

Petra Ceglarek

**DER WORTASSOZIATIONSVERSUCH
ALS WISSENSDIAGNOSTISCHES INSTRUMENT
IM ARBEITSPSYCHOLOGISCHEN KONTEXT
EINE BEFUNDINTEGRATION ZUR VERFAHRENSVALIDIERUNG**

Für meine Eltern

vorgelegt als Dissertationsschrift zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.) im Fach Psychologie
eingereicht bei der Humanwissenschaftlichen Fakultät der Universität Potsdam

Gutachter: Prof. Dr. Heinz-Jürgen Rothe (Universität Potsdam)
Prof. Dr. Niclas Schaper (Universität Paderborn)

Datum der Disputation: 25. Februar 2009

Online veröffentlicht auf dem
Publikationsserver der Universität Potsdam:
<http://opus.kobv.de/ubp/volltexte/2009/3049/>
<urn:nbn:de:kobv:517-opus-30499>
[<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-30499>]

TEIL A - EINLEITUNG	1
TEIL B - THEORIE	3
1. Der Wissensbegriff	3
1.1. Die Repräsentation von Wissen in semantischen Netzen	5
1.2. Deklaratives versus prozedurales Wissen	6
1.3. Implizites versus explizites Wissen	7
1.4. Allgemeinwissen versus Fachwissen	8
1.5. Wissen und Kompetenz	9
1.6. Auffassung des Wissensbegriffes in der vorliegenden Arbeit	10
2. Wissensdiagnosen	10
2.1. Wissensdiagnosen in der Arbeits- und Organisationspsychologie	11
2.1.1. Wissensdiagnosen zur Planung und Evaluation von Maßnahmen zur Sicherung des Unternehmenserfolges	12
2.1.2. Wissensdiagnosen im Kontext von Wissensmanagement	13
2.1.3. Wissensdiagnosen im Rahmen von Evaluationsmaßnahmen	13
2.1.4. Wissensdiagnosen zur Ermittlung von Ausbildungsinhalten	14
2.1.5. Wissensdiagnosen bei Personalentscheidungen	15
2.1.6. Wissensdiagnosen zur Entwicklung wissensbasierter Systeme	16
2.1.7. Wissensdiagnosen bei der Beantwortung von Forschungsfragen	17
2.1.8. Fazit für den Wortassoziationsversuch	18
2.2. Methoden zur Wissensdiagnose	18
2.2.1. Anforderungen an die Konstruktion wissensdiagnostischer Verfahren	19
2.2.2. Anforderungen an die Durchführung einer Wissensdiagnose	20
2.2.3. Beispiele für methodische Zugänge ohne Anforderungen an das Vorwissen des Untersuchers	22
2.2.3.1. Dokumentenanalysen	22
2.2.3.2. Methode des lauten Denkens	23
2.2.4. Beispiele für methodische Zugänge, welche Grundkenntnisse des Untersuchers erfordern	24
2.2.4.1. Befragungsverfahren	24
2.2.4.2. Lückentexte	25
2.2.5. Beispiele für methodische Zugänge, welche differenzierte Kenntnisse des Untersuchers erfordern	25
2.2.5.1. Assoziationsverfahren	25
2.2.5.2. Ratingverfahren	26
2.2.5.3. Strukturlege- und Mappingtechniken	26
2.2.5.4. Assoziations-Struktur-Test	28
2.2.6. Vergleich verschiedener methodischer Zugänge	28
3. Der Wortassoziationsversuch	30
3.1. Historische Einordnung von Assoziationsversuchen	31
3.2. Zielsetzung des Wortassoziationsversuchs in der Arbeitspsychologie	32
3.3. Theoretischer Hintergrund des Wortassoziationsversuchs	33
3.4. Konstruktion des Wortassoziationsversuchs	35
3.5. Durchführung des Wortassoziationsversuchs	36

3.6. Auswertung des Wortassoziationsversuchs	39
3.7. Gütekriterien	41
3.7.1. Hauptgütekriterien	41
3.7.1.1. Validität	42
3.7.1.2. Reliabilität	44
3.7.1.3. Objektivität	45
3.7.2. Nebengütekriterien	46
3.7.2.1. Ökonomie	47
3.7.2.2. Nützlichkeit	47
3.7.2.3. Normierung	48
3.7.2.4. Vergleichbarkeit	49
3.7.2.5. Sensitivität	49
3.8. Verhältnis zwischen Wortassoziationsversuch und Strukturlegetechnik	49
TEIL C - FRAGESTELLUNG	53
TEIL D - EMPIRISCHE BEFUNDE	57
4. Einführung in die empirischen Studien	57
4.1. Studie 1: Modus der Erhebung	61
4.1.1. Anliegen der Studie und Untersuchungsfeld	61
4.1.2. Untersuchungsaufbau	62
4.1.3. Wortassoziationsversuch	63
4.2. Studie 2: Psychotherapeutische Ambulanz	64
4.2.1. Anliegen der Studie und Untersuchungsfeld	64
4.2.2. Untersuchungsaufbau	65
4.2.3. Wortassoziationsversuch	67
4.3. Studie 3: Sparkassenakademie (A)	69
4.3.1. Anliegen der Studie und Untersuchungsfeld	69
4.3.2. Untersuchungsaufbau	69
4.3.3. Wortassoziationsversuch	70
4.4. Studie 4: Sparkassenakademie (B)	71
4.4.1. Anliegen der Studie und Untersuchungsfeld	71
4.4.2. Untersuchungsaufbau	71
4.4.3. Wortassoziationsversuch	72
4.5. Studie 5: Biopsychologie	73
4.5.1. Anliegen der Studie und Untersuchungsfeld	73
4.5.2. Untersuchungsaufbau	73
4.5.3. Wortassoziationsversuch	74
4.6. Studien 6 bis 9: Klausuren in verschiedenen Fächern der Psychologie	75
4.6.1. Anliegen der Studien und Untersuchungsfeld	75
4.6.2. Untersuchungsaufbau	76
4.6.3. Wortassoziationsversuch	77
4.7. Studie 10: Gebäudereiniger	79
4.7.1. Anliegen der Studie und Untersuchungsfeld	79
4.7.2. Untersuchungsaufbau	80

4.7.3. Wortassoziationsversuch	80
4.8. Studie 11: Bauklempner	81
4.8.1. Anliegen der Studie und Untersuchungsfeld	81
4.8.2. Untersuchungsaufbau	81
4.8.3. Wortassoziationsversuch	82
4.9. Studie 12: Handelsunternehmen	82
4.9.1. Anliegen der Studie und Untersuchungsfeld	82
4.9.2. Untersuchungsaufbau	83
4.9.3. Wortassoziationsversuch	85
4.10. Studie 13: Stoffdruckerei	86
4.10.1. Anliegen der Studie und Untersuchungsfeld	86
4.10.2. Untersuchungsaufbau	87
4.10.3. Wortassoziationsversuch	88
4.11. Studie 14: Seminar Wissensmanagement	89
4.11.1. Anliegen der Studie und Untersuchungsfeld	89
4.11.2. Untersuchungsaufbau	90
4.11.3. Wortassoziationsversuch	90
4.12. Studie 15: Wissen über Arbeitsschutz	91
4.12.1. Anliegen der Studie und Untersuchungsfeld	91
4.12.2. Untersuchungsaufbau	91
4.12.3. Wortassoziationsversuch	93
4.13. Studie 16: Call Center (A)	94
4.13.1. Anliegen der Studie und Untersuchungsfeld	94
4.13.2. Untersuchungsaufbau	95
4.13.3. Wortassoziationsversuch	95
4.14. Studie 17: Call Center (B)	96
4.14.1. Anliegen der Studie und Untersuchungsfeld	96
4.14.2. Untersuchungsaufbau	97
4.14.3. Wortassoziationsversuch	97
4.15. Zusammenfassung und erstes Fazit	99
5. Vorstudien zur Durchführung des Wortassoziationsversuchs	100
5.1. Durchführungsmodus des Wortassoziationsversuchs in Studie 1: Modus	101
5.2. Anordnung von Stimuli verschiedener Abstraktionsebenen in Studie 2: psychotherapeutische Ambulanz ..	103
5.3. Zusammenfassung der Befunde aus den Vorstudien zur Durchführung des Wortassoziationsversuchs	107
6. Hauptgütekriterien	108
6.1. Validität	110
6.1.1. Inhaltsvalidität	111
6.1.2. Kriteriumsvalidität	111
6.1.2.1. Klausurleistung als Validitätskriterium	112
6.1.2.1.1. Klausurleistung in Studie 3: Sparkassenakademie (A)	112
6.1.2.1.2. Klausurleistung in Studie 4: Sparkassenakademie (B)	113
6.1.2.1.3. Klausurleistung in Studie 5: Biopsychologie	114
6.1.2.1.4. Klausurleistung in Studien 6 bis 9: Klausuren in der Psychologie	115
6.1.2.1.5. Klausurleistung in Studie 10: Gebäudereiniger	117

6.1.2.1.6. Klausurleistung in Studie 11: Bauklempner	119
6.1.2.1.7. Zusammenfassung und Befundintegration der Studien zum Validitätskriterium Klausurleistung	120
6.1.2.2. Geschäftsführerurteil als Validitätskriterium.....	122
6.1.2.2.1. Geschäftsführerurteil in Studie 12: Handelsunternehmen.....	122
6.1.2.2.2. Geschäftsführerurteil in Studie 13: Stoffdruckerei.....	123
6.1.2.2.3. Zusammenfassung und Befundintegration der Studien zum Validitätskriterium Geschäftsführerurteil.....	124
6.1.3. Konstruktvalidität.....	125
6.1.3.1. Konvergente Konstruktvalidität.....	126
6.1.3.1.1. Lückentext als konstruktverwandter Test in Studie 2: psychotherapeutische Ambulanz.....	126
6.1.3.1.2. Tiefeninterview als konstruktverwandter Test in Studie 12: Handelsunternehmen	128
6.1.3.1.3. Strukturlegetechnik als konstruktverwandter Test	129
6.1.3.1.3.1. Strukturlegetechnik in Studie 5: Biopsychologie	131
6.1.3.1.3.2. Strukturlegetechnik in Studie 12: Handelsunternehmen	132
6.1.3.1.3.3. Strukturlegetechnik in Studie 13: Stoffdruckerei	135
6.1.3.1.3.4. Zusammenfassung und Befundintegration zur Vorhersage des Strukturlegens	137
6.1.3.1.4. Assoziations-Struktur-Test als konstruktverwandter Test.....	138
6.1.3.1.4.1. Assoziations-Struktur-Test in Studie 5: Biopsychologie.....	140
6.1.3.1.4.2. Assoziations-Struktur-Test in Studie 10: Gebäudereiniger	141
6.1.3.1.4.3. Assoziations-Struktur-Test in Studie 11: Bauklempner	143
6.1.3.1.4.4. Assoziations-Struktur-Test in Studie 14: Seminar Wissensmanagement.....	144
6.1.3.1.4.5. Zusammenfassung und Befundintegration zur Vorhersage des Assoziations-Struktur-Tests	146
6.1.3.1.5. Multiple-Choice-Tests als konstruktverwandter Test.....	147
6.1.3.1.5.1. Multiple-Choice-Test in Studie 14: Seminar Wissensmanagement	147
6.1.3.1.5.2. Multiple-Choice-Test in Studie 15: Wissen über Arbeitsschutz	148
6.1.3.1.5.3. Zusammenfassung und Befundintegration zur Vorhersage des Multiple-Choice-Tests	150
6.1.3.2. Diskriminante Konstruktvalidität	151
6.1.3.2.1. Wortflüssigkeit in Studie 1: Modus.....	152
6.1.3.2.2. Wortflüssigkeit in Studie 2: psychotherapeutische Ambulanz.....	153
6.1.3.2.3. Wortflüssigkeit in Studie 5: Biopsychologie.....	154
6.1.3.2.4. Zusammenfassung und Befundintegration der Studien zum Einfluß der Wortflüssigkeit	154
6.1.4. Modifikationen des Assoziationskennwertes	155
6.1.4.1. Gewichtung unterschiedlicher Qualitäten von Assoziationsinhalten.....	156
6.1.4.1.1. Gewichtung in Studie 13: Stoffdruckerei.....	156
6.1.4.1.2. Gewichtung in Studie 15: Wissen über Arbeitsschutz	158
6.1.4.1.3. Gewichtung in Studie 16: Call Center (A)	159
6.1.4.1.4. Gewichtung in Studie 17: Call Center (B)	161
6.1.4.1.5. Zusammenfassung und Befundintegration der Studien zur Gewichtung von Assoziationsinhalten	162
6.1.4.2. Assoziationskennwert auf Basis der hochfrequenten Assoziationen	164
6.1.4.2.1. Auszählung hochfrequenter Assoziationen in Studie 3: Sparkassenakademie (A)	164
6.1.4.2.2. Auszählung hochfrequenter Assoziationen in Studie 4: Sparkassenakademie (B).....	166

6.1.4.2.3. Auszählung hochfrequenter Assoziationen in Studie 12: Handelsunternehmen.....	167
6.1.4.2.4. Auszählung hochfrequenter Assoziationen in Studie 13: Stoffdruckerei	168
6.1.4.2.5. Auszählung hochfrequenter Assoziationen in Studie 14: Seminar Wissensmanagement.....	169
6.1.4.2.6. Zusammenfassung und Befundintegration der Studien zur Auszählung hochfrequenter Assoziationen	172
6.1.4.3. Beispiel für einen qualitativen Zugang bei der Auswertung von Assoziationsdaten	174
6.1.4.3.1. Einführung in die praxisrelevante Fragestellung von Studie 15: Wissen über Arbeitsschutz ...	175
6.1.4.3.2. Quantitativer Zugang in Studie 15: Wissens über Arbeitsschutz.....	175
6.1.4.3.3. Qualitativer Zugang in Studie 15: Wissens über Arbeitsschutz.....	177
6.1.4.3.4. Zusammenfassung der Befunde zur qualitativen Auswertung in Studie 15: Wissens über Arbeitsschutz.....	178
6.1.4.4. Zusammenfassung der Befunde zu den Modifikationen des Assoziationskennwertes	179
6.1.5. Zusammenfassung der Befunde zur Validität.....	179
6.2. Reliabilität	182
6.2.1. Retest-Reliabilität des Assoziationskennwertes in Studie 2: psychotherapeutische Ambulanz.....	183
6.2.2. Qualitative Aspekte der Testwiederholung in Studie 2: psychotherapeutische Ambulanz.....	184
6.2.3. Stimuli verschiedener Abstraktionsebenen bei Testwiederholung in Studie 2: psychotherapeutische Ambulanz	185
6.2.4. Zusammenfassung der Befunde zur Reliabilität	186
6.3. Objektivität.....	186
6.3.1. Durchführungsobjektivität.....	186
6.3.2. Auswerteobjektivität.....	187
6.3.3. Interpretationsobjektivität.....	188
6.3.4. Zusammenfassung der Befunde zur Objektivität.....	190
6.4. Zusammenfassung der Befunde zu den Hauptgütekriterien.....	190
7. Nebengütekriterien	192
7.1. Veränderungssensitivität	192
7.1.1. Veränderungssensitivität in Studie 3: Sparkassenakademie (A).....	193
7.1.2. Veränderungssensitivität in Studie 4: Sparkassenakademie (B).....	195
7.1.3. Veränderungssensitivität in Studie 14: Seminar Wissensmanagement.....	198
7.1.4. Zusammenfassung der Befunde zur Veränderungssensitivität	199
7.2. Akzeptanz.....	200
7.2.1. Akzeptanz in Studie 1: Modus.....	200
7.2.2. Akzeptanz in Studie 2: psychotherapeutische Ambulanz	202
7.2.3. Zusammenfassung der Befunde zur Akzeptanz.....	204
7.3. Ökonomie	205
7.3.1. Anmerkungen zur Vorbereitungsökonomie.....	205
7.3.2. Anmerkungen zur Durchführungsökonomie	206
7.3.2.1. Durchführungsökonomie in Studie 1: Modus	206
7.3.2.2. Durchführungsökonomie in Studie 2: psychotherapeutische Ambulanz	207
7.3.2.3. Durchführungsökonomie in Studie 13: Stoffdruckerei	208
7.3.3. Anmerkungen zur Auswerteökonomie	208
7.3.4. Zusammenfassung der Befunde zur Ökonomie	209
7.4. Nützlichkeit	211

7.5. Normierung.....	211
7.6. Vergleichbarkeit.....	212
7.7. Zusammenfassung der Befunde zu den Nebengütekriterien	213
TEIL E - ZUSAMMENFASSUNG UND DISKUSSION DER ERGEBNISSE.....	215
8. Methodische Aspekte des Wortassoziationsversuchs.....	215
8.1. Zusammenfassung der empirischen Befunde.....	215
8.2. Diskussion der Befunde	216
9. Gütekriterien des Wortassoziationsversuchs.....	217
9.1. Zusammenfassung der empirischen Befunde.....	218
9.2. Diskussion der Befunde	222
9.2.1. Anforderungen an die Testgüte	226
9.2.2. Was kann der Wortassoziationsversuch leisten, was kann er nicht leisten?.....	228
10. Ausblick und offene Fragen.....	231
11. Fazit der vorliegenden Arbeit	234
TEIL F - LITERATURVERZEICHNIS	237
TEIL G - DANKSAGUNG	249
TEIL H - ANHANG	251

Teil A - EINLEITUNG

Die vorliegende Arbeit möchte einen Beitrag leisten für die arbeitspsychologische Praxis in Felduntersuchungen, sie trägt den Titel „Der Wortassoziationsversuch als wissensdiagnostisches Instrument im arbeitspsychologischen Kontext. Eine Befundintegration zur Verfahrensvalidierung“. Im Folgenden sollen die wesentlichen Aspekte dieser Arbeit, welche sich auch aus ihrem Titel ableiten lassen, kurz skizziert werden und dem Leser einen Ausblick über die zu erwartenden Inhalte geben.

Der Wortassoziationsversuch

Der Wortassoziationsversuch ist ein relativ einfach einzusetzendes Verfahren: auf einen Stimulus hin sollen Probanden stichwortartig alles nennen, was ihnen dazu einfällt. Die dabei bestehende Grundannahme ist, daß all diejenigen Inhalte assoziiert werden, welche im Individualgedächtnis mit diesem Stimulusbegriff verknüpft sind. Je mehr von einer Person assoziiert wird, desto größer ist – so die Annahme – ihr Wissensbesitz in der jeweiligen Domäne, da viele Verknüpfungen i.S. eines semantischen Netzwerkes aktiviert und benannt wurden.

Als wissensdiagnostisches Instrument

Die Assoziationstechnik hat eine lange Tradition in der Psychologie, angefangen bei Francis Galton oder Hermann Ebbinghaus im Kontext von Lernprozessen. Die Einsatzmöglichkeiten dieser Technik sind vielfältig, neben den konkreten Durchführungsbedingungen und Einsatzfeldern machen sich die Unterschiede aber v.a. in den sehr verschiedenen Zielrichtungen fest. In Abgrenzung zu solchen Zielstellungen wie dem Verständnis von Lernmechanismen oder dem Erfassen von Wünschen, Motivationen oder Einstellungen beschäftigt sich die vorliegende Arbeit ausschließlich mit der Erfassung von Wissen. Dabei sollen diagnostische Aussagen getroffen werden über den Wissensbesitz einer Person oder einer Gruppe von Personen in einer umschriebenen Wissensdomäne, wobei hier ausdrücklich auf das explizite, verbalisierbare Wissen zurückgegriffen wird. In Vorwegnahme des folgenden Aspektes kann der Wissensgegenstand weiter eingegrenzt werden auf das explizite Fachwissen in einer umschriebenen Wissensdomäne.

Im arbeitspsychologischen Kontext

Wie bereits erwähnt liegt der Fokus dieser Arbeit auf der arbeitspsychologischen Praxis. In diesen Untersuchungsfeldern besteht Bedarf, das Fachwissen von Personen zu diagnostizieren, z.B. zur Evaluation von Weiterbildungsveranstaltungen (zum Nachweis von Wissenszuwachs) oder im Rahmen von Wissensmanagementprozessen (zur Erfassung des Ist-Wissens vor der Wissensmanagementmaßnahme und anschließend zur Erfolgskontrolle). Das betriebliche Umfeld stellt dabei besondere Bedingungen dar: arbeitspsychologische Untersuchungen haben sich gewissermaßen dem Betriebsablauf unterzuordnen oder sollten ihn zumindest so wenig wie möglich stören, hier ist aus betriebswirtschaftlicher Sicht v.a. der zeitliche Aufwand für die Durchführung der Wissensdiagnose bedeutsam (siehe z.B. Amelang und Schmidt-Atzert, 2006). Auch die besondere Situation der Probanden ist zu berücksichtigen, insbesondere ihre Rolle als

Arbeitspersonen und ihre Einstellung gegenüber der Untersuchung. Dieser Zielstellung des Praxiseinsatzes wurde in der vorliegenden Arbeit dadurch Rechnung getragen, daß die empirischen Daten überwiegend im Feld erhoben worden sind, so wie es auch Oesterreich (2005) für dringend erforderlich hält.

Eine Befundintegration

Die betriebliche Praxis bietet dem Arbeitspsychologen eine unüberschaubare Vielzahl an verschiedenartigen Einsatzfeldern, die in Hinblick auf viele Aspekte variieren können, z.B. hinsichtlich der Branche, Unternehmensform, Arbeitsgegenstände oder Arbeitsprozesse. Der Rückschluß von empirischen Ergebnissen nur einer einzigen Studie auf die Gesamtheit der arbeitspsychologischen Felder wäre äußerst bedenklich. In den letzten Jahren wurden an der Universität Potsdam und auch an der Humboldt-Universität zu Berlin viele Studien durchgeführt, welche in verschiedenen Arbeitsfeldern den Wortassoziationsversuch als Instrument zur Diagnose von Fachwissen eingesetzt haben. Diese Daten sollen nun in der vorliegenden Arbeit zusammengeführt werden, so daß i.S. der Generalisierbarkeit ein höherer Sicherheitsgrad bei der Ableitung von Erkenntnissen auf künftige Einsatzfelder erreicht werden kann.

Zur Verfahrensvalidierung

Wie bereits erwähnt hat die Assoziationstechnik eine lange Geschichte, gleichsam liegen aber kaum Erkenntnisse zur Güte des Wortassoziationsversuchs als Instrument zur Wissensdiagnose vor (Rothe, 2003). Der zentrale Anspruch der vorliegenden Arbeit ist daher, diesem Mangel empirische Belege entgegenzustellen. Wie soeben beschrieben soll dabei auf mehrere, verschiedene Studien zurückgegriffen werden, um die Haupt- und Nebengütekriterien bestimmen zu können. Das Verhältnis von Kosten (welchen Aufwand erfordern Vorbereitung, Durchführung, Auswertung des Verfahrens) und Nutzen (wie sicher können welche Aussagen über den Wissensbesitz von Personen getroffen werden) beim Einsatz des Wortassoziationsversuchs ist insbesondere für den Einsatz im arbeitspsychologischen Feld bedeutsam.

Teil B - THEORIE

Dieses Kapitel widmet sich der einschlägigen Literatur, welche im Zusammenhang mit dem Wortassoziationsversuch herangezogen werden kann. Zunächst wird der Begriff des Wissens aufgegriffen sowie seine für die vorliegende Arbeit relevanten Facetten, anschließend werden die Diagnose von Wissen thematisiert und deren spezielle Anwendungsgebiete in der Arbeits- und Organisationspsychologie dargelegt. Im weiteren werden beispielhaft verschiedene methodische Zugänge der Wissensdiagnose vorgestellt, in die auch der Wortassoziationsversuch eingeordnet wird. Dieses Kapitel schließt ab mit ausführlichen Darstellungen der bisherigen Erkenntnisse und Befunde zum Wortassoziationsversuch: zur Historie dieser Technik, zum theoretischen Hintergrund des Assoziierens, zum praktischen Einsatz sowie schließlich zu den Gütekriterien.

1. Der Wissensbegriff

Es besteht weitgehend Einigkeit darüber, daß Wissen das im Langzeitgedächtnis eines Individuums Gespeicherte repräsentiert (z.B. Kluwe, 1988; Rothe und Timpe, 1997; Strube und Schlieder, 1998, Rothe, 2006). Tergan (1989a) definiert Wissen als die Gesamtheit aller in irgendeiner Form mental repräsentierter Informationen eines Individuums, welche im Kontext von Aufgabenbearbeitungen zum Tragen kommen können. Hacker (1999) betont allerdings, daß Informationen per se nicht mit Wissen gleichzusetzen sind. Vielmehr repräsentiere nur derjenige Teil der Informationsflut, welcher im Kopf des Einzelnen gespeichert und abrufbar ist, das individuelle Wissen. Nach Rothe (2006) existiert das Wissen nicht an sich, sondern stellt eine intellektuelle Disposition dar, welche zum Lösen von Aufgaben und Problemen befähigt, und ist damit eine wesentliche Grundlage für eine bedingungsangepaßte Verhaltensregulation des Individuums in seiner Umgebung, wobei der Wahrheitsgehalt dieses Wissens keine Rolle spielt (Strube und Schlieder, 1998). Nach Opwis und Luer (1996) ist Wissen derjenige Informationsbestand, der von den informationsverarbeitenden Systemen aufgenommen, gespeichert und abgerufen werden kann, hinsichtlich der Inhalte definieren sie Wissen als die Gesamtheit der subjektiven Überzeugungen, individuellen Erfahrungen und Fertigkeiten. Klix (1996) betont dabei die Strukturiertheit und Geordnetheit im Aufbau der Wissens Elemente sowie in deren Beziehungen untereinander. Das Langzeitgedächtnis besteht aus den Grundkomponenten der Begriffe sowie den zwischen diesen bestehenden Beziehungen, den sog. semantischen Relationen (van der Meer, 1984). Hoffmann, Ziebler und Grosser (1984) definieren Begriffe als lernabhängig gebildete Klassen von Objekten und Erscheinungen in unserer Umwelt, welche mit dem Ziel der Reduktion des Informationsverarbeitungsaufwandes gebildet werden. Klix (1992) unterscheidet folgende Quellen des individuellen Wissens: das Erbgut (angeborene Wahrnehmung durch Sinnesorgane), die individuelle Erfahrung (wiederholte Handlungsausführung), die sprachliche Belehrung durch andere (Beschreibung von Objekten, Ereignissen usw.) sowie das eigene Denken, welches durch Operieren mit bereits gespeicherten Wissens Elementen und entsprechenden Schlußfolgerungen neues Wissen erzeugt.

Die Klassifikationen, die herangezogen werden, um das individuelle Wissen von Personen zu beschreiben und zu differenzieren, sind vielfältig. Opwis und Luer (1996) weisen darauf hin, daß je

nach Autor unterschiedliche Klassifikationsaspekte zugrunde gelegt wurden und es dabei dennoch große Abgrenzungsprobleme zwischen den einzelnen Taxonomien gibt. Die verschiedenen Taxonomien bewegen sich auf unterschiedlichen Dimensionen, so orientiert sich beispielsweise die Unterscheidung prozedural vs. deklarativ an den formal-abstrakten Inhalten der Wissenseinheiten, während die Ebene implizit vs. explizit die Verbalisierbarkeit des Wissens beschreibt. Die Klassifikation nach Strube und Schlieder (1998) berücksichtigt folgende verschiedene Aspekte: deklarativ vs. prozedural, allgemein vs. bereichsspezifisch, generell (semantisch) vs. fallspezifisch (episodisch) und schließlich hinsichtlich des Präzisionsgrades des Wissens die Unterscheidung nach quantitativ vs. qualitativ. Opwis und Luer (1996) führen eine ganze Reihe von verschiedenen Taxonomien von Wissen an. So lassen sich deklarativ vs. prozedural unterscheiden, epistemisch vs. heuristisch, episodisch vs. semantisch, sprachlich-begrifflich vs. praktisch-handelnd vs. bildhaft-anschaulich, Datenwissen (Was) vs. Methodenwissen (Wie) vs. Metawissen (Wann), Wissen über Sachverhalte, Fakten, Situationen vs. über die Verwendung von Wissen vs. Handlungswissen.

Eine der zentralen Wissensklassifikationen, die Mandl (1997) als die einflußreichste bewertet, betrifft die Unterscheidung in deklaratives vs. prozedurales Wissen. Strube und Schlieder (1998) definieren Wissen als diejenigen Inhalte des Langzeitgedächtnisses, die einen faktischen Charakter besitzen (deklarativ) oder handlungsrelevant sind (prozedural). Beatty und Gerace (2002) bezeichnen das prozedurale auch synonym als stilles (tacit) Wissen. Auch Knopf, Mack und Kressley-Mba (2005) verwenden in ihrer Taxonomie von Gedächtnisformen die Kategorien deklarativ vs. nicht-deklarativ synonym mit der Einteilung explizit vs. implizit. Danach läßt sich der deklarative (explizite) Teil des Gedächtnisses weiter differenzieren in semantische und episodische Inhalte, während das nicht-deklarative (implizite) Gedächtnis sich zusammensetzt aus motorischen Fertigkeiten und Gewohnheiten, Priming, klassischer Konditionierung sowie dem nichtassoziativen Lernen.

Rothe (2006) schlägt zur Charakterisierung von Wissen folgende Klassifizierung vor: nach formalen Merkmalen des Wissens, nach strukturellen Merkmalen sowie nach semantischen Merkmalen. In der erstgenannten Klasse lassen sich z.B. vages vs. exaktes Wissen voneinander abgrenzen, wobei es um die Frage geht, wie vollständig Angaben über die Merkmale von Objekten, Ereignisse, Sachverhalten bzw. deren Beziehungen untereinander sind. Rothe (2006) weist allerdings darauf hin, daß der Untersucher für diese Art der Beschreibung von individuellem Wissen selbst über Kenntnis darüber verfügen muß, welche Wissenskomponenten notwendigerweise vorhanden sein müssen, damit der gesamte Wissenskörper als vollständig zu bezeichnen ist. Der andere formale Aspekt, welchen Rothe (2006) als Ordnungskriterium heranzieht, betrifft die Verbalisierbarkeit von Wissen. Danach kann individuelles Wissen entweder verbalisierbar sein vs. nur unter bestimmten Bedingungen verbalisierbar vs. nur schwer oder kaum verbalisierbar (sog. „schweigendes“ Wissen). In der zweiten von Rothe (2006) benannten Klasse, in der Wissen nach strukturellen Merkmalen beschrieben wird, finden sich die Kategorien stationäres vs. deklaratives vs. prozedurales Wissen, welche auf Klix (1992) zurückgehen. Unter stationärem Wissen werden dabei Objekt- und Ereignisbegriffe verstanden sowie relationale Begriffe mit spezifischen Bedeutungen, Inhalten, Strukturen, außerdem Ereignisfolgen aufgrund konditionaler, kausaler oder konsekutiver Zusammenhänge. Rothes (2006) Definition des deklarativen Wissens ist nicht weit davon entfernt: es bezeichnet Wissen über Begriffe, Zustände, Ereignisse, die

gesetzmäßigen Zusammenhänge zwischen Zuständen, ihre Bedingungen sowie deren Relevanz. Prozedurales Wissen hingegen zeichnet sich aus durch das Wissen über Techniken und Methoden zur Lösung von Aufgaben. Rothe und Timpe (1997) sowie Rothe (2006) differenzieren das prozedurale Wissen dabei noch weiter nach problembereichsspezifischen Handlungssequenzen vs. Wissen über allgemeine Lösungsstrategien (den sog. Metaprozeduren). Die dritte Klasse, mit der Rothe (2006) Wissen anhand seiner semantischen Merkmale beschreibt, ist bewußt offen gehalten, da die Inhalte von Wissen sehr verschieden sein können und entsprechend beliebig viele Differenzierungen nach den unterschiedlichen Sachgebieten möglich sind.

Im Folgenden soll kurz auf die Modellvorstellung eines semantischen Netzwerkes eingegangen werden, da diese für die vorliegende Arbeit die zentrale theoretische Annahme darstellt. Anschließend werden diejenigen Klassifikationen, die im Zusammenhang mit dem Wortassoziationsversuch bedeutsam sind, diskutiert: gemäß der strukturellen Merkmale deklaratives vs. prozedurales Wissen, hinsichtlich der Verbalisierbarkeit das implizite vs. explizite Wissen sowie bzgl. der semantischen Inhalte das Allgemeinwissen vs. fach- oder arbeitsbezogenes Wissen. Zudem wird die Abgrenzung zum Konstrukt der Kompetenz vorgenommen.

1.1. Die Repräsentation von Wissen in semantischen Netzen

Mentale Modelle von Wissen sind interne Repräsentationen im individuellen Gedächtnis und somit nach Wender (1988) ein nicht-beobachtbares hypothetisches Konstrukt. Kluwe (1988) ordnet das mentale Modell als eine spezifische Variante von Vorwissen ein und beschreibt es als ein stabiles Gefüge von Wissen, welches eine Person hinsichtlich eines ganz konkreten Realitätsausschnittes ausgebildet hat und auf diesen Ausschnitt anwendet; nach Tergan (1989a) beziehen mentale Repräsentationen sowohl dem Bewußtsein zugängliche als auch unzugängliche Wissens Elemente mit ein, ebenso dauerhaft repräsentierte wie auch flüchtige Informationen. Formale Beschreibungen solcher kognitiven Modelle können z.B. semantische Netzwerke sein (Tergan, 1989a), welche der Enkodierung konzeptuellen Wissens dienen (Anderson, 2001), Wender (1988) beschreibt sie als sprachlich-abstrakte Form der Repräsentation. Semantik definieren Strube und Schlieder (1998) als die Zuweisung von logischen Bedeutungen (Wahrheitswerten) zu Sprachzeichen.

Netzwerktheorien vereinigen die Vorstellung, daß Begriffe oder Wissens Elemente im Individualgedächtnis netzwerkartig miteinander verbunden sind (Schermer, 1991). Genauer gesagt sind semantische Netze graphische Modellannahmen, in denen die Knoten Begriffe oder sprachliche Einheiten repräsentieren (z.B. Personen, Objekte, Ereignisse) und die sog. Kanten zwischen zwei Knoten semantische Relationen darstellen, wobei beliebig viele Pfade zwischen den Knoten konstruiert werden können (Wender, 1988; Strube und Schlieder, 1998). Nach Klix (1991, 1996) oder auch Oeser (1996) stellen Begriffe die zentralen Einheiten menschlichen Wissensbesitzes dar, semantische Relationen definiert Klix (1991, 1996) als Wissenskomponenten, welche die Zusammenhänge zwischen Wissens Elementen anzeigen oder sie herstellen. Somit steht im Zentrum eines Netzwerkes die Struktur zwischen den Begriffen (Bredenkamp und Erdfelder, 1996). Nach Klix (1996) weist das menschliche Gedächtnis qualitativ verschiedenartige Vernetzungen auf: stabile oder aber zeitweilige (sog. flüchtige) Verbindungen. Dabei bilden gerade diese zeitlich variablen, dynamischen Strukturen die Voraussetzung für das Denken, sind aber damit nicht in die Netzwerkmetapher integrierbar (Klix, 1996).

Opwis und Luer (1996) weisen darauf hin, daß semantische Netzwerke eher ein Sammelbegriff sind, unter dem verschiedene Modellvorstellungen subsumiert werden können, gemein ist allen jedoch die Modellierung sprachlicher Gedächtnisinhalte als Repräsentation der Bedeutung von sprachlich-begrifflichen Strukturen im menschlichen Gedächtnis. Nach Schermer (1991) stellen semantische Netzwerke den Versuch dar, Gedächtnisstrukturen über die rein quantitative Charakterisierung hinaus sowohl qualitativ als auch formal zu beschreiben und daraus Schlußfolgerungen über Gedächtnisprozesse ziehen zu können. Opwis und Luer (1996) kritisieren eine nur geringe Ausdrucksfähigkeit der semantischen Netze, da damit kaum komplexe strukturierte Wissensbausteine abgebildet werden können, auf der anderen Seite aber verweisen sie aufgrund der enormen Flexibilität semantischer Netze auf die Unmöglichkeit, sie experimentell zu überprüfen.

Nach Hacker (1992, 2005b) ist das Faktenwissen auch im Arbeitskontext am ehesten durch Netzwerke beschreibbar, auch die Arbeitsprozesse selbst lassen sich durch Relationen abbilden, denn das Wissen um eine Kette von Signalen, Zielen, Ursachen und Maßnahmen ist Voraussetzung für eine effektive Bewältigung der Arbeitsaufgaben. Die Darstellung prozeduralen Wissens hingegen bewerten Hacker (2005b) wie auch Opwis und Luer (1996) als kaum praktikabel darstellbar. Somit sind semantische Netze in erster Linie zur Darstellung deklarativer Wissenshalte geeignet. Auf weitere theoretische Modellannahmen wird im Zusammenhang mit dem theoretischen Hintergrund des Wortassoziationsversuchs eingegangen werden (Kapitel 3.3, Seite 33 ff.).

1.2. Deklaratives versus prozedurales Wissen

Eine der zentralen Wissensklassifikationen betrifft die Unterscheidung in deklaratives vs. prozedurales Wissen, Oswald und Gadenne (1984) und auch Mandl (1997) halten diese Unterscheidung für zentral zur Erklärung menschlicher Informationsverarbeitung. Unter deklarativ wird das im Individualgedächtnis gespeicherte Faktenwissen verstanden (sog. knowing that), welches eine Person sich bewußt machen und verbalisieren kann (z.B. Oswald und Gadenne, 1984; Anderson, 2001; Beatty und Gerace, 2002), während das prozedurale Wissen (sog. knowing how) diejenigen kognitiven Mechanismen beschreibt, welche eine Person in die Lage versetzen, komplexe motorische und kognitive Handlungen durchzuführen. Dieses prozedurale Wissen kann zwar in Handlung umgesetzt, aber weder ins Bewußtsein gerufen noch verbalisiert werden (z.B. Oswald und Gadenne, 1984; Anderson, 2001; Beatty und Gerace, 2002). Knopf, Mack und Kressley-Mba (2005) verwenden die Kategorie des deklarativen Wissens synonym mit explizitem Wissen, dabei verstehen sie unter deklarativ das verbalisierbare Wissen über allgemeine Fakten, Ereignisse und persönliche Erlebnisse.

Rothe (1994a) charakterisiert diese beiden Klassen (deklarativ vs. prozedural) anhand von drei phänomenologischen Aspekten: anhand der Qualität, bzgl. des Erwerbs und bzgl. der Verbalisierbarkeit des Wissens. So ist auf der Ebene der Qualität des Wissens festzuhalten, daß deklaratives Wissen die Kenntnis von Sachverhalten beschreibt, wobei dieses Wissen (nur) richtig oder falsch oder gar nicht gewußt werden kann. Prozedurales Wissen hingegen umfaßt die Fähigkeit, bestimmte Handlungen auszuführen, wobei viele graduelle Abstufungen möglich sind zwischen Gar-nicht-wissen und Perfekt-können. Die zweite Betrachtungsweise betrifft den Erwerb des Wissens: während deklaratives Wissen durch Lesen, Beobachtung oder Edukation erworben wurde, bedarf prozedurales Wissen neben den Sinneswahrnehmungen der wiederholten

Handlungsausführung (Rothe, 1994a). Hacker (1992) macht allerdings darauf aufmerksam, daß nicht alles Handlungswissen auf Erfahrung beruht, denn Wissen reguliere – auch im Arbeitsprozeß – das Handeln und entstehe gleichsam in ihm. Als letzte Ebene ist die Verbalisierbarkeit des Wissens zu betrachten. Danach besteht zumindest die Annahme, daß deklaratives Wissen weitestgehend verbalisierbar ist, während das prozedurale Wissen kaum oder gar nicht beschrieben werden kann. Rothe (1994a) und auch Mazur (2006) warnen allerdings davor, von der Nicht-Verbalisierbarkeit auf ein Nicht-Wissen zu schließen. Nach Strube und Schlieder (1998) sowie Rothe (1994a) läßt sich das prozedurale Wissen durch Regeln oder Produktionssysteme erfassen, während das deklarative Wissen durch semantische Netze darstellbar ist.

Cranach und Bangerter (2000) hingegen sehen lediglich die Dimensionen des Inhalts sowie des Bewußtseins als Unterscheidungskriterium, wobei letzteres vernachlässigbar sei. Statt dessen führen Cranach und Bangerter (2000) als dritte Kategorie neben prozedural und deklarativ das episodische Wissen ein, welches Wissen über Ereignisse sowie vergangene Handlungen und ihre Folgen einschließt und in unterschiedlichem Ausmaß bewußt repräsentiert sein kann. Mandl (1997) weist darauf hin, daß das prozedurale Wissen (Regeln, nach denen Datenstrukturen manipuliert werden) dem deklarativen Wissen (Datenstrukturen, die nach Regeln manipuliert werden) nachgeordnet sei, da es dazu diene, das deklarative Wissen zum Laufen zu bringen. Auch Cianciolo, Matthew, Sternberg und Wagner (2006) sehen das prozedurale Wissen als eine Art Endstadium eines Lernprozesses, bei dem bestimmte Aufgaben (z.B. psychomotorische Fertigkeiten) zunächst auf deklarativer Ebene erarbeitet und dann durch Übung automatisiert werden.

1.3. Implizites versus explizites Wissen

Die Frage nach der Verbalisierbarkeit individuellen Wissensbesitzes schlägt sich in der Unterscheidung explizit vs. implizit nieder. Explizites Wissen umfaßt das bewußt verbalisierbare Wissen einer Person, während das implizite Wissen sich durch seine Nicht-Verbalisierbarkeit auszeichnet, aber in der Handlungsausführung sichtbar wird (Anderson, 2001). Nach Büssing, Herbig und Latzel (2004) enthält das implizite genauso wie das explizite Wissen sowohl deklarative wie auch prozedurale Anteile. Die Unterschiede liegen dagegen im Erwerb des Wissens und dessen Reflexion: während implizites Wissen durch konkrete Erfahrungen erworben und weiterentwickelt wird, was unabhängig von Aufmerksamkeit oder Bewußtsein geschehen kann und die Wissensinhalte dabei nicht reflektiert werden (Mazur, 2006, charakterisiert den Prozeß des impliziten Lernens als beiläufig und unaufgefordert), wird explizites Wissen bewußt gelernt und kann reflektiert werden, seine Aneignung ist aber nicht auf eine konkrete Erfahrung mit dem Gegenstand angewiesen. Hacker (2005b) betont, daß der Erwerb expliziten Wissens damit auch außerhalb des Arbeitsprozesses möglich ist, eine Lernabsicht sei aber dennoch Voraussetzung. Implizites Wissen hingegen bedarf nicht dieser Lernabsicht und ist v.a. an den Kontext der Arbeitssituation gebunden (Hacker, 2005b). Im Zusammenhang mit der Bewußtseinsfähigkeit des Wissens macht Hacker (2005b) auch darauf aufmerksam, daß explizites Wissen aufgrund der reflexiven Auseinandersetzung jederzeit prüfbar und ggf. korrigierbar ist, während implizites Wissen zunächst eine Phase der Explikation durchlaufen muß, um geprüft und modifiziert werden zu können.

Büssing, Herbig und Ewert (2001) betonen die gegenseitige Beeinflussung und Ergänzung von implizitem und explizitem Wissen, da diese Unterschiedliches leisten. Anderson (2001) dagegen setzt explizites Wissen mit deklarativem gleich. Die Nichtverbalisierbarkeit wird zumeist dem Handlungswissen unterstellt (Anderson, 2001; Rothe, 2006). Die Nicht-Bewußtseinsfähigkeit des impliziten Wissens ist gemäß Büssing, Herbig und Latzel (2004) umstritten, es könne zwar teilweise verbalisiert werden, sei aber nicht vollständig replizierbar, während das explizite Wissen umfassend verbalisierbar ist (Hacker, 2005b).

1.4. Allgemeinwissen versus Fachwissen

Rothe und Schindler (1996) sowie Rothe (2006) unterscheiden hinsichtlich der semantischen Merkmale von Wissen zwischen Allgemeinwissen vs. spezifischem Wissen. Dabei wird unter Allgemeinwissen all jenes Wissen verstanden, welches sich ein Individuum im Laufe seiner Entwicklung aneignet durch eigene Erfahrungen sowie schulische Belehrungen, während spezifisches Wissen dasjenige Wissen beschreibt, welches im Rahmen beruflicher Ausbildung, Studium oder der eigenen beruflichen Tätigkeit erworben wurde. Da die vorliegende Arbeit ausschließlich das berufsbezogene Wissen thematisiert, soll im Folgenden auch nur dieses weiter erörtert werden.

Rothe und Schindler (1996) sowie Rothe und Timpe (1997) fassen die von Hacker (1992; 1993; 2005b) beschriebenen Aspekte arbeitsbezogenen Wissens wie folgt zu drei Kategorien zusammen: zum einen Wissen über Situationsmerkmale, welche Arbeitshandlungen auslösen, Wissen über Ursachen für zu verändernde Gegebenheiten, über Eigenschaften und Gesetzmäßigkeiten von technologischen Prozessen, Materialien oder Werkzeugen, Hacker (1993) betont in diesem Zusammenhang das sog. Zielwissen als Wissen über die Sollbeschaffenheit des Arbeitsgegenstandes. Nach Rothe und Timpe (1997) kann dieses Faktenwissen größtenteils durch Belehrung und nur in geringem Maße durch die eigene Erfahrung erworben werden. Die zweite Kategorie betrifft das Wissen über Arbeitsverfahren und Techniken sowie Wissen über die Ziele und Folgen von Arbeitshandlungen sowie deren Nutzen und Nebenwirkungen. Der Erwerb dieses Handlungswissens kann zwar durch sprachliche Belehrung unterstützt werden, primär ist aber die wiederholte eigene Handlungsausführung notwendig. Da gemäß Bergmann (2001) aufgrund der Schnellebigkeit des Wissens nicht mehr alles im Beruf notwendige Wissen durch die Bildungssysteme bereitgestellt werden kann, wird der Wissenserwerbsprozeß immer mehr an den Arbeitsplatz verlagert. Berufsspezifisches Wissen ist also zunehmend auch bereits in der Phase des Erwerbs eng an den Kontext der Arbeitstätigkeit und des Arbeitsplatzes gebunden. Die dritte Kategorie umfaßt schließlich Metaprozeduren, welche das Organisieren und Planen von Arbeitshandlungen betreffen. Rothe (1995) sowie Rothe und Timpe (1997) weisen darauf hin, daß zwar einerseits das durch sprachliche Belehrung erworbene Wissen bei wiederholtem Einsatz seine Bewußtseinspflichtigkeit verliert, auf der anderen Seite aber fast alle Arbeitstätigkeiten auf die Beteiligung bewußter Prozesse der Wissensverarbeitung angewiesen sind. Nach Cianciolo, Matthew, Sternberg und Wagner (2006) ist das im Arbeitsprozeß benötigte Wissen einerseits überwiegend als deklaratives Wissen konzipiert und üblicherweise explizit vermittelt, auf der anderen Seite erfordert das arbeitsbezogene Wissen mehr als nur die Implementierung von Fakten,

sondern umfaßt auch das sog. schweigende Wissen, welches während des Arbeitsprozesses erworben, aber nicht explizit vermittelt wurde.

Pleiss und Oesterreich (2003) machen eine interessante Unterscheidung zwischen aufgabenbezogenem Wissen, produktbezogenem Wissen und Fachwissen. Danach ist das aufgabenbezogene Wissen unmittelbar handlungsbezogen und umfaßt Kenntnisse über Bearbeitungswege und deren Regeln, während das produktbezogene Wissen Kenntnisse über das jeweilige Produkt oder die Dienstleistung beschreibt (z.B. Produktmerkmale, Zielgruppen, Marktstellung), an welcher der Beschäftigte oder seine Arbeitsgruppe arbeitet. In Abgrenzung dazu bezieht sich das Fachwissen nach Pleiss und Oesterreich (2003) auf den Inhalt der Arbeitstätigkeit bzw. im Zusammenhang mit den Arbeitsmitteln und wurde im Rahmen der Ausbildung erworben, damit gehen diese Kenntnisse über die konkrete Arbeitsaufgabe hinaus und behalten auch dann Gültigkeit, wenn der Beschäftigte beispielsweise das Unternehmen verläßt. So ist zwar der Bezug zur konkreten Arbeitstätigkeit geringer als z.B. beim produktbezogenen Wissen, Pleiss und Oesterreich (2003) bewerten das Fachwissen aber als organisationsunabhängig und transferierbar, da es nicht an unternehmensspezifische Ausführungsbedingungen gebunden ist. In der vorliegenden Arbeit sollen die Begriffe berufsspezifisch, arbeitsbezogen oder fachspezifisch synonym verwendet werden, da sie alle die Spezifität einer definierten Wissensdomäne sowie ihren Bezug zur Arbeitstätigkeit beschreiben und sich damit vom Allgemeinwissen abgrenzen.

1.5. Wissen und Kompetenz

Der Kompetenzbegriff ist nicht nur alltagssprachlich, sondern innerhalb der psychologisch-pädagogischen Forschungen unscharf definiert, nach Lang-von Wins (2003) haben sich dabei v.a. die Aspekte Wissen, Fertigkeit und Fähigkeit herausgeschält, welche als Leistungsvoraussetzungen für Berufe verstanden werden. Weinert (2001) sieht Kompetenz als ein Wechselspiel zwischen Wissen, Überzeugungen und Handlungstendenzen, Rothe (2003, 2006) versteht unter Kompetenz die Befähigung eines Individuums zur Selbstorganisation seines Handelns, Ley und Albert (2003) definieren Kompetenz als die zugrundeliegenden Fähigkeiten einer Person, welche in konkreten Arbeitssituationen verhaltenswirksam werden, was sowohl deklaratives Wissen als auch prozedurale Fertigkeiten einschließt. Um dabei das Ziel, innerhalb eines spezifischen Tätigkeitsbereichs Aufgaben oder Probleme effizient zu lösen, erreichen zu können, ist das fachspezifische Fakten- und Handlungswissen dazu eine notwendige, aber keine hinreichende Bedingung (Rothe, 2003, 2006; Cranach und Bangerter, 2000). Schaper (2003) weist auf die Tätigkeitsgebundenheit beruflicher Kompetenzen hin, Bergmann (2000) benennt als deren Komponenten neben der Motivation zum Kompetenzerwerb und emotionaler Faktoren v.a. das differenzierte (und explizit vermittelbare) Faktenwissen, implizites Wissen über die Arbeitstätigkeiten sowie die Prozeduralisierung, welche sie als eine Feinabstimmung des Wissens im Handeln beschreibt. Somit sollte die Auseinandersetzung mit Kompetenz im betrieblichen Kontext auch immer eng verbunden sein mit der Erfassung des dazu notwendigen berufsspezifischen Wissens, ein gelungenes Beispiel dafür ist die Studie von Schaper (2004), der zum Nachweis des Kompetenzerwerbs durch arbeitsplatzintegrierte Lernorte u.a. die arbeitsbezogenen beruflichen Kompetenzen anhand von Arbeitsproben, situativen Fragen sowie Wissensfragen erhob.

Rothe (2006) merkt noch weiteren Aufklärungsbedarf an hinsichtlich des Zusammenhangs zwischen fachspezifischem Wissen und Arbeitsleistung (z.B. welche fördernden vs. hemmenden Faktoren die Umsetzung von Wissen in kompetentes Handeln determinieren), Mandl (1997) spricht in diesem Zusammenhang von einer Kluft zwischen Wissen und Handeln. Hacker (1999, 2005b) erinnert daran, daß nur ein Teil des individuellen Wissens handlungswirksam wird, und unterscheidet handlungsleitendes von trägem Wissen. Als eine wichtige Determinante bei der Tätigkeitsregulation benennt er dabei die Intention, denn Wissen alleine bewirke nichts und könne nur in Verbindung mit einem klaren Ziel erfolgreich sein. Die Beziehung zwischen Wissen und Leistung ist nicht eindeutig, da ein umfassendes Wissen nicht zwangsläufig zu einer höheren Leistung führt (Hacker, 2005b). Büssing, Herbig und Ewert (2001, 2002) merken an, daß in ihrer Studie zum Einfluß des Wissens auf Arbeitshandeln zwar Expertenwissen extrahiert werden konnte, dieses jedoch nicht unbedingt das handlungsleitende war. Letztlich konnten sie keinen Zusammenhang nachweisen zwischen dem expliziten Wissen und dem beobachtetem Verhalten in kritischen Arbeitssituationen, wohl aber zwischen Handeln und implizitem Wissen. Auch Bergmann (2001) betont, daß berufliche Kompetenz weit über die Aneignung von explizitem Wissen hinausgeht und verweist auf die fehlende oder negative Korrelation zwischen explizitem Wissen und Performanz. Klocke (2004) hingegen konnte einen Einfluß von domänenspezifischem deklarativem Wissen auf die Performanz in komplexen Problemlösesituationen belegen.

