Mit $p = .030$ ist jedoch auch diese noch signifikant geringer, als die Hughes-Korrektklassifikation von .730 (berechnet nach Angaben in Hughes, 1968). Es sei nochmals darauf hingewiesen, dass solche Vergleiche für heuristische Suche nicht zur Verfügung stehen. Da die Umfänge ausgewählter Items zwischen einzelnen Durchgängen der Kreuzvalidierung variieren, lässt sich eine Hughes-Korrektklassifikation hier nicht eindeutig zum Vergleich angeben. Unter Verwendung von Sequential Forward Selection wurde ein Bayes-Klassifikator zur Vorhersage von Muskel-Skelett-Beschwerden induziert, der mit .705 die höchste CV-Korrektklassifikation aufweist. Diese Korrektklassifikation liegt noch unterhalb üblicher Hughes-Korrektklassifikationen. Es kann daher davon ausgegangen werden, dass alle Klassifikatoren die Hughes-Korrektklassifikation nicht übertreffen.

Beim Vergleich von mittlerer Trainings- und CV-Korrektklassifikation sind einige systematische Sachverhalte erkennbar. Beginnend mit der heuristischen Suche ist zunächst festzustellen, dass sich Trainings- und CV-Korrektklassifikation für keinen der ermittelten Klassifikatoren signifikant unterscheiden. Lediglich der unter Verwendung von Sequential Forward Floating Selection induzierte Bayes-Klassifikator zur Vorhersage von Herz-Kreislauf-Beschwerden stellt eine Ausnahme dar. Dessen CV-Korrektklassifikation von .651 ist auf dem Niveau $p < .05$ signifikant höher als dessen mittlere Trainings-Korrektklassifikation von .613. Bei genauerer Betrachtung der Unterschiede wird darüber hinaus jedoch deutlich, dass sich Trainings- und CV-Korrektklassifikation für binärlogistische Regressions-Modelle am wenigsten unterscheiden. Die größten Unterschiede sind für Bayes-Klassifikatoren unter Verwendung von Sequential Forward Floating Selection feststellbar. Diese Regelmäßigkeit trifft in mehr als 50 % der 36 Fälle (18 Kriterien x 2 Instrumente) zu. Bereits weiter oben wurde dargelegt, dass bei binärlogistischer Regression mit einer hohen Übereinstimmung zwischen

Trainings- und CV-Korrektklassifikation zu rechnen ist. Bei vollständiger Suche ist eine Beobachtung zu machen, die für alle 18 Kriterien und je beide Instrumente (vollständige EEB und WSIB) zutrifft. Mit steigender Messkomplexität (bestes Einzelitem, bestes Paar, bestes Tripel) nimmt sowohl die Trainings-Korrektklassifikation als auch die Differenz zwischen Trainings- und CV-Korrektklassifikation zu. Dabei ist die Trainings-Korrektklassifikation immer größer oder gleich der CV-Korrektklassifikation. Für das beste Tripel ist diese Differenz immer auf dem Niveau $p < .001$ signifikant. Da es sich bei den induzierten Klassifikatoren um Bayes-Klassifikatoren handelt, ist diese Beobachtung leicht verständlich. Wie weiter oben ausgeführt, ist ein Bayes-Klassifikator mit zunehmender Messkomplexität immer besser an die Trainingsdaten angepasst. Da außerdem bei vollständiger Suche die Rohdaten der EEB-Items mit sieben Antwortkategorien verwendet wurden, ist die optimale Messkomplexität bei zwei Items ($7 \times 7 = 49$) bereits überschritten. Bei drei Items ($7 \times 49 = 343$) ist sie bereits so stark überschritten, dass ein Großteil der möglichen Zellen von den Trainingsdaten nicht mehr besetzt wird. Das führt zu den vergleichsweise geringen CV-Korrektklassifikationen, weil viele Instanzen der Testdaten nicht mehr sinnvoll den beiden Klassen zugeordnet werden können. Der Klassifikator mit der besten CV-Korrektklassifikation weist in den meisten Fällen noch keinen signifikanten Unterschied zur Trainings-Korrektklassifikation auf. Ausnahmen stellen Kommunikation und Kooperation (vollständige EEB und WSIB), Regulationsbehinderungen (vollständige EEB), Zusammenarbeit (vollständige EEB) und Herz-Kreislaufbeschwerden (WSIB) dar. Für diese fünf Klassifikatoren ist die Differenz auf einem Niveau von $p < .05$ signifikant.

An dieser Stelle kann ein erstes Zwischenfazit gezogen werden. Bei den Erläuterungen des methodischen Vorgehens wurden verschiedene Bedingungen bzw. Erwartungen bezüglich der Vorhersagegenauigkeiten bzw. Korrektklassifikationen formuliert. CV-Korrektklassifikationen sollten (a) zwischen den Korrektklassifikationen bei konstanter Zuordnung und den Hughes-Korrektklassifikationen liegen und (b) Trainings- und Test-Korrektklassifikationen sollten sich nicht signifikant voneinander unterscheiden. Bedingung (a) ist mit den folgenden beiden Ausnahmen erfüllt: Für Zusammenarbeit konnte unter vollständiger Suche kein Klassifikator bezüglich des WSIB induziert werden, dessen CV-Korrektklassifikation signifikant besser ist als die Korrektklassifikation bei konstanter Zuordnung. Für Kommunikation und Kooperation konnte bezüglich beider Instrumente (vollständige EEB und WSIB) unter vollständiger Suche kein entsprechender Klassifikator induziert werden. Auch für diese beiden Kriterien konnte jedoch unter heuristischer Suche jeweils mindestens ein Klassifikator induziert werden, der Bedingung (a) erfüllt. Bedingung (b) wird unter vollständiger Suche in den oben genannten

fünf Ausnahmen gebrochen. Unter heuristischer Suche ist Bedingung (b) nur für einen einzigen Klassifikator zur Vorhersage von Herz-Kreislauf-Beschwerden nicht erfüllt. Für jedes Kriterium konnte jedoch mindestens ein Klassifikator induziert werden, der Bedingung (b) erfüllt. Betrachtet man die berichteten Ausnahmen in der Zusammenschau, so stellen sich Zusammenarbeit und Kommunikation und Kooperation als zwei besonders problematische Kriterien heraus. Für beide Kriterien gelang es unter vollständiger Suche nicht, Klassifikatoren zu induzieren, die zugleich die Bedingungen (a) und (b) erfüllen. Nur unter heuristischer Suche gelingt dies auch für diese beiden Kriterien.

Die Bootstrapping-Experimente zur endgültigen Auswahl von Klassifikatoren führten zu weiteren Korrektklassifikationen, die nun diskutiert werden sollen. Vorgabe zur Auswahl eines Klassifikators war, dass die entsprechende Bootstrapping-Korrektklassifikation maximal, jedoch kleiner oder gleich der Hughes-Korrektklassifikation von .75 bei unendlich vielen Messungen sein sollte (Hughes, 1968). Diese Bedingung wird von allen ausgewählten Klassifikatoren erfüllt. Die Bootstrapping-Korrektklassifikationen sind kleiner als .75 und unterscheiden sich von diesem Wert nur für zwei Kriterien signifikant: Kommunikation und Kooperation und Bedeutung von Anerkennung und Wertschätzung. D. h. für diese beiden Kriterien sind die Bootstrapping-Korrektklassifikationen bedeutend kleiner als .75. Eine weitere Regel zur Auswahl eines Klassifikators wurde wie folgt festgelegt: Ist die Bootstrapping-Korrektklassifikation eines ausgewählten Klassifikators bezüglich des WSIB größer oder gleich jener eines ausgewählten Klassifikators bezüglich der vollständigen EEB, so wird nur der Klassifikator bezüglich des WSIB ausgewählt. Trifft diese Regel bei einem Kriterium zu, so ist demnach nur ein Klassifikator aufgeführt. Dieser wird sowohl auf Erhebungen mittels der vollständigen EEB als auch auf Erhebungen mittels des WSIB angewendet. Diese Regel wurde damit begründet, dass die Items des WSIB eine Teilmenge der Items der vollständigen EEB darstellen. Sie konnte auf fünf Kriterien angewendet werden: Leistungs- und Zeitvorgaben, ganzheitliche Führung, Muskel-Skelett-Beschwerden, unspezifische Beschwerden sowie Präsentismus. Für die verbleibenden Kriterien wurden die Bootstrapping-Korrektklassifikationen bezüglich der beiden Instrumente (vollständige EEB und WSIB) auf signifikante Unterschiede hin überprüft. Ein solcher signifikanter Unterschied konnte für keines der verbleibenden 13 Kriterien festgestellt werden. Die beiden Bootstrapping-Korrektklassifikationen weisen für jedes Kriterium je einen vergleichbaren Wert auf. Dieses Ergebnis deutet darauf hin, dass das nur neun Items umfassende WSIB für hier behandelte Klassifikationsanwendungen vergleichbar gut geeignet ist, wie die 25 Items umfassende

vollständige EEB. Auf die Bedeutung dieses Ergebnisses für die praktische Anwendung wird weiter unten eingegangen.

Im vorliegenden Abschnitt wurde deutlich, dass auch die Vorhersagegenauigkeiten bzw. Korrektklassifikationen den theoretischen Erwartungen voll und ganz entsprechen. Systematische Muster beim Vergleich der Korrektklassifikationen können theoretisch schlüssig erklärt werden und die erhaltenen Korrektklassifikationen erfüllen weitgehend die zuvor theoretisch abgeleiteten Bedingungen. Hierbei ist selbstverständlich zwischen festgelegten Normen (z. B. Beachtung von Signifikanzniveaus) und quantitativen Tatsachen zu unterscheiden. Der Klassifikator mit der geringsten CV-Korrektklassifikation von .502 wurde unter vollständiger Suche als bestes Tripel bezüglich des WSIB zur Vorhersage von Zusammenarbeit induziert. Diese CV-Korrektklassifikation unterscheidet sich zwar nicht signifikant von der Korrektklassifikation bei konstanter Zuordnung. Dennoch ist sie um die Differenz .002 größer als jene. Die weiter oben benannten Ausnahmen sind insofern nur normativ zu verstehen. Aus rein quantitativer Perspektive ist oben genannte Bedingung (a) bei angemessener Messkomplexität immer erfüllt. Und aus theoretischer Sicht sollte sie das auch sein. Des Weiteren liefert der Nachweis der Erfüllung theoretischer Erwartungen einen weiteren Hinweis auf die Korrektheit der implementierten Klassifikator-Induktoren.

7.2.3 Methodische Diskussion und Interpretation der ausgewählten Items

Im vorliegenden Abschnitt steht die methodische Diskussion und Interpretation ausgewählter Item-Kombinationen im Mittelpunkt. Ihre inhaltliche Interpretation erfolgt weiter unten. Aus methodischer Perspektive ist zu untersuchen, inwiefern die Umfänge ausgewählter EEB-Items theoretischen Erwartungen entsprechen und inwiefern relevante EEB-Items ausgewählt wurden. Auch die Stabilität der Item-Kombinationen soll hier diskutiert werden.

Zunächst erfolgt die Diskussion in Bezug auf vollständige Suche. Die Umfänge ausgewählter EEB-Items sind in diesem Falle vorgegeben. Es wurde bereits erörtert, dass die optimale Messkomplexität bereits bei zwei Items überschritten ist. Es ist daher zu erwarten, dass die beste CV-Korrektklassifikation durch das beste Einzelitem oder das beste Paar, seltener jedoch durch das beste Tripel erzielt wird. Nur in zwei Ausnahmen erzielt das beste Tripel die höchste CV-Korrektklassifikation: Bei Kommunikation und Kooperation sowie bei Regulationsbehinderungen, jeweils bezüglich der vollständigen EEB. Kommunikation und Kooperation wurde bereits weiter oben als problematisch in Bezug auf vollständige Suche identifiziert. Für Regulationsbehinderungen ist festzustellen, dass das beste Einzelitem mit .619 eine fast identische CV-Korrektklassifikation im Vergleich zu .620 beim besten Tripel erzielt.

Dieser geringe Unterschied kann als Zufallsergebnis interpretiert werden. Bezüglich des EEB erzielt neun Mal das beste Einzelitem und sieben Mal das beste Paar die höchste CV-Korrektklassifikation. Bezüglich des WSIB fällt die höchste CV-Korrektklassifikation elf Mal auf das beste Einzelitem und sieben Mal auf das beste Paar. Mit 20 : 14 : 2 dominiert somit das beste Einzelitem gegenüber dem besten Paar und dem besten Tripel. Die vorgenommene starke Einschränkung der vollständigen Suche auf nur drei Items stellt sich angesichts dieses Ergebnisses als effiziente methodische Herangehensweise heraus und gründete auf theoretischen Überlegungen zur Messkomplexität. Auf die praktische Relevanz dieses Ergebnisses wird weiter unten vertieft eingegangen.

Bei 16 der 18 Kriterien wird bei vollständiger Suche das EEB- bzw. WSIB-Item, das die höchste Korrelation¹⁴ mit dem Kriterium aufweist, in die Item-Kombination mit der höchsten CV-Korrektklassifikation aufgenommen. In den meisten Fällen erfolgt dies jeweils in allen zehn Durchgängen der Kreuzvalidierung. In sechs der insgesamt neun Ausnahmen kommt mindestens eines der beiden Items mit den höchsten Korrelationen in den Item-Kombinationen vor. Die verbleibenden drei Ausnahmen beziehen sich auf Kommunikation und Kooperation, Leistungs- und Zeitvorgaben und Zusammenarbeit, jeweils bezüglich der vollständigen EEB. Zur Vorhersage von Kommunikation und Kooperation werden durchgehend Items ausgewählt, die eine zu vernachlässigende Korrelation mit diesem Kriterium aufweisen. Zur Vorhersage von Leistungs- und Zeitvorgaben wird durchgehend das Item „leicht“ als bestes Einzelitem gewählt, das mit -.014 eine ebenfalls zu vernachlässigende Korrelation mit diesem Kriterium aufweist. Zur Vorhersage von Zusammenarbeit wird zwar das Item „zufrieden“ mit der höchsten Korrelation vier Mal in das beste Paar aufgenommen. Durchgehend kommt jedoch das Item „leicht“ im besten Paar vor. Dieses Item weist mit .154 nur die viertgrößte Korrelation mit dem Kriterium auf. Zusammenfassend lässt sich festhalten: Versteht man die Relevanz eines Items in einer ersten Näherung als Korrelation des Items mit dem jeweiligen Kriterium, so wird in 33 von 36 Fällen mindestens eines der beiden relevantesten Items in die Item-Kombinationen aufgenommen (Blum & Langley, 1997; Caruna & Freitag, 1994; Hua et al., 2005).

Bei Betrachtung der besten Paare tritt ein weiteres Phänomen zu Tage, das theoretisch erörtert wurde. Bezüglich der vollständigen EEB entsprechen nur 13 der insgesamt 180 besten Paare der Kombination der beiden EEB-Items, die die höchsten Korrelationen mit dem jeweiligen Kriterium aufweisen. Bei vier Kriterien wurden diese beiden Items jedoch

¹⁴ Gemeint ist jeweils die *absolut* höchste Korrelation, also jene Korrelation, die den höchsten Absolutbetrag aufweist.

abwechselnd als bestes Einzelitem gewählt. Bezüglich des WSIB sind es 40 der 180 besten Paare. Hier wurden diese beiden Items bei fünf Kriterien abwechselnd als beste Einzelitems gewählt. Diese Betrachtungen bestätigen, dass zwei beste Einzelitems nur in wenigen Fällen auch das beste Paar zur Vorhersage eines Kriteriums bilden (Blum & Langlay, 1997; Caruna & Freitag, 1994; Cover & Van Campenhout, 1977; Elashoff et al., 1967; Guyon & Elisseeff, 2003; Toussaint, 1971; Cover, 1974). Dass bezüglich des WSIB mehr solcher Fälle auftreten, liegt an der gegenüber der vollständigen EEB stark verringerten Anzahl zur Verfügung stehender Items.

Die hier implementierte vollständige Suche kann als sehr stabiles Verfahren betrachtet werden. In 21 Fällen weist die Item-Kombination mit der höchsten CV-Korrektklassifikation eine optimale Stabilität von 1 auf. D. h. in diesen Fällen sind die entsprechenden Item-Kombinationen jeweils über alle zehn Durchgänge der Kreuzvalidierung hinweg identisch (Kuncheva, 2007). In weiteren sieben Fällen sind die Item-Kombinationen in jeweils neun der zehn Durchgänge identisch. In diesen Fällen wird die Stabilität mit .800 beziffert. Darüber hinaus werden nie mehr als zwei unterschiedliche Item-Kombinationen zur Vorhersage eines Kriteriums ausgewählt. Kommen zwei unterschiedliche Paare zur Vorhersage eines Kriteriums vor, so ist immer ein Item bei allen zehn Durchgängen in den Item-Kombinationen vertreten. Ausnahme dieser Regel ist wieder das Kriterium Kommunikation und Kooperation. Hier werden bezüglich des WSIB drei unterschiedliche Paare ausgewählt und das Item „aufmerksam“ kommt in nur neun der zehn Durchgänge vor. Die geringste Stabilität von .444 weist das beste Einzelitem bezüglich der vollständigen EEB zur Vorhersage von Herz-Kreislauf-Beschwerden auf. Hier wird in fünf der zehn Durchgänge der Kreuzvalidierung das Item „körperlich unwohl“ gewählt. In den anderen fünf Durchgängen wird das Item „körperlich verspannt“ gewählt.

Die Befunde zu Item-Kombinationen bei vollständiger Suche lassen sich wie folgt zusammenfassen. Es genügen ein bis zwei EEB- bzw. WSIB-Items (Kodierung in Original-Skala), um mittels Bayes-Klassifikatoren die betrachteten Kriterien vorherzusagen. Unter Beachtung der Messkomplexität lässt sich die vollständige Suche ohne wesentliche Verluste effizient beschränken. Vollständige Suche führt zu der Aufnahme der relevantesten Items. Bei vollständiger Suche werden mehrere Items jedoch nicht ausschließlich entsprechend ihrer Korrelationen sondern im Sinne einer verbesserten Vorhersagegenauigkeit kombiniert (Blum & Langley, 1997; Caruna & Freitag, 1994; Guyon & Elisseeff, 2003; Hua et al., 2005). Unter Kreuzvalidierung erweist sich vollständige Suche als ein sehr stabiles Verfahren. D. h. über verschiedene Durchgänge der Kreuzvalidierung hinweg werden weitgehend identische Item-Kombinationen ausgewählt.

Bei heuristischer Suche sind die Umfänge der auszuwählenden Item-Kombinationen nicht mehr vorgegeben. Dies macht sich unmittelbar in den Stabilitäten (Kuncheva, 2007) bemerkbar. Insbesondere bei der Induktion von binärlogistischen Regressions-Modellen wird der gesamte Wertebereich $[0, 1]$ für Stabilitäten abgedeckt. Beträgt die Stabilität Null, so werden Kombinationen von 18 (Gesamtbedarf für Arbeitsgestaltung), 21 (Kommunikation und Kooperation) oder gar 24 (Partizipative Unternehmenskultur) EEB-Items ausgewählt. Beträgt die Stabilität Eins, so wird immer ein Einzelitem ausgewählt. Bei Aufgabenanforderungen das Item „konzentriert“, bei Tätigkeitsspielraum „einflussreich“, bei autoritäre Führung „körperlich unwohl“, bei Bedeutung von Anerkennung und Wertschätzung „leistungsbereit“, bei Zusammenarbeit „körperlich unwohl“ und bei Muskel-Skelett-Beschwerden das Item „körperlich verspannt“. Ausgenommen Zusammenarbeit handelt es sich dabei immer um das EEB-Item, das die höchste Korrelation mit dem entsprechenden Kriterium aufweist. Binärlogistische Regression in Verbindung mit Sequential Forward Selection muss daher als sehr instabiles Verfahren betrachtet werden. Dies liegt sicherlich an dem hier verwendeten Abbruchkriterium, das gleichermaßen bei der Induktion von binärlogistischen Regressions-Modellen und Bayes-Klassifikatoren implementiert wurde. Wird Bayes-Klassifikation mit Sequential Forward Selection (SFS) kombiniert, so reicht die Stabilität bezüglich der vollständigen EEB von .022 (Partizipative Unternehmenskultur, Kombinationen von 1 bis 7 Items) bis .467 (autoritäre Führung, Kombinationen von 1 bis 6 Items). Bezüglich des WSIB reicht sie in diesem Falle von .044 (ganzheitliche Führung, Kombinationen von 1 bis 7 Items) bis .489 (Muskel-Skelett-Beschwerden, Kombinationen von 1 bis 5 Items). Bei Kombination von Bayes-Klassifikation mit Sequential Forward Floating Selection (SFFS) reicht die Stabilität bezüglich der vollständigen EEB von .022 (Kommunikation und Kooperation, Kombinationen von 1 bis 6 Items) bis .800 (ganzheitliche Führung, 9 Mal Item „zufrieden“, und Muskel-Skelett-Beschwerden, 9 Mal Item „körperlich verspannt“). Bezüglich des WSIB reicht sie in diesem Falle ebenfalls von .022 (Magen-Darm-Beschwerden, Kombinationen von 1 bis 7 Items) bis .800 (Muskel-Skelett-Beschwerden, 9 Mal Item „körperlich verspannt“). Maximal werden bei Bayes-Klassifikation acht Items ausgewählt. Dies geschieht genau zwei Mal und ausschließlich in Kombination mit SFFS (Gesamtbedarf für Arbeitsgestaltung bezüglich WSIB und Zusammenarbeit bezüglich vollständiger EEB). Überwiegend (in 88 % aller Fälle) werden bei der Induktion von Bayes-Klassifikatoren jedoch höchstens fünf Items ausgewählt. Dies entspricht der Item-Anzahl, bei der die optimale Messkomplexität gerade überschritten wird. Für binärlogistische Regressions-Modelle trifft dies nur in 78 % der Fälle zu. Die hier beschriebene geringe Stabilität ist bei heuristischer Suche zu erwarten. Die Ergebnisse zur

binärlogistischen Regression machen jedoch auch deutlich, dass eine angemessene Definition der Abbruchbedingung mit entscheidend ist für einen guten Klassifikator-Induktor. In der vorliegenden Arbeit wurde die Abbruchbedingung durchgängig festgelegt, um genau solche Vergleiche vornehmen zu können.

Inwiefern bei heuristischer Suche relevante Items in die Item-Kombinationen aufgenommen wurden, wird wieder durch die Betrachtung der Items deutlich, die die höchsten Korrelationen mit den entsprechenden Merkmalen aufweisen. In Bezug auf die vollständige EEB wird in 81 % der Fälle mindestens eines dieser beiden Items aufgenommen. In 28 % der Fälle werden beide Items mit den höchsten Korrelationen aufgenommen. In Bezug auf das WISB wird zu 98 % mindestens eines der beiden Items aufgenommen und in 34 % der Fälle beide. Dieser Unterschied ist auf die geringere Anzahl auswählbarer Items zurückzuführen. Außerdem ist festzustellen, dass unter Verwendung von SFFS am seltensten beide Items zugleich in eine Item-Kombination aufgenommen werden. Im Vergleich zu vollständiger Suche kommt es unter heuristischer Suche häufiger zu Fällen, in denen die Items mit hohen Korrelationen nicht in die Item-Kombinationen aufgenommen werden. Dies kann mehrere Gründe haben. Zum einen wurden bei heuristischer Suche dichotomisierte Daten verwendet, so dass die Korrelationen hier nur unter Vorsicht als Indikatoren für Relevanz verwendet werden können. Zum anderen werden unter heuristischer Suche in der Regel mehr als zwei Items kombiniert. Diese müssen nicht zwingend mit einem Kriterium hoch korrelieren, um dieses im Rahmen der Bayes-Klassifikation vorhersagen zu können. Insofern ist die Korrelation tatsächlich nur als ein erster Anhaltspunkt für Relevanz zu betrachten (Caruna & Freitag, 1994; Cover, 1974; Guyon & Elisseeff, 2003; Hua et al., 2005).

Zusammenfassend lässt sich über Item-Kombinationen bei heuristischer Suche Folgendes feststellen. Heuristische Suche führt im Vergleich zu vollständiger Suche zu einer wesentlich instabileren Auswahl von Item-Kombinationen. Dies gilt bei der gewählten Abbruchbedingung insbesondere für die Induktion binärlogistischer Regressions-Modelle. Dennoch werden in den überwiegenden Fällen höchstens fünf dichotomisierte Items für die Vorhersage eines Kriteriums kombiniert. Hiermit wird die optimale Messkomplexität gerade überschritten. In die meisten Item-Kombinationen wird mindestens eines der beiden Items aufgenommen, die die höchsten Korrelationen mit dem vorherzusagenden Kriterium aufweisen. Bei heuristischer Suche wird jedoch auch deutlich, dass die Korrelation zwischen einem Item und einem Kriterium höchstens als Indikator für die Relevanz eines Items betrachtet werden kann.

Insgesamt wurden 31 Item-Kombinationen zur Vorhersage der 18 Kriterien endgültig ausgewählt, wobei 13 Mal bezüglich der vollständigen EEB und des WSIB unterschiedliche Item-Kombinationen gewählt wurden (s. o.). Die Umfänge dieser Item-Kombinationen liegen zwischen drei und fünf Items. Kombinationen mit vier Items kommen mit 23 Mal am häufigsten vor. Vier dichotomisierte Items entsprechen einer Messkomplexität von $2^4 = 16$ Zellen. Dies liegt noch unterhalb der optimalen Messkomplexität, die bei etwa 23 Zellen liegt. Kombinationen von drei und fünf Items werden jeweils vier Mal endgültig ausgewählt. Besteht die Kombination aus drei Items, so wurde diese Kombination auch unter vollständiger Suche als bestes Tripel identifiziert. Dies betrifft die Kriterien Kommunikation und Kooperation, Bedeutung von Anerkennung und Wertschätzung, Muskel-Skelett-Beschwerden und Absentismustage. Von den beiden Items, die die höchste Korrelation mit dem entsprechenden Kriterium aufweisen, kommt in 18 Item-Kombinationen jeweils genau eines vor. Weitere zwölf Mal kommen beide Items mit hoher Korrelation in der endgültig ausgewählten Item-Kombination vor. Lediglich die Item-Kombination bezüglich der vollständigen EEB zur Vorhersage von Zusammenarbeit weist keines dieser Items auf. Endgültig ausgewählte Item-Kombinationen weisen somit überwiegend eine möglichst optimale Messkomplexität auf und enthalten in den meisten Fällen mindestens ein Item, das eine hohe Korrelation mit dem entsprechenden Kriterium aufweist.

Auch nach der eingehenden Diskussion der ausgewählten Item-Kombinationen wird die Übereinstimmung mit theoretischen Erwartungen deutlich. Die Messkomplexität erweist sich sowohl hinsichtlich der vollständigen als auch der heuristischen Suche als nützlich methodisches Konzept. Sie gibt Anhaltspunkte über die Umfänge auszuwählender Item-Kombinationen, was die vollständige Suche vereinfacht und die Prüfung von Selektions-Verfahren erlaubt. Das gewählte Selektions-Verfahren sowie die Kodierung der Daten nehmen Einfluss auf die Klassifikator-Induktion. Vollständige Suche führt unter Verwendung der Rohdaten zu einer stabilen Auswahl von nur wenigen Items. Die Auswahl mehrerer dichotomisierter Items ist unter heuristischer Suche etwas instabiler. Am Beispiel der binärlogistischen Regression konnte gezeigt werden, dass neben der Festlegung eines Selektions-Verfahrens und eines Klassifikations-Algorithmus auch die Abbruchbedingung entscheidend für die Qualität eines Klassifikator-Induktors ist. Dies wurde an der besonders geringen Stabilität binärlogistischer Regressions-Modelle deutlich. Zudem wurde deutlich, dass Item-Kombinationen nicht zwingend die Korrelationen zwischen Items und vorherzusagenden Kriterien abbilden. Diese Korrelationen sind höchstens als Hinweis für die Relevanz eines Items zu betrachten.

7.2.4 Vergleich der Klassifikator-Induktoren

In den beiden Experimentalreihen wurden drei verschiedene Selektions-Verfahren eingesetzt und zwei verschiedene Arten von Klassifikatoren induziert. Vollständige Suche, Sequential Forward Selection und Sequential Forward Floating Selection kamen als Selektionsverfahren zum Einsatz. Es wurden Bayes-Klassifikatoren und binärlogistische Regressions-Modelle induziert. Für den Vergleich dieser verschiedenen Verfahren genügt es zunächst, die weiter oben im Ergebnisteil und in den vorangegangenen Abschnitten genannten Erkenntnisse zusammenzufassen. Hinsichtlich der CV-Korrektklassifikation ist keines der beiden Arten von Klassifikatoren überlegen. Bayes-Klassifikatoren und binärlogistische Regressions-Modelle erzielen etwa gleich häufig die höchsten CV-Korrektklassifikationen. Allerdings weisen Bayes-Klassifikatoren einen entscheidenden Vorteil gegenüber binärlogistischen Regressions-Modellen auf: Sie erfordern sowohl bei der Induktion als auch bei der Anwendung wesentlich weniger Rechenaufwand (Doumpos & Slappa, 2005), was sich bei umfangreichen Experimenten auch praktisch bemerkbar macht. Darüber hinaus ist die Definition der Abbruchbedingung entscheidend. Die hier bevorzugte Definition erwies sich erwartungsgemäß als nicht geeignet in Verbindung mit binärlogistischer Regression. Dies machte sich vor allem in instabilen, z. T. sehr umfangreichen, manchmal jedoch auch sehr sparsamen Prädiktor-Teilmenen bemerkbar. Die unzureichende Kalkulierbarkeit der Umfänge von Prädiktor-Teilmenen bei Induktion binärlogistischer Regressions-Modelle ist im vorliegenden Fall vor allem auf die Definition der Abbruchbedingung zurückzuführen.

Vollständige Suche und heuristische Suche können auf Grund unterschiedlicher Kodierungen der verwendeten Daten nicht fair verglichen werden. In Bezug auf die beiden heuristischen Verfahren wurde im Ergebnisteil bereits festgestellt, dass SFFS hinsichtlich der erzielten CV-Korrektklassifikationen SFS klar überlegen ist. Diese Überlegenheit spiegelt sich jedoch nicht in den endgültig ausgewählten Klassifikatoren wider. Unter den 31 endgültig ausgewählten Induktoren finden sich vier wieder, die als bestes Tripel unter vollständiger Suche identifiziert wurden. 19 dieser Klassifikatoren wurden unter Verwendung von SFS ermittelt und acht unter Verwendung von SFFS. Dieser Vergleich schmälert die Überlegenheit von SFFS gegenüber SFS allerdings nicht. Die endgültige Auswahl der Klassifikatoren erfolgte im Rahmen von Bootstrapping. Dabei wurden die beiden heuristischen Verfahren selbst gar nicht eingesetzt.

Für Problemstellungen, die in ihrer Komplexität der vorliegenden entsprechen, können somit folgende Empfehlungen ausgesprochen werden. Bayes-Klassifikatoren sind auf Grund ihrer Effizienz bei vergleichbarer Klassifikationsgüte binärlogistischen Regressions-Modellen

überlegen (Doumpos & Slappa, 2005). Sind nur wenige Prädiktoren auszuwählen, so ist eine auf die optimale Messkomplexität zugeschnittene vollständige Suche heuristischen Verfahren vorzuziehen. Sind vergleichsweise viele Prädiktoren auszuwählen, so werden mit SFFS häufig bessere CV-Korrektklassifikationen erzielt, als mit SFS. Der Rechenaufwand erhöht sich hierbei in einem vertretbaren Maße. Diese Empfehlungen decken sich weitgehend mit Angaben aus der Fachliteratur (Jain & Zongker, 1997; Jain et al., 2000; Kudo & Sklansky, 2000; Molina et al., 2002; Pudil et al., 1994).

7.2.5 Diskussion von Kalibrierung und Validierung

Die Kalibrierung der endgültig ausgewählten Klassifikatoren diene dazu, systematische Fehler zu dokumentieren, um diese später bei der Anwendung der Klassifikatoren berücksichtigen zu können. Im Rahmen zweier Reihen von Bootstrapping-Experimenten wurden drei systematische Fehler ermittelt. Zum einen die Überrepräsentation einer der beiden Klassen 0 und 1 bei der Vorhersage, zum anderen mittlere Fehler bei der Vorhersage von Klassen-Wahrscheinlichkeiten und Erwartungswerten.

Es ist festzustellen, dass bei 22 der 31 Klassifikatoren die Klasse 1 gegenüber der Klasse 0 überrepräsentiert ist. D. h. Instanzen werden durch diese Klassifikatoren vorzugsweise der Klasse 1 zugeordnet. Der Klassifikator zur Vorhersage von Absentismustagen auf Grundlage von WSIB-Daten weist hierzu eine besonders hohe Neigung auf. Er ordnet gut 27 % der Instanzen fälschlicher Weise der Klasse 1 zu. Bei neun Klassifikatoren ist Klasse 0 gegenüber Klasse 1 überrepräsentiert. Hier ist der Klassifikator zur Vorhersage von Gesundheitskompetenz auf Grundlage von WSIB-Daten hervorzuheben. Er ordnet 24 % der Instanzen fälschlicher Weise der Klasse 0 zu. Bei allen weiteren Klassifikatoren ist dieses Verhalten zwischen diesen beiden Extremen ausgeprägt. Besonders gering ist dieses Verhalten bei dem Klassifikator zur Vorhersage des Gesamtbedarfs für Arbeitsgestaltung auf Grundlage von WSIB-Daten ausgeprägt. Er weist 0,2 % aller Instanzen fälschlicher Weise der Klasse 1 zu. Um solche Überrepräsentationen berücksichtigen zu können, wurden Korrekturfaktoren ermittelt, die im Anschluss bei der Anwendung der Klassifikatoren eingesetzt wurden. Dies geschah erstmalig bei der Ermittlung mittlerer Fehler, die im Folgenden diskutiert werden sollen.

Werden die beschriebenen Überrepräsentationen berücksichtigt, so sind die mittleren Fehler sowohl bei der Vorhersage von Klassen-Wahrscheinlichkeiten als auch bei der Vorhersage von Erwartungswerten sehr gering ausgeprägt. Der mittlere Fehler zur Vorhersage von Klassen-Wahrscheinlichkeiten liegt für alle Klassifikatoren zwischen 0,0 und 0,6 %. Den höchsten mittleren Fehler von 0,6 % weist hier der Klassifikator zur Vorhersage von Zusammenarbeit

auf Grundlage von WSIB-Daten auf. Insgesamt machen sechs der 31 Klassifikatoren diesbezüglich einen (gerundeten) mittleren Fehler von 0,0 %. Die mittleren Fehler bei der Vorhersage von Erwartungswerten machen sich – mit drei Ausnahmen – für alle Klassifikatoren erst bei der dritten Nachkommastelle bemerkbar. Die drei Klassifikatoren zur Vorhersage von Absentismus- und Präsentismustagen bilden die Ausnahmen. Bei ihnen macht sich der mittlere Fehler bereits an der zweiten Nachkommastelle bemerkbar. Diese Ausnahmen sind jedoch unter Berücksichtigung der verwendeten Skalen erklärbar. Absentismus- und Präsentismustage wurden als Anzahl von Tagen in den letzten zwölf Monaten abgefragt. Vor diesem Hintergrund können auch diese Fehler als vergleichsweise klein eingeschätzt werden.

Die Standardabweichungen der Fehler bei der Vorhersage von Klassen-Wahrscheinlichkeiten liegen zwischen 4,6 und 6,9 %. Die 99,8 %-Vertrauensintervalle decken damit einen Bereich zwischen 28 und 44 % ab. Bei Absentismus- und Präsentismustagen sind hier wieder die höchsten Werte von 44 bzw. 41 % zu verzeichnen. Diese hohen Werte sind der Verteilung und hohen Spannweite möglicher Messwerte geschuldet. Dies wird weiter unten genauer erläutert. Auch der Klassifikator zur Vorhersage von Gesundheitskompetenz auf Grundlage der WSIB-Daten weist mit 41 % ein sehr großes Vertrauensintervall auf. Wird Gesundheitskompetenz auf Grundlage vollständiger EEB-Daten vorhergesagt, so findet man das kleinste Vertrauensintervall von 28 % vor. Das nächst größere Vertrauensintervall von 29 % weist der Klassifikator zur Vorhersage ganzheitlicher Führung auf Grundlage von WSIB-Daten auf. Die meisten Vertrauensintervalle bezüglich der Vorhersage von Klassen-Wahrscheinlichkeiten weisen eine Breite um etwa 33 % auf, also etwa $1/3$ der möglichen Messwerte. Im Vergleich hierzu sind die 99,8 %-Vertrauensintervalle bezüglich der Vorhersage von Erwartungswerten vergleichsweise klein. Sie decken alle um die 15 % (gut $1/7$) der jeweiligen Gesamtskala ab. Arbeitsplatzmerkmale, gemessen mit dem SynBA, werden fast durchgehend mit 15 % umfassenden Vertrauensintervallen vorhergesagt. Bei dem hieraus aggregierten Merkmal Gesamtbedarf für Arbeitsgestaltung ist dieses Vertrauensintervall mit 8,4 % wesentlich kleiner. Bei den körperlichen Beschwerden, gemessen mit dem HkB, umfassen diese Vertrauensintervalle durchgehend 18 % der Gesamtskala. Bei den Merkmalen von Führung und Zusammenarbeit sind die Vertrauensintervalle kleiner als 15 %. Ausnahme ist hier das Merkmal Bedeutung von Anerkennung und Wertschätzung, dessen Vertrauensintervall 16 % der Gesamtskala beträgt. Ein besonders geringes Vertrauensintervall von 9 % wird für die Vorhersage von Gesundheitskompetenz ermittelt. Die Vertrauensintervalle zur Vorhersage von Absentismus- und Präsentismustagen betragen durchgehend 11 Tage. Im Vergleich zu den anderen berichteten Vertrauensintervallen sind sie als besonders hoch einzuschätzen. Ihre

Breite ist Resultat von Ausreißern. Obwohl im Mittel 7,49 Absentismustage angegeben werden, liegt das Maximum bei 215 Tagen. Bei Präsentismustagen liegt das Mittel bei 6,27 Tagen im Vergleich zu 110 Tagen maximal. Es ist davon auszugehen, dass diese Vertrauensintervalle wesentlich enger wären, wenn Instanzen mit mehr als 30 Absentismus- oder Präsentismustagen ausgeschlossen worden wären. Diese Grenze entspricht sechs Wochen Arbeitsunfähigkeit. Ab dieser Grenze übernimmt der Krankenversicherer die Entgeltfortzahlung; der Arbeitgeber ist ab dieser Grenze verpflichtet, betroffenen Arbeitnehmern ein betriebliches Eingliederungsmanagement anzubieten. Ungeachtet dieser Ausnahme ist die Enge der 99,8 %-Vertrauensintervalle zur Vorhersage von Erwartungswerten hervorzuheben. Sie wird deutlicher, wenn man sie mit den Standardabweichungen der Merkmale innerhalb einer Stichprobe vergleicht. Eine Standardabweichung eines Merkmals nimmt bereits zwischen 17 und 30 % der entsprechenden Gesamtskala ein.

Die Validierung der Klassifikatoren wurde bereits eingehend im Ergebnisteil diskutiert. Daher werden hier nur noch die wesentlichen Erkenntnisse zusammengefasst. Die folgenden Aussagen gelten nur für Klassifikatoren bezüglich des WSIB. Für die Validierung der Klassifikatoren bezüglich der vollständigen EEB lag keine geeignete Stichprobe vor. Die Validierungsstichproben flossen nicht in die Induktion der Klassifikatoren ein. Die Daten der Validierungsstichproben waren den Klassifikatoren also vollkommen unbekannt. Die Stichprobe des Wasserversorgers wies im Vergleich zu den Induktionsstichproben durchgehend günstigere Mittelwerte bei den Kriterien auf. Bei acht Merkmalen waren diese Unterschiede sogar signifikant. Dennoch lagen die gemessenen Klassen-Wahrscheinlichkeiten und arithmetischen Mittelwerte bei 16 der insgesamt 18 Kriterien innerhalb der vorhergesagten 99,8 %-Vertrauensintervalle. Dies entspricht 89 % der Kriterien. Die Stichprobe des Technologieunternehmens ist in 11 Bereiche mit Stichproben von 7 bis 22 Teilnehmern unterteilt. Unter Berücksichtigung des funktionalen Zusammenhanges zwischen dem Vorhersagefehler für Erwartungswerte und dem Stichprobenumfang lagen die empirischen arithmetischen Mittelwerte der Kriterien in 184 von 192 Fällen innerhalb der vorhergesagten bzw. berechneten 99,8 %-Konfidenzintervalle. Unter Berücksichtigung der Ergebnisse für den Wasserversorger gelingt somit die Vorhersage von Erwartungswerten mit einer Quote von 95,7 %. Den Klassifikatoren kann daher eine starke Vorhersagekraft attestiert werden. Eine Interpretation berechneter Profile sollte jedoch immer unter Berücksichtigung der 99,8 %-Vertrauensintervalle erfolgen. Einen weiteren Hinweis auf eine gute Vorhersagekraft liefert die Anwendung der Klassifikatoren auf die Anwendungsstichprobe. Auch diese war den Klassifikatoren vollkommen unbekannt. Es konnten drei Personengruppen unterschieden

werden: Gesamtstichprobe, Beschäftigte der Produktion und Führungskräfte. Für die drei Gruppen wurden qualitativ unterschiedliche Profile ermittelt: Für die Gesamtstichprobe wurden Erwartungswerte nahe den arithmetischen Mittelwerten der Induktionsstichproben vorhergesagt. Für Beschäftigte in der Produktion wurde ein wesentlich schlechteres Profil vorhergesagt, als für Führungskräfte. Es konnte dargelegt werden, dass diese Unterschiede sowohl im Vergleich mit Befunden aus der Literatur als auch unter Berücksichtigung der betrieblichen Umstände zum Erhebungszeitpunkt plausibel erscheinen.