Zusammenfassend läßt sich festhalten: Wissen ist nicht deckungsgleich mit Kompetenz, sondern vielmehr eine Facette von Kompetenz, welche vorauslaufende notwendige Bedingung für verstanden werden kann und deren Determinanten hin zu einer handlungsleitenden Funktion des Wissens weiter aufgeklärt werden müssen.

1.6. Auffassung des Wissensbegriffes in der vorliegenden Arbeit

Gemäß der Auffassung von Hacker (1999) versteht auch die vorliegende Arbeit individuelles Wissen als diejenigen Informationen oder Fakten, welche im individuellen Gedächtnis gespeichert und abrufbar sind. Dabei wird eine Struktur innerhalb dieser Wissens Elemente unterstellt, welche einer semantischen Beziehung entspricht (z.B. Klix, 1996; van der Meer, 1984). Im Kontext des Wortassoziationsversuchs erweisen sich insbesondere die Klassifikationen explizit vs. implizit sowie deklarativ vs. prozedural als relevant. Der Wortassoziationsversuch ist ein Verbalisierungsverfahren, somit ist es geeignet zur Erfassung expliziter (d.h. verbalisierbarer) Wissensinhalte aus dem Individualgedächtnis. Damit liegt der Fokus auf deklarativem Wissen, da das prozedurale Wissen nicht immer verbalisiert werden kann. Hinsichtlich der Inhalte des Wissens beschäftigt sich die vorliegende Arbeit ausschließlich mit dem berufsbezogenen Wissen.

2. Wissensdiagnosen

Wissensdiagnosen dienen der Ermittlung sowie der anschließenden qualitativ-quantitativen Beschreibung von Inhalt und Organisation der gedächtnismäßigen Repräsentation von individuellen Wissensbeständen (z.B. Gruber, 1996; Rothe, 2003), wobei hauptsächlich personenzentrierte Methoden eingesetzt werden (Hacker, 2005b). Nach Tergan (1988) ist das Ziel einer v.a. qualitativen Wissensdiagnose, eine diagnostische Aussage treffen zu können über Wissenslücken in

einer bestimmten Domäne oder aber, Fehlvorstellungen über bestimmte Begriffe oder deren Zusammenhänge zu identifizieren. Tergan (1988) unterscheidet dabei in eine Statusdiagnose, bei der der aktuelle individuelle Wissensstand in einer umschriebenen Domäne erfaßt wird, und in eine Prozeßdiagnose, bei der der Fokus auf Erwerb und Veränderung des individuellen Wissens liegt. Dabei liegt einer Wissensdiagnose immer das Prinzip zugrunde, aus der Performanz bestimmter kognitiver Leistungen auf die Strukturen der zugrundeliegenden individuellen Wissensrepräsentation rückzuschließen, welche dann gemäß Tergan (1988) in Form eines hypothetischen Wissensmodells beschrieben und schließlich bewertet werden können.

Rothe (1994a) grenzt die Wissensdiagnose von einer Wissenserfassung ab. Danach soll eine Wissenserfassung den Anwender, also z.B. einen Versuchsleiter, (lediglich) dazu befähigen, sich Wissen aus einer ihm bisher unbekanntem Domäne auf rationale Weise anzueignen und formalisiert (z.B. als hypothetisches Wissensmodell) darzustellen. Die Wissensdiagnose hingegen prüft immer, in welchem Ausmaß Wissen bei Individuen oder Gruppen ausgeprägt ist, mit dem Ziel einer qualitativ-quantitativen Beschreibung von Inhalt und Organisation der gedächtnismäßigen Wissensrepräsentationen. In der vorliegenden Arbeit jedoch soll diese Unterscheidung nicht weiter berücksichtigt werden, da in den empirischen Studien die eingesetzten wissensdiagnostischen Verfahren z.T. beide Funktionen – Erfassung und Diagnose des Wissens – gleichzeitig erfüllten.

2.1. Wissensdiagnosen in der Arbeits- und Organisationspsychologie

Nach der Einschätzung von Schraagen (2006) wurden Wissensdiagnosen bedeutsam, als in den frühen 1980er Jahren die Entwicklung der „Künstlichen Intelligenz“ begann und damit einhergehend Expertensysteme auch kommerziell an Interesse gewannen. Als eines der zentralen künftigen Betätigungsfelder, denen sich arbeitspsychologische Aktivitäten verstärkt widmen sollten, sieht Hacker (2005a) das wachsende Feld der Wissenstechnologien und schlußfolgert daraus, daß eine zukunfts-gestaltende Arbeitspsychologie auch eine Wissenspsychologie sein sollte. Daher fordert er von der Arbeitspsychologie eine Erweiterung ihres Analyseinventars, um sowohl auf der Ebene von Individuen als auch von Gruppen wissensbasierte Entwurfstätigkeiten untersuchen und letztlich auch gestalten zu können. Auch Rothe (2006) proklamiert als Aufgabe der Arbeitspsychologie die Bereitstellung geeigneter Methoden und Instrumentarien, um fachspezifisches Wissen angemessen erfassen und diagnostizieren zu können. Die vorliegende Arbeit möchte einen Beitrag dazu leisten.

Wissensanalysen besitzen gemäß Timpe, Rothe und Seifert (1994) sowie Rothe (2006) eine arbeitspsychologische Relevanz, da die Regulation von Arbeitstätigkeiten typischerweise wissensgestützt verläuft (Rothe, 1995), auch nach der Einschätzung von Bergmann (2000) ist das Beherrschen eines gesicherten Basiswissens eine notwendige Voraussetzung für berufliche Tätigkeit. Ein zentrales Anwendungsfeld von Wissensanalysen ist die Entwicklung wissensbasierter Systeme, ein zweiter Ansatz ist die Ermittlung von Aus- und Weiterbildungsinhalten für berufliche Tätigkeiten (Rothe, 2006), wobei Stracke (2004) darauf hinweist, daß bei Einsatz visualisierender Verfahren wie Concept Maps oder Strukturlegetechniken die dabei entstehenden Wissensnetze auch gezielt als Strukturierungshilfe genutzt werden können und somit didaktischen Nutzen haben. Als drittes großes Anwendungsgebiet benennt Rothe (2006) die Planung und Evaluation von Maßnahmen zur Sicherung des Unternehmenserfolges.

Als wichtiges Entscheidungskriterium bei der Wahl der eingesetzten Instrumentarien unterstreicht Schuler (1989) die Ökonomie, mit der die Daten erhoben werden können, Sonntag und Schaper (1997) verweisen in diesem Zusammenhang auf die hohen Belastungen v.a. für Produktionsbetriebe, die der zeitliche Aufwand komplexer Wissensanalysen mit sich bringen kann. Ein weiterer Aspekt bei der Auswahl von Analyseverfahren ist die Akzeptanz durch die Beurteilten (Schuler, 1989), darüber hinaus sollten die Probanden von Wissensanalysen gemäß Sonntag und Schaper (1997) gute bis sehr gute fachliche und kognitive Voraussetzungen aufweisen, um eine erfolgreiche Wissensanalyse zu gewährleisten.

Im Folgenden sollen kurz die wichtigsten Anwendungsgebiete für Wissensdiagnosen im arbeitspsychologischen Kontext skizziert werden. Dabei stehen die verschiedenen Zielrichtungen bzw. der Nutzwert einer solchen Diagnose im Fokus der Betrachtung, z.B. zum Zwecke der Evaluation oder im Rahmen von Wissensmanagement.

2.1.1. Wissensdiagnosen zur Planung und Evaluation von Maßnahmen zur Sicherung des Unternehmenserfolges

Wissen ist ein Wertschöpfungsfaktor, erbringt Wettbewerbsvorteile und ist damit eine der wichtigsten Ressourcen eines Unternehmens (z.B. Mandl, 1997; Hacker, 1999; Bergmann, 2001; Ley und Albert, 2003; Mieg, 2006), daher benennt Rothe (2006) als bedeutsames Anwendungsfeld für Wissensdiagnosen die Planung und Evaluation von Maßnahmen zur Sicherung des Unternehmenserfolges. Der ökonomische Erfolg eines Unternehmens ist u.a. determiniert durch die Leistungsfähigkeit und auch Leistungsbereitschaft der einzelnen Beschäftigten, daher zielen personenbezogene Interventionsmaßnahmen wie z.B. Weiterbildungsveranstaltungen direkt auf die Verhaltensänderungen der Arbeitspersonen ab. Auch bedingungsbezogene Maßnahmen, die z.B. das organisationale Wissensmanagement betreffen, beeinflussen die Arbeitssituation und die Ausführungsbedingungen der Tätigkeiten, so daß sich diese Interventionen ebenfalls in einer Verhaltensänderung der einzelnen Beschäftigten niederschlagen sollten (Rothe, 2006). Da die Grundlage jeder Verhaltensänderung die Modifikation der vorhandenen bzw. der Aufbau neuer handlungsleitender Wissensstrukturen darstellt, ist eine individuelle Wissensdiagnose bei den Beschäftigten notwendig, um gezielt Interventionen planen zu können. Nur durch den Vergleich zwischen dem erwarteten Soll-Wissen und dem diagnostizierten Ist-Wissen einer Arbeitsperson können Schlußfolgerungen abgeleitet werden hinsichtlich der durchzuführenden Interventionen (Rothe, 2006). Stracke (2004) weist darauf hin, daß – bei Einsatz entsprechend differenzierender Diagnoseinstrumente – diese auch in der Lage sind, Fehlkonzepte oder Wissenslücken aufzudecken, welche dann gezielt bearbeitet werden können. Ebenso ermöglicht der Vergleich zwischen Soll- und Ist-Wissen i.S. einer Post-Erhebung die Evaluation der Effizienz bereits erfolgter Maßnahmen (Rothe, 2006). Nach Bergmann (1999) läßt sich der Wirkungsnachweis einer Intervention dadurch feststellen, inwieweit das Ziel erreicht wurde – dies setzt voraus, daß die Zielgrößen zuvor definiert wurden. Da aber Arbeitsfelder oftmals schlecht definiert sind und daher eine vollständige Rekonstruktion der erforderlichen Wissensbasis kaum möglich ist, beschreibt Bergmann (1999) als einen Ausweg aus dieser Problematik, die Analyseeinheiten auf die Personen zu verlagern anstatt auf die Arbeitsaufgaben. So werden dann also Experten oder auf dem jeweiligen Gebiet erfahrene Arbeitspersonen analysiert, um über deren Wissen und Strategien die Zielgrößen für Interventionen

bestimmen zu können. Bergmann (1999) sieht allerdings als methodisches Problem dabei die Definition dessen, was ein Experte ist bzw. deren teilweise auch unkorrektes Wissen. Dennoch erweist sich also die individuelle Wissensdiagnose als gangbarer Weg, um organisationale Maßnahmen planen und evaluieren zu können.

2.1.2. Wissensdiagnosen im Kontext von Wissensmanagement

Als eine spezifische Form von Maßnahmen zur Sicherung des Unternehmenserfolges kann die Einführung eines Wissensmanagementsystems betrachtet werden. Nach Hacker (1999) kann Wissensmanagement verstanden werden als ein Prozeßbegleiter von der Sammlung von Informationen über ausgewähltes handlungsrelevantes Wissen hin zu innovativem Handeln. Vergegenständlichtes Wissen, z.B. in Form von Ablaufplänen, ist dabei ein wichtiger Schritt, entscheidend für den Unternehmenserfolg ist aber dessen psychische Verarbeitung in den Köpfen der Mitarbeiter und die Umsetzung in entsprechendes Handeln (Hacker, 1999).

Reinmann-Rothmeier und Mandl (2000) unterscheiden im individuellen Wissensmanagement die Phasen der Wissensrepräsentation, der Wissensgenerierung, der Wissenskommunikation sowie der Wissensnutzung. Der erstgenannte Abschnitt ist dabei zentral hinsichtlich der Wissensdiagnose, denn in dieser Phase wird Wissen diagnostiziert und identifiziert, um das unternehmensrelevante Wissen darstellen und letztlich bewahren zu können. Neben der Suche nach geeigneten Wissensquellen steht dabei die Vorwissensanalyse sowie die Bedarfsanalyse im Zentrum. Dies dient zum einen dazu, mittels geeigneter Techniken (z.B. visualisierende Methoden) das Vorwissen einer Person darzustellen, zum anderen wird dieses Vorwissen in einem Soll-Ist-Vergleich in Beziehung gesetzt zu demjenigen Wissen, welches für die Erfüllung bestimmter Arbeitsaufgaben wahrscheinlich erforderlich sein wird. Nach Reinmann-Rothmeier und Mandl (2000) ist dabei das Aufzeigen von Wissenslücken und die Ableitung des (qualitativen und quantitativen) Bedarfs an weiteren Informations- und Wissensbausteinen zentral.

Auch in der Konzeption von organisationalem Wissensmanagement, wie Schüppel (1996) sie versteht, nimmt die Wissensdiagnose eine zentrale Rolle ein und geht dem eigentlichen Managen von unternehmensrelevantem Wissen voraus. Danach ist eine Rekonstruktion der Wissensbasis unerlässlich, bei der eine systematische Auseinandersetzung mit dem Kernwissen der Organisation erfolgt und idealerweise das explizite individuelle und kollektive Wissen visualisiert. Die diagnostische Auseinandersetzung mit der Ist-Situation des Unternehmens ist also eine wichtige Voraussetzung für die Bedarfsermittlung und die Ableitung spezifischer, an die Organisation angepaßte Maßnahmen zum Wissensmanagement (Schüppel, 1996).

2.1.3. Wissensdiagnosen im Rahmen von Evaluationsmaßnahmen

Auch Evaluationsmaßnahmen dienen letztlich der Sicherung des Unternehmenserfolges und verfolgen die Frage, ob die Investitionen in eine Interventionsmaßnahme zum gewünschten Ziel geführt haben. Ein häufig evaluiertes Feld sind z.B. Weiterbildungsmaßnahmen, welche sich durch das primäre Ziel auszeichnen, Wissen an die Lernenden vermitteln zu wollen. Der Erfolg einer solchen Lehrveranstaltung kann demnach u.a. daran gemessen werden, inwieweit dieser Anspruch umgesetzt werden konnte, ob also der individuelle Wissensbesitz der Teilnehmer verändert bzw. erweitert werden konnte. Hasebrook (2003) plädiert für eine Evaluation von Bildungsmaßnahmen,

die sich neben (oder vielmehr: statt) der Bilanzierung der entstandenen Kosten in erster Linie solcher Erfolgsmaße wie Praxistransfer oder Integration des erworbenen Wissens in den Arbeitsprozeß zuwendet. Zur Beurteilung eines gelungenen Transfers sind Wissenstests unmittelbar nach der Bildungsmaßnahme dabei ein Baustein neben Mitarbeitergesprächen, Lernpatenschaften oder Follow-up-Erhebungen (Hasebrook, 2003).

Nach Häcker, Leutner und Amelang (1998) ist das Testen im Bildungs- bzw. Ausbildungsbereich weit verbreitet, es kann von einem fortlaufenden Beurteilungsprozeß gesprochen werden, bei dem die Lernfortschritte beobachtet werden. Somit braucht es geeignete Methoden, um solche Veränderungsprozesse abzubilden (beispielsweise die Kombination aus Assoziieren und Strukturlegen, welche Rothe, 2003, als sehr nützlich für die Evaluation innerbetrieblicher Bildungsmaßnahmen bewertet). Dabei steht der Ansatz der Individualdiagnose im Mittelpunkt: im Prä-Post-Design sollte sich eine Zunahme des individuellen Wissensbesitzes zeigen oder eine Veränderung von Inhalten bzw. Struktur des Wissens. Ist der Nachweis einer Veränderung möglich, kann auch die Lehrveranstaltung als erfolgreich in Hinblick auf ihre Zielerreichung angesehen werden. Das ermittelte Wissen kann schließlich mit dem vom Ausbilder angestrebten Wissensmodell verglichen werden (Rothe, 2003). Wegwarth (2002) beispielsweise setzte den Wortassoziationsversuch ein, um ein Weiterbildungsprogramm zu evaluieren, dabei konnte eine Veränderung des individuellen Wissensbesitzes im Laufe der Weiterbildung nachgewiesen werden. Aber auch die Ebene der Gruppendiagnose ist lohnend für die Evaluation: zeigt sich am Ende der Lernphase ein über die Teilnehmer hinweg einheitlicherer Wissensbesitz als es zu Beginn der Fall ist (i.S. von Vorwissen), so kann diese Umstrukturierung des individuellen Wissens ebenfalls auf den Einfluß der Lehrveranstaltung zurückgeführt und somit als Erfolg interpretiert werden.

2.1.4. Wissensdiagnosen zur Ermittlung von Ausbildungsinhalten

Ein weiteres großes Anwendungsgebiet von Wissensdiagnosen ist die Ermittlung von Aus- und Weiterbildungsinhalten für berufliche Tätigkeiten (Rothe, 2006). Sonntag und Schaper (1997) sehen in der Beschreibung von Lernerfordernissen eine grundlegende Voraussetzung zur Gestaltung von Qualifizierungsprozessen. Dabei ist eine Wissensdiagnose unter Einbeziehung der betrieblichen Experten nützlich, um Art und Umfang der Ausbildungsinhalte zu bestimmen. Der Fokus liegt dabei insbesondere auf der Vollständigkeit des Wissens, denn alle für den Arbeitsalltag relevanten Wissenskomponenten sollten im Zuge der Edukation berücksichtigt werden (Rothe, 2006). Hacker (2005b) hält die Kenntnis über das für eine Arbeitstätigkeit erforderliche und auch das bereits vorhandene Wissen der Arbeitspersonen für eine unerläßliche arbeitspsychologische Aufgabe, um zum einen den Ausbildungsbedarf und zum anderen die relevanten Ausbildungsinhalte zu ermitteln. Auch Sonntag und Schaper (2006) erachten es für notwendig, für die Ermittlung des Bedarfs an Lern- und Trainingserfordernissen (z.B. im Rahmen von Personalentwicklungsmaßnahmen) nicht nur die Arbeitsaufgaben und –anforderungen zu analysieren, sondern auch immer Wissensanalysen (insbesondere der betrieblichen Experten) durchzuführen. So nutzte beispielsweise Schindler (2002) die Verfahren Wortassoziationsversuch sowie Strukturlegetechnik, um dasjenige Wissen zu erfassen, welches Führungskräfte während der ersten hundert Tage ihrer Tätigkeit erwerben sollten. Dabei wurden diejenigen Assoziationen, die von vielen der beteiligten Führungskräfte benannt wurden und daher als besonders relevant für diese Wissensdomäne erachtet werden konnten, im weiteren Verlauf im Gruppenversuch zu einem Wissensnetz integriert. So dienten Methoden der

Wissensdiagnose dazu, die relevanten Inhalte zu erheben, welche schließlich die inhaltliche Basis für ein Selbstlernprogramm für Führungskräfte darstellten.

2.1.5. Wissensdiagnosen bei Personalentscheidungen

Leistungsbeurteilungen aller Art haben in der Arbeitswelt vielfältige und wichtige Funktionen: so können sie beispielsweise Basis für Personalentscheidungen oder Planungsinstrument in der Personalentwicklung sein, ein Feedback für die Beurteilten hinsichtlich ihrer Leistungsverbesserungen darstellen, bei der Gestaltung von Arbeitsorganisationen behilflich sein, zur Vergütungsbestimmung herangezogen oder bei der Evaluation von Maßnahmen eingesetzt werden (Schuler, 1989). Welches Leistungskriterium (oder welche) dabei herangezogen wird, ist abhängig von der konkreten Zielstellung und leitet sich zudem aus den jeweiligen Arbeitsanforderungen ab. Nach Schuler (1989) sollte es sich dabei um Daten handeln, die entweder als direkte Leistungsmeßwerte oder aber als Indikatoren für die interessierende Leistung interpretiert werden können. Wissensdiagnosen haben das Ziel, das fachspezifische Wissen von Arbeitspersonen abzubilden. Dahinter steht die Annahme, daß Arbeitsleistung immer auch durch handlungswirksam gewordenes Wissen determiniert wird (Rothe, 2006). Man erhofft sich also, auf Grundlage des derzeitigen individuellen Wissensstandes eine Prognose künftiger beruflicher Leistung treffen zu können. Hacker (1995, 1996, 2005b) warnt allerdings davor, im Umkehrschluß das ausgesagte Wissen einfach als handlungsleitend zu interpretieren. Seiner Ansicht nach entsteht die handlungsleitende Qualität von Wissen erst durch die Kopplung an eine Intention, da Können nicht nur etwas anders machen als die meisten, sondern dies auch noch zielgerichtet tun.

Nach Häcker, Leutner und Amelang (1998) werden Tests in der Berufseignungsdiagnostik v.a. eingesetzt zur Auswahl oder Beförderung von Mitarbeitern oder zur Klassifikation in Organisationen. Dabei bewerten diese Autoren Tests im Vergleich mit den anderen verfügbaren Alternativen bei Personalentscheidungen als die zuverlässigsten und am wenigsten diskriminierendsten Instrumente. Ziel der Diagnostik in diesem Kontext ist festzustellen, inwieweit der (potentielle) Mitarbeiter in der Lage ist, die Anforderungen des künftigen Aufgabenfeldes erfolgreich zu bewältigen (Lang-von Wins und von Rosenstiel, 2000). Dabei steht den Entscheidern in der Praxis ein umfangreiches Instrumentenangebot zur Verfügung, welches gemäß Lang-von Wins und von Rosenstiel (2000) meist in Kombination eingesetzt wird: neben der Analyse der Bewerbungsunterlagen und der Referenzen sind Gespräche mit der Personalabteilung gängiges Vorgehen, welche die Eignung auf einer eher allgemeinen Ebene prüfen. Darüber hinaus richten Methoden wie strukturierte oder unstrukturierte Interviews mit den Fachabteilungen, Arbeitsproben, Gruppendiskussionen oder Assessment Center ihren Fokus wesentlich zielgerichteter auf die Inhalte der Tätigkeiten und damit auch auf das dazu notwendige Fachwissen der Kandidaten.

Schmidt und Hunter (1998, 2000) untersuchten in einer breitangelegten Metaanalyse u.a. den Nutzen von Tests zum Berufswissen zur Prognose künftiger beruflicher Leistungen, welche i.d.R. durch Vorgesetztenurteile erhoben wurden. Dabei konnten sie eine gute Vorhersagegüte solcher Wissenstests feststellen ($r = .48$), damit erlauben diese ähnlich gute Prognosen wie Arbeitsproben, Vorstellungsgespräche oder die Probezeit und übertreffen die Validität von Assessment Centern oder gar der Berufserfahrung. Ihr Einsatz in deutschen Unternehmen ist aber mit etwa 5% nachrangig gegenüber v.a. Interviews und Referenzen, aber auch im Vergleich zu Arbeitsproben,

Assessment Centern oder biographischen Fragebögen (Amelang und Schmidt-Atzert, 2006). Schmidt und Hunter (2000) weisen allerdings darauf hin, daß Berufswissenstests – wie Arbeitsproben auch – nicht geeignet sind, um unerfahrene Kandidaten zu beurteilen. Schließlich muß beim Untersuchten eine Basismenge an berufsspezifischem Wissen und Können vorhanden sein, um Performanz in dieser Art von Verfahren zeigen zu können, während andere Instrumente wie Bewerbungsgespräche oder Assessment Center (auch) eher allgemeine, von der konkreten Arbeitstätigkeit unabhängige Merkmale der Bewerber thematisieren. Nach Schmidt und Hunter (2000) sind berufsspezifische Wissenstests i.d.R. von den Unternehmen selbst entworfen worden im Ergebnis von Arbeitsanforderungsanalysen. Dies sei zwar zeitaufwendiger und teurer als z.B. strukturierte Interviews, könne aber gleichsam auf den konkreten Arbeitsplatz zugeschnitten werden. Nach der Einschätzung von Schmidt und Hunter (2000) können Tests zur Erfassung des beruflichen Wissens im Zusammenhang mit Personalentwicklungsmaßnahmen also einen erheblichen praktischen Nutzen für die Unternehmen darstellen.

2.1.6. Wissensdiagnosen zur Entwicklung wissensbasierter Systeme

Wissen ist elementare Grundlage für die Regulation von Arbeitstätigkeiten (Rothe, 1995), zur Unterstützung des Arbeitsprozesses sowohl lernender als auch arbeitender Personen ist man daher auf das handlungsleitende Wissen angewiesen (Großmann und Teske-El Kodwa, 1995). Wissensorientierte Unterstützungssysteme sind nach Ziegler (1995) Systeme, welche die Speicherung, Verarbeitung und Vermittlung von Wissen unterstützen. Dies kann beispielsweise in Form von Online-Dokumentationen geschehen, als programmiertes Lernsystem, Hypertext, Tutorensystem, Multimedia-Präsentation oder als Expertensystem. Letztlich können solche Unterstützungssysteme eingesetzt werden zur Vermittlung beruflicher Lerninhalte, bei der Lösung von spezifischen Arbeitsaufgaben oder auch für den Wissenstransfer aus einem Arbeitsbereich in einen anderen (Ziegler, 1995).

Neben den wissensbasierten Systemen lassen sich auch sog. Expertensysteme nennen, die Abgrenzung zwischen beiden besteht gemäß Puppe (1996) lediglich darin, daß letztgenannte Systeme sich im Wesentlichen auf das Wissen ausgewiesener Fachspezialisten stützen. Wissensbasierte Expertensysteme sind also Programme, die das spezifische Wissen von Fachleuten nachbilden und letztlich den Anwender bei seiner Entscheidungsfindung unterstützen sollen (Rothe, Timpe und Warning, 1991; Buchanan, Davis und Feigenbaum, 2006), Puppe (1996) betont daher zu Recht, neben dem zu selektierenden Wissen auch die Schlußfolgerungsfähigkeit der Experten als wichtige Komponente bei der Problemlösung in das System zu integrieren. Expertensysteme werden zunehmend eingesetzt für die informationstechnische Unterstützung der Diagnose und Behebung von technischen Störungen (Timpe und Rothe, 1997). Um beispielsweise Diagnosesysteme zu konstruieren, welche Instandhalter bei der Wartung und Reparatur technischer Anlagen unterstützen, müssen neben den allgemeinen technischen Regelhaftigkeiten der Anlage auch Informationen bereitstehen über typische Fehler der Anlage sowie deren Häufigkeit und Strategien zur Erkennung dieser Fehler. Dazu muß das Wissen (Fakten, Erfahrungen, Strategien) in einer nutzbaren Form im Hilfesystem repräsentiert sein (Timpe und Rothe, 1997) All dieses Erfahrungswissen, welches nicht in Bedienungsanleitungen o.ä. abgelegt ist, muß aus dem Individualgedächtnis der jeweiligen Experten (Rothe, 2006) bzw. ergänzend durch eine kognitive Aufgabenanalyse (Schraagen, 2006) erhoben und in entsprechende Algorithmen umgesetzt werden.

Dabei ist man i.d.R. an der optimalen Problembewältigung interessiert, also auf der Ebene von Experten des Faches, deren Wissensbasis extrahiert, in Regeln gefaßt und kodiert wird (Rothe, Timpe und Warning, 1991; Buchanan, Davis und Feigenbaum, 2006). Da aber viele Befunde dafür sprechen, daß Wissen von Experten oftmals impliziter Natur ist, werden gemäß Rothe (1994a) auch Arbeitspersonen in die Wissensdiagnosen einbezogen, welche noch nicht als Experte bezeichnet werden können, aber deren Wissensbestände vermutlich leichter mittels Verbalisierungsmethoden zugänglich sind.

Die Fakten eines Expertensystems werden u.a. aus dem deklarativen Wissen der Untersuchten gewonnen, während deren prozedurales Wissen in die Regeln des Systems eingeht. Als zentrale Methoden für die Wissensermittlung in diesem Zusammenhang benennen Großmann und Teske-El Kodwa (1995) Dokumentenanalysen, Beobachtungen, Befragungen, Strukturlegetechniken oder Gruppendiskussionen. Sonntag, Rothe und Schaper (1994) machen aber auch auf die Grenzen aufmerksam: in ihren Untersuchungen, welche in einem mehrstufigen Analyseprozeß das Wissen erfahrener Instandhalter bei ausgewählten Störungen technischer Anlagen abbilden sollten, bewerteten sie die vollständige Erfassung der relevanten Wissensgrundlagen als nicht gelungen, was sie auf die Komplexität und Dynamik der technischen Systeme zurückführen. Großmann und Teske-El Kodwa (1995) beanstanden überdies die mangelnde Flexibilität solcher Unterstützungssysteme, welche der Vielfalt der Arbeitstätigkeiten nicht gerecht werden (können). Rothe (1994a) sieht diese Ermittlung des Fachwissens als das zentrale psychologische Problem bei der Entwicklung von Expertensystemen. Ist ein Expertensystem erst einmal installiert, so hängt seine Effektivität zentral davon ab, wie die darin verankerte Wissensbasis ständig aktualisiert und auf aktuellem Stand gehalten wird (Rothe, 1994a).

2.1.7. Wissensdiagnosen bei der Beantwortung von Forschungsfragen

Auch die arbeitspsychologische Forschung benutzt Methoden zur Wissensdiagnose, beispielsweise um die Determinanten von kompetentem Arbeitshandeln zu extrahieren, Rothe (2006) sieht hier weiterhin großen Forschungsbedarf, oder um den Einfluß von Wissen auf das Wohlbefinden der Arbeitsperson aufzudecken. Beispielhaft soll an dieser Stelle kurz eines der umfangreicheren Forschungsgebiete angesprochen werden, in welchem Wissensdiagnosen wertvolle Daten liefern: der Expertiseforschung. Im Zusammenhang mit der soeben erörterten Entwicklung von Expertensystemen stellt sich u.a. die Frage, was einen Experten überhaupt ausmacht, oft betrachtet im Kontrast zum sog. Novizen, wobei Hacker (2005b) vorschlägt, im Kontext der Arbeitspsychologie doch eher von Durchschnittsvertretern als von Novizen zu sprechen. Nach Gruber und Mandl (1996) ist ein Experte jemand, der dauerhaft eine herausragende Leistung auf einem bestimmten Gebiet erbringt, während der Novize nur eine geringe Leistungsstärke in dieser Domäne zeigt. Zwar ist das zentrale Merkmal der Expertise die gesteigerte Performanz (Hacker, 1995, 1996, 2005b), aber gemäß Gruber und Mandl (1996) befassen sich die meisten Untersuchungen v.a. aus methodischen Gründen eher mit dem Wissen und Gedächtnis von Experten und Novizen als mit deren Handlungen.

Rothe und Schindler (1996) sowie Chi (2006) weisen darauf hin, daß die Grundlage für die Expertise Art und Umfang des hinter der Performanz stehenden Wissens ist, wobei die überdurchschnittlichen Leistungen in erster Linie durch eigenes (schlußfolgerndes) Denken

ermöglicht werden, andere Wissensquellen wie Erbgut, individuelle Erfahrung oder die sprachliche Belehrung sind dagegen nachrangig. Auch Chi (2006) sieht die Analyse des Wissensbestandes und seiner Repräsentation als Schlüssel zum Verständnis des unterschiedlichen Handelns von Experten und Novizen. Viele empirische Befunde legen dar, daß Experten nicht nur mehr über eine spezifische Domäne wissen, sondern ihr Wissen auch in anderer Weise strukturiert ist als bei Novizen (z.B. Mandl und Ballstaedt, 1986; Kluwe, 1988; Rothe und Warning, 1991; Gäbler, 1994; Gruber und Mandl, 1996; Rothe und Schindler, 1996; Beatty und Gerace, 2002; Stracke, 2004). Experten verfügen außerdem über größere Wissensseinheiten (Chunks) als Novizen (Kluwe, 1988; Gruber und Mandl, 1996; Rothe und Schindler, 1996; Chi, 2006), Davis, Curtis und Tschetter (2003) fordern daher, genau die Methoden einzusetzen, die eine Abbildung dieser Wissensorganisation und ihrer Strukturen ermöglichen, wobei die Erhebung des deklarativen Wissens nicht vernachlässigen sollte, da dieses eine wichtige Bedingung für die Ausbildung strukturierten Wissens sei.

2.1.8. Fazit für den Wortassoziationsversuch

Wie soeben im Kapitel 2.1 dargestellt, bietet der Kontext der Arbeitspsychologie eine Vielzahl von Anwendungsfeldern, in denen Wissensdiagnosen einen wertvollen Beitrag leisten können. Es besteht ein praktisches Bedürfnis danach, die Wissensbestände von Individuen oder Gruppen abzubilden. Der Wortassoziationsversuch kann zu allen angesprochenen Einsatzgebieten wissensdiagnostischer Methoden einen Beitrag leisten, wenngleich er nicht immer als alleiniges Instrument ausreichend ist (beispielsweise kann er das deklarative Wissen von Experten erfassen, aber nicht dessen spezifische Struktur). Daher sollen im folgenden Kapitel weitere diagnostische Zugänge kurz vorgestellt und diskutiert werden.

2.2. Methoden zur Wissensdiagnose

Ebenso, wie es eine Vielzahl verschiedener Wissensbegriffe gibt, bietet auch hinsichtlich der einzusetzenden Methoden zur Erhebung oder Diagnose des Wissens ein breites Spektrum, wobei Gruber (1996) betont, daß es – nicht zuletzt aufgrund der verschiedenen Modellannahmen – nicht die einzige optimale Methode geben kann. Nach Einschätzung von Rothe (1994a) sowie Rothe, Timpe und Warning (1991) gibt es keine einzige Methode, mit der das deklarative und prozedurale Wissen einer Domäne gleichermaßen und vollständig erfaßt werden kann. Gemäß Kluwe (1988) setzen Methoden zur Gewinnung von Daten über Wissen voraus, daß die eingesetzten Methoden geeignet sind, jeweils spezifische Ausschnitte einer individuellen Wissensstruktur aus dem Langzeitgedächtnis zu aktivieren. Die Anforderungssituation muß demnach so gestaltet sein, daß eine Verarbeitung dieser Informationen im Kurzzeitgedächtnis erforderlich wird.

Grundsätzlich können bei Wissensdiagnosen zwei Zielrichtungen verfolgt werden (z.B. Tergan, 1989b): zum einen kann die Wissensrepräsentation eines einzelnen Individuums im Mittelpunkt stehen, zum anderen kann das Ziel einer Untersuchung sein, Aussagen über den Wissensbesitz einer Gruppe vergleichbarer Individuen zu treffen. Der individualdiagnostische Ansatz legt eine qualitative Beschreibung der Wissensstruktur nahe, was insbesondere in der Lehr- und Lernforschung genutzt wird zur Identifizierung von Wissensdefiziten (Tergan, 1989b), aber die quantitative Analyse des individuellen Wissensbesitzes ist ebenso gebräuchlich. Die Auswahl eines Verfahrens für einen Untersuchungsansatz sollte neben der Zielstellung (über welche Inhalte sollen

welche Aussagen getroffen werden) auch abhängen von der angenommenen Art des Wissens: für die Erhebung von explizitem und verbalisierbarem Faktenwissen beispielsweise eignen sich eher Verbalisierungen als Indikatoren für Wissen, für die Erfassung impliziten Wissens ist der Einsatz von Verfahren mit Entscheidungsaufgaben denkbar (z.B. Sortierverfahren), bei denen anhand der Aufgabenbewältigung auf den Wissensbesitz geschlossen wird, während bei prozeduralisiertem Wissen eher die Methode des lauten Denkens, Beobachtungen oder Problemlöseaufgaben angezeigt sind, da dabei das in Handlung umgesetzte Wissen erfaßt wird.

Letztlich haben alle Methoden zur Wissensdiagnose das gleiche Grundproblem: der „wahre Wert“ des interessierenden Wissensbestandes ist nicht direkt zugänglich, sondern kann immer nur durch Indikatoren erschlossen werden. Solche Indikatoren können zum einen z.B. Verbalisierungen von Wissensbeständen sein, zum anderen läßt sich auch aus dem Handeln oder aufgrund von Entscheidungen auf den dahinterstehenden Wissensbesitz schließen. Tergan (1989b) sieht genau in dieser Bestimmung geeigneter Indikatoren ein grundlegendes Problem aller Wissensdiagnosemethoden. Gemäß Grabowski (2005) ist die wissensdiagnostische Forschung i.d.R. mit sprachbasierten Verfahren verknüpft, wobei eine Beziehung zwischen dem Resultat der Sprachproduktion und dem Wissensinhalt vorausgesetzt wird. Nach Kluwe (1988) werden in den meisten Verfahren der Datengewinnung über individuelles Wissen Verbaldaten herangezogen, bei denen darauf vertraut wird, daß Personen ihr Wissen auch verbalisieren können, Gruber (1996) bewertet die Frage der Verbalisierbarkeit von Wissen als generelles Problem bei allen diesbezüglichen Diagnosemethoden. Die innewohnende Problematik bringt Mazur (2006) auf den Punkt: in seiner Kritik an Tests zum explizitem Wissen warnt er davor, von einem Nicht-Reden auf ein Nicht-Wissen zu schließen. Diese Warnung trifft insbesondere Verfahren der freien Reproduktion. Auch Kluwe (1988) und Rothe (1994a) verweisen auf diese Problematik, und Kluwe (1988) leitet daraus das Vorgehen ab, aus dem *Handeln* des Menschen auf das dahinterstehende Wissen zu schließen. Analysiert wird dabei also nicht das Wissen selbst, sondern es werden Indikatoren für das Wissen erhoben, wie es beispielsweise bei Arbeitsbeobachtungen realisiert wird. Da arbeitsbezogenes Wissen bei wiederholtem Einsatz seine Bewußtseinspflichtigkeit verliert, Arbeitshandeln aber in großem Umfang von solchem unbewußten Wissen reguliert wird, ist nach Rothe (1995) der Einsatz von beobachtungsgestützten Analysemethoden unerlässlich. Rothe und Schindler (1996) warnen allerdings davor, allein aus dem beobachteten Verhalten auf das zugrundeliegende deklarative bzw. prozedurale Wissen zu schließen – es besteht zunächst nur die Annahme, daß die Verhaltensdaten das individuelle Wissen widerspiegeln (Tergan, 1989b). Unter Berücksichtigung dieser Argumente wäre also ein multipler methodischer Zugang, welcher sowohl Verbal- als auch Beobachtungsdaten integriert, zu bevorzugen gegenüber nur einer einzigen Wissensdiagnosemethode.

2.2.1. Anforderungen an die Konstruktion wissensdiagnostischer Verfahren

Nach Eckert (1998) sowie Tergan (1989b) erfordert die Konstruktion eines wissensdiagnostischen Instruments die Wahl eines adäquaten hypothetischen Modells der individuellen Wissensrepräsentation, so daß diejenigen kognitiven Prozesse, die der Proband zur Beantwortung der Testaufgaben einsetzt, die dahinterstehenden Wissensstrukturen enthüllen können. Dabei kommt der Wahl des Aufgabenformates eine zentrale Bedeutung zu. So lassen sich mit Multiple-Choice-Aufgaben eher Wiedererkennensleistungen abbilden, während durch offene Antworten eher

Erinnerungsleistungen abgefordert werden (Eckert, 1998; Davis, Curtis und Tschetter, 2003; Kubinger, 2006). Bei der Konstruktion von Multiple-Choice-Aufgaben empfehlen Michelsen und Cordes (2005), mindestens vier verschiedene Antwortalternativen vorzugeben, um wissensbasierte Entscheidungen der Probanden besser vom Raten differenzieren zu können. Wagener (2003) weist allerdings auf ein Problem hin, was insbesondere bei der Formulierung eines Wissenstests im Multiple-Choice-Format relevant sein kann: die Items neigen dazu, schnell zu veralten (natürlich in Abhängigkeit von der Wissensdomäne) – in dem Maße, wie das Wissen selbst zunehmend seine Halbwertszeit verliert (z.B. Bergmann, 2001), veralten auch Tests zur Erhebung eben dieses Wissens schneller.

2.2.2. Anforderungen an die Durchführung einer Wissensdiagnose

Jede Wissensdiagnose erfordert in der Vorbereitung eine mehr oder weniger intensive Auseinandersetzung des Versuchsleiters mit der zu untersuchenden Wissensdomäne. Rothe (1994a) weist darauf hin, daß verschiedene Methoden der Wissensdiagnose nicht beliebig miteinander kombinierbar sind, da erst der Einsatz bestimmter Methoden (z.B. Dokumentenanalyse) dem Versuchsleiter die notwendigen Vorkenntnisse verschafft, auf die ein anderes Verfahren dringend angewiesen ist. Daher ist durch eine entsprechende Reihenfolge der gewählten Methoden zu gewährleisten, daß eine Methode nicht mehr Vorkenntnisse erfordert als bisher bereits gewonnen werden konnten (Rothe, 1994a; Hacker, 1995, 1996). So erarbeiteten beispielsweise Rothe und Warning (1991) in ihrer Feldstudie ein Wissensmodell auf Grundlage von Dokumentenanalysen, Arbeitsbeobachtungen, Wortassoziationsversuchen, Aufzeichnungen von Verbaläußerungen (lautes Denken) sowie mündlicher bzw. schriftlicher Befragungen anhand eines Leitfadens. Ergebnis all dieser Vorarbeiten war ein hypothetisches Wissensmodell in Form eines Netzwerkes, aus dem schließlich die konkreten Bedingungen für die Durchführung eines Strukturlegeversuchs als Kern der Untersuchung abgeleitet werden konnten. Nach Rothe (1994a) sowie Hacker (2005b) sollten die Prozesse von Wissenserfassung und Wissensmodellierung immer parallel erfolgen, da sie wechselseitig ineinander verzahnt sind, das hypothetische Wissensmodell kann dabei schrittweise entwickelt und präzisiert werden (Rothe, Timpe und Warning, 1991). Ein solches Vorgehen ist in seiner Durchführung sehr anspruchsvoll und aufwendig, sichert aber andererseits die praktische Relevanz der gewählten Wissensdomäne und die Validität des hypothetischen Wissensmodells. Bei der Wahl eines geeigneten Methodeninventars verweist Rothe (2006) u.a. auf folgende Merkmale von Verfahren: so unterscheiden sich verschiedene Methoden zur Wissensdiagnose z.B. hinsichtlich des zeitlichen Aufwandes bei der Durchführung der jeweiligen Techniken, der dazu erforderlichen methodischen Hilfsmittel, aber es lassen sich auch Unterschiede hinsichtlich der Quantifizierbarkeit der erhobenen Daten sowie bzgl. der damit extrahierten Wissensarten benennen. Auch Tergan (1988) macht darauf aufmerksam, daß bei der Wahl einer diagnostischen Methode zu bedenken ist, für welche Art des Wissens (z.B. deklaratives vs. prozedurales) dieses Verfahren sensibel sein soll (siehe dazu auch Kapitel 2.2.6 auf Seite 28).

Wie bereits erwähnt, erfordert jegliche Wissensanalyse eine Auseinandersetzung des Versuchsleiters mit der jeweiligen Wissensdomäne, die Anforderungen an das (domänenspezifische) Vorwissen des Versuchsleiters sind jedoch verschieden in Abhängigkeit der eingesetzten Verfahren. Rothe (2006) unterscheidet dabei zwischen Verfahren, die so gut wie keine Vorkenntnisse benötigen, solchen, die vom Versuchsleiter Grundkenntnisse in der betreffenden

Domäne verlangen, und Verfahren, für die der Untersucher differenzierte Kenntnisse des Untersuchungsgegenstandes benötigt (siehe auch Abbildung 1, in der Rothe, 2006, die am häufigsten eingesetzten Methoden entsprechend klassifiziert hat). Dabei sind Verfahren, welche der erstgenannten Kategorie zuzuordnen sind, sehr geeignet, um einen Einblick in die relevante Wissensdomäne zu gewinnen und die entsprechenden Befunde als Ausgangsbasis für ein zu entwickelndes hypothetisches Wissensmodell und die Auswahl weiterer Instrumente zu nutzen, so wie es z.B. Rothe und Warning (1991) eindrucksvoll demonstriert haben.

kein Vorwissen erforderlich	Grundkenntnisse erforderlich	differenzierte Kenntnisse erforderlich
<ul style="list-style-type: none"> • Dokumentenanalyse • Beobachtung • lautes Denken • Gedankenstichprobe 	<ul style="list-style-type: none"> • Interview • schriftliche Befragung • Satzergänzung • Gruppendiskussion 	<ul style="list-style-type: none"> • Wortassoziationsversuch • Sortiermethoden • Strukturlegetechnik • Grid-Technik

Abbildung 1: Erfordernisse verschiedener Wissensanalyse-Techniken an das Vorwissen des Versuchsleiters (vgl. Rothe, 2006)

Rothe (1994b) bewertet eine Dokumentenanalyse zu Beginn jeglicher Wissenserfassung als eine unabdingbare Voraussetzung, um seitens des Versuchsleiters ein Mindestmaß an domänenspezifischen Kenntnissen zu sichern und so den Einsatz weiterer Methoden überhaupt erst zu ermöglichen. Es ist einzuräumen, daß Verfahren wie die Beobachtung von Arbeitstätigkeiten zwar aufgrund der passiven Rolle des Versuchsleiters zwar keine Vorkenntnisse während der Beobachtung selbst erfordern, aber u.U. die Interpretation des Beobachteten nicht ohne zumindest Grundkenntnisse in der jeweiligen Domäne möglich ist. Im Gegensatz dazu benötigen Verfahren wie das Interview oder Satzergänzungstechniken bereits für ihre Durchführung ein entsprechendes Vorwissen des Untersuchers, da diesem während der Erhebung eine aktivere Rolle zukommt in dem Sinne, als daß er dem Probanden Vorgaben machen muß (z.B. Fragen formulieren), um die zentralen Elemente der interessierenden Wissensdomäne aufzugreifen. Vereinfacht gesagt: der Versuchsleiter sollte ungefähr wissen, wonach er fragt. Die dritte von Rothe (2006) formulierte Kategorie von Verfahren erfordert vergleichsweise differenzierte Kenntnisse des Untersuchers über die Domäne, da Verfahren wie z.B. der Wortassoziationsversuch oder Strukturlegetechniken nicht nur zur Formulierung der Aufgabenstellungen, sondern auch hinsichtlich der erwarteten Probandenäußerungen eine Vorstellung über den Wissensraum auf seiten des Versuchsleiters verlangen. Aber auch hinsichtlich der Methoden selbst sind mehr oder weniger umfangreiche Vorarbeiten notwendig (Rothe, Timpe und Warning, 1991; Rothe, 1994a): Interviews benötigen Interviewleitfäden, schriftliche Befragungen ein Fragebogenschema, für den Wortassoziationsversuch müssen Stimulusbegriffe erarbeitet werden, bei Sortiermethoden sind Vorstellungen darüber erforderlich, was zu sortieren ist und in welcher Beziehung diese Begriffe zueinander stehen, bei Gruppendiskussionen muß der Versuchsleiter die Interaktion führen können aufgrund seiner fachlichen und sozialen Kompetenz.

Bei der Anwendung wissensdiagnostischer Instrumente empfiehlt Eckert (1998), auch dem Kontext der Testbearbeitung Beachtung zu schenken, da dieser möglicherweise Auswirkungen auf das Testergebnis hat. So können z.B. der Wortlaut der Instruktion oder auch die Technologie der Itempräsentation (oder der Aufzeichnung der Antworten) den Probanden beeinflussen. Auch die Entscheidung Paper-Pencil vs. computerunterstützt hat möglicherweise Einfluß auf die Erhebung,

da manche Probandengruppen dem Computer grundsätzlich mehr Objektivität oder Professionalität zubilligen und ein positiveres Image damit verknüpfen als mit dem identischen Instrument in Papierform (Kubinger, 2006). Zudem ist sicherzustellen, daß der Proband die Regeln des Wissensdiagnoseinstrumentes verstanden hat und mit dem Testmaterial umgehen kann (Eckert, 1998), im Umgang mit Tastatur und Maus ungeübte Probanden können beispielsweise in computerunterstützten Verfahren benachteiligt werden, worin Kubinger (2006) die Testfairneß beeinträchtigt sieht.

Rothe (1994b) konnte belegen, daß Probanden mehr Wissens Elemente verbalisieren konnten, wenn sie kurz zuvor aktiv das relevante Wissen genutzt hatten, z.B. bei der Bearbeitung einer entsprechenden Problemlöseaufgabe. Nach Strube (1984) sowie Ericsson (2006) sind die kontextuellen Aspekte bei der Enkodierung von Lernmaterial aller Art nicht wegzudenken. Auch Hacker (1995, 1996) betont die Situationsgebundenheit insbesondere von Handlungswissen und fordert daher, dies bei der Diagnostik zu berücksichtigen. Eine situative Einbettung der Wissensdiagnose in den konkreten Arbeitskontext ist also auch unter diesem Gesichtspunkt einer laborähnlichen Untersuchungssituation vorzuziehen.

Im Anschluß soll ein kurzer Überblick gegeben werden über die gebräuchlichsten Methoden zur Wissensdiagnose. Dabei werden einige Verfahren vorgestellt, welche beispielhaft sind für die von Rothe (2006) vorgenommene Klassifikation entsprechend des Vorwissens des Versuchsleiters, zudem sollen auch die in der vorliegenden Arbeit herangezogenen Verfahren in diese Systematik eingeordnet und kurz skizziert werden.

2.2.3. Beispiele für methodische Zugänge ohne Anforderungen an das Vorwissen des Untersuchers

Im Folgenden sollen zwei Beispiele für Methoden skizziert werden, welche vom Versuchsleiter kaum Vorkenntnisse über die interessierende Wissensdomäne verlangen: die Dokumentenanalyse sowie die Methode des lauten Denkens.

2.2.3.1. Dokumentenanalysen

Vorgehen: Dokumentenanalysen stellen ein bedingungsbezogenes Vorgehen dar, bei dem betriebliche Unterlagen (z.B. Ausbildungsinhalte, Ablaufpläne) in Form von Daten oder Texten ausgewählt werden, auf Grundlage derer und in Absprache mit betrieblichen Experten schließlich die Wissensdomäne für die Untersuchung eingegrenzt werden kann (Großmann und Teske-El Kodwa, 1995). Hacker (2005b) spricht hier vom vergegenständlichten gemeinsamen Handlungswissen der Gruppe, welches z.B. in Arbeitsanweisungen, Qualitätshandbücher oder Software eingeflossen ist. Auch das Studium von Fachliteratur im Zusammenhang mit der Wissensdomäne ist hier einzuordnen (Rothe, Timpe und Warning, 1991).

Grundannahmen: Nach Ballstaedt (1994) zeichnen sich Dokumente v.a. dadurch aus, daß sie nicht explizit zum Zwecke der Forschung produziert wurden, sondern in natürlichen Situationen entstanden. Die Analyse solcher Dokumente ist daher als nonreaktives Verfahren einzuordnen.

Auswertung: Die Auswertung von Dokumentenanalysen ist in erster Linie qualitativer Natur und kann z.B. mittels Inhaltsanalysen geschehen, aber auch Quantifizierungen wie z.B. Statistiken oder

Graphiken können im Ergebnis einer solchen Analyse stehen (Ballstaedt, 1994). Im Kontext von Wissensanalysen können aufgrund von Dokumentenanalysen die ersten hypothetischen Wissensmodelle generiert werden.

Fazit: Als Vorteil von Dokumentenanalysen bewerten Großmann und Teske-El Kodwa (1995) die Objektivität der vorliegenden Daten sowie ihre grundsätzliche Zugänglichkeit, auch Ballstaedt (1994) hebt die Unabhängigkeit des Untersuchers von konkreten Probanden sowie die hohe ökologische Validität von Dokumentenanalysen hervor. Da sie dem Untersucher das Grundwissen über die Domäne verschaffen, sollten Dokumentenanalysen immer am Beginn jeglicher Erhebungen stehen.

2.2.3.2. Methode des lauten Denkens

Vorgehen: Bei der Methode des lauten Denkens (sog. „thinking aloud“) wird ein Proband gebeten, während er eine Aufgabe bearbeitet, laut auszusprechen, was er dabei gerade denkt. Damit ist die Datenerhebung als online-Verfahren zu charakterisieren (Kluwe, 1988). Diese Methode wird v.a. im Kontext von Lern- oder Lösungsprozessen eingesetzt, um Daten über das aktivierte Wissen zu gewinnen, wobei v.a. prozedurales Wissen im Vordergrund steht. Nach Ansicht von Kluwe (1988) sind die so gewonnenen Daten in Form von Verbalisierungen eine reichhaltige Datenquelle und erlauben sowohl Aussagen über Qualitäts- als auch Strukturmerkmale des Wissensbesitzes. Eckert (1998) betont den Einzelfallcharakter dieser Methode.