7.3 INHALTLICHE INTERPRETATION DER KLASSIFIKATOREN

Die inhaltliche Interpretation der endgültig ausgewählten Klassifikatoren erfolgt vor dem Hintergrund der Struktur der EEB bzw. des WSIB sowie theoretischer Zusammenhänge zwischen psychischer Belastung, psychischer Beanspruchung und langfristigen Folgen. Während die Kriterien psychische Belastung und langfristige Folgen abbilden, wird mit den EEB-Items die psychische Beanspruchung gemessen. Ausgewählte Items sollten daher Beanspruchungsmuster widerspiegeln, die auf Grund der theoretischen Zusammenhänge zu erwarten sind. Die Diskussion beginnt mit einem Überblick aller ausgewählten Items in Abhängigkeit der Kriterien. Im Anschluss werden die ausgewählten Items genauer hinsichtlich ihrer Plausibilität geprüft. Bereits an dieser Stelle soll darauf hingewiesen sein, dass diese Plausibilitätsbetrachtungen zweierlei Einschränkungen unterliegen. Zum einen handelt es sich um empirisch nicht zu belegende Aussagen, sondern lediglich um den Versuch, die jeweilige Item-Auswahl zu verstehen. Zum anderen ist die Interpretation dadurch erschwert, dass die Klassifikatoren keinerlei semantische Wertungen vornehmen. Sie basieren ausschließlich auf Verteilungen von Daten und berücksichtigen keine inhaltliche Bedeutung der Kriterien und Items.

7.3.1 Übersicht über ausgewählte Items

Die endgültig ausgewählten Klassifikatoren verwenden zwischen drei und fünf EEB- bzw. WSIB-Items. Eine Fragestellung der vorliegenden Arbeit bestand darin, ob die EEB bzw. das WSIB ergänzt werden müsse, um alle Kriterien hinreichend gut vorhersagen zu können. Diese Frage kann vor dem Hintergrund der Validierungsbefunde negativ beantwortet werden. Bereits die WSIB-Items sind ausreichend, um eine hinreichend gute Klassifikation zu erzielen. Da sie in der vollständigen EEB enthalten sind, gilt dies auch für die vollständige EEB. Eine weitere Fragestellung bestand darin, ob die EEB bzw. das WSIB hinsichtlich der Klassifikation überflüssige Items enthält. Um diese Frage beantworten zu können, werden die ausgewählten Items hier noch einmal in Form eines Überblicks zusammengetragen. Da bei diesem Überblick

auf die Nummerierungen der EEB-Items sowie ihre Zuordnung zu den theoriegeleiteten Dimensionen zurückgegriffen wird, seien diese zunächst in Tabelle 109 nochmals aufgeführt.

Tabelle 109. Nummerierungen der Items der vollständigen EEB. Angegeben sind auch die theoriegeleiteten Dimensionen, denen diese Items zugeordnet sind.

Nr.	EEB-Item	Dim.	Nr.	EEB-Item	Dim.	Nr.	EEB-Item	Dim.
1	konzentriert ^(a)	MEN	2	aggressiv	ÄRG	3	nervös ^(a)	EMO
4	körperlich verspannt ^(a)	PHY	5	selbstsicher	KNT	6	zornig	ÄRG
7	aufmerksam ^(a)	MEN	8	leistungsbereit ^(a)	MOT	9	entspannt	EMO
10	körperlich unwohl ^(a)	PHY	11	energiegeladen ^(a)	MOT	12	aufgeregt ^(a)	EMO
13	gelangweilt	MON	14	beschwingt	PG	15	fröhlich	PG
16	ärgerlich	NG	17	zufrieden	PG	18	ängstlich	NG
19	lustig	PG	20	leicht	PG	21	deprimiert	NG
22	bedrückt	NG	23	einflussreich ^(a)	KNT	24	vergnügt	PG
25	betrübt	NG						

^(a)Dieses Item gehört auch zum WSIB. ÄRG = Ärger, EMO = Nervöse Anspannung/Emotionale Beanspruchung, KNT = Kontrollerleben, MEN = Mentale Beanspruchung, MON = Monotonie, MOT = Motivationale Beanspruchung, NG = Negatives Gefühl, PG = Positives Gefühl, PHY = Körperliche Beschwerden.

In Tabelle 110 erfolgt nun der Überblick über ausgewählte Items aus dem WSIB, in Tabelle 111 der Überblick über ausgewählte Items aus der vollständigen EEB zur Vorhersage der 18 Kriterien. In diesen Tabellen werden die ausgewählten Items durch Symbole markiert. Die Symbole repräsentieren die Richtungen der Korrelationen zwischen den EEB-Items und dem jeweiligen Kriterium. Hierbei wird zwischen positiven, negativen und Nullkorrelationen unterschieden. Als Nullkorrelationen werden Korrelationen bezeichnet, deren Betrag kleiner als .100 ist, die also als praktisch nicht relevant einzustufen sind. Negative Korrelationen sind kleiner oder gleich -.100, positive Korrelationen sind größer oder gleich +.100.

Tabelle 110. Ausgewählte Items aus dem Wuppertaler Screening Instrument psychische Beanspruchung (WSIB) zur Vorhersage der 18 Kriterien. Symbolisiert sind jeweils die Richtungen der Korrelationen ρ zwischen den Items und den Kriterien.

Kriterium	WSIB-Item								
	1	3	4	7	8	10	11	12	23
Aufgabenanforderungen				(-)		(0)	(-)	(0)	
Kommunikation und Kooperation			(-)	(-)					(0)
Leistungs- und Zeitvorgaben	(+)		(0)		(0)				(0)
Regulationsbehinderungen		(+)	(+)			(+)	(-)		
Tätigkeitsspielraum	(-)	(0)	(-)						(-)
Gesamtbedarf für Arbeitsgestaltung		(0)		(-)		(+)			(-)
Autoritäre Führung	(0)	(+)		(0)	(0)	(+)			
Bedeutung von Anerkennung und Wertschätzung			(0)		(+)		(0)		
Ganzheitliche Führung		(0)	(-)				(+)		(+)
Partizipative Unternehmenskultur	(+)	(0)				(-)			(+)
Zusammenarbeit	(+)				(+)	(-)			(+)
Gesundheitsbezogene Selbstwirksamkeit			(-)	(+)	(+)	(-)			
Herz-Kreislauf-Beschwerden	(0)	(+)	(+)			(+)			
Magen-Darm-Beschwerden	(0)		(+)		(0)	(+)			
Muskel-Skelett-Beschwerden			(+)				(-)		(-)
Unspezifische Beschwerden		(+)				(+)	(-)		(-)
Absentismustage			(+)	(0)	(0)	(+)			
Präsentismustage			(+)	(0)	(0)	(+)			

(0): $|\rho| < .100$, (-): $\rho \leq -.100$, (+): $\rho \geq +.100$, fett formatiert: Korrelation ist signifikant ($p < .001$),
leere Zelle: Item wurde nicht zur Vorhersage des Kriteriums ausgewählt

Tabelle 110 ist zu entnehmen, dass für die Vorhersage aller 18 Kriterien auch alle Items des WSIB benötigt werden. Das WSIB kann hinsichtlich der Klassifikation nicht weiter reduziert werden. Das Item „aufgeregt“ (Dimension „Emotionale Beanspruchung (EMO)“) wird nur einmal, zur Vorhersage von Aufgabenanforderungen, ausgewählt. Alle anderen Items werden zwischen sechs und zwölf Mal ausgewählt. Die beiden Items der Dimension „Körperliche Beschwerden (PHY)“ („körperlich unwohl“ und „körperlich verspannt“) werden mit jeweils zwölf Mal am häufigsten gewählt. Für die Vorhersage aller 18 Merkmale wird mindestens eines dieser beiden Items ausgewählt, für sechs Kriterien sogar beide. Es folgt das Item „einflussreich“ (Dimension „Kontrollerleben (KNT)“), das insgesamt neun Mal ausgewählt wird. Für die weiteren Dimensionen lässt sich folgende Beobachtung machen. Abgesehen von zwei Ausnahmen wird zur Vorhersage eines Kriteriums jeweils höchstens eines der beiden Items aus den Dimensionen „Mentale Beanspruchung (MEN)“, „Motivationale Beanspruchung (MOT)“ oder „Emotionale Beanspruchung (EMO)“ ausgewählt. Obwohl das andere Item ggf. eine vergleichbar hohe Korrelation mit dem Kriterium aufweist, ist es für die Klassifikation redundant und wird ausgeschlossen. Die Items „körperlich unwohl“ und „körperlich verspannt“ scheinen hingegen nicht redundant zu sein. Sie scheinen zwei unterschiedliche Kategorien körperlicher Beschwerden abzubilden. Dies wird weiter unten näher erläutert.

Diskussion, Anwendungsbezug und Ausblick

Tabelle 111. Ausgewählte Items aus Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben (EEB) zur Vorhersage der 18 Kriterien. Symbolisiert sind jeweils die Richtungen der Korrelationen ρ zwischen den Items und den Kriterien.

Kriterium	Items der vollständigen EEB																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Aufgabenanforderungen	(-)					(0)	(-)					(0)				(0)									
Kommunikation und Kooperation	(-)					(0)	(-)								(0)										(0)
Leistungs- und Zeitvorgaben	(+)			(0)				(0)																(0)	
Regulationsbehinderungen		(+)		(+)													(-)				(+)				
Tätigkeitsspielraum		(0)			(-)		(-)	(-)													(0)				
Gesamtbedarf für Arbeitsgestaltung	(-)								(0)												(+)		(-)		
Autoritäre Führung												(0)				(+)	(-)			(0)					
Bedeutung von Anerkennung und Wertschätzung					(0)			(+)	(+)									(0)							
Ganzheitliche Führung				(0)	(-)						(+)												(+)		
Partizipative Unternehmenskultur				(0)										(0)		(-)							(+)		
Zusammenarbeit													(-)		(+)		(+)				(+)				
Gesundheitsbezogene Selbstwirksamkeit	(+)			(-)	(+)					(-)															
Herz-Kreislauf-Beschwerden	(0)										(+)							(+)		(0)					
Magen-Darm-Beschwerden				(+)	(+)													(+)		(+)					
Muskel-Skelett-Beschwerden				(+)							(-)												(-)		
Unspezifische Beschwerden				(+)						(+)	(-)												(-)		
Absentismustage				(+)						(0)					(0)										
Präsentismustage				(+)			(0)	(0)		(+)															

(0): $|\rho| < .100$, (-): $\rho \leq -.100$, (+): $\rho \geq +.100$, fett formatiert: Korrelation ist signifikant ($p < .001$), leere Zelle: Item wurde nicht zur Vorhersage des Kriteriums ausgewählt

Tabelle 111 ist zu entnehmen, dass die Items „lustig“, „vergnügt“ (Dimension „Positives Gefühl (PG)“) und „bedrückt“ (Dimension „Negatives Gefühl (NG)“) für keines der Kriterien zur Vorhersage ausgewählt werden. Die vollständige EEB lässt sich also hinsichtlich der Klassifikation um diese drei Items reduzieren. Für die Klassifikation braucht die EEB nur die verbleibenden 22 Items umfassen. Die Items „gelangweilt“ (Dimension „Monotonie (MON)“), „beschwingt“ (Dimension „Positives Gefühl“) und „betrübt“ (Dimension „Negatives Gefühl“) werden jeweils nur einmal für die Vorhersage eines Kriteriums ausgewählt. Mit acht Mal wird das Item „körperlich verspannt“ (Dimension „Körperliche Beschwerden“) am häufigsten gewählt. Ihm folgen mit jeweils sechs Mal die Items „konzentriert“ (Dimension „Mentale Beanspruchung“) und „einflussreich“ (Dimension „Kontrollerleben“). Ähnlich, wie bei der Zugrundelegung des WSIB findet man auch hier das Prinzip des Ausschlusses redundanter Items weitgehend bestätigt. In den meisten Fällen wird je Kriterium höchstens ein Item einer jeden EEB-Dimension für die Vorhersage ausgewählt. Ausnahmen dieser Regel findet man wieder bei der Dimension „Körperliche Beschwerden“. Außerdem bei der Dimension „Mentale Beanspruchung“ und bei der Dimension „Positives Gefühl“. Es werden je zwei Mal zwei Items aus den Dimensionen „Mentale Beanspruchung“ und „Körperliche Beschwerden“ für die Vorhersage eines Kriteriums ausgewählt. Aus der Dimension „Positives Gefühl“ werden einmal zwei und einmal sogar drei Items zur Vorhersage eines Kriteriums kombiniert.

Das Prinzip des Ausschlusses redundanter Items hat seine Ursache darin, dass die Selektions-Verfahren möglichst relevante Items auswählen und zu ihnen redundante Items ausschließen. Die Beobachtung spricht einerseits für die weitgehend gelungene theoretische Strukturierung der EEB, wirft jedoch andererseits die Frage auf, ob diese Strukturierung in jeder Hinsicht bestens gelungen ist. In Bezug auf die Dimension „Positives Gefühl“ kann man sicher konstatieren, dass mit den verbleibenden Items „beschwingt“, „fröhlich“, „zufrieden“ und „leicht“ unterschiedliche Kategorien positiver Gefühle zusammengefasst werden. Für die beiden Items der Dimension „Körperliche Beschwerden“ finden sich entsprechende Hinweise bereits in den Übersichtstabellen (Tabelle 110 und Tabelle 111). Das Item „körperlich verspannt“ wird als einziges aus dieser Dimension für die Vorhersage von Muskel-Skelett-Beschwerden ausgewählt. Muskel-Skelett-Beschwerden umfassen eine ganze Reihe von Beschwerden, die den Stütz- und Bewegungsapparat betreffen und die z. T. auf klare körperliche Ursachen zurückgeführt werden können. Das Item „körperlich unwohl“ wird als einziges aus dieser Dimension für die Vorhersage unspezifischer Beschwerden ausgewählt. Unter unspezifischen Beschwerden sind solche zusammengefasst, die in der Regel nicht eindeutig einer körperlichen Ursache zugeschrieben werden können. Darüber hinaus gehen

diese Beschwerden häufig mit psychischen Beeinträchtigungen bzw. emotionalen Verstimmungen einher. Da die theoretische Strukturierung der EEB jedoch im Wesentlichen als gelungen zu bewerten ist, werden im Folgenden die Interpretationen der Klassifikatoren auf dieser Grundlage fortgeführt.

Mit dem Einbezug von Items, die eine Nullkorrelation mit dem vorherzusagenden Kriterium aufweisen, unterscheiden sich Klassifikatoren gegenüber schrittweise erzeugten Regressionsmodellen in spezifischer Weise. In schrittweise erzeugte Regressionsmodelle werden in der Regel nur Merkmale eingeschlossen, die eine substantielle Korrelation mit der zu erklärenden Variable aufweisen. Bei der Induktion von Bayes-Klassifikatoren muss dies nicht zwingend gegeben sein. Nullkorrelierte Items werden als Prädiktoren für Bayes-Klassifikatoren aufgenommen, wenn sie in Interaktion mit den anderen Prädiktoren zu einer besseren Trennung der beiden Klassen beitragen (Guyon & Elisseeff, 2003; Ng, 1998). Sie werden für die folgenden inhaltlichen Interpretationen der Klassifikatoren eine eher untergeordnete Rolle spielen.

7.3.2 Inhaltliche Plausibilitätsbetrachtungen bezüglich der ausgewählten Items

Merkmale der Arbeitsgestaltung, die mit der Synthetischen Beanspruchungs- und Arbeitsanalyse (Wieland-Eckelmann et al., 1999) gemessen wurden, sind so kodiert, dass hohe Werte einen hohen Gestaltungsbedarf bedeuten. Dies ist bei der Interpretation von Korrelationen zu berücksichtigen. Es können hier Merkmale der Arbeitsgestaltung unterschieden werden, die als Ressourcen und die als Risikofaktoren wirksam werden. Als Ressourcen wirken Arbeitsanforderungen (im Sinne vollständiger, abwechslungsreicher Aufgaben; vgl. Hacker & Sachse, 2014; Ulich, 2011; Wieland-Eckelmann et al., 1999), Tätigkeitsspielraum und Anforderungen zu Kommunikation und Kooperation. Zur Vorhersage dieser Merkmale werden mit „aufmerksam“ und „konzentriert“ Items aus der theoriegeleiteten Dimension der mentalen Funktionsbeanspruchung („MEN“; vgl. Wieland-Eckelmann, 1992) verwendet. Darüber hinaus werden Arbeitsanforderungen und Tätigkeitsspielraum mit je einem Item aus der theoriegeleiteten Dimension der motivationalen Beanspruchung („MOT“) gewählt: „energiegeladen“ (für Arbeitsanforderungen) und „leistungsbereit“ (für Tätigkeitsspielraum). In der Arousal-Uhr entsprechen diese Items der Verhaltens-Kategorie „Konsumieren“. D. h. die Kriterien sind mit instrumentellen, zielbezogenen Verhaltensweisen verknüpft. Dies ist vollkommen in Übereinstimmung mit den Aussagen der Handlungsregulationstheorie (Hacker & Sachse, 2014), in der die Realisierung des aufgabenbezogenen Ziels handlungsleitend ist. Diesen drei Kriterien stehen die Regulationsbehinderungen als Risikofaktoren gegenüber. Sie wurden weiter oben als mögliche Quellen vermeidbarer

Selbstkontrollanforderungen beschrieben. Auch über die Gesundheitsrisiken von Regulationsanforderungen wurde dort berichtet (vgl. Greiner et al., 1987; Lüders, 1999; Semmer et al., 1999; Ulich, 2011; Wieland-Eckelmann et al., 1999). Bei der Vorhersage von Regulationsbehinderungen werden entsprechend die Items „körperlich unwohl“ und „körperlich verspannt“ verwendet. Daneben werden die Items „deprimiert“ und „nervös“ als zwei weitere, in positiver Korrelation mit Regulationsbehinderungen stehende Items verwendet. Dies bestätigt, dass auch Beeinträchtigungen des psychischen Wohlbefindens mit Regulationsbehinderungen in Zusammenhang stehen (a. a. O.). Entsprechend steht „zufrieden“ in einem negativen Zusammenhang mit Regulationsbehinderungen. Ganz im Sinne von ego-depletion (Baumeister & Leary, 1995) als Folge anhaltender Selbstkontrollanforderungen ist das ebenfalls zur Vorhersage von Regulationsbehinderungen ausgewählte Item „energiegeladen“ negativ mit diesem Kriterium korreliert.

Weniger eindeutig, als bei den zuvor genannten Merkmalen der Arbeitsgestaltung lässt sich das Ergebnis für „Leistungs- und Zeitvorgaben“ inhaltlich sinnvoll interpretieren. Hier wurde als einzig bedeutsam korreliertes Item „konzentriert“ ausgewählt. Die Korrelation ist positiv und mit .101 gerade noch praktisch bedeutsam. Eine sinnvolle Erklärung für dieses Ergebnis ist kaum zu erbringen. Bereits weiter oben wurde darauf hingewiesen, dass dieses Kriterium auch bei der Klassifikatorinduktion als besonders problematisch aufgefallen ist.

Mit dem Gesamtbedarf für Arbeitsgestaltung liegt ein Kriterium vor, dass die oben genannten fünf Merkmale der Arbeitsgestaltung summarisch aggregiert. Entsprechend finden sich auch zu seiner Vorhersage Items aus dem Bereich der mentalen Beanspruchung („aufmerksam“ und „konzentriert“) sowie aus den Bereichen körperlicher Beschwerden und psychischer Beeinträchtigung („körperlich verspannt“ und „deprimiert“). Neben diesen Items kommt noch das Item „einflussreich“ aus der theoriegeleiteten Dimension Kontrollerleben („KNT“) zur Vorhersage des Gestaltungsbedarfs hinzu. D. h. weist ein Arbeitsplatz geringen Gestaltungsbedarf auf, so geht das mit einem hohen Kontrollerleben der Beschäftigten einher.