Grundannahmen: Die zentrale Grundannahme dieser Methode ist, daß auf Denkprozesse und Wissensstrukturen bewußt zugegriffen werden kann und deshalb eine Verbalisierung der Wissensinhalte möglich ist (Eckert, 1998). Nach Kluwe (1988) sowie Großmann und Teske-El Kodwa (1995) können diejenigen Inhalte verbalisiert werden, die sich gerade im Kurzzeitgedächtnis befinden und zur weiteren Verarbeitung anstehen. Unter der Annahme, daß prozedurales Wissen in deklaratives Wissen überführbar ist, bewertet Eckert (1998) die Methode des lauten Denkens v.a. als geeignet für die Erhebung von Daten über prozedurales Wissen.

Auswertung: Die Verbalisierungen des Probanden werden protokolliert, anschließend die einzelnen Aussagen kodiert. Dem sollte ein differenziertes und der Untersuchungsabsicht entsprechendes Kodierschema zugrunde liegen, welches eine möglichst objektive Zuordnung der Aussagen zu den Kategorien erlaubt (Kluwe, 1988; Eckert, 1998).

Fazit: Kluwe (1988) bewertet diese Methode als vielfältig einsetzbar, als kritisch bewertet er jedoch die hohe Reaktivität des Verfahrens und die eingeschränkten Möglichkeiten der Reliabilitätsprüfung sowie eine nur schwer zu erreichende Objektivität. Daher empfiehlt sich ein hypothesengeleitetes Vorgehen, welches die Gefahr der Beliebigkeit im Umgang mit den Verbalaussagen und des willkürlichen Schließens reduziert (Großmann und Teske-El Kodwa, 1995). Darüber hinaus benennen Kluwe (1988) und Eckert (1998) einen Einwand bzgl. der Validität der Aussagen: Probanden verbalisieren einerseits nicht alles Wissen, was sie zur Problemlösung tatsächlich einsetzen, und andererseits verbalisieren sie z.T. auch irrelevantes Wissen, aus diesem Grunde bewerten Hoffmann und Lintern (2006) die Methode des lauten Denkens als ineffektiv. Gemäß Eckert (1998) erschwert dies den Vergleich zwischen Probanden deutlich. Überdies kritisiert Eckert (1998) die ungewohnte Situation für viele Probanden, gleichzeitig zur Problemlösung auch zu

sprechen, Großmann und Teske-El Kodwa (1995) sowie Weidle und Wagner (1994) sprechen von einer hohen Kapazitätsbelastung des Arbeitsgedächtnisses der Probanden.

2.2.4. Beispiele für methodische Zugänge, welche Grundkenntnisse des Untersuchers erfordern

Im Folgenden werden zwei methodische Ansätze zur Wissensdiagnose vorgestellt, welche gemäß Rothe (2006) zumindest Grundkenntnisse des Versuchsleiters über die interessierende Wissensdomäne verlangen. Dies sind zum einen Interviews oder Befragungsverfahren, zum anderen sei an dieser Stelle auch das Format des Lückentextes vorgestellt, bei dem Sätze ergänzt werden sollen.

2.2.4.1. Befragungsverfahren

Vorgehen: Befragungen erfolgen erst nach konkreten Tätigkeiten oder gänzlich unabhängig davon (Kluwe, 1988). Schriftliche Befragungen werden i.d.R. als Fragebogen umgesetzt, Interviews repräsentieren die mündliche Form der Befragung (Schwarzer, 1983; Hron, 1994; Bortz und Döring, 2003). Die Instruktion verlangt vom Probanden eine Wissensaktivierung. Kluwe (1988) sieht dabei als ein wesentliches Merkmal die Spezifität der Instruktion, welche das Ausmaß oder die Eingrenzung des aktivierten Wissens bestimmt: möglich sind geschlossene Fragen mit mehrfach gestuften Antworten, aber auch offene Fragen. Eckert (1998) weist darauf hin, daß die Anwendung um so unproblematischer ist, je strukturierter und standardisierter die Befragung durchgeführt wird.

Grundannahmen: Das Befragen zielt direkt auf gespeicherte Wissensinhalte, es dient der Erfassung von deklarativem Wissen (Eckert, 1998) und damit der verbalisierbaren, bewußtseinsfähigen Teile des Wissens und Könnens, während nach Hacker (2005b) implizites Wissen oder Routinewissen dem Interview unzugänglich bleiben. Je nach Format der Fragen wird eher das Erinnern (bei offenen Fragen) oder das Wiedererkennen (bei geschlossenen Fragen) von Wissensinhalten gefordert (Eckert, 1998).

Auswertung: I.d.R. liegen als Ergebnis von Befragungen Verbalprotokolle oder schriftliche Aussagen vor, die entsprechend eines vorgegebenen Schemas inhaltlichen Kategorien zugeordnet werden müssen, um die Aussagen quantifizieren und vergleichbar machen zu können. Nach Eckert (1998) erfolgt die Auswertung meist über die Aufsummierung der einschlägig beantworteten (geschlossenen) Fragen, was die Auswertung recht ökonomisch ausfallen läßt, aber andererseits die Freiheiten der Probanden erheblich einschränkt. Wurden offene Fragen verwandt, so werden die entsprechenden Antworten oftmals mittels einer Inhaltsanalyse kategorisiert, die Aussagen lassen sich aber auch als Netzwerkstruktur oder Propositionssystem darstellen (Eckert, 1998).

Fazit: Bungard, Holling und Schultz-Gambard (1996) sowie Ericsson (2006) bewerten Interviews als die wichtigste Methode der Datenerfassung in der arbeitspsychologischen Forschung, da nach Seiwald (2003) sowie Sedlmeier und Renkewitz (2007) insbesondere offene Frageformate einen hohen diagnostischen Informationsgehalt aufweisen. Nach Einschätzung von Kluwe (1988) hat die Art der Fragestellung entscheidenden Einfluß auf die Güte des Verfahrens, eine zentrale Problematik dabei ist, ob der Proband die Frage auch versteht, da ansonsten eine Beantwortung schlecht möglich ist. Nach Eckert (1998) ist bei falschen Antworten nicht eindeutig klärbar, ob dies

auf fehlerhaftes Wissen oder auf Verständnisschwierigkeiten zurückzuführen ist. Der Versuchsleiter selbst scheint aufgrund der Gesprächssituation ebenso eine kritische Größe zu sein, insbesondere beim strukturierten Interview muß dieser selbst über ein hinreichendes Fachwissen verfügen (Kluwe, 1988; Großmann und Teske-El Kodwa, 1995). Hoffmann und Lintern (2006) bewerten Interviews, v.a. unstrukturierte, als relativ uneffektive Methoden zur Wissenserfassung.

2.2.4.2. Lückentexte

Vorgehen: Der Lückentext ist den sog. Ergänzungsaufgaben zuzuordnen, bei der den Probanden keine festen Kategorien für die Antwort vorgegeben sind (Bühner, 2006), sondern innerhalb eines Textes Lücken selbständig vervollständigt werden sollen durch sinnvolle Satzteile. Als Grundlage für die Konstruktion eines geeigneten Lückentextes können z.B. Textstellen aus der jeweiligen Fachliteratur der relevanten Wissensdomäne herangezogen werden.

Grundannahmen: Lückentexte fordern vom Probanden eine Reproduktion von Wissensinhalten, welche nach Kluwe (1988) ein explizites Erinnern von zuvor explizit eingprägten Inhalten voraussetzt. Bühner (2006) weist auf die Suggestivwirkung der Formulierungen (also der zu vervollständigenden Sätze) hin, welche den Reproduktionsprozeß idealerweise steuern.

Auswertung: Die Auswertung von Lückentexten erfolgt auf Basis der Auszählung richtig ergänzter Begriffe, welche den Satz im Ganzen zu einer sinnvollen und inhaltlich richtigen Aussage werden lassen.

Fazit: Das freie Antwortformat des Lückentextes läßt Raten oder Zufallslösungen unwahrscheinlich werden. Bühner (2006) weist allerdings gleichsam auf die Gefahr von sog. Ketteneffekten hin, wonach das Nichtwissen des einen Begriffes u.U. auch das Wissen des nächsten Begriffes unwahrscheinlich werden läßt. Zudem kann die Auswerteobjektivität eingeschränkt werden, wenn mehrere Begriffe den Satz sinnvoll ergänzen – Bühner (2006) empfiehlt daher, a priori sämtliche Musterlösungen festzulegen.

2.2.5. Beispiele für methodische Zugänge, welche differenzierte Kenntnisse des Untersuchers erfordern

Viele der zur Verfügung stehenden wissensdiagnostischen Verfahren erfordern vom Versuchsleiter eine relativ differenzierte Kenntnis der zu untersuchenden Domäne, um dieses Verfahren überhaupt einsetzen zu können. Im Folgenden soll eine kleine Auswahl der gebräuchlichsten und für die vorliegende Arbeit relevanten Methoden vorgestellt werden: das Assoziationsverfahren wird an dieser Stelle kurz eingeordnet, ferner werden Ratingverfahren, Strukturlege- und Mappingtechniken sowie der Assoziations-Struktur-Test vorgestellt.

2.2.5.1. Assoziationsverfahren

Vorgehen: Der Assoziationsversuch als ein Verfahren der freien Reproduktion ist hinsichtlich der Domäne sowie des Antwortformates eingegrenzt: es werden dem Probanden Stimulusbegriffe einer umschriebenen Wissensdomäne vorgegeben, auf die hin er diejenigen Begriffe nennen soll, die in seinem Gedächtnis durch semantische Relationen mit diesem Stimulus verbunden sind (Rothe,

1994b). Eckert (1998) umschreibt dieses Vorgehen als eine durch Stimuli angeregte freie Reproduktion (sog. „cued recall“).

Grundannahmen: Das Assoziieren beruht auf der Annahme einer netzwerkanalogen Repräsentation von Wissenselementen, so daß sich zumindest deklarative Inhalte damit abbilden lassen (Rothe, 1994b). Nach Eckert (1998) soll Aufschluß über die Organisation von deklarativen Wissensinhalten gewonnen werden, indem die ordnungsstiftenden Merkmale der Begriffe analysiert werden. Für weitergehende Ausführungen zu den theoretischen Grundannahmen sei auf Kapitel 3.3 (Seite 33 ff.) verwiesen.

Auswertung: Das Erheben mittels Assoziationstechniken erbringt in erster Linie qualitative Daten, wobei in der weiteren Datenverarbeitung eine Quantifizierung angezeigt sein kann (Eckert, 1998), z.B. durch das Auszählen von Assoziationen, die einer bestimmten inhaltlichen Kategorie angehören. Der gebräuchlichste Kennwert ist dennoch rein quantitativer Natur: anhand der mittleren Anzahl assoziierter Begriffe kann zwischen Probanden differenziert werden (Rothe, 2003, 2006). Für weitergehende Ausführungen zur Auswertung des Wortassoziationsversuchs sei auf Kapitel 3.6 (Seite 39 ff.) verwiesen.

Fazit: Rothe (1994b) gesteht der Assoziationstechnik eine herausragende Rolle innerhalb der gedächtnispsychologischen Methoden zu, insbesondere kann sie einen bedeutsamen Beitrag leisten zur Vervollständigung des hypothetischen Wissensmodells. Eckert (1998) bewertet das Assoziationsverfahren als sehr ökonomisch.

2.2.5.2. *Ratingverfahren*

Vorgehen: Beim Ratingverfahren werden paarweise vorgegebene Begriffe hinsichtlich ihrer Ähnlichkeit verglichen. Dabei soll der Proband diese Ähnlichkeit quantitativ einschätzen, indem er sie auf einer mehrstufigen Skala beurteilt (Eckert, 1998).

Grundannahmen: Nach Eckert (1998) soll damit Aufschluß über die Organisation von deklarativen Wissensinhalten gewonnen werden.

Auswertung: Ratingverfahren erbringen direkt quantitative Daten, da ausdrücklich vom Probanden ein metrisches Urteil über das Ausmaß der Ähnlichkeit abgefordert wird (Eckert, 1998). Der Rückschluß von den Ratingmaßen auf die vermutete zugrundeliegende Wissensstruktur der interessierenden Begriffe und deren inhaltliche Deutung obliegt jedoch im Wesentlichen beim Untersucher.

Fazit: Eckert (1998) kritisiert v.a. die hohe Belastung der Aufmerksamkeit der Probanden durch die paarweise Beurteilung der Begriffe, wenn der Begriffspool einen gewissen Umfang hat, und plädiert daher eher für einen geringeren Umfang des Reizmaterials. Zudem sieht Eckert (1998) insbesondere beim Ratingverfahren die ggf. komplexen Überlegungen zu stark verdichtet, auf der anderen Seite bewertet er dieses Verfahren als sehr ökonomisch.

2.2.5.3. *Strukturlege- und Mappingtechniken*

Vorgehen: Die Strukturlegetechnik beschreibt eine Klasse von Visualisierungsverfahren, bei denen Wissenselemente bzw. Wissenszusammenhänge vom Probanden in einer netzwerkartigen Struktur

angeordnet werden sollen. Dabei gibt es eine Vielzahl von Variationen, die sich auf die Gewinnung der Wissens Elemente (Konzepte) sowie der anzuwendenden Regeln beim Verbinden dieser Konzepte beziehen (ein kurzer Abriß verschiedener Variationen findet sich z.B. bei Dann, 1992; Bungert, 1992; Stracke, 2004), wobei sich für die Erhebung der zu legenden Konzepte insbesondere qualitative Methoden wie z.B. das Assoziieren, lautes Denken oder Interviews eignen (Groeben, 1992). Dann (1992) weist darauf hin, daß es kein einheitliches Strukturlegeverfahren für sämtliche Fragestellungen geben kann, vielmehr sollten die konkreten Durchführungsbedingungen spezifisch entsprechend des Untersuchungszweckes gewählt werden. Stracke (2004) definiert sog. Concept Maps als Netzwerke, welche in graphischer Form die Zusammenhänge von Schlüsselbegriffen eines Wissensgebietes darstellen. Nach Stracke (2004) kann somit das Concept Mapping synonym zum Strukturlegen genutzt werden, lediglich das Mind Mapping grenzt sie ab, da dabei die Knoten streng hierarchisch und ohne Querverbindungen untereinander angeordnet sind.

Grundannahmen: Bei Strukturlege- bzw. Mappingverfahren werden die Begriffe (Wissenselemente) als Knoten innerhalb eines Netzwerkes aufgefaßt, während die Verbindungen (sog. Kanten) zwischen diesen Konzepten die semantischen Relationen zwischen ihnen darstellen. Die dahinterstehende Annahme ist die des semantischen Netzwerkes (Eckert, 1998): das Wissen um Sachverhalte oder Zusammenhänge ist in Form von untereinander vernetzten Begriffen kodiert und kann folglich durch graphische Verfahren abgebildet werden (Mandl und Fischer, 2000).

Auswertung: Als wichtigen Aspekt bei der Interpretation von Strukturen benennt Birkhan (1992) die Widerspruchsfreiheit. Die Auswertung graphischer Strukturen beinhaltet überwiegend eine formale Analyse der Netze (Mandl und Fischer, 2000), die mathematische Graphentheorie von Bonato (1990) lieferte dazu Strukturparameter wie z.B. Dichte, Durchmesser oder Umfang der gelegten Strukturen (siehe auch Eckert, 1998). Klix (1988) tritt für eine qualitative Auswertung ein, da seiner Ansicht nach die Komplexität und auch die Kompliziertheit solcher Begriffsnetze abhängig ist von der Verschiedenartigkeit sowohl der Knoten als auch der Kanten – eine rein quantitative Beschreibung würde dem nicht gerecht werden. Auch Rothe (2003) kritisiert, daß den Inhalten der Knoten zu wenig Beachtung geschenkt wird zugunsten dieser rein formal abgeleiteten quantitativen Kenngrößen.

Fazit: Hoffmann und Lintern (2006) bewerten Mapping-Techniken als die effizientesten Methoden für die Wissenserfassung, welche v.a. nützlich sind, um Wissenslücken aufzuzeigen. Allerdings ist die Erhebung mittels Strukturlegetechnik o.ä. Verfahren sehr zeitaufwendig (Eckert, 1998), ebenso die Auswertung (z.B. Stracke, 2004). Das Verfahren stellt dabei hohe Anforderungen an den Probanden: neben der Verbalisierungsfähigkeit betont Dann (1992) v.a. die Abstraktionsfähigkeit und fordert daher, den Schwierigkeitsgrad (d.h. die Komplexität der Strukturregeln) entsprechend an die Probanden anzupassen. Nach Bungert (1992) ist wesentliche Voraussetzung der Strukturlegetechnik, daß die Probanden die Regeln für die Strukturierung verstanden haben, Eckert (1998) empfiehlt daher ein einführendes Training. Rothe (2003) beschreibt v.a. bei Probanden aus dem nicht-akademischen Bereich Verständnisprobleme bzgl. der semantischen Relationen. Auch der quantitative Umfang der einbezogenen Konzepte kann nach Groeben (1992) eine kognitive Überforderung für den Probanden darstellen. Als weiteres Problem sieht Groeben (1992) die Größe der Wissens Einheiten, die inter- und auch intraindividuell variieren kann. Nach Birkhan (1992) unterliegen Strukturlegeverfahren einer gewissen Reaktivität durch die Verfahrensanwender.

Stracke (2004) betont die breite Einsatzfähigkeit dieser Verfahren: neben der Diagnose von aktuellem Wissen können Wissensnetze auch als Lehrmittel oder als Lernhilfe genutzt werden.

2.2.5.4. Assoziations-Struktur-Test

Vorgehen: Der von Meyer (2007) entwickelte computerbasierte Assoziations-Struktur-Test verbindet die Assoziationstechnik mit dem Ratingverfahren zu einem Verfahren, mit dem das strukturelle Wissen erfaßt und dargestellt werden kann. Dabei assoziieren Probanden zunächst zu vorgegeben Stimuli einer Wissensdomäne, anschließend werden diese Assoziationen paarweise präsentiert und sollen anhand einer Skala hinsichtlich ihrer Ähnlichkeit bewertet werden.

Grundannahmen: Die theoretischen Grundannahmen des Wortassoziationsversuchs sowie hinsichtlich des Ratingverfahrens können auch für den Assoziations-Struktur-Test geltend gemacht werden. Insbesondere die Befunde der Ratings zieht Meyer (2007) heran, um v.a. die nicht verbalisierbare Dimension (sog. tacit knowledge) des individuellen Wissens zu explizieren.

Auswertung: Im Ergebnis des Assoziations-Struktur-Tests steht ein vom Versuchsleiter auf Basis der Assoziation sowie der Paarvergleich-Ratings und mit Hilfe von Pathfinder-Algorithmen rekonstruiertes Netz, welches das implizite strukturelle Wissen des Probanden repräsentiert. Die graphentheoretischen Strukturparameter dieses Netzes sind analog zu denen der Netze im Strukturlegen (siehe auch Bonato, 1990; Eckert, 1998).

Fazit: Bislang liegen nur wenige Erkenntnisse über die Bewährung des Assoziations-Struktur-Tests vor, die bisherigen Befunde legen aber nahe, daß dieses Tool wertvolle Aussagen liefern kann über die Struktur der individuellen Assoziationen (Meyer, 2007).

2.2.6. Vergleich verschiedener methodischer Zugänge

Die soeben vorgestellten Methoden zur Wissensdiagnose stellen nur eine kleine Auswahl möglicher Instrumente dar, weitere Ansätze werden beispielsweise bei Kluwe (1988), Mandl und Huber (1994) oder Eckert (1998) erörtert. Die Entscheidung für einzelne Instrumente bzw. deren Anordnung im Untersuchungsdesign ist nicht nur abhängig von praktischen Erwägungen (z.B. der Ökonomie eines Verfahrens oder dessen organisatorischer Umsetzung im betrieblichen Rahmen) und den bereits besprochenen unterschiedlichen Anforderungen der Verfahren an das Vorwissen des Untersuchers, darüber hinaus ist auch zu berücksichtigen, welche Art von Wissen im Zentrum des Interesses steht (z.B. deklaratives vs. prozedurales Wissen) bzw. welche Funktion die Wissensdiagnose erfüllen soll (z.B. zur Individualdiagnose).

Rothe (1994b) unterscheidet in eine globale und eine differenziertere Wissenserfassung. Auf der globalen Ebene sollen die elementaren Wissensbausteine ermittelt werden, Rothe (1994b) sowie Timpe, Rothe und Seifert (1994) halten dafür die Methode des lauten Denkens oder der Gedankenstichprobe für geeignet, ebenso Interviews, schriftliche Befragungen sowie Gruppendiskussionen. Dokumentenanalysen sollten dabei immer am Beginn jeglicher Erhebungen stehen, um dem Untersucher das Grundwissen über die Domäne zu verschaffen. Die Funktion einer differenzierten Analyseebene sieht Rothe (1994b) in der Aufdeckung von Wissensstrukturen, v.a. der semantischen Relationen zwischen den jeweiligen Wissenseinheiten, so daß daraus ein hypothetisches Wissensmodell konstruiert werden kann. Dabei können Verfahren wie der

Wortassoziationsversuch, Strukturlegetechniken oder Analogieexperimente (nach dem Paradigma des analogen Schließens) zum Einsatz kommen, welche aber immer auf Seiten des Versuchsleiters eine entsprechende Modellvorstellung des Wissenskörpers voraussetzen (Rothe, 1994b; Timpe, Rothe und Seifert, 1994).

Hacker (2005b) differenziert zwischen dyadischen Zugängen in der Wissensdiagnose und Gruppenprozessen. Das Spektrum der letztgenannten Kategorie beschränkt sich auf Gruppendiskussionen in Form von Befragungen (als aufgabenbezogener Informationsaustausch) sowie auf die Korrektur, Ergänzung oder Bestätigung von vorgegebenen Wissenskörpern oder Teilen davon mit Hilfe einer Diskussion. Großmann und Teske-El Kodwa (1995) bewerten die durch Gruppenprozeduren erzeugten Wissensmodelle aufgrund der konsensualen Validierung als valider als bei individualisierten Methoden. Die dyadischen Ansätze reichen von mündlichen oder schriftlichen Befragungen mit unterschiedlich ausgeprägten Strukturierungshilfen, Ergänzungs- und Sortieraufgaben bis hin zur experimentellen Abklärung von Einzelaspekten (z.B. durch Informationsentzug), die Methode des lauten Denkens bewertet Hacker (2005b) dabei als diejenige Methode mit dem geringsten Ausmaß an Eingriff durch den Versuchsleiter.

Als weitere Möglichkeit der Klassifikation verschiedener Methoden schlägt Immenroth (2003) die Unterscheidung in eine offline- vs. online-Erhebung vor. Dabei umfassen online-Methoden solche, die während der Tätigkeitsausübung ihre Daten erheben, z.B. die Methode des lauten Denkens, Verhaltensbeobachtungen oder physiologische Messungen, während offline-Methoden nicht auf die Ausübung einer konkreten Tätigkeit angewiesen sind. Immenroth (2003) ordnet dieser Gruppe beispielsweise die Strukturlegetechnik zu, Computersimulationen, Szenariotechniken, tutorielle Dialoge, Interviews oder Verfahren der freien Reproduktion.

Über diese konzeptuellen Überlegungen hinaus beschäftigte sich Rothe (1994b) systematisch mit der Ergiebigkeit verschiedener Methoden zur Wissensdiagnose. Hinsichtlich der Erfassung von *deklarativem*, verbalisierbarem Wissen konnte er belegen, daß die vorherige Bearbeitung einer Problemlöseaufgabe, bei der das relevante Wissen vom Probanden aktiv genutzt wurde, zu einer Zunahme der verbalisierbaren Wissensseinheiten im Interview führte, in Abgrenzung zu der Situation, wenn eine solche unmittelbar vorherige Auseinandersetzung nicht stattfand (Rothe, 1994b), das Interview selbst erweist sich laut Hacker (1995, 1996) als deutlich ergiebiger, wenn Strukturierungshilfen dabei eingesetzt werden. Einerseits war das Interview hinsichtlich der Menge der erfaßten Wissensseinheiten der Methode des lauten Denkens überlegen, aber andererseits weit weniger ergiebig als eine schriftliche Befragung. Rothe (1994b) führt den letztgenannten Befund auf den Wegfall der sozialen Faktoren bei der Wissenserhebung zurück sowie auf die mögliche Entlastung des Arbeitsgedächtnisses bei der schriftlichen Formulierung von Wissen. Nach den Erkenntnissen von Rothe (1994b) erbringt die Gruppendiskussion keinen Mehrgewinn gegenüber der schriftlichen Befragung, kann aber zur Absicherung der bisherigen Analysen eingesetzt werden.

Mit Blick auf das *prozedurale* Wissen jedoch war die Methode des lauten Denkens während der Bearbeitung einer Problemlöseaufgabe ergiebiger als ein Interview, woraufhin Rothe (1994b) das laute Denken als unentbehrlich für die Erfassung prozeduralen Wissens bewertet. Er verglich außerdem das laute Denken, welches eine kontinuierliche Kommentierung durch die Probanden erlaubt, mit der Technik der Gedankenstichprobe, bei der im Handlungsvollzug Pausen eingelegt werden, in denen der Proband seine Gedanken zusammenfaßt. Danach konnte beim lauten Denken

mehr Wissen erfaßt werden als durch die Gedankenstichprobe, Rothe (1994b) führt dies auf die Unkontrollierbarkeit von Anzahl und Dauer der Verbalisierungsphasen bei der Gedankenstichprobe zurück. Im Vergleich von lautem Denken und der Arbeitsbeobachtung konnte Rothe (1994b) zwar bei der Beobachtung mehr prozedurales Wissen erfassen, er sieht darin aber noch keine Überlegenheit dieses Verfahrens, da die Ergebnisse nicht unerheblich von der Differenziertheit der Protokolle abhängig sind sowie vom Typ der zu lösenden Aufgaben. Zudem ist bei der Arbeitsbeobachtung nur die motorische Performanz zugänglich, nicht die dahinterstehenden kognitiven Elemente (Rothe, 1994b). Rothe und Timpe (1997) merken außerdem an, daß sich Handlungssequenzen zwar am exaktesten durch Beobachtung ermitteln ließen, dies aber ein langwieriger Prozeß sei und aufgrund der vielen unvorhersehbaren Störungen ein in der Praxis oft schwer zu realisierender Ansatz. Leider liegen keine Erkenntnisse dahingehend vor, wie der Wortassoziationsversuch sich hinsichtlich seiner Ergiebigkeit gegenüber anderen Verfahren behaupten kann.

Für die Modellierung des hypothetischen Wissensmodells bewertet Rothe (1994b) den Wortassoziationsversuch als geeignet, um die für eine Wissensdomäne relevanten Begriffe zu erfassen oder im Wissensmodell zu ergänzen. Die Strukturlegetechnik kann zur weiteren Präzisierung des Wissensmodells dienen, während Analogieexperimente ihrerseits das Strukturlegen ergänzen können, indem sie die gelegten semantischen Strukturen verifizieren und damit ggf. Verständnisschwierigkeiten seitens der Probanden erkennen lassen. Rothe (1994b) betont, daß sich durch eine entsprechende Kombination dieser Methoden eine höhere Effektivität und Sicherheit bzgl. des hypothetischen Wissensmodells erreichen läßt als durch den Einsatz der einzelnen Verfahren, weist aber gleichsam darauf hin, daß solche Methoden zur differenzierten Wissenserfassung immer mit einem hohen Zeitaufwand verbunden sind.

3. Der Wortassoziationsversuch

Hinsichtlich der methodischen Einordnung des Wortassoziationsversuchs sowie seiner Durchführung gibt es noch viele offene Fragen (Rothe, 2006). Zwar wurde die Assoziationstechnik vielfach in der gedächtnispsychologischen Forschung genutzt, aber gemäß Rothe (2006) sind kaum Anwendungen für praktisch relevante Wissensdiagnosen bekannt. Die Methoden innerhalb eines Assoziationsexperimentes sind vielfältig. Woodworth und Schlosberg (1965) unterscheiden zwei Ebenen: zum einen die freie vs. kontrollierte (gebundene) Assoziation, zum anderen die diskrete (einzelne) vs. fortlaufende Assoziation. Aus der Kombination dieser beiden Ebenen ergeben sich vier Experimentalansätze: bei einer diskreten freien Assoziation wird der Proband instruiert, lediglich das erste Wort zu nennen, welches ihm auf einen Stimulus hin einfällt. Bei der diskreten kontrollierten Assoziation hingegen soll ebenfalls nur ein einziger Begriff benannt werden, dessen Charakteristik ist aber durch die Instruktion eingeschränkt auf eine spezifische Art (z.B. soll sie in einer bestimmten semantischen Relation zum Stimulus stehen). Eine inhaltliche Einschränkung ist auch bei der kontinuierlichen kontrollierten Assoziation gegeben, aber es dürfen vom Probanden so viele Begriffe benannt werden, wie ihm spontan einfallen. Bei der kontinuierlichen freien Assoziation hingegen bestehen keine Einschränkungen: der Proband darf alle und so viele Begriffe benennen, wie ihm auf den Stimulus hin einfallen. Strube (1984) benennt zudem noch die

Unterscheidung in eine fortgesetzte vs. fortlaufende Assoziation: während bei fortgesetzter Assoziation alle genannten Konzepte einen Bezug zum Stimulusbegriff aufweisen sollen, ist es bei der fortlaufenden Assoziation gestattet, sich vom ursprünglichen Stimulus zu entfernen, indem geäußerte Konzepte als neues Reizwort für weitere Assoziationen dienen.

Der Wortassoziationsversuch, so wie er in der vorliegenden Arbeit und für den praktischen Einsatz im Arbeitskontext verstanden wird, ist entsprechend dieser Kategorisierung als kontinuierliche Assoziation zu charakterisieren, da vom Probanden so viele Begriffe benannt werden sollen wie möglich, und ebenso auch als freie Assoziation, da sicherlich eine Eingrenzung des Suchraumes in der Form besteht, daß nur fachspezifisches oder berufsrelevantes Wissen assoziiert werden soll, aber ansonsten keine Einschränkungen hinsichtlich der Merkmale der Begriffe gemacht werden.

3.1. Historische Einordnung von Assoziationsversuchen

Die Technik der Wortassoziation ist innerhalb der psychologischen Wissenschaften schon sehr frühzeitig eingesetzt worden, wobei zunächst das Verständnis von Lernprozessen sowie die psychologisch-psychiatrische Diagnostik im Vordergrund standen – der Einsatz zur Wissensdiagnose war eher nachrangig. Eine ausführliche Darstellung zur Historie findet sich bei Strube (1984) oder auch Woodworth und Schlosberg (1965). Die ersten Assoziationsexperimente gegen Ende des 19. Jahrhunderts gehen auf Francis Galton und Hermann Ebbinghaus zurück (siehe Woodworth und Schlosberg, 1965; Hiltmann und Luhr, 1982). So beschäftigte sich z.B. Ebbinghaus (1885 / 1992) mit dem Auswendiglernen von Listen mit sinnlosen Silben und konnte bereits damals belegen, daß Silbenpaare leichter produziert werden können, wenn zwischen ihnen eine assoziative Verknüpfung besteht. Während sich die ersten systematischen Untersuchungen also mit dem Assoziationslernen im Zusammenhang mit dem Verständnis von Lernmechanismen auseinandersetzten (z.B. Wreschner, 1904 / 2004), gewann die Technik der Assoziation zunehmend an Bedeutung bei der klinisch-psychiatrischen Diagnose von Persönlichkeitsveränderungen oder anderen pathologischen Zuständen i.S. eines differentialdiagnostischen Instrumentes. Zu Beginn des 20. Jahrhunderts wurden Assoziationstechniken sogar in der forensischen Psychologie eingesetzt, und in der Tiefenpsychologie fand die freie Assoziation durch Sigmund Freud eine große Verbreitung als projektives Verfahren (siehe Strube, 1984). Zwar hat der Wortassoziationsversuch seinen Stellenwert als projektive Methode eingebüßt (Hiltmann und Luhr, 1982), dennoch wird er nach wie vor in der kognitionspsychologischen oder linguistischen Forschung eingesetzt (z.B. Gegner, 1993).

Große Verbreitung erlebte in jüngerer Zeit der sog. Implizite Assoziationstest (IAT), welcher von Greenwald, McGhee und Schwartz (1998) entwickelt wurde. Dabei wird anhand eines computerbasierten Paradigmas die Assoziationsstärke (operationalisiert über Reaktionszeiten) zwischen Konzepten ermittelt, so daß diese Methode zur Messung von Einstellungen, impliziter sozialer Kognitionen oder Stereotypen herangezogen werden kann. So wurden beispielsweise Vorurteile zwischen Bayern und Norddeutschen geprüft (Neumann, Ebert, Gabel, Gülsdorff, Krannich, Lauterbach & Wiedl, 1998), die Eigengruppenbewertung bei Ost- bzw. Westdeutschen (Kühnen, Schießl, Bauer, Paulig, Pöhlmann & Schmidhals, 2001) oder die Einstellung gegenüber der Benutzung von Kondomen (Marsh, Johnson & Scott-Sheldon, 2001). Allerdings ist die Brauchbarkeit dieses Paradigmas nicht vollends belegt (siehe z.B. Banse, Seise & Zerbes, 2001;

Gawronski & Conrey, 2004; Rothermund & Wentura, 2001; Steffens & Plewe, 2001), und wie bereits angedeutet geht es dabei nicht um die Diagnose von Wissen.

Im Kontext der Markt- und Werbepsychologie wird die Assoziationstechnik nach wie vor eingesetzt, um Motivationen, Emotionen oder Einstellungen gegenüber Produkten o.ä. zu erfassen. So nutzten beispielsweise Schenk und Ganschow (2002) bzw. Wetzstein, Nitschke, Polkehn und Wandke (2003) in experimentellen Studien die Techniken der freien Assoziation sowie des Strukturlegens, um die mit verschiedenen Musikstilen verbundenen Begriffe und Emotionen zu erfassen und so Assistenzsysteme zu gestalten, welche die Bedienung von Unterhaltungselektronik unterstützen sollen. In der Arbeitspsychologie hingegen sind relativ wenig praktische Anwendungen des Wortassoziationsversuchs bekannt (Rothe, 2006).

3.2. Zielsetzung des Wortassoziationsversuchs in der Arbeitspsychologie

Gemäß der Taxonomie von Tests nach Lienert und Ratz (1998) ist der Wortassoziationsversuch einzuordnen als nicht standardisiertes, verbales Verfahren, welches aufgrund seines freien Antwortformates als niedrigstrukturiert zu beurteilen ist. Henning (1999) bewertet den Wortassoziationsversuch als eine relativ einfache Form der Informationsgewinnung und hebt als Vorteil hervor, daß die Probanden mit ihren eigenen Worten antworten können und so selbst Inhalt und Ausmaß ihrer Äußerungen bestimmen. Nach Bredenkamp und Erdfelder (1996) setzen Reproduktionsaufgaben wie das Assoziieren die Verbalisierbarkeit der relevanten Gedächtnisinhalte voraus. Der Wortassoziationsversuch ist also in erster Linie geeignet, das deklarative Wissen eines Individuums zu erfassen (Meyer, Scholl und Zhang, 2007). Rothe (2006) spezifiziert seine Anwendung auf die Diagnose von verbalisierbarem, fachspezifischem Fakten- und Handlungswissen.

Der Wortassoziationsversuch kann als ein Screening-Verfahren verstanden werden. Nach Häcker, Leutner und Amelang (1998) beschreibt ein Screening-Verfahren einen Test, der eine relativ grobe Analyse darstellt und somit nur eine unzureichende Charakterisierung von Personen liefern kann. Im Falle des Wortassoziationsversuchs wird also die Menge an assoziierten Konzepten interpretiert i.S. eines umfangreichen individuellen Wissensbesitzes. Ob der Proband die von ihm benannten Begriffe tatsächlich verstanden hat oder er lediglich in der Lage war, ein Schlagwort zu nennen, kann der Wortassoziationsversuch nicht aufklären. Dabei unterstellt man Folgendes: a) die Assoziation steht in einer sinnvollen semantischen Relation zum Stimulus, und b) der Proband kann die von ihm assoziierten Begriffe nicht nur benennen, sondern dieses Wissen auch sinnvoll anwenden. Ob dem so ist, kann und will der Wortassoziationsversuch nicht klären. Er unterstellt diese Annahmen zugunsten der Ökonomie, für die Überprüfung dieser Unterstellungen müssen andere, differenziertere und damit aufwendigere Verfahren herangezogen werden (z.B. die Strukturlegetechnik). Woike (2003) charakterisiert Screening-Verfahren durch eine i.d.R. eingeschränkte Analysetiefe zugunsten einer relativ (zeit-)ökonomischen Durchführung, nach Amelang und Schmidt-Atzert (2006) stellt die Durchführungszeit eines Verfahrens ein zentrales Kriterium für dessen Einsatz in der Praxis dar. Die vorliegende Arbeit wird zeigen, daß der Wortassoziationsversuch diesem Anspruch gerecht wird. Zumeist stehen Screening-Verfahren am Beginn eines mehrstufigen sequentiellen Diagnoseprozesses (z.B. Häcker und Sapf, 2004; Kubinger, 2006).

Die Ergebnisse eines Wortassoziationsversuchs reichen dennoch häufig aus, um rein praktische Fragestellungen zu beantworten, beispielsweise um Aussagen über den Erfolg einer Trainingsmaßnahme treffen zu können. Für die detailliertere Charakterisierung des individuellen Gedächtnisbesitzes jedoch sind Daten über die Organisation bzw. Struktur der Wissensbestandteile notwendig, welche der Wortassoziationsversuch nur in Ansätzen bieten kann, z.B. anhand der Position einzelner Assoziationen innerhalb einer Assoziationskette oder anhand der Pausenzeiten während des Assoziationsprozesses (Rothe, 2006). Ein solches Mehr an Strukturdaten kann beispielsweise durch den fortführenden Einsatz der Strukturlegetechnik gewonnen werden. Diese erfaßt über die deklarativen Wissensbestandteile (welche der Wortassoziationsversuch hervorbringt) hinausgehend die relationalen Beziehungen zwischen diesen und kann damit differenziertere Aussagen über das Verständnis des Wissensgegenstandes machen. So konnten beispielsweise Beatty und Gerace (2002) die konzeptuelle Wissensstruktur von Physik-Studierenden abbilden, indem sie die Probanden im Wortassoziationsversuch zu verschiedenen Stimuli assoziieren und anschließend in einer Concept Map die Querverbindungen zwischen diesen so erfaßten deklarativen Wissens-elementen konstruieren ließen. Nach Häcker, Leutner und Amelang (1998) ist eine solche weitergehende, differenziertere Diagnostik v.a. dann angezeigt, wenn weitreichende Entscheidungen von einer solchen Bewertung des Wissens abhängen (wie z.B. bei Personalentwicklungsmaßnahmen).

Zusammenfassend kann also der Wortassoziationsversuch im Rahmen von arbeitspsychologischen Untersuchungen beschrieben werden als ein Screening-Verfahren, mit dem mittels freier Reproduktion das explizite, in erster Linie deklarative fach- oder arbeitsbezogene Wissen einer Arbeitsperson erfaßt werden soll. Dabei kann das Verfahren – entsprechend der von Rothe (1994a) getroffenen Unterscheidung – dienlich sein sowohl zur Wissenserfassung (also um dem Untersucher einen Einblick in die Wissensdomäne zu verschaffen) als auch zur Diagnose (um den individuellen Wissensstand eines Individuums zu bestimmen), wobei in der vorliegenden Arbeit nicht weiter nach diesen beiden Zielrichtungen differenziert wird. In jedem Falle ist zu berücksichtigen, daß der Einsatz des Wortassoziationsversuchs nur bei solchen Berufen oder Arbeitstätigkeiten sinnvoll ist, wo das explizite Wissen eine wesentliche Größe für berufliche Leistung darstellt.

3.3. Theoretischer Hintergrund des Wortassoziationsversuchs

Beim Wortassoziationsversuch werden von einem Probanden auf einen Stimulus hin all diejenigen Begriffe genannt, die mit diesem im Individualgedächtnis verknüpft sind. Dabei ist die Annahme einer netzwerkartigen Repräsentation zumindest für Fakten- und Handlungswissen interpretierbar (Rothe, 2003, 2006). Der Wortassoziationsversuch ist als Reproduktionsaufgabe zu charakterisieren, bei der – in Abgrenzung zu einer Wiedererkennensleistung – ein Proband die im Gedächtnis gespeicherten Informationen aktiv benennen muß (Bredenkamp und Erdfelder, 1996). Kluwe (1988) postuliert, daß Methoden zur Gewinnung von Daten über das Wissen geeignete Maßnahmen voraussetzen, welche jeweils spezifische Ausschnitte einer individuellen Wissensstruktur im Langzeitgedächtnis aktivieren. Bezogen auf das Netzwerkmodell bedeutet dies eine Erregungsausbreitung in den Segmenten des semantischen Netzwerkes (Kluwe, 1988). Beim Wortassoziationsversuch kommt diese Aufgabe dem Stimulusbegriff zu: er soll das interessierende

Netzwerk in Schwingung versetzen, diese über die bestehenden Kanten weiterleiten und so möglichst viele weitere Knoten erreichen im Sinne des sog. Aktivationsausbreitungsmodells (z.B. Schermer, 1991; Plötzer, 1996; Anderson, 2001). Wessells (1994) vergleicht dies damit, einen Stein auf eine ruhige Wasseroberfläche zu werfen, woraufhin sich die Wellen zunächst gleichmäßig in alle Richtungen ausbreiten. Die Ausbreitung der Schwingungen und damit die Aktivierung weiterer Begriffe ist aber nicht gleichmäßig, sondern die Stärke dieser Aktivierung verringert sich mit zunehmender Entfernung vom Stimulusknoten (sog. Dämpfungsannahme; Schermer, 1991).

Zudem besitzen die Kanten eines Netzwerkes unterschiedliche Zugänglichkeit und Stärke, diese hängen in erster Linie von der Nutzungshäufigkeit der Kanten ab (Schrameier, 1990; Schermer, 1991; Wessells, 1994; Anderson, 2001). Somit sind zwei Begriffe, die häufig gemeinsam auftreten, durch eine starke Kante miteinander verbunden, d.h.: wenn einer von ihnen aktiviert wird (Stimulus), dann ist auch der andere Begriff (Assoziation) leichter zugänglich und aktivierbar, nach Klix (1991) ist diese Strukturbildung auch im menschlichen Nervensystem nachweisbar, die gegenseitige Anregung passiert also auch auf neuronaler Ebene. Schermer (1991) bezeichnet dies als semantischen Priming-Effekt, Anderson (2001) als assoziatives Priming: die Aktivierung eines Konzeptes erleichtert den Zugriff auf ein semantisch verwandtes Konzept, wenn eine Verbindung zwischen beiden besteht. Die Stärke einer Verbindung ändert sich jedoch nur allmählich und steigt jedesmal ein wenig an, wenn diese Gedächtnisspur benutzt wird, ist also auch durch Übung trainierbar, was in unzähligen Paar-Assoziations-Lernaufgaben belegt werden konnte (Anderson, 2001).

Ulrich, Stapf und Giray (1996) weisen darauf hin, daß auch der Bekanntheitsgrad sowie die Worthäufigkeit in Texten oder der gesprochenen Sprache die Assoziationswahrscheinlichkeit für einen Begriff beeinflussen, denn diese objektive Häufigkeit sei ein Anhaltspunkt für den Grad der Erfahrung mit diesem Begriff. Übertragen auf den Wortassoziationsversuch und seine Anwendung im Kontext von Arbeitstätigkeiten bedeutet dies: ein Begriff, der im Tätigkeitsvollzug häufig genutzt wird (gelesen, gehört, gesprochen oder geschrieben), wird auch mit einer erhöhten Wahrscheinlichkeit benannt, wenn ein Stimulus aus diesem Tätigkeitsbereich präsentiert wird. Hinzu kommt die Bedeutungshaltigkeit dieses Begriffes für die konkrete Arbeitstätigkeit. Im Umkehrschluß kann man Assoziationsleistungen eines Probanden wie folgt interpretieren: Begriffe, die jemand in der Testsituation auf einen Stimulus hin assoziiert, sind auch im Arbeitsalltag eng mit diesem Stimulus verknüpft, d.h. werden von einer Arbeitsperson häufig wahrgenommen oder aktiv genutzt. Die individuelle Assoziationsleistung kann somit als ein Indikator angesehen werden für die im Arbeitsalltag relevanten Verknüpfungen zwischen Begriffen bzw. den dahinterstehenden Objekten, Zuständen, Merkmalen usw.. Ulrich, Stapf und Giray (1996) weisen darauf hin, daß die (subjektive) Bedeutungshaltigkeit i.d.R. einen größeren Einfluß hat als die objektive Häufigkeit an sich. Zusammenfassend ist die Wahrscheinlichkeit für einen Begriff, auf einen Stimulus hin assoziiert zu werden, also davon abhängig, wie weit (d.h. über wieviele Kanten hinweg) er vom Stimulus entfernt liegt bzw. wie stark die Kanten zwischen Stimulus und Assoziationsknoten ausgebildet sind, was wiederum dadurch beeinflusst wird, wie häufig diese Verbindung genutzt wurde oder wie subjektiv bedeutsam sie im individuellen Wissensbesitz ist. Anderson (2001) benennt außer der Nutzungshäufigkeit auch den Zeitpunkt des letzten Abrufs einer Verbindung als Determinante, welche die Höhe der Aktivationsausbreitung beeinflusst. Erst kürzlich genutzte

Verbindungen haben also eine erhöhte Wahrscheinlichkeit, bei erneuter Aktivierung des Stimulusbegriffes wieder benutzt zu werden, Anderson (2001) bezeichnet dieses Phänomen als Bahnung oder Priming von Verbindungen, weist aber auch darauf hin, daß Gedächtnisspuren ihre Aktivierung aber auch sehr schnell wieder verlieren können.

Hinsichtlich der Mechanismen beim Einprägen von Gedächtnisinhalten, welche sich dann auch beim Abruf niederschlagen, unterscheiden Ulrich, Stapf und Giray (1996) zwischen assoziativem und kategorialem Clustering, wobei die Clusteringtendenz in der freien Reproduktion Aufschluß gibt über die Organisation des Lernmaterials im Gedächtnis. Beim assoziativen Clustering werden Itempaare nach semantischer Zusammengehörigkeit wiedergegeben (beispielsweise zwei Begriffe, die immer in Kombination gebraucht werden, z.B. „Belastung – Beanspruchung“ oder „implizit – explizit“), während beim kategorialen Clustering die Begriffe entsprechend ihrer inhaltlichen Kategorien aufgezählt werden (z.B. zum Stimulus „Eßstörungen“: „Anorexie, Bulimie“ als Subtypen, „Heißhungerattacken, Erbrechen, Gewichtskontrolle“ als Phänomenologie, „Eduktion, Reizverhinderung, Kontingenzmanagement“ als Therapie usw.). Lernprozesse und somit die Wissenszunahme resultieren in einer zunehmend dichteren Vernetzung von Konzepten (Klix, 1991). Somit wird auch die Menge von aktivierten und benannten Knoten zum Indikator für einen umfangreichen Wissensbesitz – je dichter das Netz ist, desto höher ist die Wahrscheinlichkeit, daß ein bestimmtes Konzept über den einen oder anderen Pfad aktiviert wird.

3.4. Konstruktion des Wortassoziationsversuchs

Die zentrale Aufgabe im Vorfeld eines Wortassoziationsversuchs ist die Bestimmung der Stimulusbegriffe, denn Assoziationsergebnisse hängen immer direkt davon ab, welche Begriffe als Stimuli vorgegeben wurden (Rothe, 2006). Dazu sind die zentralen Bestandteile einer Wissensdomäne zu extrahieren, dabei kann es sich z.B. um Sachverhalte, Objekte oder Vorgänge handeln (Rothe, 2003, 2006). Um diesen Schritt zu leisten, ist es notwendig, daß sich der Versuchsleiter hinreichend mit der ausgewählten Domäne vertraut gemacht hat, z.B. durch die Analyse von Lehrbuchtexten, betrieblicher Unterlagen oder Tiefeninterviews mit Experten (Rothe, 2003, 2006). Im Rahmen von Weiterbildungs Evaluationen mag diese Domäne u.U. einen relativ geringen Umfang haben und daher mit wenigen Stimuli erfassbar sein, im Kontext der Erfassung von berufsspezifischem Wissen an sich jedoch ist das Wissensgebiet vergleichsweise umfangreich, so daß entsprechend mehr Stimuli notwendig scheinen, um die Repräsentativität für die Domäne zu gewährleisten. Kersting (2006) erinnert an das Dilemma, daß mit der Verlängerung eines Tests zwar i.d.R. die Reliabilität verbessert wird, dies aber zu Lasten der Testökonomie sowie der Akzeptanz durch die Versuchsteilnehmer geht. Je mehr Stimuli einbezogen werden, desto mehr Ausschnitte aus der Wissensdomäne werden im Wortassoziationsversuch berücksichtigt und reichen an den „wahren“ Wert des individuellen Gedächtnisbesitzes heran, allerdings steigt damit natürlich auch der zeitliche Aufwand eines Wortassoziationsversuchs und hat so möglicherweise auch Auswirkungen auf die Konzentrationsleistungen des Probanden oder seine Bereitschaft, sich aktiv am Versuch zu beteiligen (je weniger Assoziationen man auf einen Stimulus hin produziert, desto schneller ist dieser Teilversuch beendet). Es gilt also, mindestens so viele Stimuli einzubeziehen, um einen repräsentativen Ausschnitt der Wissensdomäne abbilden zu können, aber nur so viele wie nötig, um den Probanden nicht zu überfordern. Rothe (2003) empfiehlt die

Extraktion von zehn bis zwanzig Stimulusbegriffen. Eine empirische Absicherung einer minimal notwendigen bzw. maximal hinreichenden Menge an Stimuli liegt in der einschlägigen Literatur jedoch nicht vor.

Im Zusammenhang mit der Konstruktion psychometrischer Tests weist Yousfi (2005a) darauf hin, daß die klassische Testtheorie zwar empfiehlt, den Test aus möglichst vielen Items zusammenzustellen, daß mit zunehmender Testlänge aber nur dann eine Steigerung von Validität und Reliabilität erreicht wird, wenn die Parallelität der Items gewährleistet werden kann – andernfalls ziehe ein Mehr an Items eine sinkende Testgüte nach sich. Auf der anderen Seite empfehle die klassische Testtheorie, Tests aus möglichst trennscharfen Items zu konstruieren, was Yousfi (2005b) nur dann als einen guten Indikator für eine Erhöhung der Testgüte bewertet, wenn die Beziehungen zwischen den wahren Werten der Items klar sind. Überträgt man diese Überlegungen auf den Wortassoziationsversuch, so läßt sich daraus folgender Konstruktionsansatz ableiten: je mehr parallele Stimuli, desto sicherer wird die diagnostische Aussage, Trennschärfe zwischen den Stimuli ist in diesem Fall nicht unbedingt erforderlich.

Aus praktischer Sicht ist außerdem Folgendes zu bedenken: die Parallelität von Stimuli wäre aufgrund ihrer Bedeutungshaltigkeit nur zu realisieren durch den Gebrauch von Synonymen – echte Synonyme jedoch sind im Fachwissensbereich kaum zu finden. Trennscharfe Items hingegen, welche „gute“ von „schlechten“ Probanden unterscheiden können (z.B. Lienert und Raatz, 1998; Zöfel, 2003), scheinen aus praktischer Sicht ein lohnender Ansatz. Bewertet der Versuchsleiter oder der betriebliche Experte einen Teilbereich der untersuchten Wissensdomäne als basal für die Arbeitstätigkeit, so wird erwartet, daß *alle* Arbeitspersonen über dieses Wissen verfügen. Besteht ein Wortassoziationsversuch allerdings nur aus solchen Stimuli, so kann im Ergebnis u.U. nur ausgesagt werden, daß die Befragten eine „kritische“ Grenze i.S. von Mindestanforderungen erreicht haben oder nicht (wie z.B. bei der schriftlichen Führerscheinprüfung) – eine Differenzierung zwischen den Probanden ist in diesem Falle aufgrund der geringen Streuung der Assoziationskennwerte deutlich erschwert (bestenfalls noch im unteren Leistungsbereich möglich). Daher hat sich in der Praxis bewährt, zusätzlich noch einige Stimuli einzufügen, welche eine gewisse Trennschärfe besitzen, also von denen erwartet wird, daß nicht alle Probanden gleichermaßen viel dazu wissen, sondern v.a. die Leistungsstarken. Auf diese Weise kann letztlich eine Varianz innerhalb der Stichprobe gewährleistet werden, auch nach Lienert und Raatz (1998) differenziert ein Test am besten, wenn er insgesamt eine mittlere Schwierigkeit aufweist.