Mit dem Bedürfnis nach sozialer Teilhabe und Zugehörigkeit wurde im Theorieteil die Bedeutung sozialer Beziehungen in der Arbeit begründet. Dies umfasst sowohl die Beziehungen zwischen Kolleginnen und Kollegen als auch die Beziehungen zwischen Führungskraft und Geführten. Darüber hinaus sind Werte und Normen angesprochen, die der Unternehmenskultur zu Grunde liegen. Führungsverhalten kann als Ressource oder als Risikofaktor wirksam werden (Gregersen et al., 2011; Judge et al., 2004; Neyberg et al., 2005; Rowolt & Heinitz, 2008; Stadler & Spieß, 2005; Wieland & Scherrer, 2007). Dieser Dualismus

wird auch bei der Betrachtung der ausgewählten Items zur Vorhersage des Führungsverhaltens deutlich. Sowohl ganzheitliche Führung (als Kombination von aufgaben- und mitarbeiterorientierter Führung; Wieland & Scherrer, 2006) als auch als gehaltvoll erlebte Rückmeldung der Führungskraft (Bedeutung von Anerkennung und Wertschätzung) werden mit „energiegeladen“, „leistungsbereit“ und „einflussreich“ von Items vorhergesagt, die den theoriegeleiteten Dimensionen „motivationale Beanspruchung (MOT)“ und „Kontrollerleben (KNT)“ zuzuordnen sind. D. h. sie gehen mit hoher Autonomie und Aktivität der Beschäftigten einher. Darüber werden diese Führungsverhaltensweisen durch die Klassifikatoren mit geringer Anspannung assoziiert, da sie von den Items „körperlich verspannt“ (negativ korreliert) und „entspannt“ (positiv korreliert) vorhergesagt werden. Diese Führungsverhaltensweisen können also als Ressourcen für das persönliche Wohlbefinden und die Gesundheit betrachtet werden (Gregersen et al., 2011; Judge et al., 2004; Neyberg et al., 2005; Rowolt & Heinitz, 2008; Wieland & Scherrer, 2007) und mobilisieren Energie für die weitgehend autonome und stressfreie Bewältigung von Arbeitsaufträgen. Nach einem ähnlichen Muster werden partizipative Unternehmenskultur und Zusammenarbeit vorhergesagt. Mit diesen Kriterien stehen Kontrollerleben („einflussreich“) und mentale Funktionsbeanspruchung („MEN“, „konzentriert“) und darüber hinaus bei Zusammenarbeit ebenfalls motivationale Beanspruchung („leistungsbereit“) positiv in Zusammenhang. Gut funktionierende Zusammenarbeit zwischen Kolleginnen und Kollegen werden darüber hinaus mit positiven Gefühlen („PG“), nämlich „fröhlich“, „zufrieden“ und „leicht“, assoziiert. Dies kann vermutlich einerseits auf die Bedeutung sozialer Unterstützung zurückzuführen sein (Stadler & Spieß, 2005). Zum anderen auf die dadurch ausgelösten positiven sozialen Emotionen (Baumeister & Leary, 1995). Partizipative Unternehmenskultur und Zusammenarbeit erscheinen darüber hinaus als Ressource für das persönliche Wohlbefinden und die Gesundheit. Beide Kriterien werden durch das negativ korrelierte Item „körperlich unwohl“ vorhergesagt. Außerdem werden sie von den negativ korrelierten Items „ärgerlich“ (partizipative Unternehmenskultur) und „gelangweilt“ (Zusammenarbeit) vorhergesagt. Autoritäre Führung wird durch ein umgekehrtes Muster vorhergesagt. Weiter oben wurde autoritäre Führung als mögliche Quelle vermeidbarer Selbstkontrollanforderungen beschrieben. Wie zu erwarten, ist dieses Kriterium zum einen in positiver Korrelation mit körperlichen Beschwerden („PHY“, „körperlich unwohl“), zum anderen mit emotionaler Selbstbeanspruchung („EMO“, „ärgerlich“, „nervös“; vgl. Wieland-Eckelmann, 1992) assoziiert (vgl. auch Gregersen et al., 2011; Stadler & Spieß, 2005). In der Arousal-Uhr entspricht dies der Verhaltens-Kategorie „Bewältigung“. Vermeidbare Selbstkontrollanforderungen durch die autoritäre Führungskraft

(z. B. Bevormundung, Kränkung) müssen von den Beschäftigten zusätzlich bewältigt werden. Das Item „zufrieden“ wird ebenfalls zur Vorhersage autoritärer Führung herangezogen und ist mit dieser negativ korreliert. Auch dies erscheint vor dem Hintergrund empirischer Befunde plausibel (Gregersen et al., 2011).

In die vorliegenden Untersuchungen wurden sieben Gesundheitsindikatoren einbezogen. Zum einen die vier Beschwerdegruppen Herz-Kreislauf-, Magen-Darm-, Muskel-Skelett- und unspezifische Beschwerden. Zum anderen Absentismus (Fehlzeiten) und Präsentismus (Anwesenheit bei der Arbeit trotz Krankheit). Darüber hinaus wurde Gesundheitskompetenz (gesundheitsbezogene Selbstwirksamkeit; Wieland & Hammes, 2009) erhoben. Da es sich dabei um ein relativ stabiles Merkmal der Person handelt, das auf dem Konzept der Selbstwirksamkeit (Bandura, 1977) beruht, ist es von den übrigen Gesundheitsindikatoren konzeptionell zu unterscheiden. Dies spiegelt sich auch in den ausgewählten Items zur Klassifikation wieder. Absentismus und Präsentismus haben einen gemeinsamen Kern: Absentismus- und Präsentismustage sind Tage, an denen die betroffenen Beschäftigten ihre Gesundheit beeinträchtigt sahen (Wieland & Hammes, 2010). Dementsprechend werden zur Klassifikation auch die beiden Items „körperlich unwohl“ und „körperlich verspannt“ ausgewählt, die positiv mit diesen beiden Kriterien korrelieren. Auch zur Vorhersage der Häufigkeit der vier Beschwerdegruppen wurde mindestens eines dieser beiden Items ausgewählt. Bei Herz-Kreislauf- und Magen-Darm-Beschwerden sind es beide Items, bei Muskel-Skelett-Beschwerden ist es das Item „körperlich verspannt“ und bei unspezifischen Beschwerden ist es das Item „körperlich unwohl“. Dass diese beiden Items zur Vorhersage körperlicher Beschwerden bevorzugt werden, erscheint inhaltlich valide. Insbesondere, da bei Muskel-Skelett-Beschwerden in zwei Items explizit nach „Verkrampfungen“ und bei unspezifischen Beschwerden explizit nach „Übelkeit“ gefragt wird. Bezüglich der übrigen ausgewählten Items lassen sich die Beschwerdegruppen in zwei Gruppen einteilen. Herz-Kreislauf- und Magen-Darm-Beschwerden werden durch Befindenzustände vorhergesagt, die der theoriegeleiteten Dimension „emotionale Beanspruchung (EMO)“ bzw. der „emotionalen Selbstbeanspruchung“ (Wieland-Eckelmann, 1992) zuzuordnen sind, nämlich „nervös“ und „ängstlich“. In der Arousal-Uhr entspricht dies der Verhaltenskategorie „Bewältigen“. Herz-Kreislauf- und Magen-Darm-Beschwerden scheinen daher mit der Bewältigung individuell herausfordernder Situationen („Stress“) in Zusammenhang zu stehen, was durch die Befundlage in der Literatur bestätigt wird (z. B. Bhatia & Tandon, 2005; Mayer, 2000; Kivimäki, Leino-Arjas, Luukkonen et al., 2002; Neyberg, Fransson, Heikkilä et al., 2013). Muskel-Skelett- und unspezifische Beschwerden werden dagegen mit Items vorhergesagt, die der motivationalen

Beanspruchung („MOT“) und dem Kontrollerleben („KNT“) entsprechen, nämlich „energiegeladen“ und „einflussreich“. Häufige Beschwerden dieser Art gehen mit geringer motivationaler Beanspruchung und Kontrolle einher. Im Theorieteil dieser Arbeit wurde auf den Zusammenhang zwischen ego-depletion (Baumeister et al., 1998) und hohen Selbstkontrollanforderungen hingewiesen. D. h. eine hohe Anforderung, eigenes Verhalten an äußere Umstände anpassen zu müssen, bzw. weniger „einflussreich“ zu sein, geht mit einem erhöhten Energieverbrauch einher. Dies geht mit einer Erosion körperlicher Strukturen (z. B. des Muskel-Skelett-Systems; Carayon, Smith & Haims, 1999; Sherwood Brown et al., 2004) sowie mit Ermüdung und Erschöpfung (unspezifische Beschwerden) einher. Hammes et al. (2010) konnten zeigen, dass Regulationsbehinderungen und autoritäre Führung positiv mit Muskel-Skelett-Beschwerden korrelieren. Dabei wurden dieselben Instrumente wie in der vorliegenden Arbeit verwendet. Beide Kriterien wurden im Theorieteil als Merkmale des Arbeitsplatzes betrachtet, die zu einer erhöhten Selbstkontrollanforderung führen. Die beiden Beschwerdegruppen scheinen daher mit besonders eingeschränkter Autonomie in Zusammenhang zu stehen. Unspezifische Beschwerden werden darüber hinaus durch das Item „nervös“ vorhergesagt. Dies ist nicht verwunderlich, da in einem Item zu unspezifischen Beschwerden explizit nach „Nervosität“ gefragt wird.

Auch Gesundheitskompetenz wird mit den beiden Items „körperlich unwohl“ und „körperlich verspannt“ vorhergesagt. Dies ist inhaltlich valide, da es sich bei dem Gesundheitskompetenzfragebogen (GKF; Wieland & Hammes, 2009) um ein Instrument zur Erfassung der Selbstwirksamkeit mit der spezifischen Domäne „Gesundheit“ handelt. Die Bedeutung von Selbstwirksamkeit für die Gesundheit ist hinreichend belegt (Schwarzer, 2004). Darüber hinaus wird Gesundheitskompetenz mit Items vorhergesagt, die den theoriegeleiteten Dimensionen mentale („MEN“) und motivationale Beanspruchung („MOT“) sowie Kontrollerleben („KNT“) entsprechen: „aufmerksam“, „konzentriert“, „leistungsbereit“ und „selbstsicher“. Die Kombination von mentaler Funktionsbeanspruchung (Wieland-Eckelmann, 1992) und Kontrollerleben entspricht in der Arousal-Uhr der Verhaltenskategorie „Konsumieren“. Hiermit ist instrumentelles, zielbezogenes Verhalten gemeint. Bei der Selbstwirksamkeitserwartung handelt es sich um die persönliche Überzeugung, Herausforderungen selbstständig erfolgreich bewältigen zu können. Sie baut sich in erster Linie durch entsprechende Erfahrungen („mastery“) auf (Bandura, 1977) und erhöht die Wahrscheinlichkeit für zukünftiges instrumentelles, zielbezogenes Verhalten. Dies ist insbesondere für Gesundheitsverhalten nachgewiesen (Schwarzer, 2004).

7.3.3 Brauchen wir eine Theorie des „conceptual act“?

Die Ausführungen im vorstehenden Abschnitt fassen die inhaltliche Validität der erzeugten Klassifikatoren zur Vorhersage der 18 Kriterien zusammen. Sie stellen eine plausible Rückschau auf die ausgewählten Items zur Vorhersage der einzelnen Kriterien dar. Hierzu wurde auf die theoretischen Ausführungen zur Arousal-Uhr sowie zu Grundbedürfnissen genauso zurückgegriffen, wie auf die theoretische Fundierung der EEB. Außerdem wurde die Befundlage der Literatur in Form von Überblicksartikeln und Einzelstudien herangezogen. Wie bereits einleitend erwähnt, kann es sich jedoch nur um eine Plausibilitätsbetrachtung handeln, insbesondere, da in dieser Arbeit ausschließlich Querschnitterhebungen die Datenbasis liefern.

Die von den Klassifikatoren erlernten Muster spiegeln jene Muster wieder, die die Probanden der Erhebungen durch ihr Antwortverhalten statistisch erzeugt haben. Es handelt sich um empirische Muster, die nicht zwingend mit dem theoretischen Konzept der Erhebungen im Einklang stehen müssen. Vielmehr können diese Muster als Ausdruck von „conceptual acts“ (Feldman Barrett, 2006b, 2012) aufgefasst werden. Ähnliche Situationen (z. B. ähnliche Ausprägung verschiedener Merkmale der Arbeitsgestaltung) werden von verschiedenen Personen mit ähnlicher Bedeutung versehen. Dies führt zu ähnlichen Mustern bei der Beschreibung des persönlichen Befindens. Von der Auswahl einer Theorie über die Konzipierung einer Erhebung sowie die Validität der verwendeten Instrumente bis hin zur Beschreibung des persönlichen Befindens in einer Erhebung gibt es eine ganze Reihe methodischer und intraindividuelle Varianzquellen, die die Passung zwischen Theorie und Empirie variieren können. Die hier entwickelten Klassifikatoren nutzen Ähnlichkeiten bei „conceptual acts“ zur Vorhersage aus. Sie sind somit viele Varianzquellen entfernt von der zu Grunde gelegten wissenschaftlichen Theorie. Wenn sie dennoch vor dem Hintergrund dieser Theorie als plausibel erscheinen, spricht dies dafür, dass das zu Grunde gelegte theoretische Konzept geeignet ist, den Ausdruck persönlichen Befindens in bescheidenem Maße vorhersagen zu können. Auch, wenn die theoretischen Ausführungen und empirischen Befunde dieser Arbeit kompatibel erscheinen mögen, bleibt die Frage offen, warum beispielsweise das Item „nervös“ Regulationsbehinderungen mit einer gewissen Genauigkeit vorhersagen kann während das Item „fröhlich“ nicht zur Vorhersage von Regulationsbehinderungen geeignet scheint.

Welche Varianzquellen genau solche empirischen Muster hervorrufen und wie wissenschaftliche Theorie und „conceptual acts“ bei Befragungen in Zusammenhang stehen, scheinen interessante zukünftige Forschungslinien zu sein. Hierbei würde es um nicht weniger als die Frage gehen, in welchem Maße zu Grunde gelegte wissenschaftlichen Theorien

einerseits die beteiligten Prozesse umfassend und korrekt abbilden und abbilden können und andererseits vorläufige heuristische (und konkurrierende) Ordnungsstrukturen darstellen und darstellen müssen. Es ginge bei einer solchen Forschungslinie nicht darum, wie Probanden auf einen Fragebogen antworten und wie entsprechend der Antworten ein Persönlichkeitsmerkmal (z. B. Gesundheitskompetenz) ausgeprägt ist. Es ginge vielmehr um die Frage, warum Probanden so antworten, wie sie antworten. Die intraindividuellen Prozesse im psychologischen Messprozess würden somit zum Gegenstand der Untersuchungen. Und zwar nicht nur im Sinne von global definierten „Verfälschungen“, „Verzerrungen“ oder „Antworttendenzen“ (Bortz & Döring, 2006; Rentzsch & Schütz, 2009) mit dem Ziel, optimale Bedingungen für die genaue Messung von Persönlichkeitsmerkmalen schaffen zu können (Amelang & Schmidt-Atzert, 2006), sondern mit dem Ziel, „conceptual acts“ im psychologischen Messprozess zu verstehen. Konkreter ginge es um die Frage, welche Bedeutung die Befragten verschiedenen Items in der gegebenen Situation und vor dem Hintergrund der individuellen Erfahrung und Entwicklung sowie des aktuellen persönlichen Zustandes zuschreiben und welche Antworten sie in Folge dessen geben. Entsprechende Fragestellungen können sicherlich auf verschiedenen Untersuchungs- und Abstraktionsebenen formuliert werden.

7.4 ANWENDUNG DER KLASSIFIKATOREN

Zum Abschluss der vorliegenden Arbeit soll auf die Anwendungsmöglichkeiten der induzierten Klassifikatoren im Speziellen sowie der eingesetzten Methoden im Allgemeinen eingegangen werden. Als Anwendungskontext der induzierten Klassifikatoren ist in der Hauptsache das betriebliche Gesundheitsmanagement zu sehen. Hierfür wird ein auf dem WSIB aufbauendes Instrument vorgestellt: Das „Wuppertaler Screening Instrument psychische Beanspruchung – Profilanalyse (WSIB Pro)“. Dieses internetbasierte Instrument ist bereits voll funktionsfähig und kann in Unternehmen für die Analyse und Bewertung von WSIB-Daten sowie für die Ableitung betrieblicher gesundheitsförderlicher Maßnahmen eingesetzt werden. Im Anschluss wird an Hand einer realistischen Zukunftsvision für WSIB Pro das grundsätzliche Potenzial der eingesetzten Methoden beschrieben. Es werden Szenarien für das betriebliche Gesundheitsmanagement ausgebaut, Anwendungsmöglichkeiten als Forschungs-Plattform in den Human-, Sozial- und Wirtschaftswissenschaften sowie in wirtschaftlichen und politischen Entscheidungssituationen aufgezeigt und schließlich der mögliche Aufbau eines gewinnorientierten sozialen Netzwerkes dargelegt. In einem abschließenden Abschnitt werden diese Szenarien noch einmal kritisch beleuchtet.

7.4.1 WSIB Pro: Internetbasierte Anwendung zur Profil-Vorhersage

Für die Anwendung der Klassifikatoren liegen alle notwendigen Parameter und Daten in dieser Arbeit vor. Im Folgenden wird beschrieben, wie diese in ein automatisiertes Erhebungs- und Auswertungsverfahren integriert werden. Ergebnis dieser Implementation ist das „Wuppertaler Screening Instrument psychische Beanspruchung - Profilanalyse (WSIB Pro)“.

Zur Erhebung wird eine Eingabemaske benötigt. Diese ist als standardisierter Fragebogen gestaltet und fragt die Items der vollständigen EEB oder des WSIB ab. Außerdem muss die Personengruppe (Abteilung, Beruf oder Funktion) spezifiziert werden. Im einfachen Falle genügt eine Auswahlliste, ggf. ist jedoch auch eine Abbildung der unternehmensinternen Hierarchie (z. B. Berücksichtigung mehrere Bereiche, Abteilungen und Arbeitsgruppen) denkbar. Der Fragebogen kann technisch so gestaltet werden, dass nur vollständig ausgefüllte Fragebögen akzeptiert werden (sog. Validitätsprüfung). Mit dem Absenden des Fragebogens werden die Angaben mit einer laufenden Nummer versehen und in einer Datenbank abgelegt. Die derzeitige Implementierung des „Wuppertaler Screening Instruments psychische Beanspruchung (WSIB)“ (Wieland & Hammes, 2014) erfasst neben den Befindenzuständen die Gruppenzugehörigkeit in zwei Hierarchiegruppen. Dies kann sich entweder auf formale Hierarchien in einer Organisation oder beispielsweise auch auf die Befragung gleicher Abteilungen über mehrere Messzeitpunkte hinweg beziehen. Mittels eines anonymisierten, persönlichen Codes können Messungen von Einzelpersonen anonym über verschiedene Messzeitpunkte hinweg einander zugeordnet werden. WSIB kann um weitere Items erweitert werden. Beispielsweise werden in einem laufenden Projekt Gesundheitszustand und Arbeitsfähigkeit mit erfasst. Auch die Personengruppen können bei Bedarf näher eingegrenzt werden, z. B. in Mitarbeiter mit oder ohne Führungsverantwortung.

Für die automatisierte Auswertung mittels der hier entwickelten Klassifikatoren werden neben den erhobenen Daten je Personengruppe noch folgende Parameter und Daten benötigt:

- Rekodierungsvorschriften zur Dichotomisierung der Rohwerte je Kriterium (Anhang A2)
- Mittelwerte, Mediane und Klassenmittelwerte aller Kriterien (Deskriptive Statistiken der Kriterien im Ergebnisteil)
- Faktoren zur Korrektur der Überrepräsentation einer Klasse (Kalibrierung im Ergebnisteil)
- Mittlere Fehler bei der Vorhersage der Wahrscheinlichkeit von Klasse 1 sowie des Erwartungswertes je Kriterium (Kalibrierung im Ergebnisteil)

- 99,8 %-Vertrauensintervalle für diese Vorhersagen
(Kalibrierung im Ergebnisteil)
- Bedingte Wahrscheinlichkeit für Klasse 1 je Antwortmuster und Kriterium
(Anhang A4)

Diese Parameter und Daten können auf geeignete Weise als Datenbank für das Erhebungs- und Auswertungsverfahren hinterlegt werden. Die automatisierte Auswertung kann dann wie folgt implementiert werden:

1. Sammle alle Rohdaten für die in Frage stehende Personengruppe aus der Datenbank
2. Führe folgende Schritte für alle 18 Kriterien durch:
3. Dichotomisiere die Rohwerte entsprechend der Rekodierungsvorschriften
4. Klassifiziere jede Instanz entsprechend ihres Antwortmusters und gemäß der bedingten Wahrscheinlichkeiten für Klasse 1 je Antwortmuster; Ergebnis ist eine unkorrigierte Wahrscheinlichkeit p_1^{**} für Klasse 1 für diese Personengruppe
5. Korrigiere die Überrepräsentation einer der beiden Klassen; Ergebnis ist p_1^*
6. Korrigiere den mittleren Fehler von p_1^* ; Ergebnis ist p_1 für die Personengruppe
7. Bestimme das 99,8 %-Vertrauensintervall von p_1
8. Bestimme mit Hilfe von p_1^* , den Klassenmittelwerten sowie dem mittleren Fehler des Erwartungswertes den Erwartungswert μ für die Personengruppe
9. Bestimme das 99,8 %-Vertrauensintervall von μ . Verwende für Stichprobenumfänge $n < 100$ die Formel $\sigma(\varepsilon(\mu)) = 0,1688 \sigma(X) \ln\left(\frac{17,18}{\sqrt{n}}\right)$, sonst den empirisch ermittelten Vorhersagefehler (siehe „Finale Auswahl und Kalibrierung der Klassifikatoren“ im Ergebnisteil).
10. Erzeuge die beiden grafischen Profile (Wahrscheinlichkeiten für Klasse 1 und Erwartungswerte), ggf. unter Berücksichtigung von Vergleichswerten (Benchmarking)

Diese Auswertungsvorschriften sind einfache Abbildungsvorschriften und arithmetische Operationen. Die Auswertungen sind daher sehr schnell durchzuführen. Mit dem größten Rechenaufwand sind die Kommunikation mit der Datenbank sowie die Erzeugung der grafischen Profile verbunden. Der Rechenaufwand für eine Auswertung ist nicht zu vergleichen mit dem Rechenaufwand, der für die Ermittlung der Klassifikatoren notwendig war. Den

Anwendern eines solchen Screenings bleibt die dahinter liegende komplexe Methodik vollständig verborgen und macht sich ihnen auch nicht in einer geringen Performanz bemerkbar.

Aktuelle Internet-Technologien erlauben, dieses Erhebungs- und Auswertungsverfahren vollständig Internet-basiert zu implementieren. Darüber hinaus können solche Internet-Anwendungen so gestaltet werden, dass sie sich auf Smartphones oder Tablets wie entsprechende Anwendungen (sog. "Apps") verhalten und bedienen lassen. Ist eine Internet-Anwendung neben der PC-Darstellung zu einer für die mobile Nutzung optimierten Darstellung fähig, so spricht man von „Responsive Design“. Mit geringfügigem Mehraufwand lassen sich Erhebung und Auswertung Passwort-geschützt voneinander trennen, Zeiträume, geografische Orte und Zielgruppen für Erhebungen festlegen, Erinnerungen an Befragungen in Zielgruppen verbreiten (sog. "Push-Benachrichtigung") und vieles andere mehr. Damit eröffnet sich den Anwendern eine sehr große Flexibilität bei der Steuerung und Verwaltung verschiedener Befragungen mit dem Screening.

Bislang ist eine internetbasierte Version implementiert, die im Rahmen des Wuppertaler Screening Instruments psychische Beanspruchung funktionsfähig ist. Diese läuft unter dem Namen "Wuppertaler Screening Instrument psychische Beanspruchung - Profilanalyse (WSIB Pro)". WSIB Pro verfügt über eine automatisierte Auswertung, die auf zwei Hierarchiestufen (z. B. Messzeitpunkt/Abteilung) vorgenommen werden kann. Die Ergebnisse werden grafisch dargestellt und Vergleichsdaten (Mediane und Mittelwerte der Induktions-Stichproben) gegenüber gestellt. Jedes Merkmal ist mit einem edukativen Text versehen. Dieser gibt Hinweise über die Bedeutung des jeweiligen Merkmals in Hinblick auf die Gesundheit sowie auf Gestaltungsmöglichkeiten. Diese Form der Darstellung kann als Grundlage für eine partizipative Bewertung von personalen und organisationalen Ressourcen und Risikofaktoren genutzt werden. Nachfolgend können verhaltens- und verhältnisbezogene Maßnahmen zur betrieblichen Gesundheitsförderung abgeleitet werden. Es bietet sich an, WSIB Pro z. B. in Mitarbeiter- oder Führungskräfteworkshops für betriebliche Gesundheitsförderung einzusetzen.

WSIB Pro verfügt über ein Responsive Design. D. h. die Internet-Anwendung kann über Smartphones abgerufen werden und zeigt sich dort in einer für mobile Nutzung optimierten Darstellung. Darüber hinaus wurde die Anwendung im Hinblick auf hohe Performanz und Antwortbereitschaft optimiert. Bei durchschnittlicher Internetverbindung ist WSIB Pro in der Regel in unter sechs Sekunden mit einem vollständigen Profil aufgebaut. Der Aufbau jedes

weiteren Profils erfolgt in der Regel in unter zwei Sekunden. Bei gedrosseltem Mobilfunknetz dauert der Aufbau eines Profils in der Regel unter zwölf Sekunden. Der Vorteil des Responsive Design wird darin gesehen, dass ein für die betriebliche Gesundheitsförderung zuständiger Mitarbeiter auch in größeren Unternehmen jederzeit (z. B. bei dem Besuch einer Abteilung oder in Workshops oder Konferenzen) Informationen aus WSIB Pro abrufen kann. Ein Zugang für eine Demonstrationsanwendung kann beim Autor angefragt werden.

7.4.2 Eine realistische Zukunftsvision für WSIB Pro

In naher Zukunft soll WSIB Pro um praxisnahe Funktionen erweitert werden. Folgende Funktionen sollen eingeführt werden: Ein Hilfesystem, eine Benutzer- und Klientenverwaltung (Administrationsaufgaben für mehrere Unternehmen durch mehrere Nutzer), eine Verwaltung von datumsgesteuerten Befragungen durch registrierte Benutzer, ein Label-System zur Abbildung beliebiger Hierarchie-, Personengruppen- und Messreihen-Strukturen sowie zur Filterung von Analyseergebnissen, ein Notizen- und Prioritätensystem (z. B. zur Dokumentation von Maßnahmen und deren Dringlichkeit) und letztlich die Einführung von Rollen, um Analyseergebnisse gezielt und mit unterschiedlichen Lese- und Schreibrechten mit anderen zu teilen (z. B. die Weitergabe eines Analyseergebnisses einer Abteilung an die dortige Führungskraft mit ausschließlicher Lese-Berechtigung). Sämtliche Funktionen sollen die Praxistauglichkeit von WSIB Pro im betrieblichen Gesundheitsmanagement steigern. Insbesondere soll, z. B. über das Notiz- und Prioritätensystem, die Dokumentation (gemeinsam) abgeleiteter Maßnahmen sowie die Dokumentation ihrer Umsetzung vereinfacht und zentralisiert werden.

In ferner Zukunft soll WSIB Pro zu einer arbeits- und organisationspsychologischen Forschungsplattform ausgebaut werden. Hierzu ist als erster Schritt die Möglichkeit zu schaffen, Befragungen unabhängig von Unternehmen durchzuführen. D. h. der Fragebogen soll für private Endbenutzer kostenlos, anonym und informativ zugänglich gemacht werden. Unternehmensbezogene Befragungen im Rahmen betrieblichen Gesundheitsmanagements erfolgen ab dann über eine eindeutige Kennung der Befragung. Erste Forschungsvorhaben können in der Untersuchung von Übertragungseffekten bestehen (siehe weiter oben). D. h. das Screening zur psychischen Beanspruchung wird in einzelnen öffentlichen Befragungen mit weiteren kurzen Fragebögen gekoppelt, die Teilmengen der hier zu Grunde gelegten Kriterien (z. B. Arbeitsgestaltung, Führung, Gesundheitszustand) erfassen. Diese können in verschiedenen Versuchsreihen randomisiert *vor* oder *nach* dem Screening angezeigt werden, um Übertragungseffekte zu bestimmen. Zugleich soll der private Endbenutzer sein Antwortverhalten bezüglich des Screenings in einem Zeitverlauf verfolgen und einen Vergleich

seiner Messwerte zu einer geeigneten Bezugsgruppe abrufen können. Die durch private Anwendung gewonnenen Daten können genutzt werden, um Vergleichskennwerte (Mittelwerte, Mediane) sowie korrelative Untersuchungen auf eine größere Stichprobe zu beziehen. Denkbar sind auch längsschnittliche Untersuchungen, wenn private Endbenutzer sich bereit erklären, Daten mit einem anonymisierten persönlichen Code miteinander zu verknüpfen. Die Bayes-Klassifikatoren können auf einer größeren Datenmenge trainiert werden und so an Vorhersagegenauigkeit und Stabilität hinzugewinnen. Durch weitgehende Automatisierung der in dieser Arbeit verwendeten Algorithmen und Vorgehensweisen kann das Training der Bayes-Klassifikatoren automatisiert erfolgen. Beispielsweise, sobald so viele Daten neu erhoben wurden, so dass eine Steigerung der Vorhersagegenauigkeit wahrscheinlich wird. Sie wären dann lernende Algorithmen, die sich beständig selbst aktualisieren und an Genauigkeit gewinnen.

In weiteren Forschungsprojekten kann das Screening auch mit anderen Kriterien (z. B. Maslach Burnout Inventory, Irritations-Skala, berufliche Selbstwirksamkeit) gekoppelt werden. Darüber hinaus ist denkbar, beliebige zwei Befragungsinstrumente miteinander zu koppeln. Hiermit erweitert sich die Anwendbarkeit des Software-Systems auf nahezu beliebige psychologische, human-, sozial- und wirtschaftswissenschaftliche Untersuchungen und Anwendungen durch öffentliche Befragungen. Da die Induktoren für Bayes-Klassifikatoren unabhängig von Inhalten mit beliebigen diskretisierten Messwerten arbeiten können, muss nicht zwingend von psychischer Beanspruchung auf Kriterien geschlossen werden. Denkbare Merkmale als Grundlage für Vorhersagen können auch beliebige andere psychologische Fragebogenitems, Items aus der Marktforschung, Kaufverhalten oder gar entsprechend aufbereitete Nutzungs- und Bewegungsprofile von Smartphones und Wearables (z. B. sogenannte Smart-Watches oder Daten-Brillen) sein.

Das gesamte Programm könnte sich selbst auf zwei Wegen finanzieren. Zum einen könnten kommerzielle Befragungen bzw. professionelle Anwendungen in Unternehmen lizenziert werden. Zum anderen könnten Einblendungen passender Angebote (Advertisements bzw. „Ads“) für private Endbenutzer monetarisiert werden. Während der erste Fall lediglich einen gewissen Verwaltungsaufwand erfordert, ist der zweite Fall auf den Wachstum der Datenbasis und insbesondere eine wachsende inhaltliche Breite der Daten angewiesen. Je mehr über die Endbenutzer bekannt ist, desto besser können ihnen Angebote eingeblendet werden, die zu ihrer momentanen Lebenssituation passen. Das System könnte Endbenutzer fragen, was ihnen in bestimmten Lebenssituationen besonders geholfen hat, diese Informationen in Klassifikatoren umsetzen und diese Klassifikatoren zur Einblendung passender Angebote für andere

Endbenutzer einsetzen. Für gewöhnlich verdient ein solches System umso mehr, je besser die Klassifikatoren arbeiten, d. h. je häufiger die eingeblendeten Angebote tatsächlich zu einer Kaufentscheidung durch die Endbenutzer über dieses System führen. Auch dieses System-interne Kaufverhalten selbst kann an die Klassifikatoren zurück gespiegelt werden und ihre Vorhersagekraft steigern. Kommerzielle bzw. professionelle Anwendungen in Unternehmen könnten vom steigenden Wissen dieses Systems ebenfalls profitieren.

Ein solches System sollte unbedingt weiteres Bindungs-Potenzial aufweisen, um die Bereitschaft zur Teilnahme an Befragungen sowie die Exposition gegenüber System-internen Angeboten zu erhöhen. Hier bietet sich eine Plattform zu sozialem Austausch über die persönlichen Befragungsergebnisse an. Beispielsweise könnte eine Person, die einen bestimmten Wert in der Beanspruchungsbilanz erzielt hat, einen Hinweis darauf bekommen, wie viele Personen ebenfalls einen solchen Wert erzielt haben. In einem anonymen Chat, der über einen entsprechenden Link erreichbar ist, können sich die betreffenden Personen über ihr Ergebnis austauschen und sich wahlweise mit einander anfreunden. Das dort anfallende, anonymisierte qualitative Material sowie die entstehenden sozialen Graphen könnten ebenfalls für Forschungs- und Anwendungszwecke automatisiert analysiert werden. Beispielsweise könnte untersucht werden, inwiefern sich Menschen über dieses System anfreunden, die ähnliche Zeitprofile bezüglich psychischer Beanspruchung aufweisen. Im persönlichen Startfenster der Anwendung könnten Endbenutzer darauf hingewiesen werden, dass verschiedene befreundete Benutzer schon diese oder jene Befragung absolviert haben. Angebote an Endbenutzer könnten bei ihrem Einverständnis mit Hinweisen versehen werden, dass befreundete Benutzer in einer ähnlichen Situation dieses oder jenes Produkt positiv bewertet haben. Dies sollte das verfügbare Datenvolumen sowie die Liquidität des hier beschriebenen Systems weiter steigern.

7.4.3 ...Beware of your Data!

Liebe Leserinnen und Leser, wenn Ihnen beim Lesen des letzten Abschnittes zunehmend unwohler zu Mute wurde, so ging es Ihnen wie mir beim Schreiben. Ausgehend von einem bereits existierenden System – WSIB Pro – wurde ein realistisches Szenario aufgebaut, in dem die Bereitschaft des Teilens persönlicher, anonymisierter Daten einem human-, sozial- und wirtschaftswissenschaftlich orientierten Informationstechnologie-Unternehmen zu immensem finanziellen Wachstum und somit zu einer gewissen gestalterischen Kraft in der globalisierten Gesellschaft verhilft. Ich bin mir sehr sicher, dass ein solches System mit einigen qualifizierten Personen aus den Bereichen der Informationstechnologie, den Human-, Sozial- und Wirtschaftswissenschaften und nicht zuletzt der Rechtslehre tatsächlich umsetzbar ist.

Insbesondere, da man für gewisse Aufgaben, wie beispielsweise die Internationalisierung oder die Verbesserung der Benutzerfreundlichkeit nahezu kostenlos die Endbenutzer einbinden kann. Das Szenario selbst stellt eine vierfache Gewinn-Situation dar. Endbenutzer haben schnellen Zugriff auf Ressourcen, die ihnen in ihrer gegenwärtigen Situation hilfreich sein könnten. Ihre Daten bleiben dabei anonym bzw. anonymisiert – persönliche Angaben (Name, Geschlecht, Alter oder Beruf) sind an keiner Stelle zwingend notwendig. Die Human-, Sozial- und Wirtschaftswissenschaften können auf einen bislang unerreichten und ständig wachsenden Datenpool zurückgreifen sowie ihn mit neuen Fragestellungen anreichern. Für viele Fragestellungen wird man nur noch auf die entsprechenden Daten zugreifen und keine neuen Daten mehr erheben müssen. Längsschnittstudien stellen kein besonderes Problem mehr dar. Anwender aus Politik und Wirtschaft können diese Daten nutzen, um gestalterische gesellschaftliche oder wirtschaftliche Vorhaben auf eine breite Entscheidungsgrundlage zu stellen. Und das fingierte Informationstechnologie-Unternehmen floriert. Niemand scheint zu verlieren. Und wenn Sie die gegenwärtigen Entwicklungen in der Konsumenten-orientierten Informationstechnologie beobachten, werden Sie vielleicht feststellen, dass solche Systeme zumindest im Ansatz schon recht etabliert sind. Auch bei diesen etablierten Systemen können sich Faszination und Unbehagen in der Gesellschaft bislang nicht endgültig gegeneinander durchsetzen. Das Unbehagen hat dabei sicherlich seinen Ursprung in der Vorstellung, dass das Wissen über Milliarden von Menschen in den Händen weniger globaler Akteure liegt und diese somit eine für Einzelne unberechenbare Macht zur gesellschaftlichen Einflussnahme und Gestaltung haben.

Mit WSIB Pro wurde ein arbeitspsychologisches Instrument entwickelt, das moderne Informationstechnologien für den Zweck der praktischen betrieblichen Gesundheitsförderung einsetzt. Tatsächlich ist in Planung, die Anwendung des Instruments in Unternehmen durch beschriebene Zusatzfunktionen deutlich zu steigern. Auch ein bescheidener Ausbau von WSIB Pro in Richtung einer Forschungsplattform wird an dieser Stelle noch nicht vollständig verworfen. Mit zukünftig erhobenen Daten sollen die Vorhersagegenauigkeiten sowie die inhaltliche Bandbreite von WSIB Pro zumindest für den ursprünglichen Zweck – die betriebliche Gesundheitsförderung – entscheidend verbessert werden. Eine Monetarisierung von WSIB Pro in der beschriebenen Art und Weise ist bereits aus methodischen und forschungsethischen Gründen fragwürdig. Denn durch die Einbindung Dritter zur Schaltung von Angeboten und die damit zusammenhängenden Verhandlungen mit den Anbietern wäre eine solche Plattform bereits wirtschaftlich beeinflussbar. Ein Ausweg wären freiwillige Spenden durch die Endbenutzer. Im Gegenzug könnte den Endbenutzern ein hilfreiches,

wissenschaftlich fundiertes Feedback passend zu ihren Eingaben gegeben werden. Eine soziale Austauschplattform hätte, indem sie von den Endbenutzern angenommen würde, auch schon ihren Nutzen. Hier sollte jedoch mit Rücksicht auf die Privatsphäre auf die Analyse von Unterhaltungen und sozialen Beziehungen verzichtet werden.

Das Szenario wurde mit dem Ziel aufgebaut, das grundsätzliche Potenzial der hier eingesetzten Methoden aufzuzeigen. Eine Möglichkeit, eine neue Dynamik in der Daten-Ökonomie zu erwirken, sehe ich persönlich darin, den bereits etablierten Systemen eine Alternative mit eigenem inhaltlichem Schwerpunkt konkurrierend anbei zu stellen. Diese Alternative sollte sich mindestens dieselben Methoden zu Nutze machen und könnte sich durch einen Schwerpunkt hervorheben, der für den einzelnen Menschen eine persönliche Bedeutung hat: Die persönliche Gesundheit und Entwicklung. Gemeinsam mit der Beschreibung der Methoden und Befunde liegt nun eine Anleitung für ähnliche Vorhaben und Untersuchungen vor. Ich lade Sie, liebe Leserinnen und Leser, herzlich ein, diese Methoden in der arbeits- und organisationspsychologischen und allgemeiner in der human-, sozial- und wirtschaftswissenschaftlichen Forschung und Anwendung soweit einzusetzen, wie Sie können, wollen und dürfen.

LITERATURVERZEICHNIS

- Abele, A. E., Stief, M. & Andrä, M. S. (2000). Zur ökonomischen Erfassung beruflicher Selbstwirksamkeitserwartung: Neukonstruktion einer BSW-Skala. *Zeitschrift für Arbeits- und Organisationspsychologie*, 44, 145-151.
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken: John Wiley & Sons.
- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken: John Wiley & Sons.
- Allmer, H. (1996). *Erholung und Gesundheit: Grundlagen, Ergebnisse und Maßnahmen*. Göttingen: Hogrefe.
- Allport, G. W. (1937). The functional autonomy of motives. *American Journal of Psychology*, 50, 141-156.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4., vollständig überarbeitete und erweiterte Aufl.). Heidelberg: Springer.
- Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *Neuroscientist*, 18, 251-270.
- Antonovsky, A. (1979). *Health, stress, and coping*. San Francisco: Jossey-Bass.
- Asmundson, G. J. G., Norton, G. R., Allardings, M. D., Norton, P. J., & Larsen, D. K. (1998). Posttraumatic stress disorder and work-related injury. *Journal of Anxiety Disorder*, 12, 57-69.
- Baddeley, A. (2012) Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1-29.
- Badura, B., Walter, U. & Hehlmann, T. (2010). *Betriebliche Gesundheitspolitik: Der Weg zur gesunden Organisation* (2., vollständig überarbeitete Aufl.). Berlin: Springer.
- Baggen, R., Schwarz, R., & Wieland-Eckelmann, R. (1993). Mental workload in 'small scale' user interfaces. In H. Luczak, & A. Cakir (eds.), *Work with display units 92. Selected proceedings of the 3rd international conference WWDU '92, Berlin, Germany, 1. - 4. September 1992*. Amsterdam: Elsevier.
- Balderjahn, I. & Specht, G. (2011). *Einführung in die Betriebswirtschaftslehre* (6., überarbeitete Aufl.). Stuttgart: Schäffer Poeschel.
- Bamberg, E., Ducki, A. & Metz, A.-M. (Hrsg.). (2011a). *Gesundheitsförderung und Gesundheitsmanagement in der Arbeitswelt – Ein Handbuch*. Göttingen: Hogrefe.
- Bamberg, E., Ducki, A. & Metz, A.-M. (2011b). Handlungsbedingungen von Gesundheitsförderung und Gesundheitsmanagement. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Gesundheitsförderung und Gesundheitsmanagement in der Arbeitswelt – Ein Handbuch* (S. 109-122). Göttingen: Hogrefe.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Baumeister, R. F. (2010). The Self. In R. F. Baumeister, & E. J. Finkel (Eds.), *Social psychology: the state of the science* (S. 139-175). Oxford: Oxford University Press.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529.