In der einschlägigen Literatur liegen keine Erkenntnisse vor hinsichtlich des optimalen Umfangs eines Wortassoziationsversuchs oder hinsichtlich der Eigenschaften der Stimulusinhalte. Die vorliegende Arbeit will daher folgenden Aspekt näher beleuchten: inwiefern die Anordnung von Stimuli verschiedener Abstraktionsebenen (also mit sehr weitem vs. eingegrenztem Bedeutungshof) das Assoziationsverhalten der Probanden beeinflusst.

3.5. Durchführung des Wortassoziationsversuchs

In der Terminologie nach Strube (1984) ist der Wortassoziationsversuch, wie er im arbeitspsychologischen Kontext eingesetzt wird, strenggenommen als Technik der kontrollierten – oder wie Strube (1984) es formuliert: restringierten – Assoziation zu bezeichnen, da die inhaltliche

Vorgabe gemacht wird, sich auf das für die Erledigung der Arbeitsaufgabe notwendige Fachwissen einzugrenzen. Bei einer freien Assoziation (im klassischen Sinne) hingegen gilt jede Antwort, inhaltliche Einschränkungen werden nicht gemacht, so daß danach beispielsweise auch emotionale Äußerungen oder persönliche Erlebnisse als Assoziationen gestattet wären. Beatty, Gerace und Dufresne (2002) sprechen im Zusammenhang mit dem Wortassoziationsversuch von einer quasi-freien Assoziation, da sich die Probanden lediglich auf Begriffe aus einem definierten Themenbereich beschränken sollen.

Nach Rothe (2003, 2006) soll der Proband nur in einzelnen Begriffen oder kurzen Phrasen antworten, ganze Sätze hingegen sind nicht erwünscht. Beatty, Gerace und Dufresne (2002) definieren die Größe eines Terms oder einer Assoziation im methodischen Sinne als ein bis maximal drei Worte, mit denen eine Idee beschrieben wird. Die Probanden werden zu freier und fortgesetzter Assoziation aufgefordert, bei der sie alles Wissen benennen sollen, was ihnen unmittelbar auf einen Stimulus hin einfällt (Rothe, 2003, 2006). Dabei wird sukzessive Stimulus für Stimulus bearbeitet, Rothe (2003, 2006) empfiehlt, den Einzelversuch (also zu einem Stimulus) zu beenden, wenn der Proband nach einer längeren Pause (zwei Minuten) keine Äußerungen mehr hervorbringt oder bekundet, daß ihm nichts mehr einfallt.

Rothe (2003) schlägt vor, den Probanden auf die Erhebung vorzubereiten, indem der Wortassoziationsversuch in dessen Arbeitskontext eingebettet wird. Dazu sollte entweder vor der Erhebung die betreffende Arbeitshandlung ausgeführt worden sein, oder der Versuchsleiter sollte innerhalb einer ausführlichen Instruktion den jeweiligen Aufgabenbereich ansprechen. Für den Wortassoziationsversuch sind keine besonderen räumlichen Voraussetzungen notwendig, lediglich die Störungsfreiheit während der Erhebung ist zu gewährleisten. Die zeitliche Inanspruchnahme ist zentral abhängig von der Anzahl der präsentierten Stimuli, Rothe (2003) bietet als Richtwert etwa 30 Minuten für zehn Stimuli an.

Der Wortassoziationsversuch ist sowohl mündlich als auch schriftlich durchführbar, was sich sowohl auf den Modus der Stimuluspräsentation seitens des Versuchsleiters als auch auf den der Assoziationsproduktion durch den Probanden beziehen kann. In der vorliegenden Arbeit wird unter Durchführungsmodus aus Gründen der praktischen Relevanz immer die Art der Äußerungen des Probanden verstanden. Bei der mündlichen Variante werden die Assoziationen idealerweise mittels eines Diktiergerätes o.ä. aufgezeichnet (Rothe, 2003), so daß sie zu einem späteren Zeitpunkt vom Versuchsleiter transkribiert werden können. Diese Form der Durchführung ist nur als Einzelversuch realisierbar, zumeist werden dabei durch den Versuchsleiter die Stimuli ebenfalls mündlich dargeboten, es ist aber auch eine gleichzeitige optische Präsentation des Stimulus denkbar. Die schriftliche Form des Wortassoziationsversuchs ist sowohl als Einzel- als auch als Gruppenversuch durchführbar, zumeist werden dabei die Stimuli gleichfalls schriftlich dargeboten. Die Sicherung der Assoziationsdaten ist bereits durch die schriftlichen Aufzeichnungen der Probanden gegeben.

Die wenigen Erkenntnisse zum Durchführungsmodus von Wissenstests und im Besonderen des Wortassoziationsversuchs sind uneindeutig. Bredenkamp und Erdfelder (1996) merken zwar an, daß bei schriftlicher Durchführung von Reproduktionsaufgaben die unterschiedliche Schreibgeschwindigkeit der Probanden sowie die Möglichkeit, bereits Reproduziertes einzusehen, sich als schwer kontrollierbare Störvariablen auswirken können, bringen dazu aber keine Belege vor. Die Übertragung eines Paper-Pencil-Formates in eine Computerversion wird v.a. unter dem

Aspekt der Durchführungs- und Auswerteökonomie sowie der höheren Standardisierung angestrebt (Fisseni, 2004). Fisseni (2004) sowie Troche, Rammstedt und Rammsayer (2002) empfehlen aufgrund ihrer Untersuchungen zum Vergleich von computerbasierten Versionen des Leistungsprüfsystems und Paper-Pencil-Formaten, nicht automatisch von einer Äquivalenz beider Formate auszugehen, da die Befunde dazu sehr heterogen sind. Als mögliche Determinanten benennen sie u.a. die erhöhten Aufmerksamkeitsanforderungen der computerbasierten Version durch den ständigen Wechsel zwischen Bildschirm und Tastatur, auf der anderen Seite die größere Flexibilität einer papierbasierten Variante, die z.B. ein späteres Nachtragen von Lösungen ermöglicht, aber auch durch vorausgehende oder nachfolgende Items ablenken kann.

Grabowski (1999) führte Experimente zum Durchführungsmodus (mündlich vs. schriftlich) bei der Wiedergabe einfacher listenartiger Wissensbestände (z.B. Hauptstädte benennen) durch. Dabei ging er von der Annahme aus, daß die Verbalisierung von Wissensinhalten aus dem Langzeitgedächtnis kognitive Ressourcen beansprucht, auf der anderen Seite aber nur eine begrenzte Verarbeitungskapazität verfügbar sein kann. Grabowski (1999) variierte dazu die Aspekte der Ausführungsgeschwindigkeit, Dichte der Ressourcennutzung (durch die Selbststeuerung in der Sprachproduktion) sowie das Vorhandensein eines Protokolls i.S. eines externen Speichers. Zwar stellt das Schreiben eine erhöhte kognitive Belastung dar aufgrund der länger dauernden graphischen Ausführung, andererseits aber erleichtert es das Arbeitsgedächtnis, da der bereits geschriebene Text wieder gelesen werden kann, während beim Sprechen die bereits verbalisierte Information im Gedächtnis gehalten werden muß (Grabowski, 2005). Im Ergebnis seiner Experimente kommt Grabowski (1999; 2005) zu dem Erkenntnis, daß die schriftliche Wiedergabe der mündlichen deutlich überlegen ist. Dabei ist nachrangig, ob die bereits produzierten Wissensinhalte für den Probanden weiterhin sichtbar sind (als externer Speicher), auch ist offenbar die Modalität innerhalb der schriftlichen Ausführung (Tastatur vs. handschriftlich) unerheblich (Grabowski, 2005). Vielmehr scheint der entscheidende Faktor die Ausführungsgeschwindigkeit des Sprachproduktionsprozesses zu sein (Grabowski, 1999; 2005). Grabowski (2005) räumt ein, daß diese Überlegenheit des schriftlichen Durchführungsmodus zumindest für einfache, listenartige Wissensbestände gilt. Strube (1984) hingegen postuliert hinsichtlich der Variationen sowohl in Darbietung der Stimuli als auch bei der Antwortmodalität in Assoziationsversuchen, daß der mündliche vs. schriftliche Modus nur äußerliche Unterschiede darstelle und das Assoziationsergebnis nicht beeinflusse. Allerdings ist dabei anzumerken, daß sich Strube's Erkenntnisse – ebenso wie diejenigen von Grabowski (1999; 2005) – auf Assoziationsexperimente mit Stimuli aus dem Allgemeinwissensbereich stützen, über fachspezifische Domänen macht er keine Aussagen.

Auf der einen Seite finden sich bei Grabowski (1999, 2005) also Belege für eine Überlegenheit des schriftlichen Modus, auf der anderen Seite aber auch für die Äquivalenz beider Modi (Strube, 1984). Allerdings ist die Aufgabenstellung in beiden Quellen nicht unmittelbar miteinander vergleichbar: während die Versuchsanordnung bei Grabowski (1999, 2005) als kontrollierte oder gebundene Assoziation verstanden werden kann (nur das Benennen von Hauptstädten), verlangte Strube (1984) eine freie Assoziation, bei der sämtliche Verknüpfungen mit dem Stimulus benannt werden konnten. Somit lassen sich die Erkenntnisse von Grabowski (1999, 2005) nicht unmittelbar auf den Wortassoziationsversuch, welcher in der vorliegenden Arbeit als freie Assoziation

verstanden wird, übertragen. Aber ebenso ist die Äquivalenz-Aussage von Strube (1984) nicht ohne weiteres übertragbar, da dieser nicht fachspezifische Wissensdomänen benutzte. Somit ist in der vorliegenden Arbeit zunächst empirisch zu klären, ob bei einer freien, fortgesetzten Assoziation mit Stimuli aus spezifischeren Wissensdomänen von einer Äquivalenz der mündlichen vs. schriftlichen Durchführung ausgegangen werden kann.

3.6. Auswertung des Wortassoziationsversuchs

Im Wesentlichen lassen sich zwei Ansätze bei der Auswertung des Wortassoziationsversuchs beschreiben: zum einen die quantitative vs. qualitative Herangehensweise, zum anderen die Analyse der Leistung eines Individuums vs. der Analyse auf der Ebene der Untersuchungsgruppe. Am verbreitetsten sind der Individualansatz verbunden mit einer quantitativen Analyse. Hiltmann und Luhr (1982) benennen folgende Variablen als Ergebnis klassischer Wortassoziationsversuche: die Reaktion (also das assoziierte Wort) sowie die Reaktionszeit (zwischen Stimulus und Assoziation). Zudem kann außerdem das Verhalten des Probanden herangezogen werden, beispielsweise durch ein anschließendes gelenktes Gespräch, um die Bedeutung der assoziierten Konzepte zu erörtern. Bereits Francis Galton nutzte in seinen ersten Assoziationsexperimenten gegen Ende des 19. Jahrhunderts die Reaktionszeit, die Häufigkeit der Wiederholung der gleichen Assoziationen sowie die Klassifikation der Assoziationen durch Auszählen innerhalb der Klassen als Methoden der Datenquantifizierung (siehe Woodworth und Schlosberg, 1965).

Strube (1984) macht darauf aufmerksam, daß nur die Instruktion zu einer kontinuierlichen Assoziation (in Abgrenzung zur Einzelantwort) Aufschluß über die Gesetzmäßigkeiten des Assoziierens erbringen kann, der Schlüssel dazu ist die inhaltliche und zeitliche Reihung von Nennungen. Folgerichtig ist dann auch die von Strube (1984) präferierte Darstellungsform individueller Assoziationsergebnisse ein Diagramm mit den Koordinatenachsen Zeit sowie Assoziationsanzahl, in dem die Inhalte der Assoziationen in ihrer dargebotenen zeitlichen Abfolge abgetragen sind. Ist man an den Reaktionszeiten interessiert, bietet sich v.a. die mündliche Durchführung des Versuchs an, da anhand der Tonaufzeichnungen die Reaktionszeiten bestimmt werden können. So konnte belegt werden, daß der zeitliche Verlauf der Assoziationen ungleichmäßig ist: die Antworthäufigkeiten nehmen kontinuierlich im Zeitverlauf ab, und es werden längere Pausen deutlich, welche die Assoziationskette in Cluster unterteilen lassen (Strube, 1984). Woodworth und Schlosberg (1965) bezeichnen die Annahme, daß die Stärke der assoziativen Verbindung sich durch eine schnellere Reaktionszeit ausdrückt, als Augenscheinvalidität. Zum einen begründen sie dies mit einer abnehmenden Reaktionszeit im Laufe von Lernprozessen, zum anderen mit der Tatsache, daß ein Stimulus ggf. mehrere verschiedene Assoziationen gleichzeitig anregen kann, der Proband diese aber nur sukzessive nennen kann, er muß also eine Auswahl treffen. Dennoch ist die Analyse der Reaktionszeiten nach wie vor eine lohnende Auswerteebene. Nach Strube (1984) gibt es Belege dafür, daß etwas semantisch Ähnliches in geringem Abstand voneinander assoziiert wird, was sich z.T. auch im zeitlichen Abstand niederschlägt. In jüngster Zeit ist eine akkurate Zeitmessung auch bei schriftlicher Durchführung möglich: mit Hilfe des computerbasierten Tools Assoziations-Struktur-Test von Meyer (2007) z.B. können die Zeitabstände zwischen den Assoziationen bestimmt werden. Beatty, Gerace und Dufresne (2002) konnten so belegen, daß diese Antwortzeiten mit der

Assoziationsstärke zwischen zwei Konzepten korrelieren, so daß aus diesen Daten Erkenntnisse über die Muster und Struktur im individuellen Wissensbesitz abgeleitet werden können. Zentrales Ergebnis des Assoziationsversuchs bleibt dennoch eine Liste von Assoziationen eines Probanden.

Nach Rothe (2003, 2006) gibt es für den Wortassoziationsversuch auch auf der Ebene des Individuums einen quantitativen und einen qualitativen Zugang bei der Auswertung. Zum einen wird demnach die Anzahl von Assoziationen eines Probanden herangezogen (als individueller Assoziationskennwert), zum anderen können Assoziationsketten nach inhaltlichen Kriterien klassifiziert werden. Nach Rothe (2006) sind die individuellen Assoziationskennwerte (als quantitativer Parameter) sensible Kenngrößen für das Ausmaß des individuellen Wissens und zudem gute Prädiktoren für die Arbeitsleistung, auch Stracke (2004) bewertet die Anzahl von Assoziationen einer Person als guten Hinweis auf die Breite der dahinterstehenden Wissensbasis. Die Aussagekraft des Assoziationskennwertes kann dabei noch erhöht werden durch die Hinzunahme qualitativer Aspekte, z.B. hinsichtlich der Herkunft des Wissens im Erwerbsprozeß oder durch Bildung domänenspezifischer Inhaltskategorien (Rothe, 2006).

Nach Müller (1998) ist die häufigste angewandte Methode bei der Quantifizierung von Assoziationsverteilungen die sog. "primary response", welche diejenige Antwort bezeichnet, welche am häufigsten genannt wurde. Eine Fortführung dieses Ansatzes i.S. der Auszählung der zweithäufigsten Nennung usw. ist ebenso möglich. Auch Rothe (2006) benennt als Variante eines quantitativen Kennwertes die durchschnittliche Häufigkeit und Position von identischen Assoziationen in den Assoziationsketten verschiedener Probanden einer Stichprobe. Allerdings hat der Ansatz der Positionsbestimmung einer Assoziation seine Grenzen. Zwar gibt es Belege dafür (s.o.), daß v.a. die erstgenannte Assoziation eine starke semantische Beziehung zum Stimulus aufweist, aber die Annahme von Strube (1984) impliziert, daß sich der Proband mit Fortschreiten seiner Assoziationskette immer mehr vom Stimulus entfernt und sich daher allein aus der Position innerhalb dieser Assoziationskette Schlußfolgerungen über die Beziehungsstärke ableiten ließen. Wie Woodworth und Schlosberg (1965) schon so treffend bemerkten – der Proband kann nicht alle Assoziationen mit gleicher Beziehungsstärke gleichzeitig nennen, sondern muß dies sukzessive tun. Zudem ist denkbar, daß ein Proband – im Bild des semantischen Netzwerkes gesprochen – vom Stimulus aus nicht nur einen einzigen Pfad einschlägt, sondern sukzessive mehrere Teiläste beschreitet und sich so immer wieder dicht an den Stimulus „zurückbewegt“.

Die Auswertung auf Ebene einer Gruppe von Probanden läßt sich als quantitative Inhaltsanalyse bezeichnen, so wie sie Bortz und Döring (2003) beschreiben: inhaltsgleiche Assoziationen von mehreren Probanden werden zusammengefaßt in einer Inhaltskategorie, die Anzahl von Nennungen einer Assoziation gibt einen Hinweis darüber, wie relevant dieser Inhalt in einer Stichprobe wahrscheinlich ist. So konnten z.B. Rothe und Schindler (1996) in ihren Untersuchungen belegen, daß bei einem Wortassoziationsversuch zu Fachbegriffen der Physik die Novizen unter den Probanden eher Wissen über Merkmale benannten, während die Experten diese Inhaltskategorie vernachlässigten und statt dessen vermehrt Prinzipien, Gesetze oder Vorgehensweisen assoziierten. Es war also anhand der qualitativen Analyse der Assoziationseinheiten möglich, die beiden Probandengruppen deutlich voneinander abzugrenzen.

Es stehen also verschiedene Kennwerte als Ergebnis eines Wortassoziationsversuchs zur Verfügung, deren Berechtigung ist aber teilweise kaum überprüfbar. Es liegen in der Literatur keine

Befunde darüber vor, welcher Kennwert besonders geeignet oder valide für die Beschreibung des individuellen Wissensbesitzes ist. In der vorliegenden Arbeit wird grundsätzlich die mittlere Anzahl geäußerter Assoziationen pro Stimulus als individueller Kennwert herangezogen, da dieser von mehreren Autoren (Stracke, 2004; Rothe, 2006) als geeignete Kenngröße beschrieben wird. Zudem ist dieser Kennwert vergleichsweise einfach zu erheben (eigentlich immer). Da der Wortassoziationsversuch als Screening-Verfahren charakterisiert werden kann, geht damit einher auch der Anspruch an eine vergleichsweise einfache und ökonomische Durchführung einschließlich Auswertung. Diesem Anspruch also kann der Kennwert der mittleren Anzahl von Assoziationen pro Stimulus genügen. Ein Kennwert wie beispielsweise die Pausenzeiten zwischen einzelnen Assoziationen kann hingegen nur unter Zuhilfenahme dafür geeigneter zusätzlicher Technik bzw. Software erstellt werden. Versteht man den Wortassoziationsversuch als einen Test, so braucht es quantitative Kennwerte, um die Gütekriterien überprüfen zu können. Qualitative Aussagen wie z.B. die Analyse der jeweils erstgenannten Assoziationen oder von Assoziationsketten sind in diesem Zusammenhang wenig hilfreich, da sie keine numerischen Werte für die einzelnen Probanden erbringen können. Auch unter diesem Aspekt scheint es angemessen, die mittlere Anzahl von Assoziationen pro Stimulus als individuellen Assoziationskennwert heranzuziehen. Im Rahmen der Bestimmung der Validität wird in der vorliegenden Arbeit zudem geprüft werden, inwieweit eine Modifikation dieses individuellen Kennwertes die Aussagen über den individuellen Wissensbesitz beeinflusst.

3.7. Gütekriterien

Nach Rothe (2003) ist die Datenlage zu den Gütekriterien des Wortassoziationsversuchs insgesamt noch unzureichend, sowohl Inhaltsvalidität als auch Durchführungsobjektivität können aber als gegeben hingenommen werden. Tergan (1988, 1989b) sieht bei qualitativ ausgerichteten Methoden der Wissensdiagnose ein grundsätzliches Problem bei der Bestimmung der Gütekriterien sowie bei der Standardisierung, da die traditionellen Methoden der Gütebestimmung auf quantitativen Parametern beruhen und die qualitativ-strukturellen Aspekte des individuellen Wissens dabei wenig berücksichtigt werden. Dennoch soll in der vorliegenden Arbeit versucht werden, diesen Parametern gerecht zu werden. Im Folgenden sollen kurz die theoretischen Grundlagen zu den einzelnen Gütekriterien skizziert werden. Soweit in der Literatur zu finden, werden Erkenntnisse zu den Gütekriterien des Wortassoziationsversuchs berichtet oder zumindest Überlegungen im Zusammenhang mit Wissensdiagnosen im Allgemeinen dargestellt.

3.7.1. Hauptgütekriterien

Es herrscht weitgehende Einigkeit darüber, daß die Aspekte von Validität, Reliabilität sowie Objektivität unter dem Begriff der Hauptgütekriterien zusammenzufassen sind (z.B. Lienert und Raatz, 1998; Zöfel, 2003; Bühner, 2006). Die Objektivität läßt sich dabei differenzieren hinsichtlich der Durchführung, Auswertung oder Interpretation des Verfahrens. Die Validität kann unterschieden werden in die Aspekte der Inhalts-, Konstrukt- sowie Kriteriumsvalidität. Die Reliabilität kann durch verschiedene Zugänge gewonnen werden: als Paralleltest-Erhebung, Retest-Erhebung oder durch die Bestimmung der internen Konsistenz.

3.7.1.1. Validität

Nach Häcker, Leutner und Amelang (1998) ist die Bestimmung der Validität der wichtigste Aspekt der Testevaluation, da sie zu beurteilen erlaubt, wie angemessen Schlußfolgerungen aus den Testwerten sind. Amelang und Schmidt-Atzert (2006) weisen somit der Reliabilität und Objektivität lediglich die Aufgabe zu, günstige Voraussetzungen für die Validität zu schaffen. Bezogen auf den Wortassoziationsversuch würde man in Anlehnung an Lienert und Raatz (1998) also dann von einem validen Verfahren ausgehen, wenn Probanden mit einem umfangreichen Wissensbesitz einen hohen Assoziationskennwert erreichen, während Probanden mit geringem Wissen in der entsprechenden Domäne nur wenig assoziieren.

Häcker, Leutner und Amelang (1998) benennen als traditionelle Kategorien der Validität die inhaltsbezogene, die kriteriumsbezogene sowie die Konstruktvalidität. Der Aspekt der *Inhaltsvalidität* beschreibt, inwieweit ein Verfahren eine repräsentative Menge von Testitems aus der Gesamtmenge möglicher Items, welche das interessierende Merkmal erfassen könnten, ausgewählt hat. Eine hohe Inhaltsvalidität kann nach Amelang und Schmidt-Atzert (2006) erreicht werden, wenn die Testaufgaben völlig identisch sind mit den Anforderungen, die in einem bestimmten Bereich gestellt werden. Nach Lienert und Raatz (1998) ist beispielsweise ein Schulkenntnistest in Geographie dann inhaltsvalide, wenn die Aufgaben eine repräsentative Auswahl aus dem Curriculum abbilden. Bühner (2006) bewertet diese Validitätsart als schwierig zu ermitteln und statistisch nicht prüfbar, vielmehr werde sie über die inhaltliche psychologische Einsicht bestimmt. Lienert und Raatz (1998) billigen auch der konsensualen Einschätzung durch Experten zu, die geforderte Repräsentativität der Items für die Grundgesamtheit zu gewährleisten. Unter diesen Aspekten kann eine Inhaltsvalidität des Wortassoziationsversuchs grundsätzlich angenommen werden, da die Bestimmung der Wissensdomäne sowie der Stimuli gemäß des bei Rothe (2003, 2006) beschriebenen Vorgehens immer nach intensiven Dokumentenanalysen bzw. in Zusammenarbeit mit betrieblichen Experten vorgenommen werden sollte.

Die *Kriteriumsvalidität* ist nach Lienert und Raatz (1998) praktisch gesehen der bedeutendste Aspekt der Validität und beschreibt, inwieweit das Testergebnis mit einem Außenkriterium korreliert. Dabei ist der Validitätskoeffizient beeinflusst vom Grad der Reliabilität, sowohl des zu prüfenden Tests als auch des Kriteriums. Amelang und Schmidt-Atzert (2006) weisen darauf hin, daß das Kriterium selbst sehr verlässlich und auch praktisch gesehen entscheidungsrelevant sein soll.

Bei der *Konstruktvalidität* geht es um den Nachweis, daß mit dem Verfahren tatsächlich das interessierende Konstrukt erfaßt wird, in diesem Falle also das berufsspezifische Wissen. Dies kann durch die konvergente Validierung erreicht werden, bei der konstruktverwandte Tests zum Vergleich herangezogen werden, oder die diskriminante Validierung, bei der mit konstruktfernden Tests korreliert wird (z.B. Bühner, 2006). Nach Lienert und Raatz (1998) hat die Beurteilung der Konstruktvalidierung aufgrund empirischer und theoretischer Erwägungen zu erfolgen, so daß Kriteriumsvalidität und Inhaltsvalidität als spezielle Aspekte der Konstruktvalidität begriffen werden können, Amelang und Schmidt-Atzert (2006) sprechen bei der Konstruktvalidität sogar von einer Synthese der anderen beiden Validitätsarten. Bungard, Holling und Schultz-Gambard (1996) weisen allerdings darauf hin, daß die Konstruktvalidität in der angewandten Forschung (welche in der vorliegenden Arbeit vertreten wird) weit weniger bedeutsam ist als in der Grundlagenforschung.

Hinsichtlich einer diskriminanten Konstruktvalidierung scheint die Frage interessant, inwieweit der Wortassoziationsversuch tatsächlich den Umfang des individuellen Wissens erfaßt und nicht die generelle Fähigkeit einer Person zur verbalen Produktion. Zum Einfluß der generellen Wortflüssigkeit (nach Jäger & Althoff, 1994, definiert als die quantitative Flüssigkeit assoziativer Worteinfälle) auf die Assoziationsleistung liegen in der Literatur keine Erkenntnisse vor. Lediglich der Einfluß der Intelligenz auf den Wissensbesitz (siehe Schweizer und Koch, 2002) bzw. auf die Reproduktionsfähigkeit listenartiger Wissensbestände (Grabowski 1994, 2005) wurde thematisiert. Zwar ist die Wortflüssigkeit in vielen Intelligenzdiagnostica als eine Facette von Intelligenz konzipiert (z.B. WILDE Intelligenztest von Jäger & Althoff, 1994), dennoch können die Erkenntnisse hinsichtlich der Intelligenz nicht ohne weiteres auf das Konstrukt der Wortflüssigkeit übertragen werden. Daher soll in der vorliegenden Arbeit der mögliche Einfluß der Fähigkeit zur Wortflüssigkeit auf die Assoziationsleistung eines Probanden empirisch abgeklärt werden.

Nach Bungard, Holling und Schultz-Gambard (1996) ist die Frage der Generalisierbarkeit der gefundenen Zusammenhänge v.a. für die anwendungsorientierten Forscher bedeutsam, sie sprechen in diesem Kontext von einer externen Validität. So empfehlen sie, anstatt einer großen Studie lieber viele kleine Studien zu realisieren und deren Ergebnisse mosaikartig zu einem Gesamtbefund zusammenzuführen. Eine ähnliche Intention verfolgt der von Schober (2003) eingeführte Begriff der ökologischen Validität, welcher die Gültigkeit eines Instrumentes für einen realen Lebenskontext beschreibt. Dahinter steht die Annahme, daß in einem künstlichen Laborumfeld erhobene Daten vermutlich wenig repräsentativ sind für ein reales Setting, auch hier stellt sich die Frage nach der Generalisierbarkeit der gewonnenen Erkenntnisse.

Als zentrales Problem im Kontext der Wissensdiagnosen und Expertiseforschung beschreiben Ackermann und Beier (2006) den Rückgriff auf Experten als Untersuchungsstichprobe. Wenn in die Validierungstichprobe lediglich Experten eines Faches einbezogen werden, ergeben sich in dieser relativ homogenen Gruppe i.d.R. geringe Streuungen der Leistungswerte, was zu schlechten Koeffizienten der Gütekriterien führt. Zudem verweisen Ackermann und Beier (2006) auf die oft extremen Basisraten von Experten, die in der Gesamtheit potentieller Probanden eher selten in dieser Ausprägung vorhanden sind. Auch die Frage nach einem geeigneten Außenkriterium, an dem ein Wissensdiagnoseinstrument validiert werden sollte, ist umstritten. Gemäß Rothe (2003) ist die Prüfung der prognostischen Validität von Wissensdiagnosemethoden anhand der Vorhersage von kompetentem Handeln in den jeweiligen Wissensdomänen vorzunehmen, welches seinerseits oftmals über die Einschätzungen durch Vorgesetzte operationalisiert wird. Nach Schuler (1989) allerdings sind Vorgesetztenurteile weniger geeignet als z.B. Gehaltsindizes oder Arbeitsproben, um Diagnoseinstrumente daran zu validieren, da sie geringere Validitätsschätzungen erbringen. Auch Hacker (2005b) bewertet Vorgesetztenurteile zwar als häufigen, aber dennoch umstrittenen Ausweg, um ein Leistungskriterium zu formulieren. Andererseits ist mehrfach belegt, daß durch die Beurteilung der Gesamtleistung einer Arbeitsperson zu wesentlichen Teilen auch die tatsächliche (typische) Arbeitsleistung erfaßt werden kann (Schuler, 1989).

Für den Wortassoziationsversuch selbst liegen bisher nur wenige Erkenntnisse zur Validität vor. Nach Rothe (2006) ist es mit Hilfe des Wortassoziationsversuchs möglich, valide Aussagen über das individuelle domänenspezifische Wissen von Arbeitspersonen treffen zu können, wenngleich er noch weiteren Aufklärungsbedarf hinsichtlich des Zusammenhangs von fachspezifischem Wissen

und Arbeitsleistung sieht. Meyer, Scholl und Zhang (2007) belegen eine signifikante (aber schwache) Korrelation ($r = .17$) zwischen der Anzahl assoziierter Konzepte und dem Score in einem Wissenstest, welcher deklaratives Wissen über die identische Domäne erhob. In der vorliegenden Arbeit sollen die wenigen in der einschlägigen Literatur erwähnten Erkenntnisse zur Validität des Wortassoziationsversuchs weiter beleuchtet werden. Dazu werden Untersuchungen herangezogen, in denen i.S. der Kriteriumsvalidität ein Maß für das Wissen bzw. für die wissensbasierte Arbeitsleistung herangezogen werden kann, sowie weitere Studien, in denen i.S. der konvergenten Konstruktvalidierung andere Verfahren oder Techniken zur Erhebung des fachspezifischen Wissens als Referenz zur Verfügung standen. Zudem werden sich empirische Studien der diskriminanten Konstruktvalidität widmen.

3.7.1.2. Reliabilität

Die Reliabilität eines Verfahrens beschreibt dessen Meßgenauigkeit, genauer gesagt, wie reproduzierbar die Meßwerte eines Probanden sind (Lienert und Raatz, 1998), nach Ansicht von Ackermann und Beier (2006) hat ein Test mit geringer Reliabilität kaum einen wissenschaftlichen oder praktischen Nutzen. Zur Bestimmung der Reliabilität können verschiedene Techniken herangezogen werden (z.B. Lienert und Raatz, 1998): Retest-Reliabilität, bei der der identische Test an den gleichen Probanden wiederholt wird, das Vorlegen einer Parallelförmigkeit, die Teilung des Tests in zwei äquivalente Hälften oder die Interitemkonsistenz, bei der jedes Testitem als separater Paralleltest angesehen wird. Im Allgemeinen wird das Paralleltestverfahren als der Königsweg angesehen (z.B. Lienert und Raatz, 1998), während die Testwiederholung gemeinhin als äußerst anfällig gegenüber Fehlerquellen wie Übungs- oder Erinnerungseffekten gilt (z.B. Lienert und Raatz, 1998; Zöfel, 2003; Ackermann und Beier, 2006; Bühner, 2006). Als bedeutsam erweist sich dabei der Zeitabstand zwischen den beiden Erhebungen sowie die Stabilität des zu messenden Merkmals selbst. Amelang und Schmidt-Atzert (2006) beschreiben diese Problematik als wiederkehrendes Dilemma zwischen Merkmalsstabilität auf der einen Seite und Erinnerungs- oder Übungseffekten auf der anderen Seite. Grundsätzlich halten Lienert und Raatz (1998) für Tests, die einen umschriebenen Wissensstand abprüfen, die Reliabilität als schwer bestimmbar, da i.d.R. Probleme bei der internen Konsistenz die Erstellung von Parallelförmigkeiten unmöglich machen, während eine Testwiederholung ihrerseits sich mit dem Problem der Merkmalsfluktuation auseinandersetzen hat. Hinsichtlich des berufsbezogenen Wissens kann jedoch angenommen werden, daß es zwar grundsätzlich veränderbar ist, aber über gewisse Zeitspannen hinweg einigermaßen stabil sein kann. Wie bereits in Kapitel 1.4 (Seite 8 f.) im Zusammenhang mit dem Begriff des Fachwissens erörtert, wurde dieses berufsbezogene Wissen im Rahmen der Ausbildung oder Tätigkeitsausführung erworben (Rothe und Schindler, 1996; Rothe, 2006) und umfaßt u.a. Wissen über grundlegende Arbeitsverfahren, Arbeitsgegenstände oder damit in Zusammenhang stehende Gesetzmäßigkeiten (siehe Hacker 1992; 1993; Rothe und Schindler, 1996; Rothe und Timpe, 1997; Pleiss und Oesterreich, 2003), so daß diesem Wissen eine gewisse Halbwertszeit zugestehen ist. Nach Bortz und Döring (2003) ist in solchen Fällen, in denen eine zeitlich überdauernde Merkmalsausprägung anzunehmen ist, die Testwiederholung die angemessene Form der Reliabilitätsprüfung.

Im Kontext der Expertiseforschung definieren Ackermann und Beier (2006) einen Test dann als reliabel, wenn durch ihn eine Gruppe von Individuen über mehrere Erhebungszeitpunkte hinweg beständig in die gleiche Rangreihe gebracht werden kann. Nach Stracke (2004) oder auch Wirtz und Caspra (2002) können bei Wissenstests die Werte der Retest-Reliabilität als Indikator für die zeitliche Stabilität des Wissens angesehen werden. Hinsichtlich des Wortassoziationsversuchs im Speziellen liegen kaum Erkenntnisse hinsichtlich seiner Reliabilität vor. Gemäß Strube (1984) steigt die Anzahl produzierter Assoziationen systematisch an bei Testwiederholung, was für eine geschälerte Reliabilität spricht. Allerdings ist einzuräumen, daß sich die Untersuchungen von Strube (1984) auf Stimulusmaterial aus dem Bereich des Allgemeinwissens beziehen. Daher soll in der vorliegenden Arbeit die Frage der Retest-Reliabilität des Wortassoziationsversuchs zur Erfassung fachspezifischen Wissens aufgegriffen und empirisch belegt werden.

3.7.1.3. Objektivität

Die Objektivität eines Verfahrens beschreibt die Unabhängigkeit der Testergebnisse von der Person des Messenden und kann als Folge einer hohen Standardisierung angenommen werden (z.B. Lienert und Raatz, 1998; Bühner, 2006). Es kann dabei unterschieden werden zwischen der Durchführungsobjektivität, der Auswerteobjektivität sowie der Interpretationsobjektivität. Nach Lienert und Raatz (1998) ist v.a. dann von einer hohen Durchführungsobjektivität auszugehen, wenn eine schriftliche Instruktion vorliegt, welche die soziale Interaktion zwischen Proband und Versuchsleiter auf ein Minimum reduziert. Nach Bortz und Döring (2003) oder auch Amelang und Schmidt-Atzert (2006) kann eine hohe Durchführungsobjektivität durch die maximale Standardisierung der Testsituation gewährleistet werden, was zwangsläufig mit einer Minimalisierung der sozialen Kontakte einhergeht. Eine quantitative Bestimmung der Durchführungsobjektivität jedoch halten Amelang und Schmidt-Atzert (2006) für unmöglich. Nach Rothe (2003) ist dieser Aspekt der Objektivität für den Wortassoziationsversuch als gegeben hinnehmbar.

Die Auswerteobjektivität ist grundsätzlich als etwas geringer anzunehmen bei Verfahren mit freien Antworten, da der Auswerter in diesem Falle entscheiden muß, ob die jeweilige Antwort angemessen ist (Lienert und Raatz, 1998; Amelang und Schmidt-Atzert, 2006). Bühner (2006) fordert daher, die jeweiligen Auswertevorschriften exakt zu definieren. Für den Wortassoziationsversuch selbst liegen keine Erkenntnisse hinsichtlich der Auswerteobjektivität vor, Rothe (2003) sieht hier noch weiteren Klärungsbedarf. An dieser Stelle sollen die Befunde aus einer Studie mit vergleichbarem freien Antwortformat berichtet werden: Hauser, Spada, Rummel und Meier (2006) untersuchten das Expertenwissen von Psychologie-Studierenden und erfahrenen Psychotherapeuten anhand eines Wissenstests im Multiple-Choice-Format, eines weiteren Tests mit offenen Fragen sowie mittels der Bearbeitung einer Fallstudie. Für die Datenauswertung wurden alle offenen Antworten von mehreren Untersuchern unabhängig voneinander nach Sinneinheiten aufgeteilt und diese einzelnen Einheiten auf ihre Korrektheit hin bewertet, bevor die Anzahl richtiger Antworten ausgezählt werden konnte. Dabei berichten Hauser, Spada, Rummel und Meier (2006) von Beurteilerübereinstimmungen von $>.7$. Zwar kann ein solch hoher Wert an übereinstimmender Auswertung auch auf die Qualifikation der Auswerter zurückgeführt werden, jedoch ist dies auch erreichbar durch eine hohe Standardisierung des Auswerteprozederes. Zudem

ist entsprechend der Taxonomie von Rothe (2006) der Wortassoziationsversuch ohnehin als ein Verfahren zu charakterisieren, welches vom Untersucher bzw. Auswerter differenzierte Kenntnisse über die jeweilige Wissensdomäne erfordert. Da es sich in der berichteten Studie um vergleichsweise komplexe inhaltliche Zusammenhänge handelte, ist davon auszugehen, daß Auswerter des Wortassoziationsversuchs, welche im Wesentlichen listenartige Verbaldaten bearbeiten sollen, diese Anforderung auch bewältigen werden. Somit kann letztlich eine Auswerteobjektivität bei der Auszählung von Daten, wie sie auch der Wortassoziationsversuch produziert, angenommen werden. Dieser Befund ist um so bedeutungsvoller, wenn man bedenkt, daß es sich hierbei um eine qualitative Analyse der Verbaldaten handelt. Nach Oesterreich (2005) ist ohnehin zu bezweifeln, daß jemals eine perfekte Objektivität erreicht werden kann, welche den Einfluß des Menschen völlig ausschließt.

Der dritte Aspekt betrifft die Interpretationsobjektivität, welche nach Amelang und Schmidt-Atzert (2006) das Ausmaß beschreibt, in dem die Schlußfolgerungen aus der Datenlage unabhängig ist von der interpretierenden Person, wenn also mehrere Auswerter die gleichen Schlüsse aus ein und demselben Auswerteergebnis ziehen. Dies ist z.B. dann gewährleistet, wenn Normtabellen vorliegen, in die das individuelle Testergebnis eingeordnet werden kann, Amelang und Schmidt-Atzert (2006) verwenden in diesem Zusammenhang den Begriff der statistischen Vorgehensweise. Für den Wortassoziationsversuch selbst sind keine empirisch begründeten Erkenntnisse hinsichtlich der Objektivität und deren Aspekten bekannt. In der vorliegenden Arbeit soll die Frage der Objektivität des Wortassoziationsversuchs aufgegriffen und diskutiert werden, wobei auf eine empirische Überprüfung verzichtet wird.

3.7.2. Nebengütekriterien

Im Gegensatz zu den Hauptgütekriterien sind die Nebengütekriterien eine eher unscharfe Sammlung von Indikatoren, die nach Bühner (2006) – neben den Hauptgütekriterien – für die Beurteilung der Güte eines Verfahrens wichtig sind (woraus sich auch ihre Bezeichnung ableiten läßt). So widmen sich beispielsweise Lienert und Raatz (1998) oder Bühner (2006) den Aspekten der Normierung, Vergleichbarkeit, Ökonomie sowie Nützlichkeit, mit dem Verweis auf weitere mögliche Nebengütekriterien, Amelang und Schmidt-Atzert (2006) führen exemplarisch die Kriterien der Normierung und der Testfairneß an, Kubinger (2006) benennt außerdem noch die Aspekte Zumutbarkeit und Unverfälschbarkeit. Im Folgenden sollen die Nebengütekriterien Ökonomie, Nützlichkeit, Normierung, Vergleichbarkeit sowie Sensitivität erörtert werden, wobei neben allgemeinen Darstellungen der Zusammenhang zu Wissensdiagnosen sowie ggf. hinsichtlich des Wortassoziationsversuchs hergestellt werden soll. Für das Kriterium der Akzeptanz durch die Probanden, dem Schuler (1989) bei der Erhebung eines Leistungskriteriums wie dem des Wissens eine wichtige Rolle beimißt, waren leider keine Ausführungen in der einschlägigen Literatur zu finden, nicht zuletzt aus diesem Grund soll diese in der vorliegenden Arbeit anhand empirischer Befunde näher beleuchtet werden. Nach Lienert und Raatz (1998) ist zu beachten, daß bei allen Nebengütekriterien keine zahlenmäßigen Kennwerte vorliegen.

3.7.2.1. *Ökonomie*

Nach Schuler (1989) ist die Ökonomie (neben der Akzeptanz) ein wichtiger Aspekt bei der Auswahl eines Kriteriums zur Leistungsbeurteilung sowie des entsprechenden Instrumentes. Lienert und Raatz (1998) sowie Bühner (2006) bewerten einen Test dann als ökonomisch, wenn er eine kurze Durchführungszeit beansprucht, dabei wenig Material verbraucht, einfach zu handhaben ist, als Gruppentest durchführbar ist sowie eine einfache und schnelle Auswertung ermöglicht. Kubinger (2006) hingegen betont vielmehr das angemessene Verhältnis von diagnostischem Informationsgewinn und der Beanspruchung von Ressourcen (v.a. Zeit und Geld). In der Einschätzung von Bungard, Holling und Schultz-Gambard (1996) ist der Preis, welcher für eine anwendungsorientierte arbeitspsychologische Forschung zu zahlen ist, i.d.R. eine erhebliche Einschränkung der zentralen Prinzipien der experimentellen Methodik (z.B. Randomisierung, Experimental-Kontrollgruppen-Design). Als oberste Prämisse dabei gilt, den Betriebsablauf so wenig wie möglich zu stören (Bungard, Holling und Schultz-Gambard, 1996). Um dieser Forderung zu entsprechen, sollte ein Verfahren in der Durchführung möglichst wenig Zeit beanspruchen und auch hinsichtlich der räumlichen oder technischen Bedingungen unaufwendig sein. Je nach Beschaffenheit der jeweiligen Arbeitsabläufe kann es von Vorteil sein, das gewünschte Verfahren entweder als Gruppenversuch oder aber als Einzelerhebung einzusetzen.

Rothe (2003) orientiert den zeitlichen Aufwand für die Durchführung eines Wortassoziationsversuchs mit beispielsweise zehn Stimuli bei etwa 30 Minuten, für die Aufbereitung der Rohdaten weitere 30 Minuten sowie etwa eine Stunde für die Bestimmung der individuellen bzw. gruppenspezifischen Kenndaten. Der Wortassoziationsversuch benötigt keine besonderen räumlichen Voraussetzungen, lediglich eine gewisse Störungsfreiheit ist (wie bei allen Wissensdiagnosen) zu gewährleisten (Rothe, 2003). Zudem ist das Assoziieren sowohl als Papier- und Bleistift-Form einsetzbar (Rothe, 2003) als auch inzwischen computerbasiert möglich (z.B. das Assoziationstool im Assoziations-Struktur-Test von Meyer, 2007). Damit eignet sich der Wortassoziationsversuch ebenso für einen Gruppenversuch wie auch für die Einzeltestung. Somit kann davon ausgegangen werden, daß mit dem Wortassoziationsversuch ein relativ ökonomisches Verfahren vorliegt. Anhand der empirischen Befunde der vorliegenden Arbeit soll dies weiter untermauert werden.

3.7.2.2. *Nützlichkeit*

Nach Lienert und Raatz (1998) sowie Bühner (2006) ist ein Test dann als nützlich zu bewerten, wenn ein praktisches Bedürfnis für seinen Einsatz besteht. In Kapitel 2.1 (Seite 11 ff.) sind die wichtigsten Anwendungsgebiete für Wissensanalysen dargestellt, welche ohne Einschränkung auch für den Wortassoziationsversuch gelten können. Somit werden viele Felder aufgezeigt, in denen der Wortassoziationsversuch von praktischem Nutzen sein kann, insbesondere, wenn man das Kriterium der ökonomischen Durchführung mit heranzieht. Nach Rosenstiel (2000, 2004) ist nicht nur der Erkenntnisgewinn wichtiges Ziel arbeitspsychologischer Forschung, sondern zentral auch der Nutzen und die praktische Relevanz der Erkenntnisse – auf wissenschaftliche Weise sollte eine Frage beantwortet werden können, die aus der Praxis kommt. Nützlichkeit sei eine politische Dimension für jede angewandte Forschung (Rosenstiel, 2000).

Die DIN-Norm zu den Anforderungen an Erhebungsmethoden psychischer Arbeitsbelastungen (DIN, 2004) formuliert die Brauchbarkeit eines Verfahrens als das Ausmaß, in dem eine Methode von spezialisierten Anwendern genutzt werden kann. Dabei sollte das Verfahren effektiv sein, effizient und zufriedenstellend und insbesondere mit minimalem Aufwand die erwünschten Ergebnisse liefern. Nach Lienert und Raatz (1998) sowie Bühner (2006) ist ein Test dann wenig nützlich, wenn bereits andere Verfahren verfügbar sind, mit denen das interessierende Merkmal genauso gut erhoben werden könnte. Allerdings ist dem entgegenzusetzen, daß der Wortassoziationsversuch zunächst nur eine generelle Methode darstellt, mit der eine bestimmte Art von Wissen erhoben werden kann – die konkurrierenden Methoden (z.B. Methode des lauten Denkens, Strukturlegetechnik usw.) unterscheiden sich in der Art des zu erfassenden Wissens sowie auch unter dem Aspekt der Ökonomie. Im Praxiseinsatz sind diese Aspekte sorgfältig gegeneinander abzuwägen und sprechen aus diesem Grunde allen hier besprochenen wissensdiagnostischen Methoden eine Nützlichkeit zu. In der Literatur konnten hinsichtlich der Nützlichkeit des Wortassoziationsversuchs im Speziellen keine Aussagen gefunden werden, daher soll dies in der vorliegenden Arbeit thematisiert werden.

3.7.2.3. Normierung

Unter Normierung eines Tests ist zu verstehen, daß Normen vorliegen, die als Bezugssystem für die Einordnung eines individuellen Testergebnisses dienen können (z.B. Lienert und Raatz, 1998; Bühner, 2006). Lienert und Raatz (1998) gestehen zwar nicht-normierten Tests nur eine geringe diagnostische Brauchbarkeit zu, räumen aber ein, daß solche Verfahren dennoch als Forschungsinstrument geeignet sein können, beispielsweise, wenn es „nur“ um den Vergleich von Gruppen geht. Als Mindestgröße für eine Normierungsstichprobe gibt Bühner (2006) eine Zahl von 300 Probanden an – eine Stichprobengröße, die bei wissensdiagnostischen Erhebungen im Feld kaum erreicht werden kann.

Zur Normierung von Testwerten im Zusammenhang mit Verfahren zur Erfassung von Wissen sind keine Anwendungen bekannt und erscheinen auch wenig angebracht. So macht Wagener (2003) beispielsweise mit Blick auf den von ihm entwickelten Wissenstest über Computernutzung darauf aufmerksam, daß die Items eines solchen Verfahrens schnell veralten und somit auch die entsprechenden Normen überholt sind. Zudem sind Verfahren zur Wissensdiagnose i.d.R. keine festgeschriebenen Tests mit einem festen Itempool, sondern sind eher als Methode zu verstehen, die auf einen konkreten Untersuchungsbereich flexibel angepaßt werden kann, also in praktisch jeder beliebigen Wissensdomäne einsetzbar ist. Hinzu kommt, daß Wissensdiagnosen neben einer quantitativen Aussage in hohem Maße auch eingesetzt werden, um qualitative Daten zu erbringen, welche sich ohnehin einer Normierung entziehen. Eine Bereitstellung von Profilen oder Interpretationshilfen im klassischen Sinne ist aufgrund der bereits angesprochenen Domänenspezifität unrealistisch, allerdings kann das jeweilig entwickelte hypothetische Wissensmodell als Bezugspunkt zur Interpretation der Probandenleistung dienen. Eine Abschätzung der Bedeutsamkeit einer individuellen Leistung ist daher nur so möglich: entweder in Relation zu den anderen Probanden der Untersuchungsgruppe, die das gleiche Prozedere in der gleichen Wissensdomäne absolviert haben, oder im Vergleich zu einem hypothetischen Wissensmodell dieser Domäne, ggf. unter Hinzuziehung eines Experten, um im Einzelfall qualitative Daten auf ihre

Korrektheit hin zu bewerten. Sowohl hinsichtlich anderer wissensdiagnostischer Verfahren als auch mit Blick auf den Wortassoziationsversuch konnten aus der Literatur keine Aussagen zur Normierung dieser Verfahren abgeleitet werden, die vorliegende Arbeit soll daher dieses Nebengütekriterium weiter besprechen.

3.7.2.4. Vergleichbarkeit

Die Vergleichbarkeit eines Tests ist dann gegeben, wenn mehrere Parallelformen vorliegen, nach Bühner (2006) ist dies v.a. bei wiederholtem Testeinsatz vorteilhaft sowie in Gruppenerhebungen (um z.B. Abschreiben zu vermeiden). Wie bereits in Kapitel 3.7.1.2 (Seite 44 ff.) ausgeführt, ist das Erstellen von Parallelformen bei Wissenstests so gut wie unmöglich aufgrund der hohen Inhaltsspezifität der Items. Daher können auch für den Wortassoziationsversuch aus der Literatur keine Aussagen hinsichtlich des Nebengütekriteriums der Vergleichbarkeit berichtet werden, in der vorliegenden Arbeit soll die Vergleichbarkeit weiter thematisiert werden.

3.7.2.5. Sensitivität

Nach Oesterreich (2005) ist das Kriterium der Sensitivität v.a. in der medizinischen Forschung verankert, bei der damit das Verhältnis beschrieben wird zwischen den mit dem Diagnoseinstrument korrekterweise als krank diagnostizierten Patienten zu allen Personen, die tatsächlich krank sind. Dieses Verständnis von Sensitivität ist im Kontext von Wissensanalysen wenig zielführend, v.a. weil das Konstrukt Wissen nicht dichotom ausgeprägt ist (i.S. von Proband weiß etwas vs. weiß nichts). Die Laborforschungen zu Arbeitsbelastungen können eine weitaus angemessenere Definition beitragen, wonach Sensitivität das Ausmaß beschreibt, indem ein Instrument in der Lage ist, zwischen verschiedenen Abstufungen eines Meßgegenstandes zu differenzieren; Oesterreich (2005) weist aber gleichsam darauf hin, daß die Sensitivität bisher als Kriterium in der Psychodiagnostik ungebräuchlich ist und es zudem keine Hinweise hinsichtlich der Anzahl vorzunehmender Stufen gibt.

Die Interpretation von Stracke (2004) ist stärker praxisbezogen: sie versteht unter Sensitivität, daß sich die erhobenen Daten in einem Prä-Post-Vergleich i.S. der Instruktionsziele verändert haben, also i.d.R. ein Wissenszuwachs nachweisbar ist. Als wichtigste Voraussetzung für Sensitivität benennt Mazur (2006) die Reliabilität des Verfahrens. Diese Sichtweise verzichtet auf die Bildung von Abstufungen und stellt vielmehr den Veränderungsaspekt im Zeitverlauf in den Vordergrund. Da dies im Rahmen von Wissensanalysen praxisrelevanter ist, soll dieses Verständnis von Sensitivität i.S. von Veränderungssensitivität im Zeitverlauf auch in der vorliegenden Arbeit aufgegriffen werden. Zur Sensitivität des Wortassoziationsversuchs werden in der Literatur keine Befunde berichtet, daher wird dies in der vorliegenden Arbeit mit empirischen Befunden näher beleuchtet werden.

3.8. Verhältnis zwischen Wortassoziationsversuch und Strukturlegetechnik

Abschließend soll an dieser Stelle das besondere Verhältnis des Wortassoziationsversuchs zu einem weiteren wissensdiagnostischen Instrument, der Strukturlegetechnik, erörtert werden. Wie bereits beschrieben, empfiehlt sich für die Beantwortung praktischer Fragestellungen eine Methodenbatterie zur Erfassung des fachspezifischen Wissens von Arbeitspersonen. Dabei bieten

sich u.a. der Wortassoziationsversuch und die Strukturlegetechnik innerhalb eines Designs an, beide Verfahren wurden in einigen Studien im Rahmen eines mehrstufigen Vorgehens sukzessive bearbeitet. So modellierte beispielsweise Schindler (2000) das Wissen, welches Nachwuchsführungskräfte während ihrer ersten Zeit auf der neuen Stelle erwerben müssen, durch folgendes Vorgehen: zunächst assoziierten mehrere Führungskräfte zu relevanten Stimuli, anschließend wurden diese Assoziationen gesammelt und in der Gruppe mittels der Strukturlegetechnik zu einem Wissensnetz zusammengetragen. Auch Rothe und Warning (1991), Rothe, Timpe und Warning (1991) sowie Timpe und Rothe (1997) nutzten neben anderen wissensdiagnostischen Verfahren auch den Wortassoziationsversuch, um damit ihr hypothetisches Wissensmodell zu vervollständigen und daraus in einem späteren Schritt die Konzepte für die Knoten des Strukturlegenetzes abzuleiten. Stracke (2004) setzte umfangreiche Wortassoziationsversuche ein mit dem Ziel, eine begründete Entscheidung darüber treffen zu wollen, welche Begriffe beim Legen der Netze vorgegeben werden sollten. Dabei griff sie auf Probanden verschiedener Expertiseniveaus zurück, um die Relevanz einzelner Begriffe angemessen zu berücksichtigen. In diesen Untersuchungen also konnte der Wortassoziationsversuch einen bedeutsamen Beitrag zur Aufklärung der internen Wissensrepräsentation leisten und diente als Datenlieferant für das hypothetische Wissensmodell und somit auch für die nachfolgende Strukturlegetechnik, denn die gruppenbezogene Auswertung der Assoziationen ergab kritische Mengen – ein Begriff, der von sehr vielen Probanden assoziiert wurde, scheint eine gewissen Relevanz für die interessierende Domäne zu besitzen, so daß auch die Probanden der sich anschließenden Strukturlegetechnik diesen Begriff höchstwahrscheinlich in ihr Netz einbauen werden.

Der Wortassoziationsversuch kann aber auch direkt (d.h. ohne den „Umweg“ über das hypothetische Wissensmodell und damit sozusagen im Within-Subject-Design) mit der Strukturlegetechnik kombiniert werden. Wie bereits erörtert besagt die theoretische Grundannahme des Wortassoziationsversuchs, daß diejenigen Begriffe auf einen Stimulusbegriff genannt werden, welche mit dem Stimulus im Individualgedächtnis verknüpft sind, gemäß Rothe (2006) sind die auf Stimulusbegriffe hin geäußerten Assoziationen miteinander verbunden durch wohl definierte und logisch unterscheidbare semantische Relationen. Die Instruktion des Wortassoziationsversuchs erlaubt aber nicht, diese Relationen zu benennen – sie fordert vielmehr vom Probanden, die Knoten des semantischen Netzwerkes abzulaufen und zu nennen. Dabei wird unterstellt, daß sinnvolle semantische Relationen zwischen Stimulus und den geäußerten Assoziationen bestehen, über die (seitens des Auswerters) aber lediglich Hypothesen aufgestellt werden können. Die Aufdeckung solcher Relationen ist aber nicht Sache des Wortassoziationsversuchs, sondern wird erst durch das Strukturlegen möglich.

Die Grundannahmen von netzwerkartiger Repräsentation individueller Gedächtnisinhalte sowie deren Verknüpfung durch semantische Relationen führte Rothe (2003, 2006) zu einer Methodenkombination aus Wortassoziationsversuch und Strukturlegen, bei der zunächst der Wortassoziationsversuch die verbalisierbaren Begriffseinheiten des individuellen Gedächtnisbesitzes ermittelt, während im zweiten Schritt die Strukturlegetechnik differenzierte Informationen über die Organisation dieser Begriffseinheiten und die Art der Relationen zwischen diesen liefern kann, so daß sich aus allen diesen Daten ein Netzwerk rekonstruiert wird, ggf. können in dieser Phase noch weitere Begriffe vom Probanden in das Netz eingebracht werden. Mit anderen

Worten: der Wortassoziationsversuch ermittelt die Knoten des semantischen Netzes, und die Strukturlegetechnik die Kanten zwischen diesen Knoten, Abbildung 2 stellt die Beziehung zwischen den beiden Verfahren graphisch dar.

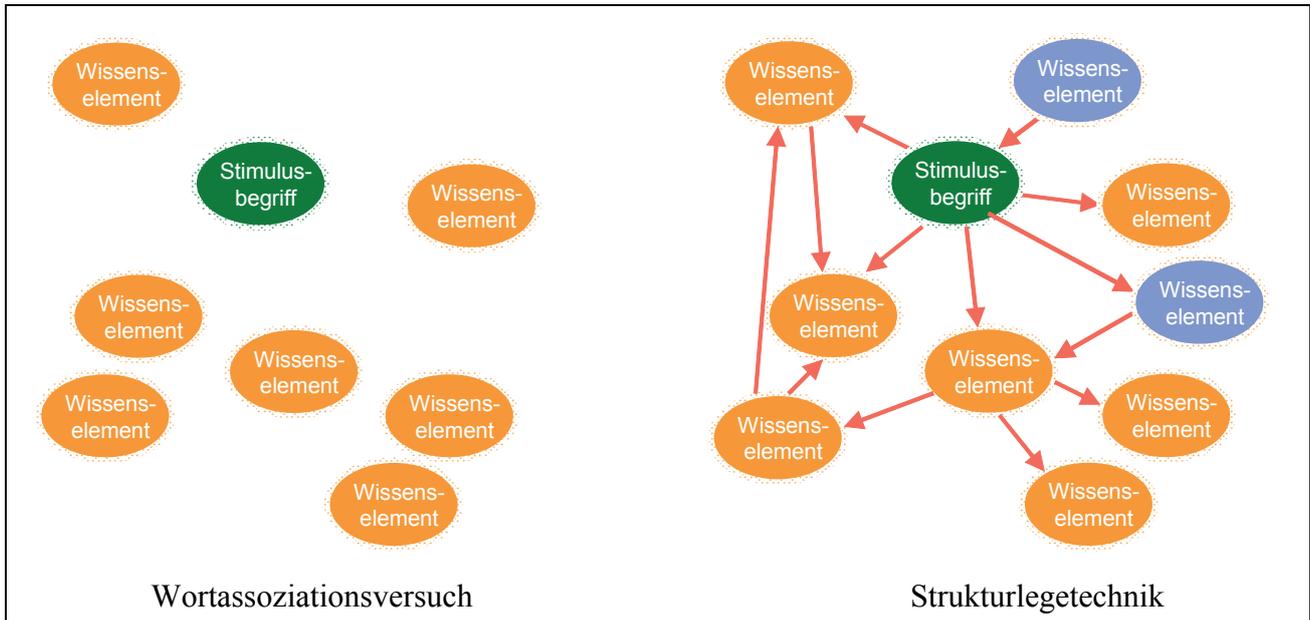


Abbildung 2: Beziehung zwischen den Ergebnissen des Wortassoziationsversuchs und des Strukturlegens im Kontext des semantischen Netzwerkes

Nach Mandl und Ballstaedt (1986) ist die Strukturlegetechnik sehr geeignet, um Wissensstrukturen abbilden zu können und Veränderungen in der Konzeptbildung im Laufe eines Lernprozesses sichtbar zu machen. Die ursprüngliche Form der Heidelberger Strukturlegetechnik nach Scheele und Groeben (1984) war zur Externalisierung sog. subjektiver Theorien von handelnden Personen entwickelt worden, nicht vordergründig zur Darstellung von Wissen (Scheele, 1992). Hinsichtlich der Durchführung der Strukturlegetechnik empfiehlt Dann (1992) die Trennung von Inhalt und Struktur, um den Probanden nicht zu überfordern. Dabei sollen zunächst die inhaltlichen Konzepte bestimmt werden, anschließend werden diese dann in einem zweiten Schritt strukturiert, d.h. die Relationen zwischen diesen Konzepten bestimmt (nach zuvor festgelegten Regeln). Genau diesem Ansatz trägt die von Rothe (2003, 2006) vertretene Modifikation der Strukturlegetechnik Rechnung: sie nutzt die Vorzüge der Strukturlegetechnik und berücksichtigt gleichzeitig diejenige Kritik, die bei der üblichen Durchführung mit einem vorgegeben Pool an zu legenden Knoten laut wird. Janetzko und Strube (2000) bescheinigen generell einem Begriffsnetz, wie es das Strukturlegen produziert, eine höhere Ausdrucksmächtigkeit als einer Merkmalsmenge, wie sie im Ergebnis des Wortassoziationsversuchs steht (wobei diese hohe Ausdruckskraft zu Lasten der Berechnungseigenschaften geht).

Rothe (2003) sieht den Nutzen dieses kombinierten Verfahrens v.a. darin, daß die Visualisierung des Wissens durch das Strukturlegen es dem Probanden erleichtert, seinen Gedächtnisbesitz zu aktivieren und schließlich zu reproduzieren, durch die separate Erhebung der zu legenden Konzepte in der vorausgehenden Assoziationsphase ist eine kognitive Entlastung der Probanden gegeben. Zudem hebt Rothe (2003) hervor, daß durch das gewählte Vorgehen die Beliebigkeit der Knoten-Inhalte deutlich reduziert wird, da der Proband nun selbstproduzierte Konzepte legt, die tatsächlich

seinem eigenen Wissensbesitz entstammen, anstatt – wie beim klassischen Vorgehen der Strukturlegetechnik – allen Probanden den gleichen Pool an zu legenden Knoten vorzugeben, welche aus dem hypothetischen Wissensmodell abgeleitet wurden. Rothe (2003, 2006) verbindet also die Vorteile des Wortassoziationsversuchs (die Produktion *individueller* Gedächtnisinhalte) mit denen der Strukturlegetechnik (differenzierterer Erkenntnisgewinn über die *Struktur* der Gedächtnisinhalte). Es ist darauf hinzuweisen, daß für das Strukturlegen – ebenso wie für den Wortassoziationsversuch – bisher nur wenig Erkenntnisse hinsichtlich der Gütekriterien vorliegen (z.B. Rothe, 2003, 2006; Stracke, 2004). Die durch die vorliegende Arbeit zu gewinnenden Erkenntnisse über die Gütekriterien des Wortassoziationsversuchs können also auch einen Beitrag leisten zur Verfahrensvalidierung der Methodenkombination Assoziieren und anschließendes Strukturlegetechnik nach Rothe (2003, 2006).

Teil C - FRAGESTELLUNG

Wie bereits geschildert lassen sich in der arbeitspsychologischen Praxis vielfältige Bereiche benennen, in denen Wissensdiagnosen eingesetzt werden können. So kann beispielsweise im Rahmen von Wissensmanagementprozessen mittels Wissensdiagnoseverfahren geprüft werden, ob und ggf. in welchen Bereichen eine Intervention notwendig ist. Nach erfolgten Maßnahmen sind dann ggf. Änderungen im individuellen Wissensbesitz zu kontrollieren. Auch für andere Bereiche wie z.B. die Evaluation von Weiterbildungsveranstaltungen oder die Entwicklung wissensbasierter Systeme lassen sich wissensdiagnostische Verfahren heranziehen. Der Wortassoziationsversuch als ein Verfahren, bei dem Arbeitspersonen ihr fachspezifisches Wissen verbalisieren, kann dazu einen Beitrag leisten. Der Einsatz im arbeitspsychologischen Feld stellt zudem eine zusätzliche Forderung an ein Diagnoseinstrument: es muß einfach und unkompliziert durchführbar sein, d.h. für die Probanden leicht verständlich, flexibel sein bzgl. räumlicher oder technischer Gegebenheiten und darf v.a. nicht zu viel Zeit beanspruchen. Gleichsam sollte ein Verfahren für die Praxis aber auch hinreichend valide zwischen bzgl. ihres Fachwissensbesitzes guten und schlechten Beschäftigten differenzieren bzw. deren qualitative Unterschiede aufdecken können.

Während sich also für den Einsatz des Wortassoziationsversuchs als wissensdiagnostisches Instrument viele Anwendungsgebiete anbieten, sind allerdings noch viele Fragen hinsichtlich der Verfahrensgüte offen. Zunächst sollen in der vorliegenden Arbeit zwei Aspekte hinsichtlich der Durchführung des Wortassoziationsversuchs aufgegriffen werden: zum einen die Abhängigkeit der Assoziationsleistung von der Präsentationsreihenfolge der Stimuli, dazu liegen in der Literatur keinerlei Hinweise vor. Zum anderen wird der optimale Durchführungsmodus des Wortassoziationsversuchs thematisiert, die Befunde dazu aus der Literatur sind rar und widersprüchlich (siehe Bredenkamp und Erdfelder, 1996; Troche, Rammstedt und Rammsayer, 2002; Grabowski, 1999, 2005; Strube, 1984). Insbesondere Grabowski (1999, 2005) hat sich experimentell mit der Durchführungsmodalität auseinandergesetzt und dabei die Überlegenheit des schriftlichen Modus gegenüber dem mündlichen Antwortformat belegt, seine Erkenntnisse beziehen sich aber lediglich auf die Wiedergabe einfacher listenartiger Wissensbestände (z.B. das Aufzählen von Hauptstädten), welche man im weitesten Sinne als gebundene oder kontrollierte Assoziation begreifen könnte. Gemäß Strube (1984) hingegen habe der Durchführungsmodus keinen Einfluß auf die (freie) Assoziationsleistung, allerdings beziehen sich seine Erkenntnisse auf Wortassoziationsversuche mit Stimuli aus dem Bereich des Allgemeinwissens. Auf Basis dieser wenigen Erkenntnisse lassen sich daher keine eindeutigen Ableitungen für den Wortassoziationsversuch mit fachspezifischen Stimuli treffen.

Da also die Befunde aus der Literatur zu den soeben angesprochenen Aspekten zur Durchführung des Wortassoziationsversuchs keine eindeutigen Hinweise geben, sollen in der empirischen Auseinandersetzung folgende grundlegende Fragen vorangestellt werden:

- 1) Welcher Durchführungsmodus des Wortassoziationsversuchs ist vorzuziehen?
- 2) Ist die Präsentationsreihenfolge der Stimuli bedeutsam für die Assoziationsleistung?

Bisher ist ungeklärt, ob mit dem Wortassoziationsversuch tatsächlich hinreichend zuverlässige Aussagen über den Wissensbesitz von Probanden getroffen werden können, Rothe (2003) bewertet die Datenlage zu den Gütekriterien als unzureichend. Nach Rothe (2006) sowie Meyer, Scholl und Zhang (2007) gibt es aber Hinweise darauf, daß sich anhand des Wortassoziationsversuchs Aussagen über den Umfang des individuellen Wissensbesitzes treffen lassen. Zur Reliabilität des Wortassoziationsversuchs liegen so gut wie keine Erkenntnisse vor, lediglich die Befunde von Strube (1984) mit Stimuli aus dem Allgemeinwissensbereich legen nahe, daß aufgrund von Lern- oder Erinnerungseffekten bei Testwiederholung eine reduzierte Reliabilität anzunehmen ist. Für die Objektivität des Wortassoziationsversuchs liegen keine Befunde vor, nach Rothe (2003) ist die Durchführungsobjektivität aber aufgrund der eingeschränkten Interaktion zwischen Versuchsleiter und Proband anzunehmen. Hauser, Spada, Rummel und Meier (2006) konnten bei der Auswertung offener Fragen eines Fachwissenstests eine hohe Beurteilerübereinstimmung belegen. Da es sich dabei um sehr komplexe inhaltliche Zusammenhänge handelte, ist anzunehmen, daß verschiedene Auswerter eines Wortassoziationsversuchs ebenso hohe Übereinstimmungen beim Auszählen der aufgelisteten (wenig komplexen) Assoziationen erzielen.

Zur Güte des Wortassoziationsversuchs liegen also kaum konkrete Erkenntnisse vor. Die primäre Zielstellung der vorliegenden Arbeit ist daher eine Verfahrensvalidierung des Wortassoziationsversuchs. Dazu sollen die testpsychologischen Hauptgütekriterien

- 3) Validität,
- 4) Reliabilität und
- 5) Objektivität

bestimmt werden. Mit Blick auf die Validitätsprüfung ist zu berücksichtigen, welche Kriterien dazu herangezogen werden können und was diese hinsichtlich der Nutzung des Wissens repräsentieren. So besteht die Anforderung für den Probanden bei einem Außenkriterium wie einer Klausur darin, Wissen in einer umschriebenen Domäne zu reproduzieren – eine adäquate Anwendung dieses Wissens ist dabei nicht Gegenstand der Betrachtungen. Anders sieht es aus, wenn als Kriterium eine Leistungsbeurteilung durch Vorgesetzte herangezogen wird. Zwar wird diese von einigen Autoren als problematisch eingeschätzt (z.B. Schuler, 1989; Hacker, 2005b), da in solche Einschätzungen neben dem fachspezifischen Wissen der Beurteilten auch noch andere Aspekte einfließen, dennoch besitzen sie eine gewisse praktische Relevanz. Ein solches Validitätskriterium erfaßt nicht direkt das Wissen, sondern letztlich das auf Wissen basierende, adäquate Handeln im Feld. Beide Arten von Kriterien besitzen im arbeitspsychologischen Kontext ihre Berechtigung, daher werden sie beide in der vorliegenden Arbeit als Validitätskriterien Berücksichtigung finden.

Im Zusammenhang mit der Bestimmung der Validität stellt sich außerdem die Frage, inwieweit der dabei herangezogene individuelle Assoziationskennwert, welcher die mittlere Anzahl von Assoziationen pro Stimulus beschreibt, tatsächlich einen geeigneten Parameter zur Bestimmung des individuellen Wissensstandes darstellt oder ob andere Kennwerte im Rahmen der Validitätsbestimmung eine zuverlässigere Vorhersage der Kriterien erlauben. Die Grundannahme des Wortassoziationsversuchs ist, daß einem Probanden, der viele Assoziationen produzieren kann, ein umfangreiches semantisches Netzwerk zur Verfügung steht, welches er abgerufen hat. Die Anzahl von Assoziationen ist somit als ein Indikator für einen elaborierten Wissensbesitz bzw.

dessen Umfang zu interpretieren (z.B. Stracke, 2004; Rothe, 2006). Dem Verständnis des Wortassoziationsversuchs als Screening-Verfahren folgend ist dieser Kennwert insofern angemessen, als daß er vergleichsweise einfach zu erheben ist. Über die Relevanz anderer Kennwerte des Wortassoziationsversuchs hingegen ist nichts bekannt, eher qualitativ ausgerichtete Parameter wie z.B. die Analyse von Assoziationsketten entziehen sich ohnehin einer numerischen Validitätsbestimmung. Verschiedene Autoren (z.B. Rothe, 2003, 2006; Müller, 1998; Bortz und Döring, 2003) geben zwar Anregungen, in welcher Weise individuelle Assoziationsdaten quantifiziert werden können, aber es lassen sich keine Hinweise darauf finden, inwieweit die so modifizierten individuellen Kennwerte die Beurteilung eines Probanden hinsichtlich seines Wissensbestandes beeinflussen. Daher soll in der vorliegenden Arbeit in Ergänzung zur Validitätsbestimmung folgende Frage thematisiert werden:

- 6) Welche Bedeutung für die Beurteilung des Wissensbestandes eines Probanden haben verschiedene Arten der Aufbereitung des Assoziationsmaterials
 - a) durch die Gewichtung qualitativ unterschiedlicher Assoziationen?
 - b) durch die Zugrundelegung lediglich der hochfrequenten Assoziationen innerhalb der Stichprobe?

Über die Bestimmung der Hauptgütekriterien hinaus setzt sich die vorliegende Arbeit mit den Nebengütekriterien

- 7) Veränderungssensitivität,
- 8) Akzeptanz,
- 9) Ökonomie,
- 10) Nützlichkeit,
- 11) Normierung und
- 12) Vergleichbarkeit

auseinander. Hinsichtlich der Veränderungssensitivität sowie der Akzeptanz des Wortassoziationsversuchs, zu denen in der Literatur keine Erkenntnisse zu finden sind, sollen empirische Befunde bereitgestellt werden. Rothe (2003) macht Aussagen zur Ökonomie des Wortassoziationsversuchs, in der vorliegenden Arbeit sollen diese vervollständigt werden um Erkenntnisse, wie sich die Ökonomie des Assoziierens im Vergleich mit anderen wissensdiagnostischen Verfahren bewährt. Zu den Nebengütekriterien der Nützlichkeit, Normierung sowie der Vergleichbarkeit sind in der Literatur keine Aussagen den Wortassoziationsversuch betreffend bekannt. In der vorliegenden Arbeit sollen diese sechs Nebengütekriterien dennoch thematisiert und – sofern möglich – mit empirischen Befunden belegt werden.

Die vorliegende Arbeit will mit der Bestimmung der Haupt- und Nebengütekriterien einen Beitrag leisten, den Wortassoziationsversuch methodisch abzusichern und damit auch für den Praxiseinsatz zur Diagnose von explizitem Fachwissen attraktiver zu machen.

Teil D - EMPIRISCHE BEFUNDE

Da die vorliegende Arbeit die Daten und Erkenntnisse aus verschiedenen Untersuchungen integriert, sollen diese Studien zunächst kurz vorgestellt werden (Kapitel 4). Diese Einführung in die einzelnen empirischen Untersuchungen ist aus Gründen der Übersichtlichkeit allen weiteren Ausführungen vorangestellt, da die nachfolgenden Kapitel in unterschiedlichem Ausmaß auf diese Studien zurückgreifen werden. Der zweite Abschnitt (Kapitel 5) thematisiert zwei Aspekte der Durchführung eines Wortassoziationsversuchs, die vorgestellten Erhebungen können als Vorstudien begriffen werden. Kapitel 6 widmet sich den Hauptgütekriterien (Validität, Reliabilität, Objektivität), der vierte Abschnitt (Kapitel 7) setzt sich mit den Nebengütekriterien auseinander (Veränderungssensitivität, Akzeptanz, Ökonomie, Nützlichkeit, Normierung, Vergleichbarkeit).

Den empirischen Befunden vorangestellt werden sollen noch einige methodische Vorbemerkungen. Aufgrund der oftmals kleinen Substichproben, v.a. beim Extremgruppenvergleich, wurden überwiegend nichtparametrische Verfahren zur Signifikanzprüfung herangezogen. Da aufgrund der in Kapitel B zusammengetragenen Befunde und theoretischen Überlegungen im Wesentlichen die Annahme bestand, daß der Assoziationskennwert linear mit einem Kriterium zusammenhängt (z.B. ein hoher Assoziationskennwert geht mit einer hohen Leistung in der Klausur einher und ist daher als ein Indikator für ein hohes Maß an fachspezifischem Wissen zu interpretieren), konnten insbesondere die korrelativen Zusammenhänge mit einseitigen Signifikanztestungen durchgeführt werden. Soweit von diesem Vorgehen abgewichen wurde, ist dies vermerkt.

4. Einführung in die empirischen Studien

Die empirischen Ausführungen integrieren Befunde, welche aus verschiedenen Studien stammen und am Institut für Psychologie der Universität Potsdam sowie der Humboldt-Universität zu Berlin durchgeführt wurden. Auf diese Weise können in der vorliegenden Arbeit die Ergebnisse von 17 Studien mit insgesamt 838 Probanden berichtet werden. Bungard, Holling und Schultz-Gambard (1996) befürworten die Strategie, anstatt einer großen Studie lieber viele kleine Studien mosaikartig zu einem Gesamtbefund zusammensetzen, da ein solches Vorgehen sich günstiger auf die externe Validität auswirke. Viele der hier zusammengetragenen Studien sind im Rahmen von Diplomarbeiten o.ä. Projekten entstanden. An dieser Stelle sei allen Versuchsleitern bzw. Autoren dieser Studien herzlichst gedankt für die Überlassung ihrer Daten und ihre Unterstützung. Soweit ihre Arbeiten in den hier folgenden Ausführungen zitiert werden, ist dies durch entsprechende Quellenangaben gekennzeichnet. Ist dies nicht der Fall, so stellen die berichteten Ergebnisse eigener Berechnungen durch die Autorin der vorliegenden Arbeit dar.

Zunächst soll ein kurzer Überblick über alle in diese Arbeit eingeflossenen Studien gegeben werden. Dazu werden in tabellarischer Form (siehe Abbildung 3) die wichtigsten, für die vorliegende Arbeit relevanten Merkmale der jeweiligen Projekte aufgeführt, nähere Erläuterungen zu den einzelnen Studien werden im Anschluß gegeben. Wie in Abbildung 3 ersichtlich, wurden für diese Arbeit überwiegend Feldstudien zusammengetragen, diese Forschungsarbeiten fanden im Kontext realer Settings statt oder, wie Lewin (1986) es beschreibt, im Lebensraum, der sich hier als

Arbeitsleben eingrenzen läßt. Drei Studien verfolgen einen experimentellen Ansatz innerhalb eines Feldes (Studie 2 – psychotherapeutische Ambulanz; Studie 4 – Sparkassenakademie B; Studie 5 – Biopsychologie), bei denen in einem natürlichen Umfeld (Sarris und Reiß, 2005) Probanden zufällig verschiedenen Versuchsbedingungen zugewiesen wurden. Nur eine der berücksichtigten Studien ist als Experiment i.e.S. einzuordnen (Studie 1 – Modus), was sich gemäß Lewin (1986) durch die gezielte Manipulierbarkeit der unabhängigen Variablen, der zufälligen Zuweisung der Probanden zu den Versuchsbedingungen sowie der Äquivalenz der Probandengruppen auszeichnet. Da die Verfahrenskonstruktion des Wortassoziationsversuchs auf den Einsatz in der arbeitspsychologischen Praxis zielt, ist es auch angemessen, daß sich die Daten zur Verfahrensvalidierung auch zentral auf Felderhebungen in realen Settings stützen. Würden ausschließlich durch Laborexperimente gewonnenen Daten herangezogen, wäre die Aussagekraft dieser Erkenntnisse für den Einsatz in der Praxis zumindest in Zweifel zu ziehen, auch Rosenstiel (2004) sieht den Nutzen solcher reinen Laboransätze kritisch für den Anwendungsbezug arbeitspsychologischer Forschung. Insbesondere der in der psychologischen Grundlagenforschung übliche Rückgriff auf ausschließlich studentische Probanden scheint für die vorliegende Arbeit wenig brauchbar, da es sich dabei i.d.R. um junge, höher gebildete Personen handelt, deren Teilnahmemotivation an der Untersuchung auf Interesse oder dem Wunsch nach Versuchspersonenstunden basiert. Die Motivationslage von Arbeitspersonen, sich an einer wissensdiagnostischen Erhebung zu beteiligen, ist vermutlich eine gänzlich andere, zudem kann die gesamte Variationsbreite soziodemographischer Merkmale zum Tragen kommen, insbesondere ist hier das Bildungsniveau zu nennen. Kurzum, es ist anzunehmen, daß sich das Test-Verhalten von Psychologiestudierenden im Laborexperiment vom Verhalten Beschäftigter im Feld unterscheidet.

Mit Feldstudien verbunden sind gewisse Besonderheiten der Stichproben, Sarris und Reiß (2005) verweisen dabei auf das Ausmaß, in dem Störvariablen kontrolliert werden können. Meist stellt die Stichprobe einer Feldstudie der Arbeitspsychologie einen Ausschnitt des Gesamtfeldes dar, der aufgrund von betrieblichen oder auch praktischen Erwägungen eingegrenzt wurde. Dieser Ausschnitt ist i.d.R. nicht alleine aufgrund der Überlegungen des Versuchsleiters gewählt, sondern meist durch die Interessen der Unternehmensleitung und oftmals auch durch die Befürchtungen der Personalvertretung determiniert. Zwar wird in Felduntersuchungen grundsätzlich die Freiwilligkeit der Teilnahme der Probanden an der Untersuchung betont, aber dennoch sehen sich Beschäftigte nicht selten einer Erwartungshaltung des Vorgesetzten gegenüber und erleben somit sozialen Druck. Entsprechend bilden auch die Untersuchungsteilnehmer selbst Hypothesen über die Zielstellung der Untersuchung und verhalten sich u.U. gemäß ihrer Vermutung. Dies kann das Spektrum abdecken von vermuteter sozialer Erwünschtheit und engagierter Beteiligung an der Erhebung über Reaktanzverhalten i.S. einer „schlechten“ Versuchsperson (siehe auch Lewin, 1986) bis hin zum gänzlichen Verweigern der Beteiligung. Bungard, Holling und Schultz-Gambard (1996) sprechen in diesem Zusammenhang vom sog. Auftraggebereffekt – dem Forscher wird unterstellt, die Interessen der Unternehmensleitung zu vertreten. Somit lassen sich diese Stichproben wohl am ehesten als Gegebenheitsstichproben einordnen (Lewin, 1986).

Nr.	Name der Studie	Art der Studie	Probanden	N	Meßzeitpunkte	Wissensdomäne	Anzahl Stimuli	Durchführungsmodus	Zeitvorgabe	Reihenfolge der Stimuli
1	Modus	Experiment	unspezifisch	30	1	Allgemein- / Spezialwissen (Bankwesen)	40	mündlich / schriftlich (Papier)	keine	fest
2	psychotherapeutische Ambulanz	Feldexperiment	Mitarbeiter, KJP-Auszubildende	20	2	psychische Störungen	14	schriftlich (PC)	keine	variabel
3	Sparkassenakademie (A)	Feldstudie	Weiterbildungsteilnehmer	38	4	Management und Führung	4	schriftlich (Papier)	2 min pr.St.	fest
4	Sparkassenakademie (B)	Feldexperiment	Weiterbildungsteilnehmer	35	3	Management und Führung	6	schriftlich (Papier)	2 min pr.St.	fest
5	Biopsychologie	Feldexperiment	Psychologiestudierende	9	1	Nervensystem	10	schriftlich (PC)	keine	fest
6	Klausur Klinische Psychologie	Feldstudie	Psychologiestudierende	38	1	Klinische Psychologie	4	schriftlich (Papier)	keine	variabel
7	Klausur Arbeitspsychologie	Feldstudie	Psychologiestudierende	13	1	Arbeitspsychologie	4	schriftlich (Papier)	keine	variabel
8	Klausur Personalpsychologie	Feldstudie	Psychologiestudierende	8	1	Personalpsychologie	4	schriftlich (Papier)	keine	variabel
9	Klausur Entwicklungspsychologie	Feldstudie	Psychologiestudierende	17	1	Entwicklungspsychologie	4	schriftlich (Papier)	keine	variabel
10	Gebäudereiniger	Feldstudie	Berufsschüler	33	1	Gebäudereinigung	10	schriftlich (PC)	keine	fest
11	Bauklempner	Feldstudie	Berufsschüler	14	1	Dachbau	10	schriftlich (PC)	keine	fest
12	Handelsunternehmen	Feldstudie	Mitarbeiter	17	1	Schweißen	26	mündlich	2 min pr.St.	variabel
13	Stoffdruckerei	Feldstudie	Mitarbeiter	21	1	Stoffdruck	28	mündlich	keine	fest
14	Seminar Wissensmanagement	Feldstudie	Psychologiestudierende	20	2	Wissensmanagement	1	schriftlich (PC)	keine	fest
15	Wissen über Arbeitsschutz	Feldstudie	Studierende	472	1	Arbeits- und Gesundheitsschutz	3	schriftlich (PC, teilweise Papier)	keine	fest
16	Call Center (A)	Feldstudie	Mitarbeiter	29	1	DSL	30	mündlich	keine	fest
17	Call Center (B)	Feldstudie	Mitarbeiter	24	1	Web-Hosting	30	mündlich	keine	fest

Abbildung 3: Übersicht über alle in der vorliegenden Arbeit herangezogenen empirischen Studien und ihre Merkmale bzgl. des Wortassoziationsversuchs (pr.St. = pro Stimulus)

Feld-Stichproben in der Arbeitspsychologie sind also oftmals eher klein und i.d.R. sehr gezielt ausgewählt, wie auch bei den in der vorliegenden Arbeit herangezogenen empirischen Studien deutlich wird. Dabei ist zu bedenken, daß der Umfang des Untersuchungsfeldes in der Arbeitspsychologie in jedem Falle nach oben hin begrenzt und insbesondere beim Umgang mit arbeitsaufgabenrelevanten Wissensdomänen nicht beliebig erweiterbar ist. Im Kontext von Wissensanalysen ist diese Begrenzung des Untersuchungsgegenstandes und somit auch der Stichprobe unumgänglich. Schließlich will man eine Aussage treffen über eine Kohorte, die sich mit identischem Arbeitsgegenstand (und damit mit einem vergleichbaren Wissensgegenstand) befaßt und so die Vergleichbarkeit der Individuen ermöglicht. Neben den beschriebenen Nachteilen für die Bildung der Stichprobe bietet die Felduntersuchung aber den großen Vorteil einer realen Untersuchungssituation, wodurch sich die externe Validität der Erkenntnisse erhöht (Sarris und Reiß, 2005).

Ebenso wird in Abbildung 3 deutlich, wie unterschiedlich der Wortassoziationsversuch eingesetzt werden kann: in den vorgestellten Studien wurde er sowohl als mündliche Variante durchgeführt wie auch als schriftliche, wobei in letzterem Fall noch zwischen Papier-und-Bleistift-Verfahren vs. computerbasierten Ansätzen zu unterscheiden ist. Auch die Anzahl der vorgelegten Stimuli innerhalb des Wortassoziationsversuchs variiert erheblich über die Studien: zwischen einem und 40 Stimuli wurden jeweils von den Probanden bearbeitet. In einigen Studien wurden die Stimuli allen Probanden in der gleichen Reihenfolge präsentiert, in anderen wurde die Reihenfolge zufällig über die Stichprobe variiert. In den meisten Studien hatten die Probanden keine zeitlichen Restriktionen bei der Bearbeitung des Wortassoziationsversuchs zu beachten, sie konnten selbst entscheiden, wie lange sie sich an einem Stimulus aufhielten. In einigen wenigen Studien bestand die Regel, nach maximal zwei Minuten zum nächsten Stimulus überzugehen. Die Anzahl der Meßzeitpunkte ist in Abhängigkeit von der jeweiligen Fragestellung der Studie zu verstehen, der längsschnittliche Ansatz beruht auf der angestrebten Veränderungsmessung von Merkmalen. Die in den Assoziationsversuchen bearbeiteten Wissensdomänen spiegeln die unterschiedlichen Untersuchungsfelder wider und zeichnen ein abwechslungsreiches Bild über die flexible Einsatzfähigkeit des Wortassoziationsversuchs.

Im Folgenden sollen die hier zusammengetragenen 17 Studien im Einzelnen vorgestellt werden. Es wird dabei nur auf diejenigen Aspekte in Design und Methodik eingegangen, welche im Zusammenhang mit dem Wortassoziationsversuch bedeutsam sind. Es sei an dieser Stelle um Verständnis gebeten, wenn die Darstellungen der Studien oder deren Ergebnisse unterschiedlich detailliert ausfallen. Aussagen wie z.B. über die Inhalte etwaiger Klausuren waren nicht immer den Berichten zu entnehmen bzw. die Autoren verfügten grundsätzlich nicht über solche Informationen. Eine Angleichung i.S. des „kleinsten gemeinsamen Nenners“ schien wenig sinnvoll, da dies in vielen Fällen bedeutet hätte, vorhandene und für das Verständnis der Ergebnisse wertvolle Informationen zu unterschlagen.

Außerdem sollen hier bereits kurz die deskriptiven Ergebnisse den Wortassoziationsversuch betreffend referiert werden, um das jeweilige Untersuchungsfeld näher zu beschreiben und die nachfolgenden Berechnungen verständlicher zu machen. Um eine Vergleichbarkeit der hier vorgestellten empirischen Studien zu ermöglichen, wurde in der vorliegenden Arbeit für alle Erhebungen eine einheitliche Auswertestrategie angewandt (z.T. in Abweichung von den in einigen

Diplomarbeiten berichteten Befunden). Danach gingen in den sog. Assoziationskennwert, welcher die Anzahl an Assoziationen eines Probanden beschreibt, *alle* vom Probanden geäußerten (sinnvollen) Assoziationen ein, lediglich Wiederholungen des Probanden innerhalb eines Stimulusbegriffes blieben in der Auszählung unberücksichtigt. Dabei ist der Assoziationskennwert eines Probanden jeweils der Mittelwert der Assoziationen über die jeweilige Anzahl der Stimuli hinweg. Wie bereits in Kapitel 3.6 (Seite 39 ff.) erörtert, scheint dieser Parameter sehr geeignet, da er – ganz im Sinne eines Screening-Verfahrens – leicht zu erheben ist und somit für alle hier vorgestellten Studien vorlag (im Gegensatz z.B. zu einem Kennwert für die Reaktionszeiten).

Neben der Betrachtung der statistischen Kennwerte für die gesamte Stichprobe soll weiterhin für jede Studie illustriert werden, wie der Wortassoziationsversuch anhand seines quantitativen Kennwertes in der Lage ist, grob zwischen den Leistungen der Probanden zu unterscheiden. Dazu wurden jeweils Extremgruppen gebildet, wobei die Gruppe der sog. „guten Assoziierer“ diejenigen Probanden zusammenfaßt, welche innerhalb der Stichprobe einen sehr hohen Assoziationskennwert erreicht haben, während die Gruppe der sog. „schlechten Assoziierer“ die Probanden mit dem jeweils geringsten Assoziationskennwert beschreibt. Das Kriterium für diese Extremgruppenbildung orientierte sich jeweils an der Abweichung der individuellen Assoziationskennwerte etwa eine Standardabweichung über- bzw. unterhalb des Stichprobenmittelwertes. In Anbetracht der oftmals sehr kleinen Stichproben konnte dieses Vorgehen jedoch nur als Richtmaß dienen, vielmehr wurde in diesen Fällen die Entscheidung aufgrund der Verteilung des Assoziationskennwertes getroffen, wobei versucht wurde, die beiden Extremstichproben nach Möglichkeit etwa gleich groß zu gestalten. Aufgrund der sehr kleinen Substichproben im Extremwertevergleich wurden nichtparametrische Verfahren zur Signifikanzprüfung herangezogen (ausgenommen bei Studie 15, s.u.). Auf das Berichten der deskriptiven Ergebnisse etwaiger anderer eingesetzter Verfahren soll an dieser Stelle aus Gründen der Übersichtlichkeit verzichtet werden. Diesbezügliche Daten sollen erst an späterer Stelle bei der Erörterung der Gütekriterien berichtet werden, wenn die entsprechenden Verfahren thematisiert werden (in den Kapiteln 6 bzw. 7).

4.1. Studie 1: Modus der Erhebung

4.1.1. Anliegen der Studie und Untersuchungsfeld

Zur Überprüfung, inwieweit der Modus der Erhebung einen Einfluß auf die Assoziationsleistung hat, wurde eine separate, experimentelle Studie dazu angelegt. Dabei wurde die mündliche Durchführung (ähnlich einer Interviewsituation) verglichen mit der schriftlichen Erhebung mittels Papier und Stift, um so einen intraindividuellen Vergleich zu erhalten. Eine computerbasierte schriftliche Erhebung wurde hier außer Acht gelassen, da zum Untersuchungszeitpunkt noch keine geeignete Software bekannt war. Da Strube (1984) zwar von einer Äquivalenz einer mündlichen und schriftlichen Durchführung berichtet, seine Aussagen sich aber auf Stimuli aus dem Allgemeinwissensbereich stützen, wurde diese Studie außerdem dazu genutzt, die Befunde von Strube (1984) das Allgemeinwissen betreffend zu replizieren und darüber hinaus deren Geltungsbereich auf spezialisiertere Wissensbestände auszudehnen.

Als Versuchspersonen wurden insgesamt 30 Personen aus dem Bekanntenkreis der Versuchsleiterin gewonnen. Dadurch konnte außer studentischen Teilnehmern auch ein Klientel höheren Lebensalters einbezogen werden: das Durchschnittsalter in dieser Stichprobe betrug 50,90 Jahre ($SD=14,45$), wobei eine Spanne zwischen minimal 24 Jahren und maximal 68 Jahren abgedeckt wurde. Gerade unter den Probanden höheren Alters befanden sich viele selbständige Handwerker und Kaufleute, drei Probanden sind bzw. waren Mitarbeiter einer Bank. Die Teilnahme an der Untersuchung wurde nicht vergütet.

4.1.2. Untersuchungsaufbau

Da, wie bereits erwähnt, diese Studie einen intraindividuellen Vergleich hinsichtlich des Durchführungsmodus anstrebte, wurde von jedem Probanden sowohl ein schriftlicher als auch ein mündlicher Wortassoziationsversuch bearbeitet. Der Ablauf der Erhebung ist in Abbildung 4 skizziert. Zu Beginn der Erhebung wurde der Wortflüssigkeitstest des WILDE-Intelligenztests in seiner Kurzform gemäß der entsprechenden Instruktionen nach Jäger und Althoff (1994) eingesetzt.

	Form A	Form B
1.	WILDE-Untertest „Wortflüssigkeit“	WILDE-Untertest „Wortflüssigkeit“
2.	Wortassoziationsversuch Block 1 mündlich	Wortassoziationsversuch Block 1 schriftlich
3.	Fragebogen zur Akzeptanz des mündlichen Teils	Fragebogen zur Akzeptanz des schriftlichen Teils
4.	Wortassoziationsversuch Block 2 schriftlich	Wortassoziationsversuch Block 2 mündlich
5.	Fragebogen zur Akzeptanz des schriftlichen Teils, vergleichende Bewertung der Modi	Fragebogen zur Akzeptanz des mündlichen Teils, vergleichende Bewertung der Modi

Abbildung 4: Schematischer Ablauf in Studie 1 – Modus der Erhebung (nach Teilstichproben)

Um mögliche Reihenfolgeeffekte der Modi auszuschließen, wurde systematisch variiert, welchen Modus die Probanden zuerst bearbeiteten, während die Reihenfolge der Stimuli stabil blieb. Dazu wurde der Pool der Stimuli in zwei gleich große Blöcke geteilt. Die Probanden wurden nach dem Zufallsprinzip zwei Gruppen zugewiesen, welche sich dahingehend unterschieden, ob sie die Stimuli zuerst mündlich (Block 1) und anschließend schriftlich (Block 2) bearbeiten (Form A, siehe Abbildung 4) oder umgekehrt (Form B). Im Anschluß an jeden Block wurde den Probanden ein kurzer Fragebogen vorgelegt, in dem sie das Verfahren bewerten sollten. Zur Bewertung ökonomischer Gesichtspunkte bei der Durchführung verschiedener Erhebungsmodi wurde die Zeit erfaßt, welche die Versuchsteilnehmer jeweils für die Durchführung eines Modus brauchten.

Fragebogen zur Akzeptanz: In einem kurzen Fragebogen wurden die Probanden gebeten, unmittelbar nach dem mündlichen bzw. schriftlichen Teil zu bewerten, wie leicht oder schwer es ihnen fiel, zu antworten und dabei nur Stichworte zu verwenden. Zudem wurde nach der Verständlichkeit der Instruktion gefragt sowie bei der mündlichen Variante nach der erlebten Unterstützungsfunktion durch die gleichzeitige auditive und schriftliche Präsentation der Stimuli durch den Versuchsleiter. Abschließend wurde dem Probanden ein Item zur Präferenz gegenüber den Verfahrensteilen vorgelegt, welches am Schluß des Arbeitsbogens positioniert war. Alle diese Items wurden im Multiple-Choice-Format vorgegeben (i.d.R. als fünfstufige Skala der Zustimmung, siehe auch Anhang 1) sowie eine Begründung für die jeweilige Entscheidung erfragt.

4.1.3. Wortassoziationsversuch

Der Wortassoziationsversuch umfaßte zu gleichen Teilen Stimuli aus dem Allgemeinwissen wie aus einem Spezialwissensbereich, insgesamt wurden 40 Stimuli dargeboten. Die Begriffe des Allgemeinwissensbereichs wurden Strube (1984) entnommen. Als Spezialwissensbereich wurde die Domäne Bankgeschäfte gewählt, da dieser Bereich hinreichend spezifisch ist und angenommen werden kann, daß auch normale Bankkunden Wissen über diese Domäne besitzen. Die Stimuli zu dieser Domäne wurden auf Basis der Kundenseiten des Internetauftritts einer Sparkasse zusammengestellt. Das den Probanden vorgelegte Untersuchungsmaterial ist in Anhang 1 in Ausschnitten dargestellt.

In Abbildung 5 finden sich alle Stimuli in der vorgelegten Reihenfolge (siehe Numerierung); außerdem ist Zuordnung der Stimuli zu den zwei Blöcken gekennzeichnet, welche mit jeweils 10 Stimuli den Bereich Allgemein- als auch Spezialwissen abdeckten. Allen Versuchsteilnehmern wurden diese insgesamt 40 Stimuli in der gleichen Reihenfolge präsentiert, variiert wurde lediglich die Reihenfolge der beiden Bearbeitungsmodi.

	Allgemeinwissen	Spezialwissen: Bankgeschäfte
<i>Block 1</i>	1 Wasser	11 Kontonummer
	2 Katze	12 Kreditkarte
	3 Schule	13 Sparkassenbrief
	4 Hirsch	14 Geldautomat
	5 Schlüssel	15 Dispositionskredit
	6 Bär	16 Überweisung
	7 Hase	17 Kontoauszug
	8 Flasche	18 Lastschrift
	9 Löwe	19 Bausparvertrag
	10 Maus	20 Kapitalanlage
<i>Block 2</i>	21 Schaf	31 Girokonto
	22 Buch	32 Scheck
	23 Kuh	33 Sparkonto
	24 Fenster	34 Zinsen
	25 Hund	35 Kontoführung
	26 Lehm	36 Festgeld
	27 Ziege	37 Buchungstag
	28 Schwein	38 Dauerauftrag
	29 Haus	39 Bankeinzug
	30 Pferd	40 Saldo

Abbildung 5: Stimuli des Wortassoziationsversuchs in Studie 1 – Modus der Erhebung

Der mündliche Teil des Wortassoziationsversuchs wurde so gestaltet, daß die Versuchsleiterin die Instruktion vorlas, anschließend wurden die Stimuli sukzessive genannt. Parallel dazu wurde der jeweilige Stimulus auf einem Kärtchen präsentiert und war für den Probanden solange sichtbar, bis der nächste Stimulus bearbeitet wurde. Die Äußerungen des Probanden wurden auf Tonband aufgezeichnet und später transkribiert. Für den schriftlichen Teil des Versuchs wurde ein Arbeitsbogen vorbereitet: nach einmaliger Instruktion wurde jeder der 20 Stimuli in den oberen Bereich des Blattes geschrieben, der Rest der Seite blieb leer, so daß für jeden Stimulus eine ganze

Seite zur Verfügung stand. Dieser Arbeitsbogen enthielt außerdem ein Deckblatt mit einigen kurzen Worten zum Zweck dieser Untersuchung, der Abfrage des Alters des Probanden sowie den Vordrucken zur Bearbeitung des WIT-Wortflüssigkeitstests, am Ende des Arbeitsbogens fand sich der Fragebogen zur Akzeptanz der Verfahrensmodi.

Ungeachtet der verschiedenen Bestandteile dieses Wortassoziationsversuchs (Darbietungsform, Wissensbereich) wird an dieser Stelle – analog zu den anderen hier berichteten Studien – der Assoziationskennwert über alle vorgelegten Stimuli hinweg herangezogen (einer differenzierten Auswertung dieser Versuchsteile widmet sich Kapitel 5.1 in einer Vorstudie zum Durchführungsmodus, Seite 55 ff.). Die Probanden assoziierten im Mittel $M=6,15$ Begriffe ($SD=3,09$), minimal wurde ein Kennwert von 2,98 erreicht, maximal ein Kennwert von 16,58. Auf Basis dieses Assoziationskennwertes konnten zwei Extremgruppen gebildet werden, welche sich signifikant voneinander abgrenzen ließen (Mann-Whitney-U: $z=-3,243$; $p=.000$). Wie auch in Abbildung 6 ersichtlich, ist die Gruppe der sieben guten Assoziierer durch einen mittleren Assoziationskennwert von 10,79 ($SD=2,90$) gekennzeichnet, die Gruppe der acht schlechten Assoziierer hingegen produzierte durchschnittlich $M=3,47$ Begriffe ($SD=0,39$).

	gesamte Stichprobe (N=30)	gute Assoziierer (N=7)	schlechte Assoziierer (N=8)
<i>M</i>	6,15	10,79	3,47
<i>SD</i>	3,09	2,90	0,39
Minimum	2,98	7,93	2,98
Maximum	16,58	16,58	3,97

Abbildung 6: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 1 - Modus

4.2. Studie 2: Psychotherapeutische Ambulanz

4.2.1. Anliegen der Studie und Untersuchungsfeld

Eine separate, quasi-experimentelle Feldstudie wurde durchgeführt, um die Retest-Reliabilität des Wortassoziationsversuchs zu ermitteln. Dazu konnten Mitarbeiter einer psychotherapeutischen Ambulanz für Kinder und Jugendliche gewonnen werden. Gleichzeitig diente diese Untersuchung dazu zu klären, ob die Position von Ober- vs. Unterbegriffen einer Wissensdomäne innerhalb der Stimulusliste einen Einfluß auf die Assoziationsleistung hat. Die Frage dabei ist: erbringt eine Anordnung von Stimulusbegriffen, die vom Besonderen (Unterbegriff) zum Allgemeinen (Oberbegriff) führen, quantitativ und qualitativ andere Assoziationsleistungen als eine Stimuliabfolge, die vom Allgemeinen zum Besonderen führt.

Als Probanden konnten Mitarbeiter einer Hochschulambulanz gewonnen werden, welche bei Kindern und Jugendlichen psychotherapeutische Leistungen (Verhaltenstherapie) erbringt. Aufgrund der organisatorischen Einbindung in den Lehrstuhl für Klinische Psychologie sowie in eine Ausbildungseinrichtung für Kinder- und Jugendlichenpsychotherapeuten arbeiten in der Ambulanz zum einen approbierte Kinder- und Jugendlichenpsychotherapeuten bzw. wissenschaftliche Mitarbeiter des Lehrstuhls, außerdem Psychologiestudenten im fortgeschrittenen Semester, die i.d.R. als studentische Hilfskraft am Lehrstuhl für Klinische Psychologie tätig sind.

Der größte Teil der Mitarbeiter der Ambulanz befindet sich in der Ausbildung zum Kinder- und Jugendlichenpsychotherapeuten (d.h. entweder Diplom-Psychologen oder Diplom-Pädagogen). In der Ambulanz leisten diese Ausbildungstherapeuten den praktischen Teil ihrer Ausbildung ab, d.h. sie behandeln Patienten unter Supervision (i.d.R. parallel zur Theorieausbildung).

Als Versuchspersonen konnten insgesamt 20 Personen gewonnen werden, die Expertenkenntnisse bzgl. der psychischen Störungen im Kindes- und Jugendalter vorweisen können. Die Teilnahme an der Erhebung war freiwillig und wurde nicht vergütet. Diese Stichprobe setzt sich zusammen aus sechs Psychologie-Studierenden, zwei Diplom-Psychologen, sieben Kinder- und Jugendlichentherapeuten in Ausbildung sowie fünf bereits approbierten Therapeuten. So konnte sichergestellt werden, daß die Probanden sich einigermaßen sicher in der Wissensdomäne bewegen können. Andererseits dient eine gewisse Heterogenität der Probanden bzgl. ihres Ausbildungsstandes dazu, die Streuung in den individuellen Wissensbeständen und damit letztlich auch der Testwerte zu gewährleisten.

4.2.2. Untersuchungsaufbau

Da die vorliegende Studie der Feststellung der Retest-Reliabilität dienen sollte, wurden zwei Meßzeitpunkte realisiert, zwischen denen etwa zwei Wochen Zeit lagen. Dieser zeitliche Abstand wurde als geeignet angesehen, da er lang genug schien, um Erinnerungseffekte bei der Testwiederholung zu vermeiden, was durch die recht umfangreiche Stimulusliste noch unterstützt wurde. Andererseits war zu beachten, daß nicht zu viel Zeit zwischen den Terminen lag, da zum einen die Terminabsprache mit den Probanden planbar und realistisch sein sollte, zum anderen der Einfluß von Lernereignissen auf den Wissensbesitz möglichst zu vermeiden war (da viele Probanden sich in einer langfristig angelegten Lernphase wie Studium bzw. Therapeutenausbildung befanden, war das Risiko solcher Ereignisse gegeben), so daß auch unter diesem Aspekt ein Abstand von zwei Wochen realistisch schien.