Literaturverzeichnis

- Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. (2011). Does conscious thoughts cause behavior? *Annual Review of Psychology*, 62, 331-361.
- Baumgart, U., Debitz, U., Metz, A.-M., Richter, P., Schulze, F. et al. (2002). *Call Center auf dem arbeitspsychologischen Prüfstand* (Teil 2, CCall-Report 11). Hamburg: Verwaltungs-Berufsgenossenschaft.
- Bayes, T. (1764). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.
- Bechmann, S., Jäckle, R., Lück, P. & Herdegen, R. (2011). *Motive und Hemmnisse für Betriebliches Gesundheitsmanagement (BGM) - Umfrage und Empfehlungen*. Initiative Gesundheit und Arbeit.
- Beck, D., Richter, G., Ertel, M. & Morschhäuser, M. (2012). Gefährdungsbeurteilung bei psychischen Belastungen in Deutschland - Verbreitung, hemmende und fördernde Bedingungen. *Prävention und Gesundheitsförderung*, 7, 115-119.
- Beesdom-Baum, K. & Wittchen, H.-U. (2011). Depressive Störungen: Major Depression und Disthymie. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie & Psychotherapie* (2. überarbeitete und erweiterte Aufl., S. 879-914). Berlin: Springer.
- Berntson, G. G., Norman, G. J., Bechara, A., Tranel, D., Bruss, J., & Cacioppo, J. T. (2011). The insula, the amygdala and evaluative processes. *Psychological Science*, 22, 80-86.
- Berufsverband Deutscher Psychologinnen und Psychologen (Hrsg.). (2008). *Psychische Gesundheit am Arbeitsplatz in Deutschland*. Berlin: Herausgeber.
- Bhatia, V., & Tandon, R. K. (2005). Stress and the gastrointestinal tract. *Journal of Gastroenterology and Hepatology*, 20, 332-339.
- Birbaumer, N. & Schmidt, R. F. (2010). *Biologische Psychologie* (7., überarbeitete und ergänzte Aufl.). Heidelberg: Springer.
- Bittner, A. & Helbig-Lang, S. (2011). Stressbewältigungs- und Problemlösetrainings. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 577-585). Berlin: Springer.
- Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97, 245-271.
- Boos, D. D. (2003). Introduction to the bootstrap world. *Statistical Science*, 18, 168-174.
- Born, J. & Plihal, W. (2000). Gedächtnisbildung im Schlaf: Die Bedeutung von Schlafstadien und Streßhormonfreisetzung. *Psychologische Rundschau*, 51, 198-208.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Aufl.). Heidelberg: Springer.
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Science*, 8, 539-546.
- Breiman, L., & Spector, P. (1992). Submodel selection and evaluation in regression: The X-random case. *International Statistical Review*, 16, 291-319.

Literaturverzeichnis

- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Science, 14*, 277-290.
- Bronstein, I. N., Semendjajew, K. A., Musiol, G. & Mühlig, H. (2006). *Taschenbuch der Mathematik* (6., vollständig überarbeitete und ergänzte Aufl.). Frankfurt a. M.: Harri Deutsch.
- Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., & McCarley, R. W. (2012). Control of sleep and wakefulness. *Physiological Review, 92*, 1087-1187.
- Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., & Sonuga-Barke, E. J. S. (2009). Default-mode brain dysfunction in mental disorders: a systematic review. *Neuroscience & Biobehavioral Reviews, 33*, 279-296.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. (2008). The brain's default network - anatomy, function, and relevance to disease. *Annals of the New York Academy of Science, 1124*, 1-38.
- Bundesministerium für Arbeit und Soziales (Hrsg.). (2013). *Sicherheit und Gesundheit bei der Arbeit 2011. Unfallverhütungsbericht Arbeit*. Berlin: Herausgeber.
- Bundesministerium für Arbeit und Soziales (Hrsg.). (2014). *Sicherheit und Gesundheit bei der Arbeit 2012. Unfallverhütungsbericht Arbeit*. Berlin: Herausgeber.
- Burisch, M. (2006). *Das Burnout-Syndrom - Theorie der inneren Erschöpfung* (3., überarbeitete Aufl.). Heidelberg: Springer.
- Burke, H. M., Davis, M. C., Otte, Ch., & Mohr, D. C. (2005). Depression and cortisol responses to psychological stress: a meta-analysis. *Psychoneuroendocrinology, 30*, 846-856.
- Busch, Ch. & Werner, D. (2011). Qualitätssicherung durch Evaluation. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Gesundheitsförderung und Gesundheitsmanagement in der Arbeitswelt – Ein Handbuch* (S. 221-234). Göttingen: Hogrefe.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Science, 4*, 215-222.
- Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50*, 191-214.
- Carayon, P., Smith, M. J., & Haims, M. C. (1999). Work organization, job stress, and work-related musculoskeletal disorders. *Human Factors, 41*, 644-663.
- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). Machine Learning - a historical and methodological analysis. *AI Magazine, 3*, 69-79.
- Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 367-379.
- Caruana, R., & Freitag, D. (1994). How useful is relevance? *AAAI Technical Report, FS-94-02*, 21-25.
- Carver, C. S., Sutton, S. K., & Scheier, M. F. (2000). Action, emotion, and personality: emerging conceptual integration. *Personality and Social Psychology Bulletin, 26*, 741-751.

Literaturverzeichnis

- Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, *80*, 265-278.
- Clarke, B., Fokoué, E., & Zhang, H. H. (2009). *Principles and theory for data mining and machine learning*. Heidelberg: Springer.
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*, 155-159.
- Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., et al. (2010). Neural systems subserving valence and arousal during the experience of induced emotion. *Emotion*, *10*, 377-389.
- Coppen, A. (1967). The biochemistry of affective disorders. *The British Journal of Psychiatry*, *113*, 1237-1264.
- Cosmar, M., Eichendorf, W. & Portuné, R. (2014). Maßnahmen für die psychische Gesundheit im Betrieb und Verantwortlichkeiten aus der Sicht der DGUV. In P. Angerer, J. Glaser, H. Gündel, P. Henningsen, C. Lahmann et al. (Hrsg.), *Psychische und psychosomatische Gesundheit in der Arbeit - Wissenschaft, Erfahrungen und Lösungen aus Arbeitsmedizin, Arbeitspsychologie und Psychosomatischer Medizin* (S. 319-325). Heidelberg: ecomed MEDIZIN.
- Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, *3*, 326-334.
- Cover, T. M. (1974). The best two independent measurements are not the two best. *IEEE Transactions on Systems, Men, and Cybernetics*, *1*, 116-117.
- Cover, T. M., & Van Campenhout, J. M. (1977). On the possible orderings in the measurement selection problem. *IEEE Transactions on Systems, Men, and Cybernetics*, *7*, 657-661.
- Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International Journal of Psychophysiology*, *73*, 88-94.
- Csernansky, J. G., Dong, H., Fagan, A. M., Wang, L., Xiong, C., et al. (2006). Plasma cortisol und progression of dementia in subjects with Alzheimer-type of dementia. *American Journal of Psychiatry*, *163*, 2164-2169.
- Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, *1*, 131-156.
- Davison, A. C., Hinkley, D. V., & Young, G. A. (2003). Recent developments in bootstrap methodology. *Statistical Science*, *18*, 141-157.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and self-determination of behavior. *Psychological Inquiry*, *11*, 227-268.
- Demerouti, E., Fergan, A., Glaser, J., Herbig, B., Hofmann, E. et al. (2011). *Psychische Belastung und Beanspruchung am Arbeitsplatz - Inklusive DIN EN ISO 10075-1 bis -3*. Berlin: Beuth.
- Denoth, J. & Stacoff, A. (1991). Belastung und Beanspruchung der Muskulatur - Einige Überlegungen aus mechanischer Sicht. *Sportverletzung Sportschaden*, *5*, 17-21.
- DeRijk, R. H., & de Kloet, E. R. (2006). Stress, glucocorticoids, and the brain. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 111-135). Totawa: Humana Press.

Literaturverzeichnis

- Devroye, L. (1988). Automatic pattern recognition: A study of the probability error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 530-543.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol response: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130, 355-391.
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*,
- Dilling, H., Mombour, W. & Schmidt, M. H. (Hrsg.). (2011). *Internationale Klassifikation psychischer Störungen - ICD-10 Kapitel V (F) - Klinisch-diagnostische Leitlinien* (8., überarbeitete Aufl.). Bern: Huber.
- Doumpos, M., & Slappa, A. (2005). Feature selection algorithms in classification problems: An experimental evaluation. *AIKED'05 Proceedings of the 4th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering Data Bases*, 36, 1-6.
- Dubnov, G., & Berry, E. M. (2006). Stress and the pathogenesis of eating disorders and obesity. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (S. 253-263). Totawa: Humana Press.
- Duda, R. O. (1970). Elements of pattern recognition. *Mathematics in Science and Engineering*, 66, 3-33.
- Duffy, E. (1962). *Activation and behavior*. Oxford: Wiley.
- Eckert, A. & Müller, W. E. (2008). Pharmakologische Grundlagen. In F. Holsboer, G. Gründer & O. Benkert (Hrsg.), *Handbuch der Psychopharmakotherapie* (S. 29-41). Heidelberg: Springer.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78, 316-331.
- Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science*, 18, 135-140.
- Eigenstetter, M., Kunz, T., Portuné, R. & Trimpop, R. (Hrsg.). (2014). *18. Workshop Psychologie der Arbeitssicherheit und Gesundheit - Psychologie der gesunden Arbeit*. Kröning: Asanger.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207-283). Lincoln, Nebraska: University of Nebraska Press.
- Elashoff, J. D., Elashoff, R. M., & Coldman, G. (1967). On the choice of variables in classification problems with dichotomous variables. *Biometrika*, 54, 668-670.
- Elkeles, T. & Beck, D. (2010). Evaluation von Betrieblicher Gesundheitsförderung – mehr als ein „Datenvergleich“. In G. Faller (Hrsg.), *Lehrbuch Betriebliche Gesundheitsförderung* (S. 156-164). Bern: Huber.
- Faller, G. (Hrsg.). (2010). *Lehrbuch Betriebliche Gesundheitsförderung*. Bern: Huber.
- Fahrenberg, J. (1994). *Die Freiburger Beschwerdeliste (FBL)*. Göttingen: Hogrefe.
- Faltermaier, T. (2005). *Gesundheitspsychologie*. Stuttgart: Kohlhammer.
- Felfe, J. (2009). Mitarbeiterführung. In H. Schuler, R. Hossiep, M. Kleinmann & W. Sarges (Hrsg.), *Praxis der Personalpsychologie: Human Resource Management Kompakt* (Band 20). Göttingen: Hogrefe.

Literaturverzeichnis

- Feldman Barrett, L. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science, 1*, 28-58.
- Feldman Barrett, L. (2006b). Solving the emotion paradox: categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20-46.
- Feldman Barrett, L. (2009). The future of psychology - Connecting mind to brain. *Perspectives on Psychological Science, 4*, 326-339.
- Feldman Barrett, L. (2012). Emotions are real. *Emotion, 12*, 413-429.
- Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*, 967-984.
- Feldman Barrett, L., & Russell, J. A. (1999). The structure of current affect: controversies and emerging consensus. *Current Directions in Psychological Science, 8*, 10-14.
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotion is not two-dimensional. *Psychological Science, 18*, 1050-1057.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 9673-9678.
- Foy, M. R., Kim, J. J., Shors, T. J., & Thompson, R. F. (2006). Neurobiological foundations of stress. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 37-65). Totawa: Humana Press.
- Frese, M., & Zapf, D. (1994). Action as the core of Work Psychology: A German approach. In H. C. Triandis & M. D. Dunnette (Hrsg.), *Handbook of industrial and organizational psychology* (pp. 271-340). Palo Alto: Consulting Psychologists Press.
- Fu, K. S. (1994). Statistical pattern recognition. In J. M. Mendel (ed.), *A prelude to neural networks: Adaptive and learning systems* (pp. 35-79). Upper Saddle River: Prentice Hall.
- Gebert, D. (2007). Organisationsentwicklung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie (4., aktualisierte Aufl., S. 601-616)*. Bern: Huber.
- Geddes, J. R., Carney, S. M., Davies, C., Furukawa, T. A., Kupfer, D. J., et al. (2003). Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *The Lancet, 361*, 653-661.
- Gerber, A. J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., et al. (2008). An affective circumplex model of neural systems subserving valence, arousal & cognitive overlay during the appraisal of emotional faces. *Neuropsychologia, 46*, 2129-2139.
- Glavas, M. M., & Weinberg, J. (2006). Stress, alcohol consumption, and the hypothalamic-pituitary-adrenal axis. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 165-183). Totawa: Humana Press.
- Görg, P. & Wieland, R. (2008). Verhältnisprävention durch Partizipation der Beschäftigten an der Gestaltung ihrer Arbeitsbedingungen. In C. Schwennen, G. Elke, B. Ludborz, H. Nold, S. Rohn, S. Schreiber-Costa & B. Zimolong (Hrsg.), *Psychologie der Arbeitssicherheit und Gesundheit. Perspektiven – Visionen* (S. 289-292). Kröning: Asanger.

Literaturverzeichnis

- Görg, P. & Wieland, R. (2009). Verhältnisprävention in der Finanzverwaltung: Ein Ansatz zur Mitgestaltung der Arbeitsbedingungen durch die Beschäftigten. In Gesellschaft für Arbeitswissenschaft (Hrsg.), *Arbeit, Beschäftigungsfähigkeit und Produktivität im 21. Jahrhundert* (S. 489-492). Dortmund: GfA-Press.
- Goldgruber, J. & Ahrens, D. (2009). Gesundheitsbezogene Interventionen in der Arbeitswelt - Review über die Wirksamkeit betrieblicher Gesundheitsförderung und Primärprävention. *Prävention und Gesundheitsförderung*, 4, 83-95.
- Gollwitzer, P. M. (1999). Implementation intentions – strong effects of simple plans. *The American Psychologist*, 54, 493-503.
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, 64, 1029-1041.
- Gregersen, S., Kuhnert, S., Zimber, A. & Nienhaus, A. (2011). Führungsverhalten und Gesundheit: Zum Stand der Forschung. *Gesundheitswesen*, 73, 3-12.
- Greif, S. (2011). Gesundheits- und Stressmanagementcoaching. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Betriebliche Gesundheitsförderung: Ein Handbuch* (S. 341-369). Göttingen: Hogrefe.
- Greiner, B., Leitner, K., Weber, W.-G., Hennes, K. & Volpert, W. (1987). RHIA – ein Verfahren zur Erfassung psychischer Belastung. In K. Sonntag (Hrsg.), *Arbeitsanalyse und Technikentwicklung* (S. 145-161). Köln: Bachem.
- Gunnar, M., & Quevedo, K. (2007). The Neurobiology of Stress and Development. *Annual Review of Psychology*, 58, 145-173.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Hacker, W. (1978). *Allgemeine Arbeits- und Ingenieurspsychologie - psychische Struktur und Regulation von Arbeitstätigkeiten* (3., überarbeitete Aufl.). Bern: Huber.
- Hacker, W. & Sachse, P. (2014). *Allgemeine Arbeitspsychologie: Psychische Regulation von Tätigkeiten* (3., vollständig überarbeitete Aufl.). Göttingen: Hogrefe.
- Hacker, W. & Richter, P. (1984). *Psychische Fehlbeanspruchung: Psychische Ermüdung, Monotonie, Sättigung und Streß* (2., veränderte und ergänzte Aufl.). Berlin: Springer.
- Hackman, J. R. (1970). Tasks and task performance in research on stress. In McGrath, J. E. (Ed.), *Social and psychological factors in stress* (pp. 202-237). New York: Holt, Rinehart & Winston.
- Hämming, O. & Bauer, G. F. (2010). Work-Life-Balance: Ein Thema der Betrieblichen Gesundheitsförderung. In G. Faller (Hrsg.), *Lehrbuch Betriebliche Gesundheitsförderung* (S. 229-239). Bern: Huber.
- Hall, P. (2003). A short prehistory of the bootstrap. *Statistical Science*, 18, 158-167.
- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Science*, 16, 458-466.

Literaturverzeichnis

- Hammes, M., Herold, K., Karakus, M., Lemke, P., Manz, R. et al. (2010). *PAKT - Programm Arbeit Rücken Gesundheit - Ein ganzheitliches Präventionsprogramm zur Reduzierung von Muskel-Skelett-Beschwerden* (Abschlussbericht). Berlin: uve GmbH für Managementberatung.
- Hammes, M. & Wieland, R. (2012a). Kontrollerleben und psychische Beanspruchung - Das Job-Strain-Control-Modell. In: R. Riemann (Hrsg.), *48. Kongress der Deutschen Gesellschaft für Psychologie* (S. 632). Lengerich: Pabst Science Publishers.
- Hammes, M. & Wieland, R. (2012b). Screening-Instrument zu Beanspruchung während der Arbeit. In G. Athanassiou, S. Schreiber-Costa & O. Sträter (Hrsg.), *Psychologie der Arbeitssicherheit und Gesundheit - Sichere und gute Arbeit erfolgreich gestalten - Forschung und Umsetzung in die Praxis* (S. 331-334). Kröning: Asanger.
- Hammes, M. & Wieland, R. (2014). Beanspruchungsbasierte Arbeitsplatzbewertung. In M. Eigenstetter, T. Kunz, R. Portuné & R. Trimpop (Hrsg.), *Psychologie der Arbeitssicherheit und Gesundheit - Psychologie der gesunden Arbeit* (S. 483-486). Kröning: Asanger.
- Hammes, M., Wieland, R. & Krajewski, J. (2011). Problemnahes Survey-Feedback durch systematische Item-Selektion - Anwendung nichtlinearer Selektions- und Klassifikationsverfahren in der arbeitspsychologischen Praxis. In F. W. Nerdinger & S. Curth (Hrsg.), *Arbeit - Organisation - Praxis - Wissenschaft für die Praxis* (S. 120). Rostock: Universität Rostock.
- Hammes, M., Wieland, R. & Winizuk, S. (2009). Wuppertaler Gesundheitsindex für Unternehmen (WGU). *Zeitschrift für Arbeitswissenschaft*, 4, 304-314.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning - data mining, inference, and prediction* (2nd ed.). Heidelberg: Springer.
- Heatheron, T. F. (2011). Neuroscience of self and self-regulation. *Annual Review of Psychology*, 62, 363-390.
- Hebb, D. O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 62, 243-254.
- Heinonen, A., Sievänen, H., Kannus, P., Oja, P., & Vuori, I. (2002). Site-specific skeletal response to long-term weight-training seems to be attributable to principal loading modality: a pQCT study of female weightlifters. *Calcified Tissue International*, 70, 469-474.
- Helbig-Lang, S. & Klose, M. (2011). Trainings der Selbstsicherheit, sozialer Fertigkeiten und der Kommunikation. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 565-576). Berlin: Springer.
- Higgins, E. T. (1996). Ideals, oughts, and regulatory focus: affect and motivation from distinct pains and pleasures. In P. M. Gollwitzer, & J. A. Bargh (eds.), *The psychology of action: linking cognition and motivation to behavior* (pp. 91-114). New York: Guilford.
- Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: a cognitive energetical framework. *Biological Psychology*, 45, 73-93.
- Holsboer-Trachsler, E. & Holsboer, F. (2008). Antidepressiva. In F. Holsboer, G. Gründer & O. Benkert (Hrsg.), *Handbuch der Psychopharmakotherapie* (S. 517-553). Heidelberg: Springer.