T1	T2
1. WIT-Untertest zur Wortflüssigkeit	1. Wortassoziationsversuch (Stimuli 1-14)
2. Wortassoziationsversuch (Stimuli 1 – 14)	2. Fragebogen zur Akzeptanz des Verfahrens Wortassoziationsversuch
3. Fragebogen zur Akzeptanz des Verfahrens Wortassoziationsversuch	3. Lückentext zu den Themen
	4. Fragebogen zur Akzeptanz des Verfahrens Lückentext

Abbildung 7: Schematischer Ablauf in Studie 2 – psychotherapeutische Ambulanz

Wie in Abbildung 7 ersichtlich, wurde zu Beginn der Erhebung (T1) der Wortflüssigkeitstest des WIT in seiner Kurzform gemäß der entsprechenden Instruktionen nach Jäger und Althoff (1994) eingesetzt. Anschließend wurde der Proband gebeten, die Instruktion für den Wortassoziationsversuch im Testmaterial zu lesen, ggf. wurden Fragen dazu beantwortet. Der Wortassoziationsversuch selbst wurde schriftlich durchgeführt, hierzu konnte auf das computergestützte Assoziationsstool des Assoziations-Struktur-Tests (Meyer, 2007) zurückgegriffen werden. Nach Bearbeitung des Wortassoziationsversuchs wurden die Probanden um eine schriftliche Einschätzung dieses Verfahrens gebeten.

Der zweite Untersuchungszeitpunkt (T2) begann mit dem Wortassoziationsversuch, bei dem die gleichen Stimuli in gleicher Weise wie beim ersten Testzeitpunkt vorgelegt wurden. Im Anschluß wurde wieder nach der Bewertung des Verfahrens gefragt. Daran schloß sich unmittelbar die Bearbeitung eines Lückentextes in einer Papier-Bleistift-Version an. Abschließend wurde auch hinsichtlich des Lückentextes um eine schriftliche Bewertung des Verfahrens sowie eine zusammenfassende Präferenz gegenüber einem der beiden Verfahren gebeten.

Die Computerunterstützung erlaubte ein gleichzeitiges Testen mehrerer Probanden, aus Rücksicht auf das individuelle Zeitmanagement der Probanden überwogen jedoch die Einzelsitzungen. Die Erhebung wurde arbeitsplatznah und innerhalb der Arbeitszeit durchgeführt. Darüber hinaus wurde zur Bewertung ökonomischer Gesichtspunkte bei der Durchführung der Wissensdiagnosen die Zeit erfaßt, welche die Versuchsteilnehmer jeweils für die Durchführung des Wortassoziationsversuchs sowie des Lückentextes benötigten. Zunächst sollen die Untersuchungsbestandteile Lückentext sowie der Fragebogen zur Akzeptanz der Verfahren kurz skizziert werden, anschließend wird das Vorgehen beim Wortassoziationsversuch ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Lückentext: Um die Ergebnisse des Wortassoziationsversuchs mit einem konstruktverwandten Verfahren vergleichen zu können, wurde zusätzlich ein Lückentext in die Untersuchung aufgenommen, welcher den Probanden zum zweiten Meßzeitpunkt vorgelegt wurde (siehe auch Ablaufplan in Abbildung 7). Die Form des Lückentextes wurde gewählt, um die Ratewahrscheinlichkeit, welche gebundenen Antwortformaten wie z.B. Multiple-Choice-Aufgaben innewohnt, zu minimieren. Ziel bei der Konstruktion des Lückentextes war eine größtmögliche Vergleichbarkeit der Inhalte von Wortassoziationsversuch und Lückentext. Da der Wortassoziationsversuch sich an den Störungsklassifikationen der ICD-10 (Weltgesundheitsorganisation, 1999) orientiert (siehe nachfolgendes Kapitel), basiert der dem Lückentext zugrundeliegende Inhalt auf den Erläuterungen, die die ICD-10 zu den einzelnen Störungsbildern liefert. Soweit möglich wurden Textpassagen unverändert übernommen, an einigen Stellen war jedoch eine sprachliche Anpassung notwendig, um den Text gleichsam lesbar und in Bezug auf den zu ergänzenden Begriff möglichst eindeutig zu gestalten.

Für den Inhalt des Lückentextes wurden diejenigen Störungsgruppen herangezogen, die bereits in den Wortassoziationsversuch eingingen. Zu jeder Störungsgruppe wurde ein kurzer Text entwickelt, in dem jeweils vier Lücken zu füllen waren, d.h. vier Punkte erreicht werden konnten. Als Referenz für die als richtig zu bewertenden Antworten dienten ebenfalls die Angaben im ICD-10 (Weltgesundheitsorganisation, 1999). Für die Gruppe F9 konnte keine Passage im Lückentext erstellt werden, da zu dieser übergeordneten Kategorie kein Text im ICD-10 (Weltgesundheitsorganisation, 1999) zur Verfügung stand. Ähnliches gilt für die Gruppe F93, hier war zwar ein kurzer Text vorhanden, aber so gestaltet, daß sich daraus kein sinnvoller Lückentext erstellen ließ. So ergab sich schließlich ein Lückentext, bei dem in insgesamt zwölf Themen jeweils vier Punkte und somit eine Gesamtpunktzahl von 48 Punkten erreicht werden konnte. Der gesamte Lückentext findet sich in Anhang 3. Die Reihenfolge der Präsentation im Lückentext bildet diejenige Reihenfolge ab, welche die ICD-10 (Weltgesundheitsorganisation, 1999) vorgibt, so daß keine Übereinstimmung mit der Reihenfolge innerhalb des Wortassoziationsversuchs besteht.

Fragebogen zur Akzeptanz des Verfahrens: Die Probanden wurden gebeten, nach dem Wortassoziationsversuch zu bewerten, wie leicht oder schwer es ihnen fiel, stichwortartig zu antworten, ebenso wurde nach der Verständlichkeit der Instruktion gefragt. Alle Items wurden als Multiple-Choice-Test vorgegeben (als fünfstufige Skala der Zustimmung, siehe auch Anhang 2 bzw. Anhang 3), zudem wurde um eine Begründung für die jeweilige Entscheidung gebeten. Dies geschah nach dem Wortassoziationsversuch sowohl beim ersten als auch beim zweiten Meßzeitpunkt. Zusätzlich wurde zum zweiten Termin erfragt, inwieweit die Probanden das Gefühl hatten, das Verfahren bilde ihr Wissen in angemessener Weise ab. Analog dazu wurde nach dem Bearbeiten des Lückentextes eine Einschätzung desselben erfragt (zweiter Meßzeitpunkt). Abschließend wurde um eine Präferenz hinsichtlich der beiden Verfahren sowie eine Begründung dazu gebeten.

4.2.3. Wortassoziationsversuch

Die Wissensdomäne in dieser Studie umfaßt Gruppen psychischer Störungen im Kindes- und Jugendalter. Als Referenz dient dabei das Internationale Klassifikationssystem psychischer Störungen nach ICD-10 (Weltgesundheitsorganisation, 1999), aus dem zwei Abschnitte ausgewählt wurden: Entwicklungsstörungen (F8) und Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend (F9). Da diese beiden Themen Störungen von Kindern und Jugendlichen beinhalten, eigneten sie sich in besonderer Weise für eine Stichprobe in einer Ambulanz für Kinder- und Jugendlichenpsychotherapie. Insgesamt wurden 14 Stimuli vorgelegt (siehe Abbildung 8), bei der Darbietung innerhalb des Versuchs wurde dabei auch die Codierungsnummer des ICD-10 präsentiert. Zu erwähnen ist, daß Wiederholungen von Assoziationen, die bereits bei einem vorangegangenen Stimulus genannt wurden, ausdrücklich gestattet (und auch erwünscht) waren.

<i>Nr.</i> Thema F8 (Abschnitt des ICD)	<i>Nr.</i> Thema F9 (Abschnitt des ICD)
1) F8 Entwicklungsstörungen	8) F90.0 einfache Aktivitäts- und Aufmerksamkeitsstörung
2) F80 Umschriebene Entwicklungsstörung des Sprechens und der Sprache	9) F91.2 Störung des Sozialverhaltens bei vorhandenen sozialen Bindungen
3) F81 Umschriebene Entwicklungsstörung schulischer Fertigkeiten	10) F93.1 phobische emotionale Störung des Kindesalters
4) F84 tiefgreifende Entwicklungsstörungen	11) F90 hyperkinetische Störungen
5) F80.1 expressive Sprachstörung	12) F91 Störungen des Sozialverhaltens
6) F81.0 Lese- und Rechtschreibstörung	13) F93 emotionale Störungen des Kindesalters
7) F84.0 frühkindlicher Autismus	14) F9 Verhaltens- u. emotionale Störungen mit Beginn in der Kindheit und Jugend

Abbildung 8: Stimuli der Studie 2 – psychotherapeutische Ambulanz

Die Reihenfolge der Stimuli war festgelegt, es existierten zwei Präsentationsvarianten, welche sich durch die Reihenfolge der Stimuli unterschieden. Der einen Hälfte der Probanden wurden die Stimuli wie in der in Abbildung 8 dargestellten Reihenfolge dargeboten, die andere Hälfte der Stichprobe bearbeitete die Stimuli in umgekehrter Reihenfolge (zuerst innerhalb von Thema F8, anschließend innerhalb von Thema F9). Den Probanden wurde der Wortassoziationsversuch zu beiden Meßzeitpunkten mit jeweils identischer Stimulusreihenfolge vorgelegt, die Zuordnung der

Probanden zu einer der beiden Gruppen erfolgte zufällig. Auf Details der Anordnung der Stimuli wird an späterer Stelle eingegangen werden (siehe Kapitel 5.2, Seite 55 ff.).

Wie in Abbildung 9 dargestellt, zeichnet sich diese Studie durch einen ausgeprägt hohen Assoziationskennwert aus: bei der ersten Erhebung assoziierten die Probanden im Mittel 21,70 Konzepte ($SD=16,67$), bei der Wiederholungsmessung $M=19,14$ ($SD=9,50$). Zu beiden Meßzeitpunkten ist der Assoziationskennwert normalverteilt. Dieser augenscheinliche Unterschied zwischen den mittleren Assoziationskennwerten ist v.a. auf die extreme Assoziationsleistung eines einzelnen Probanden zurückzuführen und daher nicht signifikant (Wilcoxon: $z=-.149$; $p>.05$). Die mittleren Assoziationsleistungen sind also zu beiden Erhebungszeitpunkten annähernd vergleichbar, wenngleich anzumerken ist, daß sich die Verteilung des Assoziationskennwertes innerhalb der Stichprobe zur Wiederholungsmessung homogener darstellt.

	gesamte Stichprobe (N=20)		gute Assoziierer (N=3)		schlechte Assoziierer (N=3)	
	T1	T2	T1	T2	T1	T2
<i>M</i>	21,70	19,14	34,50	27,10	8,91	11,19
<i>SD</i>	16,67	9,50	14,10	5,88	2,01	1,09
Minimum	6,79	10,00	23,36	20,93	6,79	10,00
Maximum	50,79	32,64	50,79	32,64	10,79	12,14

Abbildung 9: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe zu beiden Meßzeitpunkten sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 2 - psychotherapeutische Ambulanz

Des weiteren wurden anhand des Assoziationskennwertes Extremgruppen mit jeweils drei Probanden gebildet. Hierbei ist zu berichten, daß die Assoziationskennwerte der Probanden ein insgesamt relativ homogenes Feld bildeten, lediglich im oberen Leistungsbereich setzen sich einige Probanden anhand ihrer Assoziationsleistung etwas ab (wie auch anhand der Standardabweichungen der Extremgruppen in Abbildung 9 deutlich wird). Auf diesen Umstand ist dann auch zurückzuführen, daß insbesondere die Gruppe der schlechten Assoziierer zu den beiden Meßzeitpunkten nicht dieselben Probanden umfaßt – lediglich ein Proband war sowohl zum ersten als auch zum zweiten Erhebungszeitpunkt der Extremgruppe der schlechten Assoziierer zuzuordnen. Rangplatz 1 sowie 2 (in Bezug auf die Assoziationsleistung) hingegen blieben über beide Erhebungen von denselben Probanden besetzt. Lediglich ein Mitglied aus der Gruppe der guten Assoziierer wurde in der Wiederholungsmessung durch einen anderen Probanden ersetzt.

In der ersten Erhebung konnte eine Extremgruppe von Probanden mit hohem Assoziationskennwert identifiziert werden, welche mit $M=34,50$ und $SD=14,10$ eine große Streubreite aufweist. Die Gruppe der schlechten Assoziierer hingegen präsentiert sich mit $M=8,91$ Assoziationen pro Stimulus ($SD=2,01$) als sehr homogen. Die Assoziationskennwerte beider Extremgruppen unterscheiden sich signifikant voneinander (Mann-Whitney-U: $z=-1,993$; $p=.046$). In der Wiederholungsmessung nähern sich beide Extremgruppen einander an (verglichen mit der ersten Erhebung): die guten Assoziierer weisen einen reduzierten mittleren Assoziationskennwert auf $M=27,10$ ($SD=5,88$), während die schlechten Assoziierer ihrerseits mit $M=11,19$ ($SD=1,09$) ihre mittlere Assoziationsleistung steigern konnten. Der Mittelwertsunterschied zwischen beiden Extremgruppen zu T2 ist in der Tendenz signifikant (Mann-Whitney-U: $z=-1,964$; $p=.050$).

4.3. Studie 3: Sparkassenakademie (A)

4.3.1. Anliegen der Studie und Untersuchungsfeld

In der von Wegwarth (2002) vorgelegten Diplomarbeit wurde eine Weiterbildungsveranstaltung einer Sparkassenakademie evaluiert. Dabei lag ein Fokus auf der Effizienz der Weiterbildung, welche Wegwarth (2002) anhand des Wissenserwerbs und dessen Nachhaltigkeit definierte. Die Untersuchung sollte eine Bestandsaufnahme des Moduls „Management und Führung“ sein und ggf. als Grundlage für Optimierungsbemühungen dienen. Zentrales Anliegen war also eine Wissenserfassung mittels der Technik des Wortassoziationsversuchs und eine Darstellung des Wissenserwerbsprozesses und -transfers sowie der Nachweis von Nachhaltigkeit des erworbenen Wissens.

Die Sparkassenakademie ist eine Aus- und Weiterbildungseinrichtung, welche den Teilnehmern die Möglichkeit bietet, verschiedene Abschlüsse des Bankgewerbes zu erwerben (z.B. Sparkassenkaufmann, Sparkassenbetriebswirt). Diese modular angebotenen Aus- und Weiterbildungen sind konzipiert als Mischung aus Präsenzlehre sowie Selbstlernphase. Dabei zeichnet sich die fünf Wochen lange Präsenzphase i.d.R. durch Frontalunterricht aus, während der sechswöchigen Selbstlernphase haben die Teilnehmer eine Hausarbeit zu verfassen und sich außerdem auf die schriftliche Abschlußprüfung vorzubereiten, nach weiteren vier Wochen folgt eine mündliche Abschlußprüfung. Das von Wegwarth (2002) untersuchte Modul „Management und Führung“ wird obligatorisch in der Ausbildung von Nachwuchsführungskräften der Sparkassen eingesetzt. Für differenziertere Ausführungen über das Untersuchungsfeld sei auf Wegwarth (2002) verwiesen. Von den insgesamt 50 Seminarteilnehmern konnten 38 in die Erhebungen einbezogen werden, da diese Probanden an allen Meßzeitpunkten teilnahmen.

4.3.2. Untersuchungsaufbau

Wegwarth (2002) wählte ein Untersuchungsdesign mit vier Meßzeitpunkten (siehe Abbildung 10), welche sich an den für den Wissenserwerb besonders bedeutsamen Phasen Seminarbeginn (T1), Seminarende (T2) sowie schriftliche Prüfung (T3) orientieren. Zur Prüfung der Nachhaltigkeit des Wissens wurde ein vierter Meßzeitpunkt vier Monate nach der Prüfung gewählt (T4). Auf den von Wegwarth (2002) eingesetzten Fragebogen zum Lerntransfer soll hier nicht weiter eingegangen werden. In dieser Felduntersuchung wird die Leistung der Probanden in der Abschlußklausur (erreichte Punkte) als Außenkriterium herangezogen.

	T1	T2	T3	T4
	erster Tag des Seminars	letzter Tag der Präsenzphase (nach 5 Wochen)	zur schriftlichen Prüfung (nach 6 Wochen Selbststudium)	4 Monate nach der schriftlichen Prüfung
1.	Wortassoziationsversuch	Wortassoziationsversuch	schriftliche Prüfung (Klausur)	Wortassoziationsversuch
2.	Fragebogen zum Lerntransfer	Fragebogen zum Lerntransfer	Wortassoziationsversuch	
3.			Fragebogen zum Lerntransfer	

Abbildung 10: Schematischer Ablauf in Studie 3 – Sparkassenakademie (A) (vgl. Wegwarth, 2002)

4.3.3. Wortassoziationsversuch

Der von Wegwarth (2002) durchgeführte Wortassoziationsversuch erfolgte schriftlich (Papier-Bleistift) und mit einer Zeitbegrenzung von zwei Minuten pro Stimulus. Dabei wurden den Probanden im Gruppenversuch ein Blatt pro Stimulus vorgelegt (vorgegebene Linien, Stimulus auf dem Blattkopf), nach jeweils zwei Minuten forderte Wegwarth (2002) die Probanden auf, zum nächsten Stimulus überzugehen. Auf diese Weise wurden vier Stimuli präsentiert, die sich mit den während des Moduls behandelten Themen decken (siehe Abbildung 11).

Nr.	Stimulus	Nr.	Stimulus
1)	Strategische Unternehmensführung	3)	Personalentwicklung
2)	Führen von Gruppen	4)	Motivationstheorien

Abbildung 11: Stimuli der Studie 3 – Sparkassenakademie (A) (vgl. Wegwarth, 2002)

Die Assoziationskennwerte als Mittelwert über die Stichprobe sollen separat für die verschiedenen Meßzeitpunkte betrachtet werden (siehe Abbildung 12). Es wird deutlich, daß vor der Weiterbildung durchschnittlich $M=10,46$ Assoziationen ($SD=3,73$) produziert wurden, nach Ende der Präsenzlernphase $M=10,18$ ($SD=3,61$), zur Prüfung $M=14,10$ ($SD=4,05$) und in der Follow-up-Erhebung $M=10,54$ ($SD=4,18$) Assoziationen. Die Assoziationskennwerte sind zu allen vier Meßzeitpunkten normalverteilt, auf die Entwicklung des Assoziationskennwertes über die Meßzeitpunkte hinweg wird in Kapitel 7.1.1 (Seite 55 ff.) näher eingegangen werden.

	T1	T2	T3	T4
<i>M</i>	10,46	10,18	14,10	10,54
<i>SD</i>	3,73	3,61	4,05	4,18
Minimum	5,0	4,3	6,3	3,0
Maximum	22,3	18,0	25,0	20,0

Abbildung 12: Kennwerte des Wortassoziationsversuchs– Mittelwerte (M) und Standardabweichungen (SD) über die gesamte Stichprobe in Studie 3 – Sparkassenakademie (A)

	gute Assoziierer (N=8)	schlechte Assoziierer (N=8)
<i>M</i>	19,83	8,92
<i>SD</i>	2,48	1,51
Minimum	17,67	6,33
Maximum	25,00	10,67

Abbildung 13: Mittelwerte (M) und Standardabweichungen (SD) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert in T3) in Studie 3 – Sparkassenakademie (A)

In dieser Studie ließen sich Extremgruppen hinsichtlich des Assoziationskennwertes bestimmen, zugunsten der Übersichtlichkeit soll hier nur der dritte Meßzeitpunkt berichtet werden (da zu diesem Zeitpunkt auch das Außenkriterium erhoben wurde). Abbildung 13 listet die statistischen Kennwerte des Assoziationskennwertes in den beiden Extremgruppen (zu T3) auf, welche jeweils acht Probanden umfassen. Danach liegt die durchschnittliche Assoziationsleistung der Gruppe mit hohem Assoziationskennwert mit $M=19,83$ ($SD=2,48$) mehr als doppelt so hoch wie die der Gruppe

mit geringem Assoziationskennwert ($M=8,92$; $SD=1,51$). Diese beiden Extremgruppen lassen sich signifikant voneinander abgrenzen (Mann-Whitney-U: $z=-3,371$; $p= .000$).

4.4. Studie 4: Sparkassenakademie (B)

4.4.1. Anliegen der Studie und Untersuchungsfeld

Jordan (2004) ging mit seiner Diplomarbeit der Fragestellung nach, inwieweit die Technik des Strukturlegens lernunterstützend sein kann. Dies überprüfte er in einem Feldexperiment, bei dem Teilnehmern einer Weiterbildungsveranstaltung der Sparkassenakademie diese Technik vermittelt wurde. Des weiteren wollte Jordan (2004) wissen, ob die Weiterbildungsteilnehmer überhaupt Wissen während der Veranstaltung erwarben und sich dies auch in den Prüfungsleistungen widerspiegelt.

Die von Jordan (2004) durchgeführte Untersuchung fand im selben organisationalen Kontext statt wie die von Wegwarth (2002) durchgeführte Studie 3 (Sparkassenakademie A, siehe Kap. 4.3.). Auch hier wurden innerhalb der Weiterbildungsveranstaltung „Management“ an der Sparkassenakademie Probanden rekrutiert, Aufbau und Inhalt dieses Moduls sind analog zu den Ausführungen bei Wegwarth (2002). Von den insgesamt 42 Teilnehmern an der Weiterbildungsveranstaltung konnten 35 in die Auswertungen zum Wortassoziationsversuch einbezogen werden.

4.4.2. Untersuchungsaufbau

Ähnlich wie in der Erhebung von Wegwarth (2002) orientierte sich auch Jordan (2004) an den entscheidenden Phasen der Weiterbildung und realisierte drei Meßzeitpunkte: vor Beginn der Weiterbildung (T1), am Ende der Präsenzphase (T2) und schließlich am Ende der Selbstlernphase, zeitgleich mit der schriftlichen Prüfung (T3). Der Ablauf der Wissensdiagnosen ist in Abbildung 14 schematisch dargestellt.

T1	T2	T3
<i>erster Tag des Seminars</i>	<i>letzter Tag der Präsenzphase (nach 5 Wochen)</i>	<i>zur schriftlichen Prüfung (nach 7 Wochen Selbststudium)</i>
Wortassoziationsversuch	Wortassoziationsversuch	1. schriftliche Prüfung (Klausur) 2. Wortassoziationsversuch 3. Strukturlegetechnik

Abbildung 14: Schematischer Ablauf in Studie 4 – Sparkassenakademie (B) (vgl. Jordan, 2004)

Da Jordan (2004) die Erhebung als Feldexperiment anlegte, wurden die Probanden in zwei Gruppen eingeteilt: der Experimentalgruppe wurde nach dem zweiten Meßzeitpunkt, also zu Beginn der Selbstlernphase, die Strukturlegetechnik vermittelt. Die Probanden wurden dazu angeregt, diese Technik während ihres Selbststudiums einzusetzen. Der Kontrollgruppe hingegen wurde diese Technik nicht nahegebracht, statt dessen nahmen diese Probanden zu Beginn der Selbstlernphase an einer Gruppendiskussion teil. Am letzten Meßzeitpunkt (T3) wurde die Prüfungsklausur geschrieben, anschließend bearbeiteten alle Untersuchungsteilnehmer den Wortassoziationsversuch

sowie das Strukturlegen. Die Strukturlegetechnik hat in dieser Studie also eine Doppelfunktion: als Treatment (für die Experimentalgruppe) und als Diagnostikum (für alle Probanden). Aufgrund dessen und aufgrund der ungünstigen Versuchsbedingung beim Strukturlegen als Diagnoseinstrument (die Probanden sollten ein Wissensnetz selbständig zu Hause legen, der Rücklauf in diesem Untersuchungsteil war entsprechend gering) werden diese Daten in der vorliegenden Arbeit unberücksichtigt bleiben.

Auch in dieser Studie wurde die Leistung in der Klausur als Kriterium für den Lernerfolg herangezogen. Außerdem liegen die Bewertungen (Punkte) der Hausarbeiten vor, welche von den Weiterbildungsteilnehmern im Rahmen der Selbstlernphase zu erstellen waren.

4.4.3. Wortassoziationsversuch

Jordan (2004) setze den Wortassoziationsversuch schriftlich in der Papier-Bleistift-Form ein. Dabei gab er für jeden Stimulus ein separates Blatt aus und diktierte die Stimuli sukzessive, wobei er pro Stimulus zwei Minuten Zeit zur Bearbeitung gab. Auf diese Weise konnte der Wortassoziationsversuch als Gruppenerhebung gut in den Seminarbetrieb integriert werden (siehe Jordan, 2004). Der Wortassoziationsversuch umfaßt sechs Stimuli mit zentralen Begriffen aus der Domäne „Führung“, welche in Abbildung 15 dargestellt sind.

Nr.	Stimulus	Nr.	Stimulus
1)	Banksteuerung	4)	Risikomanagement
2)	Strategische Unternehmensführung	5)	Betriebswirtschaftslehre von Unternehmen
3)	Marketing und Vertriebsmanagement	6)	Kosten- und Leistungsrechnung

Abbildung 15: Stimuli der Studie 4 – Sparkassenakademie (B) (vgl. Jordan, 2004)

Die Assoziationskennwerte sind zu allen drei Meßzeitpunkten normalverteilt, die statistischen Kennwerte sind in Abbildung 16 dargestellt. Zu Beginn der Weiterbildung werden durchschnittlich $M=11,29$ ($SD=3,14$) Konzepte assoziiert, am Ende der Präsenzphase $M=13,59$ ($SD=3,03$), und am Ende der Selsbtlernphase beträgt der durchschnittliche Assoziationskennwert $M=13,39$ ($SD=3,94$). Für eine differenziertere Betrachtung der Assoziationsleistungen über die Meßzeitpunkte hinweg sei auf Kapitel 7.1.2 (Seite 55 ff.) verwiesen.

	T1	T2	T3
<i>M</i>	11,29	13,59	13,39
<i>SD</i>	3,14	3,03	3,94
Minimum	5,00	7,50	2,33
Maximum	18,50	21,33	20,33

Abbildung 16: Kennwerte des Wortassoziationsversuchs– Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 4 – Sparkassenakademie (B)

Analog zum Vorgehen in Studie 3 (Kapitel 4.3) sollen die Extremgruppen hinsichtlich ihrer Assoziationsleistungen nur für den dritten Meßzeitpunkt bestimmt werden. Es konnten zwei Extremgruppen mit jeweils sechs Probanden identifiziert werden, ihre statistischen Kennwerte sind in Abbildung 17 dargestellt. Danach assoziierte die Gruppe der guten Assoziierer mit $M=18,75$ Nennungen ($SD=1,50$) mehr als doppelt soviel wie die Gruppe der schlechten Assoziierer ($M=7,56$;

$SD=3,37$). Diese Abgrenzung zwischen den Gruppenkennwerten ist signifikant (Mann-Whitney-U: $z=-2,892$; $p= .002$).

	gute Assoziierer (N=6)	schlechte Assoziierer (N=6)
<i>M</i>	18,75	7,56
<i>SD</i>	1,50	3,37
<i>Minimum</i>	17,00	2,33
<i>Maximum</i>	20,33	10,17

Abbildung 17: Mittelwerte (M) und Standardabweichungen (SD) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert in T3) in Studie 4 – Sparkassenakademie (B)

4.5. Studie 5: Biopsychologie

4.5.1. Anliegen der Studie und Untersuchungsfeld

In einer Kooperation der Humboldt Universität zu Berlin und der Universität Potsdam widmete sich diese Studie einem Vergleich der Verfahren Wortassoziationsversuch mit anschließender Strukturlegetechnik (nach Rothe, 2003) und Assoziations-Struktur-Test (nach Meyer, 2007). Zentrale Fragestellung in dieser Erhebung war, a) inwieweit die beiden, auf unterschiedlichen Wegen erzeugten Netze miteinander vergleichbar sind, und b) inwieweit die Kenngrößen dieser Netze die Vorhersage eines Leistungskriteriums ermöglichen (Klausurleistung). Für differenzierte Erörterungen dieser Fragestellungen sei auf Kermas (2006) verwiesen sowie auf Ceglarek, Meyer, Kermas und Rothe (2006). Diese Studie ist als Feldexperiment einzuordnen: einerseits bewegt sich die Untersuchung im Feld der Biopsychologie-Lehre und greift auf eine reale Prüfungsklausur zurück, andererseits werden Versuchsbedingungen systematisch variiert.

Da als Außenkriterium die Leistung in einer Klausur zur Biopsychologie herangezogen werden sollte, kamen als Probanden nur Studenten in Frage, welche an der aktuellen Prüfungsklausur teilnahmen. So wurde gegen Ende des Semesters an beiden Universitäten Studenten für diese Erhebung geworben, v.a. durch eine kurze Projektvorstellung während der Biopsychologie-Vorlesung sowie durch Aushänge und Mail-Verteiler. Die Studenten erhielten für ihre Teilnahme an der Untersuchung sechs Versuchspersonenstunden. Aufgrund der zeitlichen Nähe zur Klausur war die Resonanz gering, es konnten letztendlich nur neun Probanden, die an allen drei Meßzeitpunkten teilnahmen, in die Auswertungen einbezogen werden.

4.5.2. Untersuchungsaufbau

Als Wissensdomäne der Untersuchung wurde in Absprache mit den Lehrstuhlinhabern der Biopsychologie die Neuroanatomie gewählt. Sie war sowohl Bestandteil von Vorlesungen und Klausur, und ist überdies basal für ein grundlegendes Verständnis der biologischen Vorgänge und kann somit als repräsentativ für die Wissensdomäne Biopsychologie angesehen werden.

Die Erhebung bestand aus drei Erhebungsblöcken, deren Bestandteile in Abbildung 18 dargestellt sind. Allerdings ist anzumerken, daß es dabei nicht um eine Meßwiederholung handelte. Vielmehr wurde das umfangreiche Untersuchungsdesign auf mehrere Termine verteilt, zum einen, um die

Probanden (v.a. zeitlich) nicht zu überfordern, zum anderen waren teilweise Vorbereitungen durch die Versuchsleiter zu erbringen, welche auf die Ergebnisse von vorhergehenden Untersuchungsteilen angewiesen waren.

Block 1	Block 2	Block 3
<ul style="list-style-type: none"> • WIT: Untertests „Wortgewandtheit“; „Räumliche Vorstellung“ • Assoziation am PC (im Rahmen des Assoziations-Struktur-Test) 	<ul style="list-style-type: none"> • Paarvergleich am PC (mit den assoziierten Konzepten aus Block 1, im Rahmen des Assoziations-Struktur-Test) 	<ul style="list-style-type: none"> • Strukturlegetechnik (mit den assoziierten Konzepten aus Block 1)

Abbildung 18: Schematischer Ablauf in Studie 5 – Biopsychologie

Im ersten Block bearbeiten die Probanden Untertests aus dem WILDE-Intelligenztest (Jäger & Althoff, 1994), welche Wortgewandtheit und räumliches Vorstellungsvermögen erfassen (auf Letztgenanntes wird in der vorliegenden Arbeit nicht weiter eingegangen). Anschließend assoziierten die Probanden am Computer zu zehn Stimuli aus der Wissensdomäne „Neuroanatomie“. Im zweiten Untersuchungsblock wurden diejenigen Assoziationen, die vom jeweiligen Probanden beim ersten Termin zum Stimulus „Nervensystem“ abgegeben wurden, paarweise vorgelegt und durch den Probanden hinsichtlich ihres inhaltlichen Zusammenhangs bewertet. Dieser Paarvergleich wurde computerunterstützt erhoben und repräsentiert den Assoziations-Struktur-Test von Meyer (2007), nähere Ausführungen zu diesem Verfahren finden sich in Kapitel 6.1.3.1.4 (Seite 55 ff.). Im letzten Block der Erhebung wurden den Probanden diejenigen Assoziationen, die sie beim ersten Termin zum Stimulus „Nervensystem“ produziert hatten, nochmals vorgelegt. Die Aufgabe bestand nun darin, als Paper-Pencil-Verfahren ein Wissensnetz zu erstellen, welches das individuelle Wissen der Probanden zum Thema „Nervensystem“ repräsentiert (Strukturlegetechnik nach Rothe, 2003). Die Klausur im Fach Biopsychologie fand i.d.R. zwischen dem zweiten und dritten Erhebungsblock statt, insgesamt wurde bei allen Terminen die größtmögliche zeitliche Nähe zur Klausur gesucht, um einen annähernd gleichen individuellen Wissensbesitz über die Messungen zu gewährleisten. Um die Güte der so erzeugten Wissensnetze beurteilen zu können, wurde als Außenkriterium die Leistung bei der Klausur im Fach Biopsychologie herangezogen.

Strukturlegetechnik: Derselbe Pool von Assoziationen, welche die Probanden in Block 1 zum Stimulus „Nervensystem“ produzierten, wurde genutzt, um sie selbst ihr Wissen mittels der Strukturlegetechnik visualisieren zu lassen. Das so gewonnene Netz wurde – im Gegensatz zu dem beim Assoziations-Struktur-Test erstellten Netz – explizit durch die Versuchsperson gelegt. Die Regeln für das Legen der Strukturen wurden in dieser Studie intraindividuell variiert, so daß je Proband mehrere Netze erzeugt wurden; auf Einzelheiten soll an dieser Stelle nicht näher eingegangen werden. Um die Vergleichbarkeit mit den anderen hier vorgestellten Studien, welche die Strukturlegetechnik nutzen, zu ermöglichen, sei hier ausschließlich auf diejenigen Daten zurückgegriffen, welche unter der Bedingung entstanden, daß alle Kanten des Netzwerkes benannt werden sollten anhand der Relationen von Klix (1992).

4.5.3. Wortassoziationsversuch

Die Funktion des Wortassoziationsversuchs in dieser Erhebung ist in erster Linie als „Datenlieferant“ für die beiden Struktur-Verfahren zu sehen, er kann aber auch isoliert ausgewertet

werden. Den Probanden wurden insgesamt zehn Stimuli aus dem Bereich Neuroanatomie vorgelegt, sie sind in der Reihenfolge ihrer Präsentation in Abbildung 19 aufgeführt. Diese Stimuli wurden anhand einer Dokumentenanalyse (Vorlesungsskript und das empfohlene Standardlehrbuch der Biopsychologie) gewonnen. Die jeweiligen Assoziationen zum Stimulus Nervensystem sind Grundlage der nachfolgenden Untersuchungsteile, daher kommt ihm eine zentrale Bedeutung zu. Dieser Stimulus wurde gewählt, da er hinreichend komplex und allgemein ist, so daß die daraus zu erstellenden Wissensnetze nicht schon allein dadurch begrenzt werden, daß der Bedeutungshof zu eng umschrieben ist. Die Darbietung der Stimuli und Erfassung der entsprechenden Assoziationen darauf erfolgte computerunterstützt, hierzu wurde die Plattform des Assoziations-Struktur-Tests (Meyer, 2007) genutzt.

Nr.	Stimulus	Nr.	Stimulus	Nr.	Stimulus
1)	limbisches System	4)	Gehirn	7)	Ganglien
2)	Nervensystem	5)	Liquor	8)	graue Substanz
3)	Cortex	6)	Rückenmark	9)	Aktionspotential
				10)	bildgebende Verfahren

Abbildung 19: Stimuli der Studie 5 – Biopsychologie

Die Assoziationsleistung in dieser Studie streute sehr stark innerhalb der kleinen Stichprobe: minimal wurden 4,0 und maximal 20,5 Assoziationen pro Stimulus geäußert. Über alle Probanden hinweg betrug der Assoziationskennwert $M=11,68$ ($SD=5,83$). Die beiden Extremgruppen mit jeweils drei Probanden unterstreichen diese Streubreite (siehe Abbildung 20), vermutlich aufgrund der sehr kleinen Substichproben jedoch wird dieser Unterschied nicht statistisch signifikant (Mann-Whitney-U: $z=-1,964$; $p>.10$).

	gesamte Stichprobe (N=9)	gute Assoziierer (N=3)	schlechte Assoziierer (N=3)
<i>M</i>	11,68	17,93	5,00
<i>SD</i>	5,83	2,60	1,65
Minimum	4,00	15,30	4,00
Maximum	20,50	20,50	6,90

Abbildung 20: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 5 – Biopsychologie

4.6. Studien 6 bis 9: Klausuren in verschiedenen Fächern der Psychologie

4.6.1. Anliegen der Studien und Untersuchungsfeld

Um seine Kennwerte an einem Außenkriterium validieren zu können, wurde der Wortassoziationsversuch im Kontext von Prüfungsklausuren eingesetzt. Dazu konnten die Dozenten von drei Vorlesungen aus dem Psychologie-Hauptstudium (Klinische Psychologie, Arbeitspsychologie, Personalpsychologie) sowie dem Psychologie-Grundstudium (Entwicklungspsychologie) an der Universität Potsdam gewonnen werden. Da den daraus entstandenen vier Validierungsstudien ein identisches Design zugrunde liegt, wird deren Darstellung hier in einem Kapitel zusammengefaßt. Ziel des im folgenden beschriebenen

Vorgehens war es, innerhalb eines Meßzeitpunktes sowohl Prädiktor (Assoziationskennwert) als auch Kriterium (Klausurleistung) zu erheben, so daß davon ausgegangen werden kann, daß die Probanden bei beiden Verfahren auf denselben Wissensbesitz zugriffen. Diese Untersuchungen sind als Feldstudien zu klassifizieren, da die Prüfungsklausuren eine reale, nicht experimentell erzeugte Leistungsanforderung mit den entsprechenden Konsequenzen darstellten.

Die Beteiligung an der Erhebung zum Wortassoziationsversuch war in den einzelnen Klausuren unterschiedlich stark ausgeprägt. In der Klausur zur Klinischen Psychologie war mit 70 % der Klausurteilnehmer die höchste Beteiligung am Wortassoziationsversuch zu verzeichnen: von 54 Klausurteilnehmern bearbeiteten 38 auch den Wortassoziationsversuch. In den anderen drei Klausuren schwankten die Beteiligungen zwischen 15 und 23 % (siehe dazu auch Abbildung 21). Dies könnte möglicherweise damit erklärt werden, daß die Klausur der Klinischen Psychologie nur das Eintragen von Lösungsworten in vorbereitete Antwortzettel erforderte, wodurch die Beanspruchung durch die manuelle Schreibearbeit relativ gering war. Im Gegensatz dazu hatten die Prüflinge in den anderen drei Klausuren längere Texte zu verfassen. Nach teilweise seitenlangem Schreiben unter Zeitdruck waren nur noch wenige dazu zu motivieren, anschließend im Wortassoziationsversuch weiter zu schreiben. Die Teilnahme an der Erhebung war freiwillig unter Zusicherung der Anonymität.

	Klinische Psychologie	Arbeitspsychologie	Personalpsychologie	Entwicklungspsychologie
Teilnehmer an der Klausur	n=54	n=56	n=52	n=95
Teilnehmer am Wortassoziationsversuch	n=38	n=13	n=8	n=17
Rücklaufquote	70%	23%	15%	18%

Abbildung 21: Beteiligung an den Erhebungen in den Studien 6 bis 9 (Klausuren in Fächern der Psychologie)

4.6.2. Untersuchungsaufbau

Im Rahmen der Studienordnung müssen Studierende des Diplomstudienganges Psychologie an der Universität Potsdam vorlesungsbegleitende Klausuren schreiben, welche den Charakter von Prüfungsleistungen tragen. Die Klausuren dauerten jeweils 1,5 Stunden und waren handschriftlich ohne jegliche Hilfsmittel zu erbringen. Durch Ankündigung unmittelbar vor der Klausur wurden die Studierenden über das Anliegen der Erhebung mittels Wortassoziationsversuch informiert und gebeten, sich an dieser Erhebung zu beteiligen. Außerdem wurde darauf hingewiesen, daß die Teilnahme an der Erhebung keinen Einfluß auf die Bewertung der Klausur hat. Die Zuordnung von Klausurleistungen und Assoziationsleistungen machte es allerdings erforderlich, die Probanden um ihre persönlichen Daten (Name, Matrikelnummer) zu bitten, welche auf einem separatem, an den eigentlichen Assoziationsbogen angehefteten Blatt erfaßt wurden. Es wurde zugesichert, daß dieses Blatt und damit auch die Angaben zur Person nach erfolgter Zuordnung der Leistungsdaten vom Assoziationsbogen entfernt und vernichtet werden.

Da es sich um Klausuren mit Prüfungscharakter handelte, hatten diese Vorrang vor dem Forschungsinteresse und wurden als erstes bearbeitet, der Wortassoziationsversuch wurde unmittelbar im Anschluß vorgelegt. Bei Abgabe der Klausurbögen wurden die Studierenden nochmals um die Teilnahme am Wortassoziationsversuch gebeten und erhielten den Arbeitsbogen.

Spätestens bei Rückgabe des Assoziationsbogens konnten die Probanden als Dank für ihre Teilnahme Schokoladentäfelchen erhalten.

Die Klausurnote bzw. die erreichte Punktzahl dient in dieser Studie als Außenkriterium. Zunächst sollen die Klausuren der einzelnen Fächer kurz kommentiert werden, anschließend wird das Vorgehen beim Wortassoziationsversuch ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Klausur im Fach Klinische Psychologie: In der Klinischen Psychologie wurde für zwei parallel stattfindende Vorlesungsteile jeweils eine Klausur geschrieben, die von allen Studierenden bearbeitet wurden. Pro Klausur lagen jeweils zwei Parallelformen A und B vor, welche jeweils 30 Fragen mit einem maximalen Punktwert von 32 enthielten. Die Aufgaben waren gestaltet als offene Fragen, das Wissen zu z.B. Prävalenzraten, Diagnosen, Kriterien für Störungen, Definitionen, Nomenklatur sollte als kurze Antwort in Stichwortform wiedergegeben werden.

Klausur im Fach Arbeitspsychologie: Die Klausur der Arbeitspsychologie umfaßte sieben Fragen mit maximal 50 erreichbaren Punkten. Die Fragen waren formuliert als Schlagworte, zu denen die Studierenden Ausführungen in Stichwortform oder kurzen Texten machen sollten, auch das Aufzeichnen von Modellen war möglich. Hauptsächlich wurde das Nennen, Aufzählen und Erläutern von Vorlesungsinhalten abgefordert, in einem kleinen Fallbeispiel wurde schließlich das Verständnis dieser Inhalte abgeprüft.

Klausur im Fach Personalpsychologie: Ähnlich zur Arbeitspsychologie war die Klausur zur Personalpsychologie aufgebaut: in sieben Fragen wurde das Nennen, Beschreiben und Erläutern zu bestimmten Vorlesungsinhalten und –themen gefordert. Das Verständnis des Gelernten wurde in einigen Fragen geprüft, welche z.B. vergleichende oder wertende Schlußfolgerungen abverlangten. Es konnten maximal 44 Punkte erreicht werden.

Klausur im Fach Entwicklungspsychologie: In der Klausur zur Entwicklungspsychologie wurden zwei Komplexe thematisiert: die frühe Kindheit sowie das (Vor-) Schulalter. Dazu wurde jeweils eine Aufgabe mit Begriffserläuterungen formuliert sowie vier weitere Fragen zum inhaltlichen Verständnis der Lehrinhalte. Zu jedem dieser Komplexe waren maximal 32 Punkte zu erreichen. In der vorliegenden Arbeit werden beide Komplexe zusammengefaßt, so daß sich eine maximal erreichbare Punktzahl von 64 ergibt.

4.6.3. Wortassoziationsversuch

Die Stimuli wurden auf Grundlage der Vorlesungsmaterialien gewonnen, welche allen Studierenden über die Lernplattform in digitaler Form zugänglich waren. Dabei handelte es sich im Wesentlichen um die in der Vorlesung gezeigten Folien in Form von Power-Point-Präsentationen. Mit Blick auf die Repräsentativität der Stimuli für die Klausur wurde nach Durchsicht dieser Materialien eine Auswahl mehrerer möglicher Stimuli getroffen, und in Absprache mit dem jeweiligen Dozenten die endgültigen Stimuli festgelegt. In Abbildung 22 finden sich die letztlich ausgewählten jeweils vier Stimuli für jedes Klausurfach. Um den Einfluß der spezifischen Inhalte der Stimuli auf die Reihenfolge zu kontrollieren, wurden vier verschiedene Formen entwickelt, bei denen die Präsentationsreihenfolge der Stimuli variiert. Abbildung 23 stellt die Systematik dar, nach der bei allen vier Klausurfächern vorgegangen wurde. Zu jeder Klausur wurden also jeweils vier

verschiedene Fragebogen entwickelt, die sich lediglich in der Reihenfolge der Stimuli unterscheiden. Die Zuweisung der verschiedenen Formen zu den Probanden erfolgte zufällig.

Klinische Psychologie	Arbeitspsychologie	Personalpsychologie	Entwicklungspsychologie
<ul style="list-style-type: none"> • Hyperkinetische Störung • Klassifikation in der Klinischen Psychologie • Suizidalität • Verhaltenstherapie bei spezifischen Störungsbildern 	<ul style="list-style-type: none"> • Erwerbslosigkeit • Veränderungen in der Arbeitswelt • Handlungsregulation • gesundheitsrelevante Wirkung von Arbeit 	<ul style="list-style-type: none"> • Human Ressource Management • Führung • Berufseignungsdiagnostik • Kompetenz 	<ul style="list-style-type: none"> • Piaget • Bindungsqualität • Längsschnittdesign • Symbolspiel

Abbildung 22: Stimuli nach Fächern in den Studien 6 bis 9 – Klausuren im Fach Psychologie

Form A	Form B	Form C	Form D
1. Stimulus a	1. Stimulus b	1. Stimulus c	1. Stimulus d
2. Stimulus b	2. Stimulus a	2. Stimulus d	2. Stimulus c
3. Stimulus c	3. Stimulus d	3. Stimulus a	3. Stimulus b
4. Stimulus d	4. Stimulus c	4. Stimulus b	4. Stimulus a

Abbildung 23: Systematik der Variationen in der Stimulusreihenfolge in den Studien 6 bis 9

Die Darbietung des Wortassoziationsversuchs erfolgte schriftlich in der Papier-und-Bleistift-Version. Die vier Stimuli wurden jeweils am Kopf einer separaten Seite mit darunterliegenden Hilfslinien dargestellt, so daß sich daraus ein Faltblatt im Format A5 ergab (siehe Anhang 4). Daran angeheftet war ein weiteres Blatt (im Format A5), auf dem das Anliegen der Studie und Untersuchungsfeld kurz geschildert wurde, um die Angabe der persönlichen Daten gebeten sowie die Instruktion zum Wortassoziationsversuch (Anhang 4) vermerkt wurde. Nach erfolgter Zuordnung von Klausur- und Assoziationsergebnissen konnte dieses Blatt und damit auch die persönlichen Angaben leicht wieder entfernt werden, so daß die Anonymität der Probanden im weiteren gewahrt wurde.

	Klinische Psychologie	Arbeitspsychologie	Personalpsychologie	Entwicklungspsychologie
<i>M</i>	10,71	10,21	9,50	5,94
<i>SD</i>	4,69	5,69	6,18	2,25
Minimum	2,25	1,50	4,75	1,75
Maximum	20,25	22,25	24,00	9,50

Abbildung 24: Statistische Kennwerte (Mittelwert, Standardabweichung, Minimal- und Maximalwerte) der Wortassoziationsversuche in den Studien 6 bis 9

Abbildung 24 stellt die statistischen Kennwerte der vier Wortassoziationsversuche als mittlere Anzahl von Assoziationen auf einen Stimulus dar. So werden bei der Klinischen Psychologie im Mittel $M=10,71$ ($SD=4,69$) Assoziationen produziert, in der Arbeitspsychologie $M=10,21$ ($SD=5,69$) und in der Personalpsychologie $M=9,50$ ($SD=6,18$). Es werden nicht nur Unterschiede zwischen den verschiedenen einzelnen Fächern sichtbar, sondern v.a. auch deutliche interindividuelle Streuungen in den einzelnen Fächern. Hierbei fällt die Klausur der Entwicklungspsychologie nicht nur durch eine im Vergleich zu den anderen Fächern deutlich

geringe mittlere Assoziationsleistung auf ($M=5,94$; $SD=2,25$), sie weist auch eine sehr geringe Streuung der Assoziationskennwerte innerhalb der Stichprobe auf.

	Klinische Psychologie		Arbeitspsychologie	
	gute Assoziierer (N=6)	schlechte Assoziierer (N=6)	gute Assoziierer (N=5)	schlechte Assoziierer (N=5)
<i>M</i>	18,67	4,92	15,85	4,75
<i>SD</i>	1,19	1,75	3,64	1,97
Minimum	17,25	2,25	13,50	1,50
Maximum	20,25	6,25	22,25	6,50

Abbildung 25: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert) in Studie 6 (Klausur Klinische Psychologie) und Studie 7 (Klausur Arbeitspsychologie)

	Personalpsychologie		Entwicklungspsychologie	
	gute Assoziierer (N=3)	schlechte Assoziierer (N=3)	gute Assoziierer (N=4)	schlechte Assoziierer (N=4)
<i>M</i>	14,50	5,33	8,94	3,13
<i>SD</i>	8,23	0,52	0,66	0,97
Minimum	9,50	4,75	8,25	1,75
Maximum	24,00	5,75	9,50	4,00

Abbildung 26: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert) in Studie 8 (Klausur Personalpsychologie) und Studie 9 (Klausur Entwicklungspsychologie)

In den einzelnen Studien läßt sich gut zwischen den Extremgruppen bzgl. der Assoziationsleistungen differenzieren (siehe Abbildung 25 bzw. Abbildung 26). Die Extremgruppen in der Studie zur Klinischen Psychologie ließen sich signifikant voneinander abgrenzen (Mann-Whitney-U: $z=-2,903$; $p=.002$), ebenso die Extremgruppen der Studie Arbeitspsychologie (Mann-Whitney-U: $z=-2,611$; $p=.008$). Trotz der vergleichsweise geringen Streuung des Assoziationskennwertes in der Studie zur Entwicklungspsychologie lassen sich Extremgruppen deutlich voneinander abgrenzen (Mann-Whitney-U: $z=-2,323$; $p=.029$): die Gruppe der fünf Probanden mit sehr hohem Assoziationskennwert assoziierte durchschnittlich $M=8,94$ ($SD=0,66$) Assoziationen, die Gruppe der fünf Probanden mit sehr niedrigem Assoziationskennwert $M=3,13$ ($SD=0,97$). Lediglich die augenscheinlich deutlichen Unterschiede zwischen den beiden Extremgruppen in der Studie Personalpsychologie lassen sich nicht durch die Signifikanzprüfung bestätigen (Mann-Whitney-U: $z=-1,964$; $p>.10$), was allerdings auch auf den sehr geringen Umfang dieser Substichprobe zurückzuführen sein dürfte.

4.7. Studie 10: Gebäudereiniger

4.7.1. Anliegen der Studie und Untersuchungsfeld

Zur Überprüfung der Verfahrensgüte des Assoziations-Struktur-Tests konzipierte Meyer (2007) eine Feldstudie, um die Ergebnisse seines Verfahrens zur Erfassung strukturellen Wissens mit

einem Außenkriterium in Beziehung zu setzen. Dazu konnte er als Probanden die Schüler einer Berufsschule im Ausbildungsgang Gebäudereinigung gewinnen. Als Außenkriterium diente eine Klausur, welche im Rahmen der Ausbildung zu absolvieren war.

Insgesamt nahmen 33 künftige Gebäudereiniger an der Erhebung teil, wobei bei fünf von ihnen keine Daten hinsichtlich der Klausurleistung vorlagen, so daß in die Validitätsberechnungen anhand der Klausurnote nur 28 Probanden dieser Stichprobe eingehen konnten, während bei allen Probanden die Assoziationskennwerte zu den Kenngrößen des Assoziations-Struktur-Tests in Beziehung gesetzt werden können. Die Stichprobe setzte sich dabei mehrheitlich aus jungen Männern im Alter von etwa 20 Jahren zusammen, über ein Drittel von ihnen besuchte zuvor eine Sonder- oder Hauptschule.

4.7.2. *Untersuchungsaufbau*

In Vorbereitung der Erhebungen wurden von Meyer (2007) gemeinsam mit dem Berufsschullehrer als Experten die zentralen Konzepte für die Wissensdomäne Gebäudereinigung identifiziert. Unmittelbar im Anschluß an die Klausur bearbeiteten die Probanden das computerbasierte Verfahren Assoziations-Struktur-Test, abschließend wurde ein kurzer Fragebogen zu den soziodemographischen Angaben vorgelegt. Zunächst sollen die Untersuchungsbestandteile Assoziations-Struktur-Test sowie die Klausur kurz skizziert werden, anschließend wird das Vorgehen beim Assoziieren ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Klausur: Die Klausur setzte sich zusammen aus einem Multiple-Choice-Test, offenen Fragen, technischen Zeichnungen, Berechnungen sowie einem Essay, welche zu einem Gesamtpunktwert zusammengefaßt wurden. Die Daten zur Klausurleistung liegen sowohl als erreichte Punkte als auch als Gesamtnote vor.

Assoziations-Struktur-Test: Der Assoziations-Struktur-Test von Meyer (2007) erfolgte im unmittelbaren Anschluß an die Klausur. Nach dem Assoziieren wurden die assoziierten Konzepte zum Stimulus Detaschur den Probanden zum Paarvergleich vorgelegt. Auf Grundlage der so gewonnenen Daten von Assoziationen und Paarvergleich rekonstruierte Meyer (2007) einen individuellen Wissensgraphen und wertete diesen nach quantitativen Kenngrößen aus.

4.7.3. *Wortassoziationsversuch*

Der Wortassoziationsversuch war wie bereits erwähnt in den Assoziations-Struktur-Test eingebunden, so daß er also schriftlich am Computer und ohne Zeitbegrenzung durchgeführt wurde. In enger Absprache mit dem Berufsschullehrer wurde eine Liste von zehn Stimuli zusammengetragen, welche in Abbildung 27 dargestellt sind.

Die Berufsschüler der Gebäudereinigung assoziierten $M=5,40$ Nennungen pro Stimulus ($SD=3,24$), siehe auch Abbildung 28. Der Assoziationskennwert verteilt sich normal über die Stichprobe. Anhand des Assoziationskennwertes lassen sich zwei Extremgruppen (mit jeweils acht Probanden) beschreiben, die sich signifikant voneinander abgrenzen lassen (Mann-Whitney-U: $z=-3,366$; $p=.000$): die Gruppe der guten Assoziierer mit durchschnittlich $M=9,93$ ($SD=2,91$) Äußerungen und die Gruppe der schlechten Assoziierer mit $M=2,09$ ($SD=0,77$), deren Kennwerte auch in Abbildung 28 gegenübergestellt sind.

Nr.	Stimulus	Nr.	Stimulus	Nr.	Stimulus
1)	Floor	4)	Farbecht	7)	Flecken
2)	Detaschur	5)	Herstellungsart	8)	Kundenzufriedenheit
3)	Sisal	6)	Naßshampoonierung	9)	Flächenberechnung
				10)	Mischungsverhältnis

Abbildung 27: Stimuli der Studie 10 – Gebäudereiniger

	gesamte Stichprobe (N=33)	gute Assoziierer (N=8)	schlechte Assoziierer (N=8)
<i>M</i>	5,40	9,93	2,09
<i>SD</i>	3,24	2,91	0,77
Minimum	1,30	7,20	1,30
Maximum	14,10	14,10	3,40

Abbildung 28: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 10 – Gebäudereiniger

4.8. Studie 11: Bauklempner

4.8.1. Anliegen der Studie und Untersuchungsfeld

In Ergänzung zu Studie 10 – Gebäudereiniger – konnte Meyer (2007) zur Überprüfung der Verfahrensgüte des Assoziations-Struktur-Tests eine weitere Stichprobe im Setting Berufsschule gewinnen: für den Ausbildungsgang Bauklempnerei wurde ebenfalls das strukturelle Wissen erhoben und zur Klausurleistung in Beziehung gesetzt. In die Auswertungen dieser Studie gingen 14 Probanden ein. Die ausschließlich männlichen Probanden dieser Stichprobe waren im Mittel etwa 20 Jahre alt. Die Hälfte der Bauklempner-Berufsschüler besuchte zuvor die Realschule und in einem Fall sogar das Gymnasium, lediglich einer der Probanden war auf einer Hauptschule, alle anderen Probanden besuchten andere Schultypen.

4.8.2. Untersuchungsaufbau

In Vorbereitung der Erhebungen wurden von Meyer (2007) gemeinsam mit dem Berufsschullehrer als Experten die zentralen Konzepte für die jeweiligen Ausbildungsgänge identifiziert. Als Außenkriterium lag in dieser Studie lediglich der in der Klausur erreichte Punktwert vor, zu Aufbau bzw. Inhalt der Klausur ist nichts bekannt. Unmittelbar im Anschluß an die Klausur bearbeiteten die Probanden das computerbasierte Verfahren Assoziations-Struktur-Test, abschließend wurde ein kurzer Fragebogen zu den soziodemographischen Angaben vorgelegt. Zunächst soll der Untersuchungsbestandteil Assoziations-Struktur-Test kurz skizziert werden, anschließend wird das Vorgehen beim Wortassoziationsversuch ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Assoziations-Struktur-Test: Nach erfolgtem Wortassoziationsversuch zu zehn Stimuli wurden für den Paarvergleich die Assoziationen zum Stimulus Dach herangezogen. Auf Grundlage der so gewonnenen Daten von Assoziationen und Paarvergleich rekonstruierte Meyer (2007) einen individuellen Wissensgraphen und wertete diesen nach quantitativen Kenngrößen aus.

4.8.3. Wortassoziationsversuch

Der Wortassoziationsversuch als Teil des Assoziations-Struktur-Tests erfolgte schriftlich am Computer und ohne Zeitbegrenzung. Die in Abbildung 29 dargestellte Liste von zehn Stimuli wurde von Meyer (2007) in enger Absprache mit dem Berufsschullehrer entwickelt.

Nr.	Stimulus	Nr.	Stimulus	Nr.	Stimulus
1)	Dach	4)	Elektrolyt	7)	Winkelfunktionen
2)	Dachform	5)	Regen	8)	Temperaturunterschied
3)	Taupunkt	6)	Dachkonstruktion	9)	Anfallspunkt
				10)	Dach

Abbildung 29: Stimuli der Studie 11 – Bauklempner

In der Stichprobe der angehenden Bauklempner wurden durchschnittlich 8,91 Assoziationen produziert ($SD=3,25$), siehe auch Abbildung 30. Der Assoziationskennwert verteilt sich normal über die Stichprobe. Auf Grundlage des Assoziationskennwertes ließen sich zwei Extremgruppen mit jeweils drei Probanden identifizieren (Abbildung 30), der Unterschied zwischen beiden Gruppen hinsichtlich des Assoziationskennwertes ist zwar augenscheinlich, jedoch nicht statistisch signifikant (Mann-Whitney-U: $z=-1,964$; $p>.10$), was vermutlich auch auf den sehr geringen Stichprobenumfang zurückzuführen ist.

	gesamte Stichprobe (N=14)	gute Assoziierer (N=3)	schlechte Assoziierer (N=3)
<i>M</i>	8,91	13,97	5,17
<i>SD</i>	3,25	1,27	1,10
Minimum	3,90	12,50	3,90
Maximum	14,80	14,80	5,90

Abbildung 30: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 11 – Bauklempner

4.9. Studie 12: Handelsunternehmen

4.9.1. Anliegen der Studie und Untersuchungsfeld

Die von Hinnerichs und Rothe durchgeführte Untersuchung beschäftigt sich mit dem individuellen und organisationalen Wissensmanagement in einem mittelständischen Handels- und Dienstleistungsunternehmen (siehe auch Rothe und Hinnerichs, 2005). Ziel der Studie war das Gewinnen von Erkenntnissen über das Arbeitsverhalten und die Leistungen von Mitarbeitern, eine Ausweitung erfuhr diese Erhebung durch die Diplomarbeit von Liebs (2003). Diese beschäftigt sich mit dem Zusammenhang zwischen Wissen, Belastung bzw. Beanspruchung und der Arbeitsleistung. Ein Fokus lag auf der Beziehung zwischen Leistung und dem individuellen Wissensbesitz, der Wortassoziationsversuch diente dabei zur Ermittlung der individuellen Wissenskomponenten.

Die Felderhebung fand in einem mittelständischen Handels- und Dienstleistungsunternehmen statt, welches Handwerks- und Industriebedarf (z.B. Werkzeuge, Maschinen, Befestigungstechnik, Arbeitsschuttmittel) vertreibt und auch Wartungsarbeiten dazu tätigt. Die Beschäftigten lassen sich folgenden Arbeitsbereichen zuordnen: Außendienst (Kundenbetreuung vor Ort, d.h. Beratung,

Verkauf, z.T. Belieferung der Kunden; Akquise), Produktmanager (Kundenbetreuung am Telefon, d.h. Beratung und Verkauf; Erstellen von Angeboten, Einkauf der Artikel), Verkauf (Betreuung des Ladengeschäftes, d.h. Kundenberatung und Verkauf, Regalpflege) sowie Service (Wartung und Reparatur defekter Maschinen, Überprüfung von Sicherheitstechnik), darüber hinaus lassen sich noch Einzelarbeitsplätze für Lager und Verwaltung benennen. Für differenziertere Ausführungen über das Untersuchungsfeld sei auf Liebs (2003) oder Rothe und Hinnerichs (2005) verwiesen.

In die Untersuchungen von Rothe und Hinnerichs (2005) sowie Liebs (2003) wurden die Mitarbeiter aus den vier Arbeitsbereichen Produktmanagement, Außendienst, Verkauf sowie Service einbezogen. An den Tiefeninterviews nahmen 12 Mitarbeiter teil, der Wortassoziationsversuch sowie das Strukturlegen wurden mit insgesamt 17 Mitarbeitern durchgeführt. Die Erhebungen fanden während der Arbeitszeit statt. Den Mitarbeitern wurde vermittelt, daß die Geschäftsführung die Erhebung unterstützt bzw. gutheißt, außerdem wurde die Anonymität der Ergebnisse zugesichert.

4.9.2. Untersuchungsaufbau

Auf Grundlage von Experteninterviews (v.a. mit dem Geschäftsführer des Unternehmens), Dokumentenanalysen und Recherchen in Fachliteratur wurde zunächst exemplarisch eine Domäne herausgegriffen, welche Produktwissen repräsentiert, und anschließend dazu ein hypothetisches Wissensmodell erstellt, welches das Soll-Wissen der Beschäftigten beschreibt. Aus der Breite der Produktpalette des Unternehmens wurde eine Wissensdomäne gewählt, die für alle Mitarbeiter in allen Arbeitsbereichen gleichermaßen relevant ist und dabei hinreichend komplex. So bewegen sich alle wissensdiagnostischen Erhebungen im Themenbereich Schweißen, mit Fokus auf den beiden Schweißtechniken Elektrodenschweißen sowie MIG/MAG-Schweißen. Da für die Mitarbeiter im Service ein sehr spezielles Wissen über Schweißen relevant ist, welches von den anderen Arbeitsbereichen abweicht, wurden zwei Wissensmodelle entwickelt: zum einen für den Bereich Service, ein zweites Modell für die Bereiche Außendienst, Verkauf und Produktmanagement. Zunächst wurde aus diesen Modellen ein Interviewleitfaden entwickelt, welcher anschließend in Tiefeninterviews mit den Mitarbeitern eingesetzt wurde.

Der zweite Zugang zum individuellen Wissen bestand in der Durchführung des Wortassoziationsversuchs und anschließendem Strukturlegen, ebenfalls auf Basis des hypothetischen Wissensmodells. Die qualitativen Ergebnisse des Wortassoziationsversuchs dienten dabei zur weiteren Modifizierung des Wissensmodells (in der Stichprobe hochfrequente Assoziationsinhalte werden als besonders relevant angesehen). Das hypothetische Wissensmodell ist also letztendlich ein Ergebnis der Dokumentenanalysen, der Experteninterviews, der Tiefeninterviews sowie des Wortassoziationsversuchs. Die Wissensdomäne Schweißen war dabei der zentrale Gegenstand sämtlicher wissensdiagnostischer Instrumente. Um den individuellen Wissensbesitz in Beziehung zur Arbeitsleistung setzen zu können, wurde als Leistungskriterium das Urteil des Geschäftsführers hinsichtlich Arbeitsleistung seiner Mitarbeiter gewählt. Zunächst sollen die Untersuchungsbestandteile Tiefeninterview, Strukturlegetechnik sowie die Geschäftsführerbeurteilung kurz skizziert werden, anschließend wird das Vorgehen beim Wortassoziationsversuch ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Tiefeninterviews: Die in dieser Studie von Rothe und Hinnerichs (2005) eingesetzten Tiefeninterviews als eine Form der individuellen Wissensdiagnose erfaßten von insgesamt zwölf ausgewählten Vertretern der vier Arbeitsbereiche den individuellen Fachwissensbesitz zum Themenbereich Schweißen. Wie bereits erwähnt wurden die Ergebnisse der Interviews auch zur Modifizierung des hypothetischen Wissensmodells herangezogen. Inhaltlich orientierten sich die Interviewkomplexe an den beim Wortassoziationsversuch verwendeten Teildomänen (siehe Abbildung 31 auf Seite 55). Es wurde nach Schweißverfahren im Allgemeinen gefragt sowie nach Arbeitsschutz, Unfall- und Brandverhütung, außerdem wurden folgende Aspekte thematisiert: Bedeutung dieser Schweißart, schweißbare Materialien, Aufbau des Gerätes, Funktionsprinzip, Ablauf, Vor- und Nachteile. Diese Inhalte wurden in enger Zusammenarbeit mit dem Geschäftsführer entwickelt. Hinsichtlich der Auswertung der Tiefeninterviews sei auf Rothe und Hinnerichs (2005) verwiesen.

Strukturlegetechnik: Im Anschluß an den Wortassoziationsversuch wurde die Strukturlegetechnik eingesetzt. Dazu wurden zwei Themen für das Legen der Wissensnetze ausgewählt: a) das Schweißprinzip beim Elektroden- bzw. MIG/MAG-Schweißen, und b) der Aufbau von Elektroden- bzw. MIG/MAG-Schweißgeräten. Es waren also pro Proband zwei Netze zu legen, diese wurden als Papier-Bleistift-Version erstellt. Eine Zeitbegrenzung gab es nicht. Die Begriffe und Relationen für das Strukturlegen waren durch den Versuchsleiter vorgegeben und für alle Probanden gleich. Die Relationskarten beinhalteten 14 Relationsarten (vgl. Liebs, 2003). Die Begriffskarten leiteten sich aus dem hypothetischen Wissensmodell ab, in das wie bereits erwähnt auch zusätzliche, relevante Begriffe aus dem vorangegangenen Wortassoziationsversuch eingingen. Zusätzlich wurden den Probanden leere Kärtchen zur Verfügung gestellt, auf denen Begriffe (Knoten) oder auch Relationen ggf. ergänzt werden konnten. Die Probanden wurden dazu angehalten, aus dem umfangreichen Angebot der Knoten nur diejenigen auszuwählen, die auch tatsächlich ihrem Wissensbesitz entsprechen. Zur Ermittlung eines Korrespondenzkoeffizienten wurde von den Versuchsleiterinnen gemeinsam mit dem Geschäftsführer ein Referenznetzwerk erstellt, welches das Sollwissen repräsentiert.

Geschäftsführerurteil: Als Leistungsindikator für erfolgreiches berufliches Handeln wurde in dieser Untersuchung das Urteil des Geschäftsführers herangezogen. Aufgrund der geringen Mitarbeiterzahl kannte der Geschäftsführer jeden einzelnen Mitarbeiter hinreichend gut, um ein Urteil über dessen Arbeitsleistung abgeben zu können. Im Interview benannte der Geschäftsführer für ihn relevante Aspekte der Arbeitsleistung, beschrieb diese und deren Operationalisierung. So ergaben sich neun Leistungskriterien: Fachkompetenz, soziale Kompetenz, Teamfähigkeit Qualität der Arbeit, Quantität der Arbeit, Effizienz der Arbeit, Arbeitsengagement, Zuverlässigkeit sowie logisches Denkvermögen (für eine differenzierte Erörterung dieser Kategorien siehe Liebs, 2003). Das Rating erfolgte anhand einer fünfstufigen Skala (ähnlich den Schulnoten) von 1=„Anforderungen werden außergewöhnlich erfüllt“ bis 5=„Anforderungen werden kaum erfüllt“. Durch systematische Paarvergleiche durch den Geschäftsführer wurden Gewichte für die einzelnen Kriterien ermittelt, welche schließlich in die Gesamtbeurteilung einfließen (siehe Liebs, 2003). Für die Arbeitsbereiche Außendienst, Verkauf sowie Produktmanagement wurden identische Leistungskriterien angesetzt, für den Bereich Service wurden separate Gewichtungen erarbeitet. Die

Gesamtbeurteilung eines Mitarbeiters ist somit die Summe von Einzelleistungskriterien, welche ihrerseits ein Produkt des Ratings mit dem ermittelten Gewichtskoeffizienten darstellen.

4.9.3. Wortassoziationsversuch

Ziel des Wortassoziationsversuchs in dieser Erhebung war in erster Linie die Ermittlung des individuellen Wissensbesitzes, welcher zur Erledigung der Arbeitsaufgaben notwendig ist. Zum anderen dienten die beim Wortassoziationsversuch gewonnenen Wissenskomponenten zur Modifikation des hypothetischen Wissensmodells (a posteriori). Dabei wurden diejenigen Assoziationen, die von mindestens zwei Probanden benannt wurden und somit als relevant zu betrachten sind, in das Modell integriert (siehe Liebs, 2003; Rothe und Hinnerichs, 2005).

Die aus der Wissensdomäne Schweißen abgeleiteten 26 Stimuli des Wortassoziationsversuchs sind in Abbildung 31 dargestellt. Der Wortassoziationsversuch wurde mit 17 Mitarbeitern mündlich durchgeführt, die aufgezeichneten Äußerungen anschließend transkribiert. Für die Beantwortung eines jeden Stimulus wurde ein maximaler Zeitrahmen von 2 Minuten gewährt, nach Angaben von Liebs (2003) aber wurde diese zeitliche Obergrenze von keinem Probanden erreicht. Um einen Ermüdungseffekt im fortschreitenden Assoziationsverlauf und dessen (systematische) Auswirkungen auf die Assoziationsleistung zu vermeiden, balancierte Liebs (2003) die Abfolge der übergeordneten Themenblöcke systematisch über die Mitarbeiter, wobei eine gewisse logische Reihenfolge der Stimuli gewährleistet wurde (z.B. wurde Stimulus Nr. 1 immer an erster Stelle positioniert). Die Zuweisung von Probanden zu einer der Präsentationsvarianten erfolgte zufällig.

Nr.	Stimulus	Nr.	Stimulus	Nr.	Stimulus
1)	Schweißverfahren	10)	Metall-Schutzgas-Schweißen	21)	Arbeitsschutz
2)	Elektrodenschweißen	11)	MIG-Schweißen	22)	Brandschutz
3)	schweißbare Materialien Elektrodenschweißen	12)	MAG-Schweißen	23)	Reparatur
4)	Aufbau Elektrodenschweiß-Gerät	13)	schweißbare Materialien MIG	24)	Fehlersuche
5)	Elektrodenschweiß-Prinzip	14)	schweißbare Materialien MAG	25)	Sicherheitstechnische Überprüfung
6)	Vorbereitung Elektrodenschweißen	15)	Aufbau MIG/MAG-Gerät	26)	Meßgeräte zur Überprüfung
7)	Nachbereitung Elektrodenschweißen	16)	MIG/MAG-Schweiß-Prinzip		
8)	Vorteile Elektrodenschweißen	17)	Vorbereitung MMS		
9)	Nachteile Elektrodenschweißen	18)	Nachbereitung MMS		
		19)	Vorteile MMS		
		20)	Nachteile MMS		

Abbildung 31: Stimuli der Studie 12 - Handelsunternehmen (vgl. Liebs, 2003)

Hinsichtlich des Assoziationskennwertes berichtet Liebs (2003) über große interindividuelle Unterschiede. Abbildung 32 stellt die statistischen Kennwerte dar, danach assoziierten die Probanden im Mittel $M=4,54$ Konzepte mit einer Standardabweichung $SD=1,66$. Zwischen einer Assoziationsleistung von minimal 1,77 und maximal 8,42 Nennungen pro Stimulus beschreibt Liebs (2003) ein relativ homogenes Mittelfeld. Liebs (2003) bildete auf Basis der relevanten Assoziationen zwei Extremgruppen, welche aus jeweils vier Probanden bestand, im Folgenden die guten bzw. schlechten Assoziierer genannt (siehe auch Abbildung 32). Die guten Assoziierer weisen mit $M=7,02$ ($SD=0,97$) einen signifikant höheren Assoziationskennwert auf als die Gruppe

der schlechten Assoziierer mit $M=2,76$ ($SD=0,66$) Äußerungen (Mann-Whitney-U: $z=-2,323$; $p=.029$).

	gesamte Stichprobe (N=17)	gute Assoziierer (N=4)	schlechte Assoziierer (N=4)
<i>M</i>	4,54	7,02	2,76
<i>SD</i>	1,66	0,97	0,66
Minimum	1,77	6,19	1,77
Maximum	8,42	8,42	3,12

Abbildung 32: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 12 – Handelsunternehmen

4.10. Studie 13: Stoffdruckerei

4.10.1. Anliegen der Studie und Untersuchungsfeld

Die von Pokrandt (2004) vorgelegte Diplomarbeit thematisiert das kompetente Handeln im Beruf. In dieser Feldstudie in einem mittelständischen Produktionsbetrieb wurde untersucht, welche bedingungs- und v.a. personenbezogenen Faktoren möglicherweise eine Abweichung von der planmäßigen Auftragsabwicklung erklären können. Neben der Komponente des individuellen Wissensbesitzes wurden außerdem die Variablen Arbeitszufriedenheit sowie Commitment betrachtet. Ziel der Arbeit war, aufgrund der Analysen Aussagen über das Wissensmanagement im Unternehmen treffen zu können bzw. ggf. Modifikationen vorzuschlagen.

Die Untersuchung fand in einer mittelständischen Stoffdruckerei statt, welche im Siebdruck- und Digitaldruckverfahren Flaggen, Werbefahnen, Wimpel, Transparente u.ä. in verschiedenen Ausführungen herstellt. Zum Untersuchungszeitpunkt beschäftigte das Unternehmen etwa 80 Mitarbeiter (einschließlich Zeitarbeiter). Aufträge durchlaufen die einzelnen Abteilungen der Stoffdruckerei sukzessive, so daß die erfolgreiche Arbeit der einen Abteilung die Grundlage bildet für die erfolgreiche Weiterführung des Auftrages in der nachgelagerten Abteilung. Im folgenden sollen die einzelnen Abteilungen des Unternehmens in der Reihenfolge entlang der Prozeßkette kurz benannt werden: Verkauf (Beschaffung aller für die Auftragsausführung notwendigen Informationen), Auftragsbearbeitung (Planung der Produktionsschritte), Einkauf (Beschaffung der Bedarfsartikeln für Administration und Produktion), Filmfertigung (grafische Bearbeitung und Herstellung der Druckvorlagen), Schablonenfertigung (Herstellen der Schablonen für den Siebdruck), Farbküche (Vorbereitung der Druckpasten für den Druck), Maschinendruck (Bedrucken der Stoffbahnen mit Siebdruckschablonen), Nachbehandlung (Fixieren und Waschen der bedruckten Stoffe), Konfektion (Schneiden / Nähen der Fahnen), Versand (Zusammenstellung der Lieferung an Kunden). Außerdem sind Arbeitsplätze in der Fakturierung, Buchhaltung, Instandhaltung sowie im Lager zu nennen, desweiteren die Produktionsleitung sowie Geschäftsleitung. Für differenziertere Ausführungen über das Untersuchungsfeld sei auf Ceglarek und Rothe (2003a), Pokrandt (2004) oder Rothe und Ceglarek (2007) verwiesen.

In die Untersuchung von Pokrandt (2004) wurden insgesamt 21 Mitarbeiter aus den fünf Bereichen Verkauf, Auftragsbearbeitung, Einkauf, Filmfertigung sowie Auslieferung einbezogen, außerdem stand der Produktionsleiter als interner Experte zur Verfügung. Das Strukturlegen wurde nur von 19

Probanden absolviert, da zwei Mitarbeiter aufgrund von Überforderungsgefühlen darum baten, nicht am Verfahren teilzunehmen. Die Teilnahme an der Erhebung war freiwillig und konnte während der Arbeitszeit in einem separaten Raum absolviert werden. Die Geschäftsleitung unterstützte die Untersuchung, ebenso der Betriebsrat.

4.10.2. Untersuchungsaufbau

Auf Grundlage von Dokumentenanalysen (betriebliche Dokumente, Projektberichte aus vorangegangenen Forschungsprojekten im Unternehmen) und Interviews mit betrieblichen Experten sowie Mitarbeitern wurde von Pokrandt (2004) die Wissensdomäne erarbeitet. In Anbetracht der Ausgangsfrage, welche Variablen Abweichungen vom prototypischen Produktionsdurchlauf bedingen, wurden der Qualitätsablaufplan des Unternehmens und die damit verbundenen Arbeitsschritte in den einzelnen Abteilungen als Wissensdomäne für die Wissensdiagnosen gewählt. Auch in dieser Studie wurde das Urteil des Geschäftsführers als Leistungskriterium gewählt, zu dem die Ergebnisse der Wissenserfassung in Beziehung gesetzt werden sollen. Daher wurde gemeinsam mit Geschäftsführer und Produktionsleiter ein Leistungsbeurteilungssystem entwickelt, um anhand dessen die Arbeitsleistung der Mitarbeiter quantifizieren zu können. Daran schloß sich die Phase der Wissensdiagnose an, bei der sowohl der Wortassoziationsversuch als auch die Strukturlegetechnik eingesetzt wurden. Die Ergebnisse der Wissensdiagnosen wurden genutzt, um das hypothetische Wissensmodell zu modifizieren und weiter zu verfeinern. In der letzten Erhebungsphase wurden den Mitarbeitern Fragebögen zur Erfassung von Arbeitszufriedenheit, verschiedenen Aspekten des Commitments sowie soziodemographischer Daten vorgelegt, auf die im Weiteren nicht eingegangen werden wird. Zunächst sollen die Untersuchungsbestandteile Strukturlegetechnik sowie die Geschäftsführerbeurteilung kurz skizziert werden, anschließend wird das Vorgehen beim Wortassoziationsversuch ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Strukturlegetechnik: Die Strukturlegetechnik wurde als zweites wissensdiagnostisches Instrument neben dem Wortassoziationsversuch eingesetzt. Die Knoten wurden aus den Dokumentenanalysen sowie Expertengesprächen und dem Experten-Wortassoziationsversuch extrahiert. Somit stand allen Probanden der gleiche Pool an Begriffen zur Verfügung, sie wurden aber aufgefordert, nur diejenigen Begriffe als Knoten in ihr Netz zu integrieren, die auch ihrem eigenen Wissensbesitz entsprechen. Für die Benennung der Kanten standen 11 Relationen zur Verfügung (siehe dazu Pokrandt, 2004). Das Strukturlegen wurde nach einer kleinen Einführung in das Verfahren sowie dem gemeinsamen Bearbeiten (Versuchsleiterin und Proband) eines Übungsbeispiels als Papier-Bleistift-Version durchgeführt. Es bestand die Aufgabe, zwei Netze zu legen: erstens den Qualitätsablaufplan, zweitens den auftragsbezogenen Prozeßdurchlauf in der eigenen Abteilung. Der letztgenannte Umstand erforderte es, für jede Abteilung ein separates hypothetisches Wissensmodell zu erzeugen. Dieses Modell sowie ein entsprechendes Referenznetzwerk wurde durch den internen Experten (Produktionsleiter) erzeugt.

Geschäftsführerurteil: Ebenso wie in Studie 12 wurde auch hier das Urteil des Geschäftsführers als Indikator für erfolgreiches berufliches Handeln herangezogen. In einem Interview mit dem Geschäftsführer arbeitete Pokrandt (2004) alle relevanten Kriterien heraus, welche diesem als Grundlage für eine angemessene Beurteilung seiner Mitarbeiter bedeutsam waren. Anschließend

wurden die so erzeugten Beurteilungskriterien einem Paarvergleich unterzogen, woraus entsprechende Gewichte abgeleitet werden konnten. Im Ergebnis dieses Prozesses konnten die folgenden neun Kategorien gewonnen werden: Sozialkompetenz, Arbeitsqualität, Arbeitsengagement, Fachkompetenz, Organisations- und Planungsvermögen, Pünktlichkeit, Potential, Arbeitsquantität. Für eine differenzierte Erörterung dieser Kategorien sei auf Pokrandt (2004) verwiesen. Der Geschäftsführer setzte dabei für die Mitarbeiter aller Abteilungen die gleichen Kriterien in der gleichen Gewichtung an. Anschließend beurteilte er die ihm direkt unterstellten Mitarbeiter (Abteilungsleiter) anhand einer fünfstufigen Skala von 1=„Anforderungen werden außergewöhnlich erfüllt“ bis 5=„Anforderungen werden kaum erfüllt“. Die verbleibenden Mitarbeiter wurden durch den Produktionsleiter anhand des gleichen Kriterienkataloges beurteilt, nachdem dieser in das Prozedere eingewiesen wurde. Die Gesamtbeurteilung eines Mitarbeiters ist somit auch in dieser Studie die Summe von Einzelleistungskriterien, welche ihrerseits ein Produkt der Beurteilung mit dem ermittelten Gewichtskoeffizienten darstellen.

4.10.3. Wortassoziationsversuch

Der Wortassoziationsversuch war einer von zwei Bausteinen in der Phase der Wissensdiagnose. Neben der Erhebung einer individuellen Kenngröße für jeden Teilnehmer wurden die Ergebnisse des Assoziierens auch genutzt, um das hypothetische Wissensmodell zu modifizieren. Damit flossen sie indirekt in das Strukturlegen ein, da dessen Knoten auf dem Wissensmodell basieren, aber darüber hinaus wurde der Wortassoziationsversuch als eigenständiges Instrument begriffen.

<i>Nr.</i>	Stimulus	<i>Nr.</i>	Stimulus
1)	Qualitätsablaufplan	15)	Arbeitsschritte zwischen Bedarfsermittlung und Wareneingang im EK
2)	Verkauf	16)	Filmherstellung/ Grafik
3)	Möglichkeiten der Kundenakquise	17)	Möglichkeiten des Vorlageneingangs in der FI
4)	Wege der internen Klärung im VK / interne Klärungsprozesse im VK	18)	Bearbeiten des Vorabauftrags in der FI
5)	Arbeitsschritte, die zum Angebot gehören	19)	Arbeitsschritte, die zum Farbausdruck/ Farbplot gehören
6)	Arbeitsschritte von der Kundenakquise über das Angebot bis zur Bestellung	20)	Arbeitsschritte vom Eingang des Vorabauftrags in der FI bis zur Weiterleitung des fertigen Auftrags an Werk II bzw. DigiBEST
7)	Auftragsvorbereitung/Auftragsbearbeitung	21)	Arbeitsschritte der Schablonenabteilung
8)	Möglichkeiten des Bestellungseingangs / des Dateneingangs in der AB	22)	Arbeitsschritte der Farbküche
9)	Interne Klärungsprozesse in der AB	23)	Arbeitsschritte des Maschinendrucks
10)	Vorabauftrag	24)	Arbeitsschritte in der Nachbehandlung
11)	Prüfzettel	25)	Arbeitsschritte in der Konfektion
12)	Arbeitsschritte vom Vorabauftrag bis zur Produktion	26)	Auslieferung
13)	Einkauf	27)	Möglichkeiten & Wege des Wareneingangs in der AU
14)	Ablauf der Bedarfsmeldungen der Abteilungen	28)	Arbeitsschritte vom Wareneingang bis zum Bereitstellen der Ware für Transporteure

Abbildung 33: Stimuli der Studie 13 – Stoffdruckerei (vgl. Pokrandt, 2004)

Wie bereits erwähnt wurde als Wissensdomäne der Auftragsdurchlauf und die damit verbundenen Produktionsschritte gewählt. Der Qualitätsablaufplan ist Kernstück des sog. Integrierten Managementsystems des Unternehmens, welches für jeden Mitarbeiter jederzeit im Intranet

einsehbar ist; ein Flußdiagramm des Qualitätsablaufplans hängt sichtbar und großformatig im Eingangsbereich des Werkes. Die daraus abgeleiteten 28 Stimuli (siehe dazu Abbildung 33) wurden von Pokrandt (2004) mit dem Geschäftsführer abgestimmt und mit dem internen Experten in einer Pilotphase erprobt.

Der Wortassoziationsversuch wurde mündlich durchgeführt, dabei auf Tonband aufgezeichnet und anschließend transkribiert. Für die Äußerungen der Probanden wurde keine Zeitbegrenzung vorgegeben, der nächste Stimulus wurde auf Zeichen des Probanden präsentiert. Vor dem eigentlichen Wortassoziationsversuch wurde das Verfahren an einem Übungsstimulus kurz demonstriert. Allen Probanden wurden die Stimuli in gleicher Reihenfolge vorgelegt, so wie in Abbildung 33 aufgelistet.

Pokrandt (2004) berichtet von erheblichen interindividuellen Unterschieden hinsichtlich Qualität und Niveau der Assoziationen, in Abbildung 34 wird eine große Spanne zwischen minimal 2,82 und maximal 21,54 assoziierten Konzepten auf einen Stimulus hin deutlich. Im Mittel über die gesamte Stichprobe werden von einem Probanden durchschnittlich 6,85 Assoziationen auf einen Stimulus hin assoziiert ($SD=4,01$).

	gesamte Stichprobe (N=21)	gute Assoziierer (N=7)	schlechte Assoziierer (N=7)
<i>M</i>	6,85	10,73	3,93
<i>SD</i>	4,01	4,89	0,71
Minimum	2,82	6,68	2,82
Maximum	21,54	21,54	4,86

Abbildung 34: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 13 – Stoffdruckerei

Auf Grundlage der Assoziationsleistung bildete Pokrandt (2004) Extremgruppen, wobei jeweils sieben Probanden die Gruppe der guten bzw. schlechten Assoziierer repräsentierten. Pokrandt (2004) verweist allerdings auch auf die große Standardabweichung bei den guten Assoziierern, somit stellt sich die Gruppe der schlechten Assoziierer deutlich homogener dar als die der guten Assoziierer. Diese beiden Gruppen, deren Kennwerte in Abbildung 34 aufgeführt sind, lassen sich anhand dieses quantitativen Assoziationskennwertes signifikant voneinander unterscheiden (Mann-Whitney-U: $z=-3,130$; $p=.001$), die guten Assoziierer produzierten dabei im Mittel etwa dreimal so viel wie die schlechten Assoziierer.

4.11. Studie 14: Seminar Wissensmanagement

4.11.1. Anliegen der Studie und Untersuchungsfeld

Die von Meyer (2007) durchgeführte Erhebung diente dem Nachweis des Wissenszuwachses durch eine Seminarveranstaltung. Meyer zielte dabei weniger auf die Assoziationen i.e.S. ab, vielmehr war die Wortassoziation Teil einer Verfahrensvalidierung des Assoziations-Struktur-Tests, welcher von Meyer (2007) entwickelt wurde. Die Untersuchung fand im Rahmen einer Seminarveranstaltung des Hauptstudiums im Fach Psychologie statt. Das betreffende Seminar wurde von Meyer selbst durchgeführt und thematisierte Wissensmanagement.

Die Probanden waren die Teilnehmer des Seminars, d.h. es handelte sich um Psychologie-Studenten im Hauptstudium. Die Erhebung wurde in die Lehrveranstaltung integriert, so daß eine hohe Beteiligung gewährleistet werden konnte. Aufgrund von Fluktuation bzw. mangelnder Teilnahme liegen nicht von allen Probanden Daten zu allen Meßzeitpunkten vor. Die hier vorgestellten Ergebnisse sollen sich auf die Daten derjenigen 20 Probanden beziehen, für die zu beiden Erhebungszeitpunkten ein vollständiger Datensatz erhoben werden konnte.

4.11.2. Untersuchungsaufbau

Wie bereits erwähnt diente die Erhebung dem Nachweis von Wissenszuwachs. Dementsprechend wurde ein Prä-Post-Design gewählt, welches in Abbildung 35 dargestellt ist. Zu Beginn der ersten Seminarveranstaltung (T1) wurde ein Multiple-Choice-Test zum Thema Wissensmanagement durchgeführt, bei dem maximal 10 Punkte zu erreichen waren. Anschließend wurde am Computer der Assoziations-Struktur-Test von Meyer (2007) durchgeführt. Dabei bearbeiteten die Probanden zunächst den Wortassoziationsversuch zu einem einzigen Stimulus, unmittelbar im Anschluß wurden die darauf assoziierten Konzepte den Probanden zum Paarvergleich vorgelegt, um dann letztlich durch den Versuchsleiter in einem Wissensgraphen rekonstruiert zu werden. Zum Multiple-Choice-Test liegen leider keine detaillierteren Angaben vor. Am Ende der letzten Seminarsitzung (T2) wurde von den Probanden das identische Methodeninventar bearbeitet. Das Seminar erstreckte sich über ein Semester, d.h. zwischen den beiden Meßzeitpunkten lagen etwa vier Monate Treatment.

	Prä-Test (T1, erste Seminarsitzung)	Post-Test (T2, letzte Seminarsitzung)
1.	Assoziations-Struktur-Test	Assoziations-Struktur-Test
2.	Multiple-Choice-Test	Multiple-Choice-Test

Abbildung 35: Schematischer Ablauf in Studie 14 – Seminar Wissensmanagement

4.11.3. Wortassoziationsversuch

In dieser Studie ist der Wortassoziationsversuch als Teil des übergeordneten, computerbasierten Verfahrens Assoziations-Struktur-Test von Meyer (2007) zu verstehen, dabei wurde nur ein einziger Stimulus präsentiert: der Oberbegriff Wissensmanagement. In der vorliegenden Arbeit soll der Wortassoziationsversuch nun als separates Verfahren betrachtet werden.

	T1 (vor Seminarbeginn)	T2 (Seminarende)
<i>M</i>	10,00	16,35
<i>SD</i>	4,22	7,50
Minimum	3	7
Maximum	18	32

Abbildung 36: Kennwerte des Wortassoziationsversuchs– Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 14 – Seminar Wissensmanagement

Die Assoziationskennwerte zu beiden Meßzeitpunkten sind normalverteilt. Vor Beginn des Seminars werden mit $M=10,00$ ($SD=4,22$) signifikant weniger Assoziationen produziert als zum Ende des Seminars mit $M=16,35$ ($SD=7,50$) Assoziationen ($t(19)=-4,028$; $p=.001$). Sehr

eindrucksvoll bildet sich dies auch im Range ab, der maximal erreichte Assoziationskennwert verdoppelt sich fast bei der Meßwiederholung (siehe auch Abbildung 36).

	gute Assoziierer (N=6)	schlechte Assoziierer (N=6)
<i>M</i>	26,00	6,67
<i>SD</i>	4,24	1,51
Minimum	21	7
Maximum	32	10

Abbildung 37: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert in T2) in Studie 14 – Seminar Wissensmanagement

Der Kennwert des Wortassoziationsversuchs ist in der Lage, extreme Leistungsunterschiede zwischen den Probanden zu markieren, exemplarisch sei dies am zweiten Erhebungszeitpunkt demonstriert (siehe Abbildung 37). Es konnten Extremgruppen mit jeweils sechs Probanden gebildet werden, welche sich hinsichtlich ihres Assoziationskennwertes signifikant voneinander unterscheiden (Mann-Whitney-U: $z=-2,908$; $p=.002$). Dabei assoziieren die guten Assoziierer mit $M=26,00$ ($SD=4,24$) etwa viermal soviel wie die schlechten Assoziierer ($M=6,67$; $SD=1,51$).

4.12. Studie 15: Wissen über Arbeitsschutz

4.12.1. Anliegen der Studie und Untersuchungsfeld

Im Rahmen der Europäischen Woche 2006 (ausgerichtet von der Europäischen Arbeitsschutzagentur) wurde im Diplomstudiengang Psychologie an der Universität Potsdam ein Projektseminar initiiert, welches das Thema „Arbeitsschutz und junge Beschäftigte“ aufgriff. Dazu wurde gemeinsam mit den Studierenden im Seminar ein Fragebogen entwickelt und eingesetzt. Ausgangspunkt des Projektes war der Umstand, daß das Studium auf die künftige Berufstätigkeit vorbereiten sollte. In vielen Fällen übernehmen Akademiker später Führungsaufgaben und sind dabei verantwortlich für Gesundheit und Leben der Mitarbeiter. Im Gegensatz zur betrieblichen Berufsausbildung ist der Arbeits- und Gesundheitsschutz jedoch in der universitären Ausbildung kaum thematisiert. Ziel der Studie war eine Bestandsaufnahme, inwieweit das Studium auf diese künftigen Arbeitsanforderungen eingeht und über welchen diesbezüglichen Wissensstand die Studierenden verfügen (siehe dazu auch Ceglarek, 2007).

Insgesamt nahmen 472 Studierende aller Fakultäten an der Erhebung teil. Über die Hälfte der Probanden war der humanwissenschaftlichen Fakultät zuzuordnen (55,3 %), die wirtschafts- und sozialwissenschaftliche Fakultät war mit 17,6 % sowie die naturwissenschaftliche Fakultät mit 15,7 % in der Stichprobe vertreten. Die Probanden waren durchschnittlich 24,00 Jahre alt ($SD=3,37$) und hatten im Mittel 6,37 Semester absolviert ($SD=3,60$), wobei die Streubreite zwischen einem und 22 Semestern lag. Fast drei Viertel der Befragten (72,2 %) waren Frauen.

4.12.2. Untersuchungsaufbau

Der Entwicklung des Fragebogeninstrumentes voran gingen Dokumentenanalysen (Curricula, Studienordnungen) sowie Interviews mit Experten des universitären Sicherheitswesens bzw. aus

speziellen Fachbereichen. Die Studie selbst war als Online-Erhebung konzipiert, Gründe dafür waren die größere Erreichbarkeit potentieller Probanden sowie die zu erwartende Auswerteökonomie. Über den Mail-Verteiler der Universität Potsdam wurde für die Teilnahme an der Untersuchung geworben und dabei die Verlosung von Kinogutscheinen ausgelobt. Neben dem Aushängen von Plakate in den Gebäuden der Universität und dem Verteilen von Handzetteln warben Angehörige des Projektseminars in einigen Lehrveranstaltungen persönlich für die Studie. Darüber hinaus konnten Dozenten gewonnen werden, die sogar Zeit ihrer eigenen Lehrveranstaltung dafür einräumten, damit die Studierenden den Fragebogen ausfüllten. Dies betraf zwei Veranstaltungen im Fach Psychologie sowie eine Lehrveranstaltung im Fach Arbeitslehre / Technik. Aufgrund der räumlich-zeitlichen Einbindung in die Lehrveranstaltung war in diesen Fällen keine Online-Erhebung möglich, statt dessen wurde der Fragebogen in einer äquivalenten Papierversion vorgelegt.

Zunächst soll der eingesetzte Online-Fragebogen im Ganzen vorgestellt und der darin integrierte Wissenstest kurz skizziert werden; anschließend wird das Vorgehen beim Wortassoziationsversuch, welcher ebenfalls ein Bestandteil des Online-Fragebogens ist, ausführlich dargestellt sowie dessen deskriptive Statistik berichtet.

Abschnitt	Erläuterungen
1. Angaben zur Person	<ul style="list-style-type: none"> – soziodemographische Angaben – Angaben zum Studienfach, Semesterzahl
2. Vorerfahrungen	<ul style="list-style-type: none"> – Erwerbstätigkeit vor dem Studium, Ausbildung, Wehrdienst o.ä. – Erwerbstätigkeit während des Studiums – selbst einen (Arbeits-) Unfall gehabt – Teilnahme an arbeitspsychologischen Untersuchungen – Teilnahme am Erste-Hilfe-Kurs
3. Freie Assoziation	<ul style="list-style-type: none"> – zu drei Stimulusbegriffen
4. Einschätzungen ...	<ul style="list-style-type: none"> – ... der Relevanz des Themas Arbeitsschutz – ... der Unfallgefahren für verschiedene Arbeitnehmergruppen – ... der Campi bzgl. Behindertenfreundlichkeit, Unfallgefahren – ... der Ergonomie studentischer Arbeitsplätze – ... Kenntnis von sicherheitsrelevanten Personen / Orten der Universität
5. Wissenstest	<ul style="list-style-type: none"> – Regelungen zur studentischen Unfallversicherung (Versicherungsträger, -leistungen, versicherte Aktivitäten) – Meldekette bei Eintreten eines Unfalls – Verantwortlichkeit der Universität im Arbeitsschutz
6. Wissenserwerb im Studium	<ul style="list-style-type: none"> – selbständiges Beschäftigen mit dem Thema Arbeitsschutz – Besuch von Lehrveranstaltungen zum Thema Arbeitsschutz
7. persönliche Ressourcen	<ul style="list-style-type: none"> – zeitliche Beanspruchung durch das Studium, Schlafdauer – inhaltliche Beanspruchung durch das Studium – gesunde Ernährung – Entspannung

Abbildung 38: Bestandteile des Online-Fragebogens Studie 15 – Wissen über Arbeitsschutz

Online-Fragebogen zum Wissen über Arbeits- und Gesundheitsschutz: Der den Probanden dargebotene Fragebogen umfaßte sieben Inhaltsbereiche, welche in Abbildung 38 skizziert sind. Auf der Startseite des Online-Fragebogens wurde den Probanden kurz das Anliegen und die organisatorische Einbettung des Projektes erläutert. Technisch gesehen war der Online-Fragebogen aus drei html-Seiten aufgebaut, deren Eintragungen in einer Datenbank abgelegt wurden. Nach vollständigem Bearbeiten einer Seite konnte die nächste Seite erreicht werden. Sollte ein Proband

vergessen haben, ein Item zu beantworten, so wurde er beim Versuch, die nächste Seite zu erreichen, daran erinnert, das entsprechende Item zu vervollständigen.

Multiple-Choice-Test: Im sog. Wissenstest als Bestandteil des Online-Fragebogens (siehe Abschnitt 5. in Abbildung 38) wurden insgesamt sieben Fragen im Multiple-Choice-Format gestellt (zum ausführlichen Fragebogen siehe Anhang 5). Zu jeder Frage wurden jeweils mehrere Antworten angeboten, welche vom Probanden dahingehend beurteilt werden sollten, ob diese die Frage korrekt beantworten oder nicht. Jeder dieser Antworten stand eine dreistufige Antwortskala bei, welche die Alternativen „Ja“, „Nein“ oder „weiß nicht“ umfaßte. Um zu einem Maß für das Wissen über Regelungen des Arbeits- und Gesundheitsschutzes zu gelangen, wurde für jedes richtige „Ja“ sowie jedes richtige „Nein“ jeweils ein Punkt vergeben, für falsche Antworten oder „weiß nicht“ kein Punkt. Die richtigen Antworten wurden aufsummiert, so daß sich ein maximaler Wissensscore von 58 Punkten ergab. Die inhaltlichen Aspekte sind Abbildung 38 zu entnehmen, damit deckt dieser Test einen allgemeineren, eher formalen Aspekt des möglichen Wissens über Arbeitsschutz ab.

4.12.3. Wortassoziationsversuch

Die Stimuli des Wortassoziationsversuchs wurden in den Fragebogen integriert und dort relativ zu Beginn positioniert, wie in Abbildung 38 (siehe Abschnitt 3.) zu erkennen ist. Grund für dieses Vorgehen war, daß die Probanden nicht zu sehr geprimt werden sollten, z.B. durch die Items zum Wissenstest und die dort verwendeten Begrifflichkeiten. Die vorgeschalteten Items zu den soziodemographischen Angaben schienen diesbezüglich weniger problematisch.

Angesichts des Umfangs des gesamten Fragebogens wurde die Anzahl der Stimuli auf drei begrenzt. Da a priori die Annahme bestand, daß die Studierenden wenig über Arbeits- und Gesundheitsschutz wissen, wurden sehr allgemeine Stimuli festgelegt. Folgende Stimuli kamen zum Einsatz:

1. Arbeitsschutz,
2. Gesundheitsschutz,
3. Arbeitsunfall.

Die technische Systematik hinsichtlich des Fehlens von Eintragungen, wie sie für den gesamten Fragebogen bereits beschrieben wurde, traf auch für die Assoziationsitems zu: war in eines der Textfelder nichts eingetragen, wurde der Proband beim Seitenwechsel dazu aufgefordert, diese Items zu bearbeiten. Dem konnte allerdings auch durch das Eintragen sinnloser Zeichen entsprochen werden. Anhang 6 stellt die Ansicht des Wortassoziationsversuchs dar, wie sie innerhalb des Online-Fragebogens präsentiert wurde. Die jeweils neben dem Stimulus angeordnete Bildlaufleiste ermöglichte das Eintragen von beliebig langen Listen von Begriffen.

Die Assoziationsleistung der Erhebungsteilnehmer weist interindividuelle Unterschiede auf, zwischen minimal null und maximal 9,0 assoziierten Inhalten pro Stimulus beträgt der Mittelwert über die gesamte Stichprobe $M=2,76$ ($SD=1,56$). Der Assoziationskennwert ist normalverteilt, wenn auch linksschief (der Median liegt bei 2,67 Assoziationen). An dieser Stelle sei daran erinnert, daß in dieser Studie die Grundannahme bestand, der Wissensstand der Studierenden sei insgesamt

unzureichend ausgeprägt. Dies konnte bestätigt werden, dennoch lassen sich zwei Extremgruppen klar voneinander unterscheiden (siehe Abbildung 39): die Gruppe der guten Assoziierer mit 57 Probanden assoziiert im Mittel $M=5,67$ ($SD=1,09$), die Gruppe der schlechten Assoziierer hingegen mit 67 Probanden nur durchschnittlich $M=0,57$ ($SD=0,45$) Assoziationen. Diese Abgrenzung ist statistisch signifikant ($t(122)=35,01$; $p= .000$).

	gesamte Stichprobe (N=472)	gute Assoziierer (N=57)	schlechte Assoziierer (N=67)
<i>M</i>	2,76	5,67	0,57
<i>SD</i>	1,56	1,09	0,45
Minimum	0,00	4,67	0,00
Maximum	9,00	9,00	1,00

Abbildung 39: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 15- Wissen über Arbeitsschutz

4.13. Studie 16: Call Center (A)

4.13.1. Anliegen der Studie und Untersuchungsfeld

Studie 16 – Call Center (A) – thematisiert das individuelle Wissensmanagement an Call-Center-Arbeitsplätzen, bei denen Dienstleistungen per Telefon erbracht werden. Das dafür notwendige fachspezifische Wissen läßt sich in anhand von drei Klassen beschreiben: a) Fachwissen über die Produkte oder die Dienstleistungen, die das Call Center anbietet, b) Wissen über den Umgang mit der Computertechnik des Call Centers (Systemwissen), und c) Wissen über den Umgang mit Kunden (Kommunikationswissen). Im Rahmen eines Forschungsprojektes (Ceglarek und Rothe, 2002; Ceglarek und Rothe, 2003b) wurde das individuelle Fachwissen der sog. Call-Center-Agents erhoben. Dabei sollte auch geklärt werden, inwieweit sich die Call-Center-interne Aufteilung der Agents in verschiedene Spezialisierungslevels sich durch deren individuellen Wissensbesitz abbilden läßt, außerdem wurde der Zusammenhang von Wissen und Arbeitsbelastung thematisiert, worauf hier nicht weiter eingegangen werden soll.

Das untersuchte Call Center war eine technische Berater-Hotline für ein einziges Telekommunikationsprodukt. Dieses Produkt stand im Zusammenhang mit dem Internetzugang DSL, die Aufgabe des Call Centers war in erster Linie, den Kunden Hilfestellung zu geben bei der Installation dieses Produktes, ggf. Störungen zu erkennen und zu beheben sowie allgemeine und konkrete Beratungsleistungen zum Produkt zu geben. Das Call Center war ausschließlich als Inbound- und Frontoffice-Tätigkeit organisiert, d.h. eingehende Anrufe wurde entgegengenommen und im direkten Kundenkontakt umgehend bearbeitet.

Von den insgesamt 44 im Call Center beschäftigten Agents nahmen 29 Agents am Wortassoziationsversuch teil. Fast die Hälfte der Probanden war älter als 40 Jahre, ein knappes Drittel ist zwischen 30 und 40 Jahren alt, ein Fünftel der Agents war jünger als 30 Jahre. Die Mehrheit der Befragten hatte eine abgeschlossene Berufsausbildung, überwiegend im Bereich der Telekommunikation. Für detailliertere Ausführungen zum Untersuchungsfeld sei auf Ceglarek und Rothe (2002) verwiesen.