Literaturverzeichnis

- Holzmann, G., Meyer, H. & Schumpich, G. (2012). *Technische Mechanik Festigkeitslehre* (10., überarbeitete Aufl.). Wiesbaden: Springer.
- Hua, J., Xiong, Z., Lowey, J., Suh, E., & Dougherty, E. R. (2005). Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*, 21, 1509-1515.
- Hüther, G. (2014). *Biologie der Angst – Wie aus Streß Gefühle werden* (12. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Hüther, G., Doering, S., Rüger, U., Rüter, E., & Schüssler, G. (1999). The stress-reaction process and the adaptive modification and reorganization of neuronal networks. *Psychiatry Research*, 87, 83-95.
- Hüther, G. & Sachsse, U. (2007). Angst- und stressbedingte Störungen - Auf dem Weg zu einer neurobiologisch fundierten Psychotherapie. *Psychotherapeut*, 52, 166-179.
- Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, IT-14, 55-63.
- Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE transactions on pattern analysis and machine intelligence*, 22, 4-37.
- Jain, A., & Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 153-158.
- John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. In W. W. Cohen, & H. Hirsh (eds.), *Machine learning: Proceedings of the eleventh international conference* (pp. 121-129). San Francisco: Morgan Kaufmann.
- Joiko, K., Schmauder, M. & Wolff, G. (2010). *Psychische Belastung und Beanspruchung im Berufsleben - Erkennen - Gestalten* (5. Aufl.). Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Jones, B. E. (2003). Arousal systems. *Frontiers in Bioscience*, 8, 438-451.
- Jonge, J. de, Dollard, M. F., Dormann, Ch., Le Blanc, P. M., & Houtman, I. L. D. (2000). The demand-control-model: specific demands, specific control and well-defined groups. *International Journal of Stress Management*, 7, 269-287.
- Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiation structure in leadership research. *Journal of Applied Psychology*, 89, 36-51.
- Junge-Hoffmeister, J. (2011). Operante Verfahren. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 511-527). Berlin: Springer.
- Känel, R. v. (2008). Das Burnout-Syndrom: eine medizinische Perspektive. *Praxis*, 97, 477-487.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
- Kalousis, A., Prados, J., & Hilario, M. (2007). Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowledge and information systems*, 12, 95-116.
- Kaluza, G. (2011). *Stressbewältigung - Trainingsmanual zur psychologischen Gesundheitsförderung* (2., vollständig überarbeitete Aufl.). Berlin: Springer.

Literaturverzeichnis

- Kaluza, G. (2012). *Gelassen und sicher im Stress - Das Stresskompetenz-Buch - Stress erkennen, verstehen, bewältigen* (4., überarbeitete Aufl.). Berlin: Springer.
- Kanal, L. (1974). Patterns in pattern recognition: 1968-1974. *IEEE Transactions on Information Theory*, 20, 697-722.
- Kanal, L., & Chandrasekaran, B. (1971). On dimensionality and sample size in statistical pattern recognition. *Pattern Recognition*, 3, 225-234.
- Kanning, U. P. (2002). Soziale Kompetenz: Definition, Strukturen und Prozesse. *Zeitschrift für Psychologie*, 210, 154-163.
- Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: implications for job design. *Administrative Science Quarterly*, 24, 285-308.
- Karasek, R. A., & Theorell, T. (1990). *Healthy work: stress, productivity, and the reconstruction of working life*. New York: Basic Books.
- Kerns, J. G., Cohen, J. D., MacDonald III, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustment in control. *Science*, 303, 1023-1026.
- Kieseler, L. K. & Hammes, M. (2012). Prädiktion von Belastungen und Beschwerden auf Basis psychophysischer Beanspruchung in der Arbeit mit Hilfe von Support Vector Machines. In R. Riemann (Hrsg.), *48. Kongress der Deutschen Gesellschaft für Psychologie* (S. 344). Lengerich: Pabst Science Publishers.
- Kivimäki, M., Leino-Arjas, P., Luukkonen, R., Riihimöki, H., Vahtera, J., & Kirjonen, J. (2002). Work stress and risk of cardiovascular mortality: prospective cohort study of industrial employees. *British Medical Journal*, 325 (7369), 857.
- Klump, P. L. & Gemmiti, M. (2011). Harmonisierung von Berufs- und Privatleben: Vorsicht vor schnellen Lösungen. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Betriebliche Gesundheitsförderung: Ein Handbuch* (S. 259-277). Göttingen: Hogrefe.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273-324.
- Kohavi, R., & Sommerfield, D. (1995). Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. *KDD-95 Proceedings*, 192-197.
- Kohavi, R., & Wolpert, D. H. (1996). Bias plus variance decomposition for zero-one loss function. *Machine Learning Proceedings of the Thirteenth International Conference*.
- Köhler, Th. (2005). Psychotherapie und Pharmakotherapie. In F. Petermann & H. Reinecker (Hrsg.), *Handbuch der Klinischen Psychologie und Psychotherapie* (S. 394-401). Göttingen: Hogrefe.
- Köhler, Th. (2011). Psychopharmakologische Grundlagen. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie & Psychotherapie* (2. überarbeitete und erweiterte Aufl., S. 223-258). Berlin: Springer.
- Korczak, D. & Huber, B. (2012). Burn-out - Kann man es messen? *Bundesgesundheitsblatt*, 55, 164-171.

Literaturverzeichnis

- Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science, 1*, 111-117.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review, 26*, 159-190.
- Krajewski, J. & Wieland, R. (2003). Silent Room – die andere Mittagspause im Call Center. *Wirtschaftspsychologie aktuell, 4*, S. 50-53.
- Krishnan, V., & Nestler, E. J. (2008). The molecular neurobiology of depression. *Nature, 455*, 894-902.
- Krizek, P. (2008). *Feature selection: Stability, algorithms, and evaluation* (dissertation). Prag: Center for Machine Perception.
- Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition, 33*, 25-41.
- Kuhl, J. (1996). Who controls whom when „I control myself“? *Psychological Inquiry, 7*, 61-68.
- Kuhl, J. (2001). *Motivation und Persönlichkeit: Interaktionen psychischer Systeme*. Göttingen: Hogrefe.
- Kuncheva, L. I. (2007). A stability index for feature selection. In W. Hasselbring (ed.), *Proceedings of the 25th conference on iasted international multi-conference: Software engineering* (pp. 390-395). Anaheim: ACTA.
- Kupfer, D. J., Frank, E., & Phillips, M. L. (2012). Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *The Lancet, 379*, 1045-1055.
- Länderausschuss für Arbeitsschutz und Sicherheitstechnik (Hrsg.). (2002). Konzept zur Ermittlung psychischer Fehlbelastungen am Arbeitsplatz und zu Möglichkeiten der Prävention (LV 28). Ohne Ort: Herausgeber.
- Länderausschuss für Arbeitsschutz und Sicherheitstechnik (Hrsg.). (2003). Handlungsanleitung für die Arbeitsschutzverwaltungen der Länder zur Ermittlung psychischer Fehlbelastungen am Arbeitsplatz und zu Möglichkeiten der Prävention (LV 31). Ohne Ort: Herausgeber.
- Länderausschuss für Arbeitsschutz und Sicherheitstechnik (Hrsg.). (2009). Integration psychischer Belastungen in die Beratungs- und Überwachungspraxis der Arbeitsschutzbehörden der Länder (LV 52). Ohne Ort: Herausgeber.
- Langlay, P. (1994). Selection of relevant features in machine learning. *AAAI Technical Report, FS-94-02*, 127-131.
- Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion. Review of personality and social psychology* (No 13, pp. 25-59). Oaks: Sage.
- Latham, G. P., & Pinder, C. C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology, 56*, 485-516.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron, 73*, 653-676.
- Leitner, K., Volpert, W., Greiner, B., Weber, W. G. & Hennes, K. (1987). *Analyse psychischer Anforderungen und Belastungen in der Büroarbeit. Das RHIA/VERA-Büroverfahren* (Handbuch und Manual). Göttingen: Hogrefe.

Literaturverzeichnis

- Leontjew, A. N. (1982). *Tätigkeit, Bewusstsein, Persönlichkeit – Studien zur kritischen Psychologie*. Köln: Campus.
- Lerner, D. J., Levine, S., Malspeis, S., & D'Agostino, R. B. (1994). Job strain and health-related quality of life in a national sample. *American Journal of Public Health, 84*, 1580-1585.
- Lewin, K. (1938). *The conceptual representation and measurement of psychological forces* (Contributions to psychological theory, Vol. I, No. 4). Durham: Duke University Press.
- Lewin, K. (1939). Field theory and experiment in social psychology: concepts and methods. *American Journal of Sociology, 44*, 868-896.
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Feldman Barrett, L. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Science, 35*, 121-202.
- Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering, 17*, 491-502.
- Liu, H., & Setiono, R. (1998). Some issues on scalable feature selection. *Expert Systems with Applications, 15*, 333-339.
- Löffler, G. (2008). *Basiswissen Biochemie mit Pathobiochemie* (7., komplett überarbeitete Aufl.). Heidelberg: Springer.
- Lohmann-Haislah, A. (2012). *Stressreport Deutschland 2012 - Psychische Anforderungen, Ressourcen und Befinden*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Lord, R. G., Diefendorff, J. M., Schmidt, A. M., & Hall, R. J. (2010). Self-Regulation at Work. *Annual Review of Psychology, 61*, 543-568.
- Lüders, E. (1999). Analyse psychischer Belastungen in der Arbeit: Das RIHA-Verfahren. In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (Schriftenreihe Mensch-Technik-Organisation, Band 14, S. 365-395). Zürich: Verlag der Fachvereine.
- MacDonald, H. A., Colotla, V., Flamer, S., & Karlinsky, H. (2003). Posttraumatic stress disorder (PTSD) in the workplaces: A descriptive study of workers experiencing PTSD resulting from work injury. *Journal of Occupational Rehabilitation, 13*, 63-77.
- Maimon, O., & Rokach, L. (Eds.). (2005). *The data mining and knowledge discovery handbook*. New York: Springer.
- Malmö, R. B. (1959). Activation: a neuropsychological dimension. *Psychological Review, 66*, 367-386.
- Manz, R. (1998). *KÖPS: Ein Fragebogen für körperliche, psychische und soziale Symptome. Manual, Testbögen und Zubehör*. Frankfurt a. M.: Swet & Zeitliner.
- Marill, T., & Green, D. M. (1960). Statistical recognition functions and the design of pattern recognizers. *IRE Transactions on Electronic Computers, 4*, 472-477.
- Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory, 9*, 11-17.

Literaturverzeichnis

- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory manual* (3rd ed.). Palo Alto: Consulting Psychologists Press.
- Maslow, H. A. (1943). A theory of human motivation. *Psychological Review*, *50*, 370-396.
- Mausfeld, R. (2009). Psychologie, 'weiße Folter' und die Verantwortlichkeit von Wissenschaftlern. *Psychologische Rundschau*, *60*, 229-240.
- Mayer, E. A. (2000). The neurobiology of stress and gastrointestinal disease. *Gut*, *47*, 861-869.
- Meewisse, M.-L., Reitsma, J. B., de Vries, G.-J., Gersons, B. P. R., & Olf, M. (2007). Cortisol and post-traumatic stress disorder in adults: Systematic review and meta-analysis. *The British Journal of Psychiatry*, *191*, 387-392.
- Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying Tripel network model. *Trends in Cognitive Science*, *15*, 483-506.
- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*, 669-680.
- Michaelis, L. & Menten, M. L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, *49*, 333-369.
- Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society (A)*, *147*, 389-425.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.
- Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society B*, *357*, 1123-1136.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and structure of behavior*. New York: Holt, Rinehart and Winston.
- Mohr, T., Müller, A. & Rigotti, T. (2005). Normwerte der Skala Irritation: Zwei Dimensionen psychischer Beanspruchung. *Diagnostica*, *51*, 12-20.
- Mohr, G., Rigotti, T. & Müller, A. (2005). Irritation - ein Instrument zur Erfassung psychischer Beanspruchung im Arbeitskontext. Skalen- und Itemparameter aus 15 Studien. *Zeitschrift für Arbeits- und Organisationspsychologie*, *49*, 44-48.
- Molina, L. C., Belanche, L., & Nebot, A. (2002). Feature selection algorithm: A survey and experimental evaluation. *Second IEEE International Conference on Data Mining*, *4*, 306-313.
- Mosteller, F., & Tuckey, J. W. (1968). Data analysis, including statistics. In G. Lindzey, & E. Aronson (eds.), *Handbook of social psychology* (2nd ed., pp. 80-183). New York: Addison-Wesley.
- Mühlig, S. & Jacobi, F. (2011). Psychoedukation. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 477-490). Berlin: Springer.
- Mühlig, S. & Poldrack, A. (2011). Kognitive Therapieverfahren. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 543-564). Berlin: Springer.

Literaturverzeichnis

- Müller, T. & Paterok, B. (2010). *Schlaftraining - Ein Therapiemanual zur Behandlung von Schlafstörungen* (2., überarbeitete Aufl.). Göttingen: Hogrefe.
- Nachreiner, F. (1999). International standards on mental work-load - the ISO 10 075 series. *Industrial Health*, 37, 125-133.
- Nagy, G. (1968). State of the art in pattern recognition. *Proceedings of the IEEE*, 56, 836-857.
- Narendra, P. M., & Fukunaga, K. (1977). A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 9, 917-922.
- Nationale Arbeitsschutzkonferenz (2012). *Leitlinie Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz*. Berlin: Nationale Arbeitsschutzkonferenz.
- Nerdinger, F. W. (2012). Emotionsarbeit im Dienstleistungsbereich. *Report Psychologie*, 37 (1), S. 8-18.
- Neuberger, O. (2002). *Führen und führen lassen* (6., völlig neu bearbeitete und erweiterte Aufl.). Stuttgart: Lucius & Lucius.
- Neumann, U.-N. & Frasch, K. (2005). Biologische Mechanismen antidepressiver Wirksamkeit von körperlicher Aktivität. *psychoneuro*, 31, 513-518.
- Neyberg, A., Bernin, P., & Theorell, T. (2005). *The impact of leadership on the health of subordinates*. Stockholm: National Institute for Working Life.
- Neyberg, S. T., Fransson, E. I., Heikkilä, K., Alfredsson, L., Casini, A., et al. (2013). Job strain and cardiovascular disease risk factors: meta-analysis of individual-participant data from 47,000 men and women. *PLoS ONE*, 8 (6), e67323.
- Ng, A. Y. (1998). *On feature selection: Learning with exponentially many irrelevant features as training examples* (dissertation). Massachusetts: Massachusetts Institute of Technology.
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3, 400-424.
- Novara, C., Buodo, G., Ghisi, M., Poli, S., Bichescu, D. M., et al. (2009). Short Communication: Integrated assessment of emotional distress after work-related accidents. *Stress and Health*, 25, 195-204.
- Novovicova, J., Pudil, P., & Kittler, J. (1996). Divergence based feature selection for multimodal class densities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18, 218-223.
- Nowlis, V. (1965). Research with the mood adjective checklist. In S. S. Tompkins, & C. E. Izard (Eds.), *Affect, cognition and personality: empirical studies* (pp. 352-389). New York: Springer.
- Nowlis, V., & Nowlis, H. H. (1956). The description and analysis of mood. *Annals of the New York Academy of Sciences*, 65, 345-355.
- Ochsner, K. N., & Gross, J. G. (2005). The cognitive control of emotion. *Trends in Cognitive Science*, 9, 242-249.
- Osgood, Ch. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. Illinois: University of Illinois.

Literaturverzeichnis

- Paul, K. I. & Moser, K. (2007). Arbeitslosigkeit. In K. Moser (Hrsg.), *Wirtschaftspsychologie* (S. 285-305). Heidelberg: Springer.
- Pauwels, F. (1948). Die Bedeutung der Bauprinzipien des Stütz- und Bewegungsapparates für die Beanspruchung der Röhrenknochen. *Zeitschrift für Anatomie und Entwicklungsgeschichte*, 114, 129-166.
- Pauwels, F. (1950). Die Bedeutung der Bauprinzipien der unteren Extremitäten für die Beanspruchung des Beinskeletes. *Zeitschrift für Anatomie und Entwicklungsgeschichte*, 114, 525-538.
- Pauwels, F. (1960). Eine neue Theorie über den Einfluß mechanischer Reize auf die Differenzierung der Stützgewebe. *Zeitschrift für Anatomie und Entwicklungsgeschichte*, 121, 476-515.
- Penzlin, H. (2014). *Das Phänomen Leben: Grundfragen der Theoretischen Biologie*. Berlin: Springer.
- Peter, R. (2010). Von Handlungs- und Entscheidungsspielräumen, Belohnungen und betrieblicher Gerechtigkeit: Die Modelle Demand-Control und berufliche Gratifikationskrisen. In G. Faller (Hrsg.), *Lehrbuch Betriebliche Gesundheitsförderung* (S. 59-69). Bern: Huber.
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective, neuroscience, cognitive development and psychopathology. *Development and Psychopathology*, 17, 715-734.
- Pribram, K. H., & McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological Review*, 82, 116-149.
- Pudil, P., Novovicová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15, 1119-1125.
- Puri, P. K. (2006). Lipids and depression. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 221-230). Totawa: Humana Press.
- Rabinovitz, S. (2006). Stress and food craving. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 155-164). Totawa: Humana Press.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain functioning. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 676-682.
- Rau, R. (2011). Zur Wechselwirkung von Arbeit, Beanspruchung und Erholung. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Betriebliche Gesundheitsförderung: Ein Handbuch* (S. 83-106). Göttingen: Hogrefe.
- Raudys, S. (2006). Feature over-selection. In D.-Y. Yeung et al. (eds.), *SSPR & SPR 2006, LNCS 4109* (pp. 622-631). Heidelberg: Springer.
- Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Review*, 93, 681-766.
- Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67, 525-539.
- Renneberg, B. & Hammelstein, P. (Hrsg.). (2006). *Gesundheitspsychologie*. Heidelberg: Springer.

Literaturverzeichnis

- Renneberg, B., Heidenreich, Th. & Noyon, A. (2009). *Einführung Klinische Psychologie*. München: Reinhardt/UTB.
- Rentzsch, K. & Schütz, A. (2009). *Psychologische Diagnostik – Grundlagen und Anwendungsperspektiven*. Stuttgart: Kohlhammer.
- Reunanan, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research*, 3, 1371-1382.
- Richter, G. (2010). Stress, psychische Ermüdung, Monotonie, psychische Sättigung. In Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Hrsg.), *Psychische Belastung und Beanspruchung* (Arbeitswissenschaftliche Erkenntnisse - Forschungsergebnisse für die Praxis, Nr. 116, S. 1-13). Dortmund: Herausgeber.
- Richter, P., Buruck, G., Nebel, C. & Wolf, S. (2011). Arbeit und Gesundheit: Risiken, Ressourcen und Gestaltung. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Betriebliche Gesundheitsförderung: Ein Handbuch* (S. 25-59). Göttingen: Hogrefe.
- Rigotti, T. & Mohr, G. (2011). Gesundheit und Krankheit in der neuen Arbeitswelt. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Gesundheitsförderung und Gesundheitsmanagement in der Arbeitswelt - Ein Handbuch* (S. 61-82). Göttingen: Hogrefe.
- Rohmert, W. (1984). Das Belastungs-Beanspruchungskonzept. *Zeitschrift für Arbeitswissenschaften*, 38, 193-200.
- Rohmert, W. & Rutenfranz, J. (1975). *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen*. Bonn: Bundesminister für Arbeit und Sozialordnung.
- Rosenblatt, F. (1960). Perceptron simulation experiments. *Proceedings of the IRE*, 48, 301-309.
- Rosenstiel, L. von (2007). *Grundlagen der Organisationspsychologie* (6., überarbeitete Aufl.). Stuttgart: Schäffer Pöschel.
- Rosenstiel, L. von, Molt, W. & Rüttinger, B. (2005). *Organisationspsychologie* (9., vollständig überarbeitete und erweiterte Aufl.). Stuttgart: Kohlhammer.
- Rothe, I. (Präsidentin der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin). (2014). *Psychische Gesundheit in der Arbeitswelt - Neue Anforderungen an Wissenschaft, Politik und Praxis*. Vortrag im Rahmen des 18. Workshop Psychologie der Arbeitssicherheit und Gesundheit am 16. Juni 2014 in Dresden.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80.
- Rowolt, J. & Heinitz, K. (2008). Führungsstile als Stressbarrieren: Zum Zusammenhang zwischen transformationaler, transaktionaler, mitarbeiter- und aufgabenorientierter Führung und Indikatoren von Stress bei Mitarbeitern. *Zeitschrift für Personalpsychologie*, 7, 129-140.
- Ruhl, U., Hach, I. & Wittchen, U.-H. (2011). Entspannungsverfahren. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (2., überarbeitete und erweiterte Aufl., S. 587-599). Berlin: Springer.

Literaturverzeichnis

- Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., et al. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *American Journal of Psychiatry*, *163*, 1905-1917.
- Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, *37*, 345-356.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145-172.
- Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, *23*, 1259-1283.
- Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805-819.
- Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*, 61-97.
- Sapolsky, R. M. (2003). Stress and plasticity in the limbic system. *Neurochemical Research*, *28*, 1735-1742.
- Satzer, R. & Langhoff, T. (2010). Betriebliche Erfahrungen zur Umsetzung der Gefährdungsbeurteilung bei psychischen Belastungen. *Gute Arbeit*, *3*, S. 13-16.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*, 695-729.
- Scherrer, K. (2001). Kommunikationsarbeit im Call Center: Umfassende Arbeitsgestaltung fördert Gesundheit und Produktivität. In M. Kastner, K. Kipfmüller, W. Quaas & Kh. Sonntag (Hrsg.), *Gesundheit und Sicherheit in Arbeits- und Organisationsformen der Zukunft. Ergebnisbericht des Projektes gesina* (S. 169-189). Bremerhaven: Wirtschaftsverlag NW.
- Scherrer, K. (2002). *Kommunikationsarbeit im Teleservice: Beanspruchung und emotionale Regulation bei Call Center-Dienstleistungen* (Dissertation). Wuppertal: Bergische Universität Wuppertal.
- Scherrer, K. & Wieland, R. (2006). *Fragebogen zu Führung und Zusammenarbeit. Handbuch und Manual*. Wuppertal: Kompetenzzentrum für Fortbildung und Arbeitsgestaltung (KomFor).
- Schildkraut, J. J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *American Journal of Psychiatry*, *122*, 509-522.
- Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology*, *29*, 497-510.
- Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: a quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*, 325-345.
- Schimmack, U., & Reisenzein, R. (2002). Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion*, *2*, 412-417.
- Scholz, D. (2013). *Arbeitsfähig bis zur Rente? DGB-Index Gute Arbeit - Der Report*. Berlin: Institut DGB-Index Gute Arbeit.

Literaturverzeichnis

- Schönpflug, W. (1987). Beanspruchung und Belastung bei der Arbeit: Konzepte und Theorien. In U. Kleinbeck & J. Rutenfranz (Hrsg.), *Arbeitspsychologie* (Enzyklopädie der Psychologie, Band D-III-1, S. 130-184). Göttingen: Hogrefe.
- Schönpflug, W. (1991). Von der Aktivierungstheorie zur Ressourcentheorie: Die Regulation von Aktivierungszuständen. In J. Janssen, E. Hahn & H. Strang (Hrsg.), *Konzentration und Leistung* (S. 61-72). Göttingen: Hogrefe.
- Schulz, P. (2012). *Beanspruchung und Gesundheit - Fehlbeanspruchung, Gesundheitsrisiken und Beanspruchungsoptimierung im Arbeitsleben*. Kröning: Asanger.
- Schüpbach, H. (2013). *Arbeits- und Organisationspsychologie*. München: Ernst Reinhardt.
- Schüpbach, H. & Zölch, M. (2007). Analyse und Bewertung von Arbeitssystemen und Arbeitstätigkeiten. In H. Schuler (Hrsg.), *Organisationspsychologie* (4., aktualisierte Aufl., S. 197-220). Bern: Huber.
- Schütte, M. & Köper, B. (2013). Veränderung der Arbeit. *Bundesgesundheitsblatt*, 56, 422-429.
- Schwarzer, R. (2004). *Psychologie des Gesundheitsverhaltens - Einführung in die Gesundheitspsychologie* (3., überarbeitete Aufl.). Göttingen: Hogrefe.
- Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.
- Schwenkmezger, P. (1990). Ärger, Ärgerausdruck und Gesundheit. In R. Schwarzer (Hrsg.), *Gesundheitspsychologie* (S. 295-310). Göttingen: Hogrefe.
- Schyns, B., & Collani, G. von. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology*, 11, 219-241.
- Seegerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, 130, 601-630.
- Seiler, K. & Jansing, P.-J. (Hrsg.). (2014). *Erkrankungsrisiken durch arbeitsbedingte psychische Belastung* (transfer 4). Düsseldorf: Landesinstitut für Arbeitsgestaltung des Landes Nordrhein-Westfalen.
- Seligman, M. E. P. (1972). Learned helplessness. *Annual review of medicine*, 23, 407-412.
- Selye, H. (1974). *Stress without distress*. Philadelphia: Lippincott.
- Semmer, N. K. (1984). *Stressbezogene Tätigkeitsanalyse*. Weinheim: Beltz.
- Semmer, N. K. & Udris, I. (2007). Bedeutung und Wirkung von Arbeit. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (4., aktualisierte Aufl., S. 157-195). Bern: Huber.
- Semmer, N. K., Zapf, D. & Dunckel, H. (1999). Instrument zur Stressbezogenen Tätigkeitsanalyse (ISTA). In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (Schriftenreihe Mensch-Technik-Organisation, Band 14, S. 179-204). Zürich: Verlag der Fachvereine.

Literaturverzeichnis

- Sherwood Brown, E., Varghese, F. P., & McEwen, B. S. (2004). Association of depression with medical illness: does cortisol play a role? *Biological Psychiatry*, *2004*, 1-9.
- Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, *10*, 335-347.
- Siegrist, J. (1996a). Adverse health effects of high-effort/low-reward conditions at work. *Journal of Occupational Health Psychology*, *1*, 27-43.
- Siegrist, J. (1996b). *Soziale Krisen und Gesundheit. Eine Theorie der Gesundheitsförderung am Beispiel für Herz-Kreislauf-Risiken im Erwerbsleben*. Göttingen: Hofgrefe.
- Siegrist, J. (2002). Effort-reward imbalance at work and health, historical and current perspectives on stress and health. In P. Perrewe, & D. Gangster (Eds.), *Research in occupational stress and well being* (2nd ed., pp. 261-291). New York: Elsevier.
- Siegrist, J. & Dragano, N. (2008). Psychosoziale Belastungen und Erkrankungsrisiken im Erwerbsleben. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *51*, 305-312.
- Somol, P., & Novovicova, J. (2010). Evaluating stability and comparing output of feature selectors that optimize feature subset cardinality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 1921-1939.
- Somol, P., Novovicova, J., & Pudil, (2010). Efficient feature subset selection and subset size optimization. In A. Herout (ed.), *Pattern recognition recent advances* (pp. 75-97). Rijeka: InTech.
- Somol, P., Pudil, P., Novovicová, J., & Paclík, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, *20*, 1157-1163.
- Sonntag, Kh. & Stegmaier, R. (2007). *Arbeitsorientiertes Lernen: Zur Psychologie der Integration von Lernen und Arbeit*. Stuttgart: Kohlhammer.
- Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2008). The common basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*, 489-510.
- Stadler, P. & Spieß, E. (2005). Gesundheitsförderliches Führen – Defizite erkennen und Fehlbelastungen der Mitarbeiter reduzieren. *Arbeitsmedizin Sozialmedizin Umweltmedizin*, *40*, 384-390.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictors. *Journal of the Royal Statistical Society (B)*, *36*, 111-147.
- Ströhle, A. (2003). Die Neuroendokrinologie von Stress und die Pathophysiologie und Therapie von Depression und Angst. *Der Nervenarzt*, *74*, 279-292.
- Strunk, G. & Schiepek, G. (2006). *Systemische Psychologie - Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens*. München: Elsevier.
- Taris, T. W., & Feij, J. A. (2004). Learning and strain among newcomers: a three-way-study on the effects of job demands and job control. *Journal of Psychology*, *138*, 543-563.
- Taylor, M. J., Freemantle, N., Geddes, J. R., & Bhagager, Z. (2006). Early onset on selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Archives of General Psychiatry*,

Literaturverzeichnis

- Teather, L. A. (2006). Pathophysiological effects of inflammatory mediators and stress on distinct memory systems. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (S. 377-386). Totawa: Humana Press.
- Teigen, K. H. (1994). Yerkes-Dodson: a law for all seasons. *Theory & Psychology, 4*, 525-547.
- Tellegen, A., Watson, D., & Clark, L. E. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297-303.
- Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports, 20*, 663-678.
- Thayer, R. E. (1978). Towards a psychological theory of multidimensional activation (arousal). *Motivation and Emotion, 2*, 1-34.
- Thayer, R. E. (1996). *The origin of everyday moods: Managing energy, tension, and stress*. New York: Oxford University Press.
- Tononi, G., & Cirelli, Ch. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*, 49-62.
- Toussaint, G. T. (1971). Note on optimal selection of independent binary-valued features for pattern recognition. *IEEE Transaction on Information Theory, 17*, 618.
- Trimpop, R., Hamacher, W., Lenartz, N., Ruttke, T., Riebe, S. et al. (2012). *Sifa-Langzeitstudie: Tätigkeiten und Wirksamkeit von Fachkräften für Arbeitssicherheit* (Abschlussbericht). Dresden: Technische Universität.
- Udris, I. (2007). Gesundheits- und verhaltensbezogene Intervention. In H. Schuler & Kh. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 668-675). Bern: Huber.
- Ulich, E. (2011). *Arbeitspsychologie* (7., neu überarbeitete und erweiterte Aufl.). Zürich: vdf Hochschulverlag.
- Ulich, E. & Wiese, B. S. (2011). *Life Domain Balance: Konzepte zur Verbesserung der Lebensqualität*. Wiesbaden: Springer Gabler.
- Ulich, E. & Wülser, M. (2012). *Gesundheitsmanagement in Unternehmen: Arbeitspsychologische Perspektiven* (5., überarbeitete und erweiterte Aufl.). Berlin: Springer Gabler.
- Ursin, H., & Olf, M. (1993). The stress response. In S. C. Stanford, & P. Salmon (Eds.), *Stress. From synapse to syndrome* (pp. 3-22). London: Academic Press.
- van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional activity. *European Neuropsychopharmacology, 20*, 519-534.
- Van Vegchel, N., De Jonge, J., Bosma, H., & Schaufeli, W. (2005). Reviewing the effort-reward imbalance model: drawing up the balance of 45 empirical studies. *Social Science & Medicine, 60*, 1117-1131.
- Velázquez-Moctezuma, J., & Drucker-Colín, R. (2006). Nicotine, sleep, and depression. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 231-244). Totawa: Humana Press.
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology, 100*, 3328-3342.
- Wald, A. (1949). Statistical decision functions. *Annals of Mathematical Statistics, 20*, 165-205.
- Warr, P. (1994). A conceptual framework for the study of work and mental health. *Stress and Work, 8*, 84-97.

Literaturverzeichnis

- Warschburger, P. (Hrsg.). (2009). *Beratungspsychologie*. Heidelberg: Springer.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measure of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Weber, W. G. (1999). Organisationale Demokratie: Anregungen für innovative Arbeitsformen jenseits bloßer Partizipation? *Zeitschrift für Arbeitswissenschaft*, 53, 270-281.
- Wegge, J. & Rosenstiel, L. von. (2007). Führung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (4., aktualisierte Aufl., S. 475-512). Bern: Huber.
- Wheatly, D. (2006). Herbal products, stress, and the mind. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 137-153). Totawa: Humana Press.
- White, R. W. (1959). Motivation reconsidered: the concept of competence. *Psychological Review*, 66, 297-333.
- Wieland, R. (1999). Analyse, Bewertung und Gestaltung psychischer Belastung und Beanspruchung. In B. Badura, M. Litsch & Ch. Vetter (Hrsg.), *Fehlzeiten-Report: Psychische Belastung am Arbeitsplatz* (S. 197-211). Berlin: Springer.
- Wieland, R. (2001). Zeitarbeit beanspruchungsoptimal gestalten. In M. Kastner, K. Kipfmüller, W. Quaas & Kh. Sonntag (Hrsg.), *Gesundheit und Sicherheit in Arbeits- und Organisationsformen der Zukunft. Ergebnisbericht des Projektes gesina* (S. 148-168). Bremerhaven: Wirtschaftsverlag NW.
- Wieland, R. (2004). Arbeitsgestaltung, Selbstregulationskompetenz und berufliche Kompetenzentwicklung. In: B. Wiese (Hrsg.), *Individuelle Steuerung beruflicher Entwicklung. Kernkompetenzen in der modernen Arbeitswelt* (S. 170-197). Frankfurt am Main: Campus.
- Wieland, R. (2006a). *Eigenschaftswörterliste zur Erfassung von Emotionen und Beanspruchungserleben während der Arbeit (EEB)*. Wuppertal: Kompetenzzentrum für Fortbildung und Arbeitsgestaltung (KomFor).
- Wieland, R. (2006b). Psychische Belastung. In K. Landau (Hrsg.), *Lexikon Arbeitsgestaltung* (S. 301-304). Stuttgart: Genter.
- Wieland, R. (2008). Ansätze der betrieblichen Gesundheitsförderung zur Rückengesundheit. In Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (Hrsg.), *Rückengesundheit fördern und verbessern. Dokumentation der Fachtagung zu einem der zehn Gesundheitsziele im Land Nordrhein-Westfalen* (Gesundheitsberichte Spezial, Band 5, S. 41-48). Düsseldorf: Herausgeber.
- Wieland, R. (2009). *Barmer Gesundheitsreport 2009: Psychische Gesundheit und psychische Belastungen*. Wuppertal: Barmer.
- Wieland, R. (2010a). *Barmer GEK Gesundheitsreport 2010, Teil 1: Gesundheitskompetenz in Unternehmen stärken, Gesundheitskultur fördern*. Berlin: Barmer GEK.
- Wieland, R. (2010b). Gestaltung gesundheitsförderlicher Arbeitsbedingungen. In U. Kleinbeck & K.-H. Schmidt (Hrsg.), *Arbeitspsychologie* (Enzyklopädie der Psychologie, Themenbereich D, Serie III, Band 1, S. 869-919). Göttingen: Hogrefe.

Literaturverzeichnis

- Wieland, R. (2013). Status-Bericht: Psychische Gesundheit in der betrieblichen Gesundheitsförderung - eine arbeitspsychologische Perspektive. In H. Nold & G. Wenninger (Hrsg.), *Rückengesundheit und psychische Gesundheit* (S. 103-128). Kröning: Asanger.
- Wieland, R. (2014a). Gestaltungsfreiheit als Zweck und Mittel psychologischer Arbeits- und Organisationsgestaltung. In P. Sachse & E. Ulich (Hrsg.), *Psychologie menschlichen Handelns: Wissen und Denken - Wollen und Tun* (S. 207-242). Lengerich: Pabst.
- Wieland, R. (2014b). *Wuppertaler Inventar zur ressourcenorientierten Arbeits- und Organisationsgestaltung (WITAL): Arbeitspsychologisches Verfahren zur Analyse, Bewertung und Gestaltung der Arbeit im Rahmen des betrieblichen Gesundheitsmanagements und ressourcen-orientierter Arbeits- und Organisationsgestaltung*. Wuppertal: Bergische Universität Wuppertal.
- Wieland, R. & Görg, P. (2009). Gesundheitskompetenzentwicklung in der Finanzverwaltung durch gesundheitsförderliche Arbeitsgestaltung und Teamentwicklung. In K. Henning, I. Leisten & F. Hees (Hrsg.), *Innovationsfähigkeit stärken - Wettbewerbsfähigkeit erhalten. Präventiver Arbeits- und Gesundheitsschutz als Treiber* (Aachener Reihe Mensch und Technik, Band 60, S. 207-225). Aachen: Wissenschaftsverlag Mainz.
- Wieland, R. & Hammes, M. (2009). Gesundheitskompetenz als personale Ressource. In K. Mozygemba et al. (Hrsg.), *Nutzenorientierung - ein Fremdwort in der Gesundheitssicherung?* (S. 177-190). Bern: Huber.
- Wieland, R. & Hammes, M. (2010). *Barmer GEK Gesundheitsreport 2010, Teil 2: Ergebnisse der Internetstudie zur Gesundheitskompetenz*. Berlin: Barmer GEK.
- Wieland, R. & Hammes, M. (2014a). Psychologie für den Menschen. In R. Wieland, O. Strohm, W. Hacker & P. Sachse (Hrsg.), *Wir müssen uns einmischen: Arbeitspsychologie für den Menschen* (S. 101-114). Kröning: Asanger.
- Wieland, R. & Hammes, M. (2014b). Wuppertaler Screening Instrument Psychische Beanspruchung (WSIB) - Beanspruchungsbilanz und Kontrollerleben als Indikatoren für gesunde Arbeit. *Journal Psychologie des Alltagshandelns*, 7, 30-50.
- Wieland, R. & Krajewski, J. (2002). *Psychische Belastung und Qualifizierung in neuen Arbeitsformen I: Zeitarbeit* (Wuppertaler psychologische Berichte, Heft 1). Wuppertal: Bergische Universität Wuppertal.
- Wieland, R. & Krajewski, J. (2007). Neue Organisationsformen der Arbeit. In H. Schuler & Kh. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 210-216). Bern: Huber.
- Wieland, R., Krajewski, J. & Memmou, M. (2006): Arbeitsgestaltung, Persönlichkeit und Arbeitszufriedenheit. In L. Fischer (Hrsg.), *Arbeitszufriedenheit. Konzepte und empirische Befunde* (S. 226-242). Göttingen: Hogrefe.
- Wieland, R., Metz, A.-M. & Richter, P. (2001). *Call Center auf dem arbeitspsychologischen Prüfstand* (Teil 1, CCall Report 3). Hamburg: Verwaltungs-Berufsgenossenschaft.
- Wieland, R. & Scherrer, K. (Hrsg.). (2000). *Arbeitswelten von morgen. Neue Technologien und Organisationsformen, Gesundheit und Arbeitsgestaltung, flexible Arbeitszeit- und Beschäftigungsmodelle*. Wiesbaden: Springer VS.
- Wieland, R. & Scherrer, K. (2007). *Barmer Gesundheitsreport 2007: Führung und Gesundheit*. Wuppertal: Barmer.

Literaturverzeichnis

- Wieland, R., Winizuk, S. & Hammes, M. (2009). Führung und Arbeitsgestaltung - Warum gute Führung allein nicht gesund macht. *Zeitschrift Arbeit*, 4, 282-297.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung*. Göttingen: Hogrefe.
- Wieland-Eckelmann, R. (1996). Streß (Belastung und Beanspruchung). In G. Wenninger & C. G. Hoyos (Hrsg.), *Arbeits-, Gesundheits- und Umweltschutz. Handwörterbuch psychologischer Grundbegriffe* (S. 480-491). Heidelberg: Asanger.
- Wieland-Eckelmann, R., Allmer, H., Kallus, K. W. & Otto, J. (Hrsg.). (1994). *Erholungsforschung: Beiträge Emotionspsychologie, Sportpsychologie und Arbeitspsychologie*. Weinheim: Belz.
- Wieland-Eckelmann, R., Baggen, R. & Schwarz, R. (1991). Analysen mentaler Beanspruchungen als Mittel der Schnittstellengestaltung in Mensch-Maschine-Systemen. In M. Frese, C. Kasten, C. Skarpelis & Z. Zang-Scheucher (Hrsg.), *Software für die Arbeit von morgen* (S. 47-60). Krefeld: Vennekel und Partner.
- Wieland-Eckelmann, R. & Bösel, R. (1987). Konstruktion eines Verfahrens zur Erfassung dispositioneller Angstbewältigungsstile im Leistungsbereich. *Zeitschrift für Differenzielle und Diagnostische Psychologie*, 8, 39-56.
- Wieland-Eckelmann, R., Saßmannshausen, A., Rose, M. & Schwarz, R. (1999). Synthetische Beanspruchungs- und Arbeitsanalyse: SynBA-GA. In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (Schriftenreihe Mensch-Technik-Organisation, Band 14). Zürich: Verlag der Fachvereine.
- Wilken, B. (2006). *Methoden der Kognitiven Umstrukturierung - Ein Leitfaden für die psychologische Praxis* (3., aktualisierte Aufl.). Stuttgart: Kohlhammer.
- Wittchen, H.-U. & Hoyer, J. (Hrsg.). (2011). *Klinische Psychologie & Psychotherapie* (2. überarbeitete und erweiterte Aufl.). Berlin: Springer.
- Wundt, W. (1924). *An introduction to psychology* (R. Pintner, Trans.). London: Allen & Unwin. (Original work published 1912).
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Yehuda, S., & Mostofsky, D. I. (Eds.). (2006). *Nutrients, stress, and medical disorders*. Totawa: Humana Press.
- Yehuda, S., Rabinovitz, S., & Mostofsky, D. I. (2006). Essential fatty acids and stress. In S. Yehuda, & D. I. Mostofsky (Eds.), *Nutrients, stress, and medical disorders* (pp. 99-110). Totawa: Humana Press.
- Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current-affect: integration and beyond. *Journal of Personality and Social Psychology*, 77, 600-619.
- Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, 11, 705-731.
- Yusta, S. C. (2009). Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, 30, 525-534.
- Zhang, D., & Raichle, M. E. (2010). Disease and the brain dark energy. *Nature Reviews Neurology*, 6, 15-28.