4.13.2. Untersuchungsaufbau

Die Untersuchungen im Call Center befaßten sich sowohl mit dem organisationalen als auch dem individuellen Wissensmanagement. Dazu wurden betriebliche Unterlagen ausgewertet sowie Experteninterviews mit dem Hotline-Manager geführt. Die gesichteten betrieblichen Unterlagen umfaßten Schulungsunterlagen zum betreuten Produkt und zur eingesetzten Software sowie die Intranetseiten der Hotline (u.a. mit reichhaltigen Produktinformationen, Leitfaden zur Hotlinearbeit, Dokumente zum Qualitätsmanagement). Zu Beginn der Studie wurden Ganztagsbeobachtungen an mehreren Arbeitsplätzen durchgeführt und dabei u.a. die Nutzung von Wissensspeichern protokolliert. In der sich anschließenden Fragebogenerhebung wurden Ressourcen und Arbeitsbelastungen der Agents mit standardisierten Verfahren erfragt. Auf die Ergebnisse von Arbeitsbeobachtungen und Belastungsuntersuchungen soll hier nicht weiter eingegangen werden. Die Wissensdiagnose mittels Wortassoziationsversuch wurde am Ende der mehrwöchigen Erhebungsphase durchgeführt: zum einen wurde die Zeitspanne für den Vertrauensaufbau zwischen Beschäftigten und Versuchsleitern genutzt, zum anderen dienten insbesondere die Arbeitsbeobachtungen dazu, die Relevanz potentieller Stimuli durch die Versuchsleiter abschätzen zu können.

4.13.3. Wortassoziationsversuch

Zur Wissensdiagnose wurde der Wortassoziationsversuch eingesetzt, da dieser als besonders geeignet für Call-Center-Arbeitsplätze angesehen wurde: die Agents müssen tagtäglich ihr Wissen auf Hinweisreize des Kunden hin verbalisieren – der Wortassoziationsversuch stellt eine vergleichbare Anforderung dar. Als Wissensdomäne wurde das betreute Produkt DSL herangezogen, wobei nach folgenden drei Bereichen differenziert wurde: Produktwissen (Aufbau und Funktion der technischen Anlagen), Diagnosewissen (z.B. Wissen über Schwierigkeiten bei der Installation, über Störungen, deren Symptome sowie über das Vorgehen zur Störungsidentifikation) sowie Kommunikationswissen (Wissen über Gesprächsführung, über den Umgang mit unterschiedlichen Kundentypen). Aus jedem dieser drei Bereiche wurden zehn Stimuli zusammengetragen, so daß der Wortassoziationsversuch insgesamt 30 Stimuli umfaßte, welche in Abbildung 40 aufgelistet sind.

Nr. Stimuli Produktwissen	Nr. Stimuli Diagnosewissen	Nr. Stimuli Kommunikationswissen
1) ADSL	11) Zugangsproblem	21) Verhalten als Telefonprofi
2) Reichweiten	12) LED Leuchten	22) Gesprächsbeginn
3) kundenseitige Verkabelung	13) Ersteinrichtung eines DSL-Zugangs	23) Gesprächsende
4) Splitter	14) kein Seitenaufbau	24) Beratungstiefe
5) Modem	15) Verbindungstest	25) Kundenzufriedenheit
6) NWK	16) Geschwindigkeitsprobleme	26) Bearbeitungszeit
7) Treiber	17) Installationsprobleme	27) sprachlich unbegabter Kunde
8) Protokolle	18) Probleme mit Providerssoftware	28) technisch unbegabter Kunde
9) DFÜ- Verbindung	19) Protokollprobleme	29) Beschimpfungen
10) Provider	20) Probleme mit IMac	30) Beschwerden

Abbildung 40: Stimuli der Studie 16 – Call Center (A)

Der Versuch wurde mündlich durchgeführt, auf Tonband aufgezeichnet und später transkribiert. Allen Probanden wurden die Stimuli in einer festgelegten Reihenfolge vorgelegt, zur Unterstützung wurde der jeweilige Stimulus nicht nur genannt, sondern parallel auf einem kleinen Kärtchen präsentiert. Eine Zeitbegrenzung bestand nicht.

Im Mittel über alle Stimuli werden von den Probanden durchschnittlich etwa 4 Assoziationen auf einen Stimulus hin produziert ($M=3,95$; $SD=1,43$). Differenziert man weiter nach den drei Bereichen der Wissensdomäne (siehe auch Abbildung 41), so zeigt sich die größere Assoziationsmenge im Bereich des Diagnosewissens, welches sich mit $M=4,42$ ($SD=1,99$) signifikant von der Leistung im Produktwissen (Wilcoxon: $z=-2826$, $p=.005$) und auch im Kommunikationswissen (Wilcoxon: $z=-2,096$, $p=.036$) abhebt.

Assoziationen pro Stimulus über alle Stimuli	... im Produktwissen	... im Diagnosewissen	... im Kommunikationswissen
<i>M</i>	3,95	3,77	4,42	3,66
<i>SD</i>	1,43	1,74	1,99	1,01
Minimum	1,9	1,5	1,7	2,0
Maximum	7,6	8,2	9,7	6,6

Abbildung 41: Kennwerte des Wortassoziationsversuchs– Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe und differenziert nach Inhaltsbereichen in Studie 16 – Call Center (A)

Der Wortassoziationsversuch ist in der Lage, zwischen extremen Leistungsunterschieden zu unterscheiden (siehe auch Abbildung 42). So wurden anhand des Assoziationskennwertes zwei Extremgruppen mit jeweils fünf Probanden gebildet, welche sich signifikant voneinander abgrenzen (Mann-Whitney-U: $z=-2,611$; $p=.008$). Während die guten Assoziierer im Mittel 6,50 Nennungen produzierten ($SD=0,69$), erreichten die schlechten Assoziierer einen durchschnittlichen Assoziationskennwert von 2,38 ($SD=0,38$).

	gute Assoziierer (N=5)	schlechte Assoziierer (N=5)
<i>M</i>	6,50	2,38
<i>SD</i>	0,69	0,38
<i>Minimum</i>	5,83	1,90
<i>Maximum</i>	7,60	2,80

Abbildung 42: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert) in Studie 16 – Call Center (A)

4.14. Studie 17: Call Center (B)

4.14.1. Anliegen der Studie und Untersuchungsfeld

Die von Ruprecht (2004) vorgelegte Diplomarbeit beschäftigt sich mit der Analyse der Arbeitsbedingungen in einem Call Center. Dabei steht neben der Erfassung psychischer Belastungen und arbeitsbezogener Einstellungen der Beschäftigten die Erfassung des fachbezogenen Wissens der Mitarbeiter im Fokus, da diese eine zentrale Leistungsvoraussetzung

für den Umgang der Beschäftigten mit den Arbeitsanforderungen darstellen (Ruprecht, 2004). Untersucht wurden u.a. Hypothesen zum Zusammenhang des Fachwissens mit der psychischen Beanspruchung, mit Verhaltensmerkmalen sowie Arbeitszufriedenheit und Commitment.

Die Studie von Ruprecht (2004) stellt ein Nachfolgeprojekt der von Ceglarek und Rothe (2002) erhobenen Untersuchung dar. Beide Call Center (Studie 16 und Studie 17) weisen den gleichen organisationalen Hintergrund auf und gehören zum gleichen Telekommunikationsunternehmen, beide besitzen ähnliche Call-Center-interne Strukturen und betreuen eine ähnliche Produktparte, nämlich Dienstleistungen im Zusammenhang mit Internet. Die Arbeitsaufgabe der Untersuchungsteilnehmer bei Ruprecht (2004) bestand in der telefonischen Beratung bei technischen Problemen, im Umgang mit den Produkten sowie ggf. bei Kaufinteressen. Inhaltlich betrafen die vom Call Center bzw. der untersuchten Subeinheit betreuten Produkte v.a. das Bereitstellen von Diensten auf einem Internetserver. Das Call Center arbeitete überwiegend im Frontoffice, d.h. die Anfragen werden telefonisch just in time bearbeitet, sowie als Inbound, d.h. eingehende Anfragen werden bearbeitet, bei denen Zeitpunkt und Inhalt des Kontaktes von den Kunden bestimmt werden. Für detailliertere Angaben zum Untersuchungsfeld sei auf Ruprecht (2004) verwiesen.

Im untersuchten Unternehmensbereich des Call Centers arbeiteten zum Erhebungszeitpunkt 34 Beschäftigte, welche in mehreren Teams mit identischen Aufgaben organisiert waren (siehe auch Ruprecht, 2004). Am Wortassoziationsversuch nahmen 24 Mitarbeiter teil. Die ausschließlich männlichen Probanden waren im Mittel 31 Jahre alt ($SD= 9,15$), der jüngste Mitarbeiter war 21 Jahre, der älteste 50 Jahre alt. Zwei Drittel der Probanden besaßen die Mittlere Reife, alle verfügten über eine abgeschlossene Berufsausbildung, wobei fast alle (20 Probanden) diese im Bereich der Telekommunikation absolviert hatten.

4.14.2. Untersuchungsaufbau

Den Beginn der Erhebung von Ruprecht (2004) markierte eine intensive Phase des Vertrautmachens mit dem Untersuchungsfeld. Dabei wurden neben ersten Arbeitsbeobachtungen bzw. Beobachtungsinterviews diverse betriebliche Unterlagen wie Ablaufpläne, Stellenbeschreibungen, Arbeitsanweisungen sowie Schulungsunterlagen gesichtet. Zunächst führte Ruprecht (2004) eine systematische Arbeitsbeobachtung durch zur Erfassung der objektiven Arbeitssituation und der Bedingungen der Tätigkeitsausführung. Daran schloß sich eine Fragebogenerhebung an, welche mit Hilfe von diversen Fragebögen die arbeitsbezogenen Beanspruchungen sowie Einstellungen zur Arbeit erfassen sollten, worauf hier nicht näher eingegangen werden soll. Die Methodenbatterie schloß mit dem Wortassoziationsversuch als wissensdiagnostischem Instrumentarium ab. Ein Außenkriterium wie etwa eine Beurteilung der Arbeitsleistung durch den Vorgesetzten konnte nicht herangezogen werden, da ein solches Vorgehen durch das Unternehmen nicht gebilligt wurde.

4.14.3. Wortassoziationsversuch

Der Wortassoziationsversuch wurde als eigenständiges und einziges Instrument zur Diagnose des individuellen Fachwissens eingesetzt. Ruprecht (2004) bewertet dieses Verfahren als für die Untersuchungsstichprobe sehr geeignet, da es den täglichen Arbeitsanforderungen ähnelt: als Inbound-Frontoffice-Call-Center werden den Beschäftigten von Kunden Schlüsselworte genannt,

auf die die Mitarbeiter reagieren müssen, indem sie adäquates Wissen aktivieren und dementsprechend den Kunden instruieren oder beraten. Fachwissen muß also täglich im Rahmen von Kundengesprächen verbalisiert werden.

Der Arbeitsgegenstand in der Untersuchungsstichprobe umfaßt u.a. Dienste, welche auf einem Internetserver bereitgestellt werden, Ruprecht (2004) bezeichnet diese Wissensdomäne als Web-Hosting. Im Ergebnis der Arbeitsbeobachtungen konnte dieser Themenbereich als derjenige identifiziert werden, auf den sich die meisten Kundenanfragen bezogen, so daß dies als Domäne für die Wissensdiagnose gewählt wurde. Nach der Auswertung betrieblicher Unterlagen, Arbeitsbeobachtungen sowie Expertengesprächen extrahierte Ruprecht (2004) aus dieser Domäne 30 Stimuli, welche in Abbildung 43 dargestellt sind. Die Stimuli wurden allen Probanden in festgelegter Reihenfolge präsentiert. In Anlehnung an die Vorgängererhebung von Ceglarek und Rothe (2002) in einem ähnlichen Call Center wählte Ruprecht (2004) die Stimuli aus den Kategorien Wissen über das Produkt, Wissen über die spezifische Diagnose bei Störungen sowie Wissen über die Kommunikation mit dem Kunden, wobei alle diese Kategorien in gleichem Umfang ausgefüllt wurden. Die Stimuli der Kategorie Kommunikationswissen wurden aufgrund der bereits erwähnten identischen Call-Center-Strukturen aus der Erhebung von Ceglarek und Rothe (2002) unverändert übernommen.

<i>Nr.</i> Stimuli Produktwissen	<i>Nr.</i> Stimuli Diagnosewissen	<i>Nr.</i> Stimuli Kommunikationswissen
1) Web Easy	11) Cookies	21) Verhalten als Telefonprofi
2) MX-Record	12) Publishing Kennwort	22) Gesprächsbeginn
3) IP- Adresse	13) Ping	23) Gesprächsende
4) SMTP-Relayserver	14) Zugangsprobleme zu OBSOC	24) Beratungstiefe
5) Domain Name System	15) kein Upload möglich	25) Kundenzufriedenheit
6) Apacheserver	16) Postfächer in Outlook anlegen	26) Bearbeitungszeit
7) FTP- Upload	17) Mails kommen nicht an	27) sprachlich unbegabter Kunde
8) dedizierter Server	18) Domain Check Tools	28) technisch unbegabter Kunde
9) BIND	19) Domainweiterleitung	29) Beschimpfungen
10) Skripte	20) Contentseite	30) Beschwerden

Abbildung 43: Stimuli der Studie 17 – Call Center (B) (vgl. Ruprecht, 2004)

Die Durchführung des Wortassoziationsversuchs erfolgte in Einzelsitzungen mit den Probanden, es gab keine Zeitbegrenzungen für die Bearbeitung der einzelnen Stimuli. Der Wortassoziationsversuch wurde mündlich durchgeführt, auf Tonband aufgezeichnet und anschließend transkribiert. Neben der Instruktion zum Antworten in Stichworten oder kurzen Phrasen, ermunterte Ruprecht (2004) die Probanden, Äußerungen, die auf einen vorangegangenen Stimulus bereits genannt wurden, ggf. zu wiederholen, wenn sie bei einem weiteren Stimulus vom Probanden als relevant erachtet würden.

Ruprecht (2004) berichtet große interindividuelle Unterschiede zwischen den Probanden hinsichtlich der Anzahl als auch der Qualität der Äußerungen. Insgesamt wurden die meisten Nennungen zu den Stimuli des Diagnosewissens assoziiert, während zum Produkt- sowie Kommunikationswissen etwa gleich viel benannt wurde (siehe auch Abbildung 44), insgesamt wurden über alle Wissensbereiche hinweg im Mittel 5,26 Äußerungen pro Stimulus ($SD=2,30$) produziert.

Assoziationen pro Stimulus über alle Stimuli	... im Produkt- wissen	... im Diagnose- wissen	... im Kommunikations- wissen
<i>M</i>	5,26	4,87	6,13	4,74
<i>SD</i>	2,30	1,86	3,00	2,35
Minimum	2,23	1,90	2,70	1,90
Maximum	13,30	9,10	16,50	14,10

Abbildung 44: Kennwerte des Wortassoziationsversuchs– Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe und differenziert nach Inhaltsbereichen in Studie 17 – Call Center (B) (siehe auch Ruprecht, 2004)

Auch hier konnte die Stichprobe in zwei Extremgruppen differenziert werden, in die jeweils fünf Probanden eingingen (Abbildung 45). Die guten Assoziierer erreichten im Mittel einen Assoziationskennwert von 8,56 ($SD=2,67$), die schlechten Assoziierer hingegen einen mittleren Kennwert von 2,85 ($SD=0,47$). Diese Abgrenzung zwischen den Extremgruppen ist statistisch signifikant Mann-Whitney-U: $z=-2,611$; $p=.008$). Ruprecht (2004) weist darauf hin, daß sich die Extremgruppen auch bei differenzierter Betrachtung der einzelnen Inhaltskategorien (Produktwissen etc.) wiederfinden. Die herausgehobene Stellung des Diagnosewissens im Kontrast zu Produkt- bzw. Kommunikationswissen, welche für die gesamte Stichprobe berichtet wurde, läßt sich allerdings nicht als generelles Phänomen beschreiben, denn dieser Effekt konnte lediglich für die Gruppe der guten Assoziierer wiedergefunden werden (Ruprecht, 2004).

	gute Assoziierer (N=5)	schlechte Assoziierer (N=5)
<i>M</i>	8,56	2,85
<i>SD</i>	2,67	0,47
<i>Minimum</i>	6,97	2,23
<i>Maximum</i>	13,30	3,43

Abbildung 45: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wortassoziationsversuchs in den Extremgruppen (gem. Assoziationskennwert) in Studie 17 – Call Center(B)

4.15. Zusammenfassung und erstes Fazit

Die soeben vorgestellten Studien geben einen Überblick über das breite Spektrum an Einsatzfeldern für den Wortassoziationsversuch. In vielen dieser Untersuchungen war das Fachwissen von Beschäftigten zentraler Bestandteil der Fragestellung: so zielten beispielsweise einige Studien auf den Zusammenhang von Wissen und Arbeitsbelastung oder Arbeitsaufgabe ab, andere auf die Evaluation von Weiterbildungsveranstaltungen oder auf Analysen im Rahmen eines Wissensmanagementprozesses. Es konnte gezeigt werden, daß der Wortassoziationsversuch bei der Beantwortung solcher Fragestellungen ein nützliches Instrument darstellte, welches sich gut in das Untersuchungsdesign integrieren ließ. Dabei wurde auch die Breite verschiedener Einsatzfelder deutlich: bei Studierenden oder Akademikern wurde das Verfahren ebenso eingesetzt wie bei Berufsschülern, Beschäftigten im Dienstleistungsbereich sowie mit kaufmännischen oder produzierenden Arbeitstätigkeiten.

In engem Zusammenhang mit Fragestellung und Einsatzfeld ist auch die Wahl einer zu untersuchenden Wissensdomäne und die Ableitung geeigneter Stimuli zu sehen. Auch hier präsentiert der Wortassoziationsversuch ein breites Einsatzspektrum: er wurde beispielsweise zur Erfassung psychologischer Domänen eingesetzt, auf dem Gebiet der Internetkommunikation oder für Wissensdomänen aus dem handwerklichen Bereich. Der Umfang der eingesetzten Stimulusliste konnte dabei den Erfordernissen der Wissensdomäne sowie des Untersuchungsfeldes flexibel angepaßt werden. Zwischen einem bis hin zu 40 eingesetzten Stimuli kamen in den hier vorgestellten Studien zum Einsatz, der Median liegt bei 10 Stimuli, durchschnittlich umfaßte der Wortassoziationsversuch $M=13,41$ ($SD=12,31$) Stimuli.

Es wurde auch deutlich, daß – beim Einsatz des Wortassoziationsversuchs im Feld – die intensive Auseinandersetzung des Versuchsleiters mit dem Arbeitsgegenstand und Arbeitsinhalten unabdingbare Voraussetzung ist. Nur so läßt sich die Relevanz der Stimuli beurteilen und die Bedeutung der von den Probanden geäußerten Assoziationen begreifen. Als sehr wirkungsvoll (wenngleich zeitintensiv) haben sich dabei Beobachtungsinterviews durch die Versuchsleiter erwiesen, wie sie beispielsweise in Studie 16 oder 17 in den Call Centern systematisch eingesetzt wurden, die allerdings die Einbeziehung eines betrieblichen Experten nicht ersetzen können.

Die Assoziationskennwerte variieren sehr stark über die einzelnen Studien hinweg. Während Studie 15 (Wissen über Arbeitsschutz) mit $M=2,76$ ($SD=1,56$) Assoziationen pro Stimulus den geringsten Kennwert der hier vorgestellten Erhebungen aufweist, markiert Studie 2 (psychotherapeutische Ambulanz) mit einer durchschnittlichen Assoziationsleistung von $M=21,70$ ($SD=16,67$) beim ersten Erhebungszeitpunkt den höchsten hier berichteten Assoziationskennwert. Die Höhe des Kennwertes ist dabei immer in Zusammenhang mit dem Untersuchungsfeld zu sehen: verschiedene Wissensdomänen haben unterschiedlich starkes Potential für das Produzieren von Assoziationen, Begriffe können einen unterschiedlich großen Bedeutungshof haben. Es konnte gezeigt werden, daß die Assoziationsleistung in den jeweiligen Stichproben meist hinreichend breit gestreut waren, um zwischen Extremgruppen hinsichtlich des Assoziationskennwertes unterscheiden zu können. Die jeweils guten bzw. schlechten Assoziierer einer jeden Studie waren deutlich voneinander abzugrenzen. Selbst in Studie 15 (Wissen über Arbeitsschutz), in der a priori die Annahme bestand, daß die Probanden nicht über ein angemessenes Fachwissen verfügen, war dennoch diese deutliche Differenzierung in Leistungsgruppen möglich.

Die hier vorgestellten Studien illustrieren die breite Einsetzbarkeit des Wortassoziationsversuchs, sowohl in Hinblick auf das Untersuchungsfeld als auch hinsichtlich der Wissensdomänen. Dabei konnte die Länge der Stimulusliste sowie deren Präsentationsform den Erfordernissen der jeweiligen Fragestellungen bzw. Untersuchungsfelder angepaßt werden.

5. Vorstudien zur Durchführung des Wortassoziationsversuchs

Um die Berechtigung der nachfolgend zusammengetragenen empirischen Auseinandersetzungen abzusichern, soll zunächst die Betrachtung zweier Aspekte der Durchführung des Wortassoziationsversuchs vorangestellt werden (siehe Abbildung 46). Da – wie bereits in Auseinandersetzung mit der einschlägigen Literatur erörtert (siehe Kapitel 3.5 auf Seite 36 ff.) –

bisher ungeklärt ist, ob für das freie fortgesetzte Assoziieren innerhalb einer fachspezifischen Wissensdomäne von der Äquivalenz einer mündlichen oder aber schriftlichen Durchführungsform des Wortassoziationsversuchs ausgegangen werden kann, wurde im Rahmen einer Vorstudie dieser Frage nachgegangen (Fragestellung 1): Studie 1 wurde explizit zur Abklärung des Durchführungsmodus konzipiert. Dabei wird der schriftliche Modus dem mündlichen gegenübergestellt. Da jeder Proband sowohl die mündliche als auch die schriftliche Form bearbeitete, ist ein intraindividueller Vergleich der Darbietungsformen möglich.

Nr.	Name der Studie	Durchführungsmodus	Anordnung der Stimuli
1	Modus	x	
2	psychotherapeutische Ambulanz		x

Abbildung 46: Übersicht über alle Studien, die einen empirischen Beitrag hinsichtlich der Durchführung des Wortassoziationsversuchs leisten

Der zweite hier thematisierte Aspekt betrifft die Anordnung der Stimuli bei der Konstruktion eines Wortassoziationsversuchs. Es liegen in der Literatur keinerlei Hinweise darauf vor, in welcher Weise Stimuli angeordnet sein sollten. Insbesondere die Frage, ob dem Probanden zunächst ein Oberbegriff der interessierenden Domäne und dann im weiteren Verlauf des Versuchs immer spezifischere Stimuli vorgelegt werden sollten, oder eher umgekehrt, stellt sich dem Praktiker oft. Aus diesem Grunde wurde dieser Frage in Studie 2 – psychotherapeutische Ambulanz – mittels eines experimentellen Designs im Feld nachgegangen (Fragestellung 2).

5.1. Durchführungsmodus des Wortassoziationsversuchs in Studie 1: Modus

Wie bereits beschrieben, wurde Studie 1 – Modus der Erhebung – explizit konzipiert, um im intraindividuellen Vergleich den möglichen Einfluß einer mündlichen vs. einer schriftlichen Durchführung des Wortassoziationsversuchs auf die Assoziationsleistung des Probanden zu überprüfen. Dabei setzen sich die insgesamt 40 Stimuli gleichmäßig aus Begriffen des Allgemeinwissens sowie eines Spezialwissensbereichs (Bankgeschäfte) zusammen, welche jeweils in beiden Durchführungsvarianten bearbeitet wurden. Alle Probanden bearbeiten die Stimuli in identischer Reihenfolge, dabei wurde zufällig variiert, mit welcher der Darbietungsformen begonnen wurde (siehe dazu auch Versuchsablauf in Abbildung 4, Seite 55).

Es zeigten sich keine signifikanten Unterschiede (Mann-Whitney-U: $z=-0,332$; $p>.05$; $d=.18$) hinsichtlich der beiden Versuchsaufbau-Gruppen: die Gruppe der Probanden, welche zuerst den mündlichen und anschließend den schriftlichen Teil bearbeiteten (Form A; N=15), assoziierten im Mittel 5,87 ($SD=2,64$) Begriffe, die Gruppe mit umgekehrter Reihenfolge der Durchführungsmodi (Form B; N=15) assoziierte $M=6,42$ ($SD=3,56$) Begriffe. Es spielt also keine Rolle, in welcher Reihenfolge die Modi eingesetzt wurden, so daß diese Differenzierung in Teilstichproben in den weiteren Ausführungen ignoriert werden kann und sich alle weiteren Angaben auf N=30 beziehen.

Abbildung 47 listet die Assoziationskennwerte auf, welche sich ergeben, wenn lediglich die mündlich bzw. die schriftlich bearbeiteten Stimuli berücksichtigt werden (jeweils 20 Stimuli). Außerdem differenziert diese Darstellung innerhalb eines Durchführungsmodus hinsichtlich des

Wissensbereichs (Allgemeinwissen vs. Spezialwissen), der Assoziationskennwert basiert in diesen Subkategorien also auf jeweils 10 Stimuli. Es wird deutlich, daß insgesamt im schriftlichen Teil des Wortassoziationsversuchs mit $M=6,76$ ($SD=4,17$) etwas mehr assoziiert wird als im mündlichen ($M=5,51$; $SD=2,31$), dieser Unterschied ist statistisch signifikant (Wilcoxon: $z=-2,098$; $p=.036$), mit $d=.37$ wird allerdings nur ein Effekt geringer Stärke beschrieben. Bei genauerer Betrachtung wird jedoch deutlich, daß dies vor allem auf den Einfluß der Allgemeinwissens-Stimuli zurückzuführen ist. Der im Vergleich zum Spezialwissen deutlich höhere Assoziationskennwert im Bereich des Allgemeinwissens war zu erwarten und bildet sich in beiden Verfahrensmodi ab. Auf die Stimuli des Allgemeinwissens wurden im mündlichen Teil im Mittel $7,83$ ($SD=3,62$) Begriffe assoziiert, im schriftlichen Teil sogar $M=9,73$ ($SD=6,17$). Die abgebildete Überlegenheit des schriftlichen Modus bei der Bearbeitung der Allgemeinwissens-Stimuli ist statistisch signifikant (Wilcoxon: $z=-2,274$; $p=.023$), mit $d=.38$ wird ein Effekt geringer Stärke beschrieben. Stellt man jedoch den mündlichen Teil des Spezialwissens dem schriftlichen Teil im selben Wissensbereich gegenüber, so unterscheiden sich die Assoziationsleistungen in beiden Teilen nicht signifikant voneinander (Wilcoxon: $z=-1,493$; $p>.05$), wenngleich in der Tendenz ein Vorsprung des schriftlichen Teils ($M=3,79$; $SD=2,62$) gegenüber dem mündlichen Teil ($M=3,20$; $SD=1,79$) zu erkennen ist (eine Effektstärke von $d=.26$ weist auf einen geringen Effekt hin).

	mündlicher Teil			schriftlicher Teil		
	gesamt	Allgemeinwissen	Spezialwissen	gesamt	Allgemeinwissen	Spezialwissen
<i>M</i>	5,51	7,83	3,20	6,76	9,73	3,79
<i>SD</i>	2,31	3,62	1,79	4,17	6,17	2,62
Minimum	2,30	3,40	0,80	2,35	2,60	1,40
Maximum	13,40	21,00	7,40	19,75	28,80	10,70

Abbildung 47: Kennwerte des Wortassoziationsversuchs – Mittelwerte (*M*) und Standardabweichungen (*SD*) – nach Durchführungsmodus bzw. Wissensbereich in Studie 1 – Modus der Erhebung

Der Assoziationskennwert im mündlichen Teil des Spezialwissens läßt sich sehr zuverlässig durch den Assoziationskennwert beim schriftlichen Spezialwissen vorhersagen ($r=.734$; $p=.000$), gleiches trifft auf die Rangreihe der Probanden zu, welche in der Rangkorrelation nach Spearman durch einen ebenfalls hohen signifikanten Korrelationskoeffizienten gekennzeichnet ist ($r=.757$; $p=.000$).

Zusammenfassend ist festzustellen, daß für Stimuli aus dem Bereich des Allgemeinwissens die schriftliche Bearbeitung des Wortassoziationsversuchs zu signifikant höheren Assoziationsleistungen führt. Im Bereich des Spezialwissens hingegen konnte dieser Effekt nicht nachgewiesen werden. Es kann also davon ausgegangen werden, daß es bei der Präsentation von Stimuli aus dem Spezial- oder auch Fachwissensbereich i.e.S. keine Rolle mehr spielt, ob der Wortassoziationsversuch schriftlich (Paper-Pencil) oder mündlich durchgeführt wird. Somit kann für die in der vorliegenden Arbeit herangezogenen empirischen Studien die Äquivalenz der mündlichen bzw. schriftlichen Durchführungsform des Wortassoziationsversuchs angenommen werden.

5.2. Anordnung von Stimuli verschiedener Abstraktionsebenen in Studie 2: psychotherapeutische Ambulanz

Bei der Erstellung einer Stimulusliste für den Wortassoziationsversuch stellt sich in der Praxis oft die Frage, wie abstrakt oder konkret Stimuli sinnvollerweise formuliert bzw. in welcher Reihenfolge diese unterschiedlichen Abstraktionsebenen präsentiert werden sollten. Die einschlägige Literatur bietet zu dieser Frage keine Hinweise. Pokrandt (2004) merkt an, daß in ihrer Studie vermutlich einige Stimuli des Wortassoziationsversuchs zu allgemein gehalten waren, da einige Probanden daraufhin ihr Wissen über den gesamten Produktionsprozeß abriefen. Dadurch ergaben sich v.a. bei den sog. guten Assoziierern z.T. Wiederholungen von Äußerungen auf verschiedene Stimuli hin, wobei angemerkt werden muß, daß dies ja nicht zu verurteilen ist. Vielmehr lassen sich anhand dessen Vermutungen über die Vernetzung, Struktur oder die Größe des individuellen Suchraums aufstellen. Die Erfahrungen im Feld zeigen, daß Probanden nicht selten auf ihre Assoziationen zu bereits bearbeiteten Stimuli verweisen und diese Wissensinhalte dann bewußt nicht wiederholen. Solch ein Verhalten erwartet der Versuchsleiter intuitiv v.a. dann, wenn zuerst ein sehr konkreter Stimulus präsentiert wird und anschließend ein Stimulus einer übergeordneten Abstraktionsebene, welche diesen konkreten Term und die Menge seiner Konzepte subsumiert. Sind die Stimuli in umgekehrter Reihenfolge angeordnet, erwartet man diesen Effekt weniger, da zunächst allgemeinere Assoziationen erfolgen und anschließend die Spezifika des konkreten Stimulus herausgegriffen werden können, sozusagen die zusätzliche Menge an Konzepten, die sich von der Menge der Oberbegriffskonzepte abhebt. Systematische Erkenntnisse zu dieser Problematik liegen jedoch nicht vor. Aus diesem Grunde wurde die Anordnung von Stimuli verschiedener Abstraktionsebenen in einem Feldexperiment systematisch variiert, welches im Folgenden näher beschrieben werden soll.

Zur Klärung der Frage nach solchen Positionseffekten von Stimuli verschiedener Abstraktionsebenen wurde eine Domäne gesucht, in der die Gliederung der Begriffsebenen klar definiert ist. Dazu bot sich das Internationale Klassifikationssystem psychischer Störungen nach ICD-10 (Weltgesundheitsorganisation, 1999) an. In diesem Klassifikationssystem werden einzelne Störungsbilder differenziert anhand ihrer diagnostischen Leitlinien beschrieben und aufgrund inhaltlicher Merkmale zu einer übergeordneten Kategorie zusammengefaßt, welche ihrerseits wieder mit anderen Störungsgruppen in einer weiteren übergeordneten Kategorie vereint werden. Die Zugehörigkeit zu einer Kategorie ist darüber hinaus codiert, die Anzahl der Zeichen im Code weist auf die Abstraktionsebene hin. So beschreibt beispielsweise „F“ die Gesamtheit der psychischen Störungen, „F1“ beschreibt die Gruppe der psychischen und Verhaltensstörungen durch psychotrope Substanzen, „F10“ definiert solche Störungen genauer als durch Alkohol verursacht, die vierte und fünfte Stelle beschreibt noch näher das klinische Zustandsbild, so daß „F11.0“ eine akute Intoxikation bezeichnet, und „F11.00“ teilt mit, daß diese Intoxikation ohne Komplikationen verlief. Somit wird klar, daß Inhalte einer höheren Abstraktionsebene sich immer zusammensetzen aus der Summe der Teilmengen der Inhalte zugehöriger Unterbegriffe plus ggf. einer Anzahl Begriffe, die in gleicher Weise für alle Unterkategorien gelten.

Eine Personengruppe, die über eingehende Kenntnisse in dieser Wissensdomäne verfügt, sind Psychologen oder Psychotherapeuten. Daher bot sich an, das in einer psychotherapeutischen Ambulanz durchgeführte Feldexperiment (Studie 2, vgl. Kapitel 4.2, Seite 55 ff.) zu erweitern um

die Fragestellung der Stimuli-Abstraktionsebenen und deren Anordnung. Für diese Untersuchung wurden wie bereits beschrieben zwei Abschnitte der ICD ausgewählt: Entwicklungsstörungen (F8) sowie Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend (F9). Diese beiden Oberbegriffe der Störungsgruppen wurden als Stimuli eingesetzt sowie jeweils drei Unterbegriffe der nächsten Ebene (z.B. F91 Störungen des Sozialverhaltens) sowie der übernächsten Ebene (z.B. F91.2 Störungen des Sozialverhaltens bei vorhandenen sozialen Bindungen), die sich daraus ergebenden 14 Stimuli sind in Abbildung 8 (Seite 55) aufgeführt.

Die Anordnung dieser Stimuli erfolgte anhand des in Abbildung 48 dargestellten Schemas: zunächst wurde Thema F8 gemäß der Strategie „Vom Allgemeinen zum Besonderen“ dargeboten (deduktive Strategie, d.h. zuerst der Oberbegriff des Abschnitts, danach drei Begriffe aus der nächst niedrigeren Kategorie, und weitere drei Begriffe aus der abermals niedrigeren Kategorienebene). Anschließend wurde das zweite Thema (F9) gemäß der Strategie „Vom Besonderen zum Allgemeinen“ präsentiert, also in umgekehrter Reihenfolge der Abstraktionsebenen (induktive Strategie).

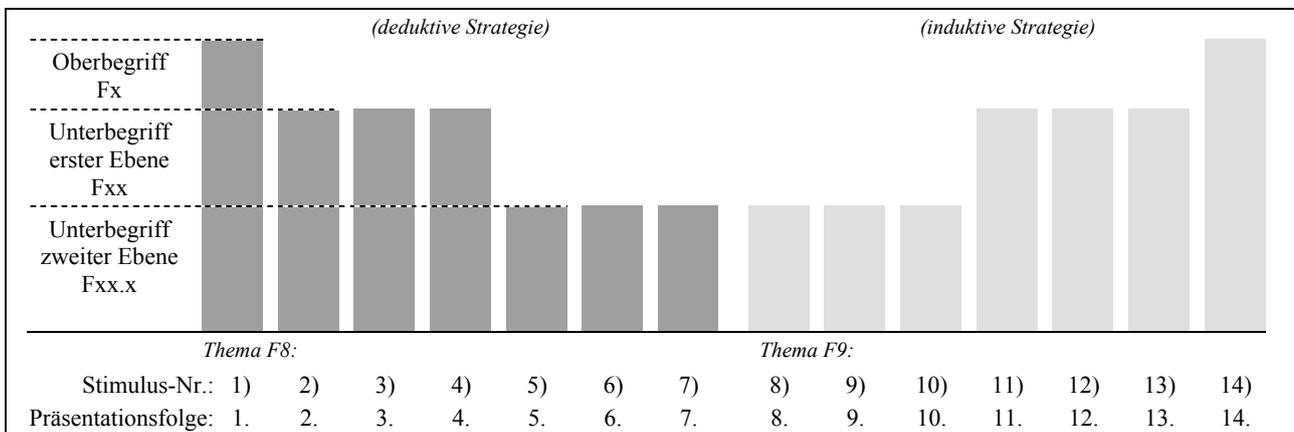


Abbildung 48: Ablaufschema des Wortassoziationsversuchs in Studie 2 (psychotherapeutische Ambulanz) gem. der Stimuluskategorien – Form A

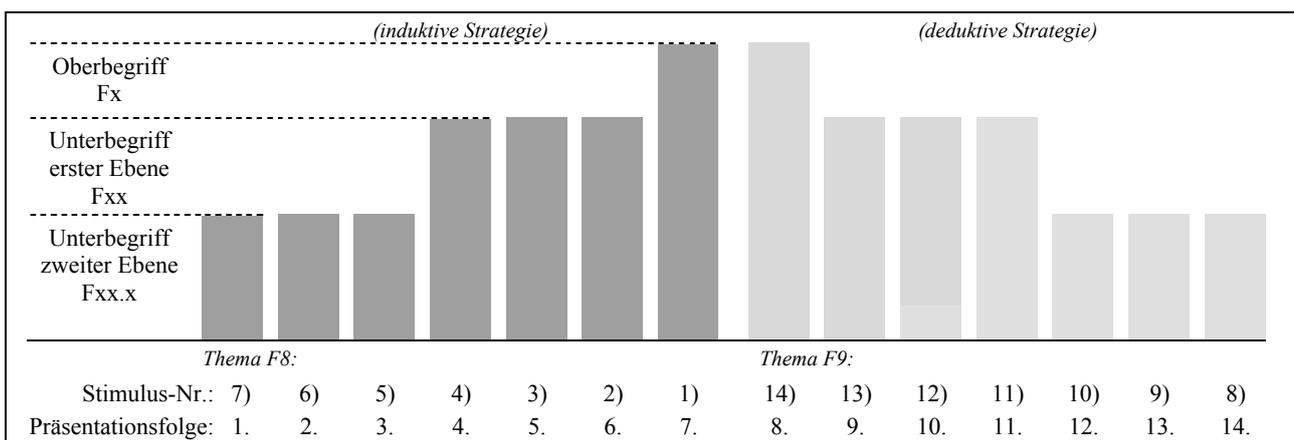


Abbildung 49: Ablaufschema des Wortassoziationsversuchs in Studie 2 (psychotherapeutische Ambulanz) gem. der Stimuluskategorien – Form B

Es wurden zwei Themen gewählt, um einen intraindividuellen Vergleich zu erlangen über die induktive vs. deduktive Strategie. Da aber bereits in anderen Studien beobachtet wurde, daß im Laufe des Wortassoziationsversuchs die Anzahl der Nennungen tendenziell abnimmt (v.a. bei sehr langen Stimuluslisten), kann durch die soeben beschriebene Anordnung der Stimuli nicht sicher

bestimmt werden, ob die Effekte allein durch den Inhalt der Themen oder durch ihre Stellung im Wortassoziationsversuch zustande kommen. Deshalb wurde die Untersuchung insofern ausbalanciert, als daß das Stimulusmaterial in zwei Formen präsentiert wurde. Der einen Hälfte der Probanden wurden die Stimuli zu beiden Testzeitpunkten in der oben beschriebenen Reihenfolge vorgelegt (Form A, siehe Abbildung 48). Die andere Hälfte der Probanden bearbeitete zu beiden Testzeitpunkten die in Abbildung 49 dargestellte Variante (Form B). Dabei wurde die Reihenfolge der Themen beibehalten, die Reihenfolge der Stimuli innerhalb eines Themas dabei aber umgekehrt, so daß zuerst Thema F8 nach der Strategie „vom Besonderen zum Allgemeinen“ und anschließend Thema F9 nach der Strategie „vom Allgemeinen zum Besonderen“ bearbeitet wird.

Die Zuweisung der Probanden zu den beiden Gruppen (Form A und B) erfolgte zufällig, wobei darauf geachtet wurde, beide Gruppen hinsichtlich der Zusammensetzung bzgl. berufsbezogener Merkmale auszubalancieren. So war dann die Zusammensetzung der Probanden in den beiden Teilstichproben annähernd gleich: in beiden Gruppen waren jeweils drei Psychologie-Studierende und ein Diplompsychologe, in Gruppe A fanden sich vier Ausbildungs- und zwei bereits approbierte Therapeuten wieder, in der Stichprobe der Form B waren diese beiden Personengruppen durch jeweils drei Probanden vertreten.

Die beiden Versuchsaufbauvarianten A und B sollen nun hinsichtlich ihres Assoziationskennwertes verglichen werden. Aus Gründen der Übersichtlichkeit wird dies hier graphisch veranschaulicht, die statistischen Kennwerte sind in Anhang 7 bzw. Anhang 8 aufgelistet. Wie in Abbildung 50 ersichtlich, assoziierte Gruppe A zum ersten Erhebungszeitpunkt im Mittel 19,77 Konzepte pro Stimulus ($SD=11,90$), während Gruppe B einen mittleren Assoziationskennwert von 14,59 erbringt ($SD=6,21$), dieser Unterschied ist aber nicht statistisch signifikant (Mann-Whitney-U: $z=-1,323$; $p>.05$), weist aber dennoch mit $d=.55$ auf einen starken Effekt hin. Da in diesem Kennwert sowohl der deduktive als auch der induktive Teil des Wortassoziationsversuchs subsumiert sind, sollen nun die beiden Themen F8 und F9 separat betrachtet werden. Dies geschieht zunächst bezogen auf alle sieben Stimuli eines Themas, im weiteren wird differenziert nach den einzelnen Stimulusebenen: Oberbegriff des Themas (jeweils ein Stimulus), Unterbegriffe der ersten Ebene (jeweils drei Stimuli; hier aus Gründen der sprachlichen Vereinfachung Zwischenbegriff genannt), Unterbegriffe der zweiten Ebene (jeweils drei Stimuli; hier aus Gründen der sprachlichen Vereinfachung Unterbegriff genannt).

Trotz augenscheinlicher Differenzen zwischen den Kennwerten beider Gruppen werden lediglich zwei von ihnen signifikant: bei den sog. Zwischenbegriffen zum Thema F8 wurde in der Gruppe A in der deduktiven Anordnung signifikant mehr ($M=20,00$; $SD=11,72$) assoziiert als in Gruppe B ($M=11,97$; $SD=8,83$), welche diese Stimuli innerhalb einer induktiven Anordnung vorgelegt bekam (Mann-Whitney-U: $z=-2,310$; $p=.019$; $d=.77$). Der zweite signifikante Unterschied betrifft die Assoziationsleistung zum Oberbegriff von Thema F9: hier assoziierten die Probanden der Gruppe A, welche diesen Stimulus am Ende des gesamten Wortassoziationsversuchs präsentiert bekam, mit $M=11,50$ ($SD=4,06$) signifikant weniger (Mann-Whitney-U: $z=-2,124$; $p=.035$; $d=1.12$) als die der Gruppe B ($M=21,80$; $SD=12,39$), bei denen dieser Stimulus in der Mitte des Versuchs positioniert war. Betrachtet man im Vergleich dazu die Befunde zum zweiten Erhebungszeitpunkt in Abbildung 51, so gleichen sich die Assoziationsleistungen der Probanden beider Gruppen noch stärker an, es sind keine signifikanten Unterschiede mehr zwischen ihnen auszumachen, die Effektstärken

reduzieren sich weiter und lassen keinerlei Unterschiede annehmen. Aus diesem Grunde sind auch die soeben berichteten zwei Unterschiede vernachlässigbar. Für die Ausprägung des Assoziationskennwertes ist es also unerheblich, ob die Reihenfolge der Stimuli einer induktiven oder einer deduktiven Anordnung entspricht.

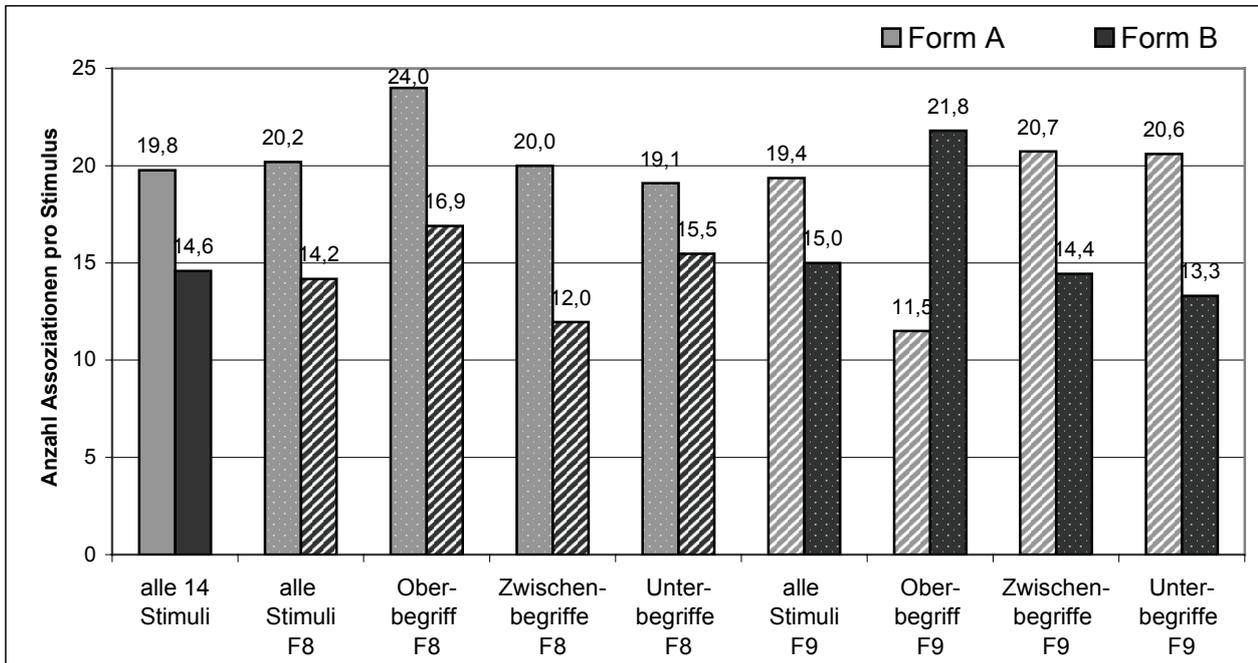


Abbildung 50: Assoziationsleistung im Vergleich der beiden Varianten des Aufbaus im Wortassoziationsversuch zum ersten Erhebungszeitpunkt in Studie 2 – psychotherapeutische Ambulanz, differenziert nach Themen sowie Stimuluskategorien (gepunktet = deduktiver Aufbau, gestreift = induktiver Aufbau)

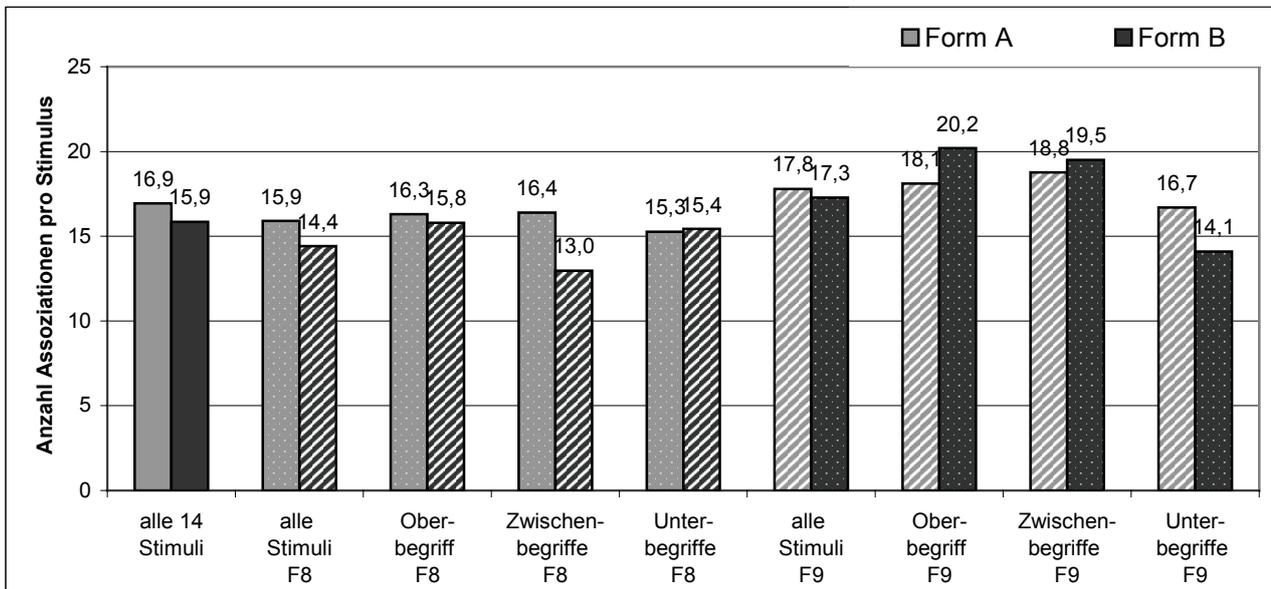


Abbildung 51: Assoziationsleistung im Vergleich der beiden Varianten des Aufbaus im Wortassoziationsversuch zum zweiten Erhebungszeitpunkt in Studie 2 – psychotherapeutische Ambulanz, differenziert nach Themen sowie Stimuluskategorien (gepunktet = deduktiver Aufbau, gestreift = induktiver Aufbau)

5.3. Zusammenfassung der Befunde aus den Vorstudien zur Durchführung des Wortassoziationsversuchs

Hinsichtlich der Durchführung des Wortassoziationsversuchs war zunächst zu klären, ob die Form des Antwortformates (mündlich vs. schriftlich) einen Einfluß auf die Assoziationsleistung der Probanden hat (Fragestellung 1). Aufgrund der empirischen Befunde dazu ist folgendes festzuhalten: gehen die Stimuli über den Bereich des Allgemeinwissens hinaus (in der vorgestellten Studie 1 als sog. Spezialwissen bezeichnet), so führen mündlicher bzw. schriftlicher Verfahrensmodus zu vergleichbaren Assoziationsleistungen (hier nachgewiesen im intraindividuellen Vergleich). Die hier vorgestellten empirischen Studien, können also hinsichtlich ihrer Assoziationsergebnisse miteinander in Beziehung gesetzt werden, unabhängig davon, in welchem Durchführungsmodus die Probanden assoziierten.

Da also der Durchführungsmodus keinen systematischen Einfluß auf die Assoziationsleistung hat, können andere Aspekte bei der Wahl des Untersuchungsdesigns in den Vordergrund gerückt werden. Zu bedenken sind dabei v.a. die unterschiedlichen Konsequenzen der verschiedenen Modi in Hinblick auf die Durchführungs- und Auswerteökonomie, wonach einem Versuchsleiter mit Blick auf die Auswertung vom mündlichen Modus doch eher abzuraten ist (bei gleichzeitig günstiger Durchführungsökonomie), während der schriftliche Einsatz mehr Zeit in der Durchführung benötigt, aber auch als Gruppentest eingesetzt werden kann und daher einen Vorteil in der Durchführungs- und zudem in der Auswerteökonomie erbringen kann. Wie in der Übersicht der hier vorgestellten Studien ersichtlich (Abbildung 3, auf Seite 55), wurde der Wortassoziationsversuch überwiegend im schriftlichen Format bearbeitet (in 12 Studien). In diesen Studien lagen pragmatische Gründe für die Erhebung des Assoziierens in schriftlicher Form vor. So war es beispielsweise im Rahmen der Lehrveranstaltungsevaluation (Studien 3, 4, und 14) relativ leicht zu realisieren, alle Probanden zeitgleich den Wortassoziationsversuch bearbeiten zu lassen, ähnliches gilt für Studien in Verbindungen mit Prüfungsklausuren (Studien 6 bis 11), die nur realisiert werden konnten, indem alle Probanden zeitgleich einbezogen wurden. Somit kam einzig die schriftliche Verfahrensform in Betracht. In den Feldstudien 12 und 13 sowie 16 und 17 hingegen wurde der Wortassoziationsversuch mündlich durchgeführt, wofür es auch hier pragmatische Gründe gab: da die Erhebungen während des laufenden Betriebs stattfanden, war es unrealistisch, alle Probanden gleichzeitig aus dem Arbeitsprozeß herausziehen zu wollen. Da in Einzelsitzungen auch die mündliche Durchführungsform möglich ist, wurde sich für diese entschieden. Dabei war v.a. die kürzere Bearbeitungszeit der mündlichen gegenüber der schriftlichen Durchführung bedeutsam, um die Arbeitshandlungen nicht zu lange zu unterbrechen.

Der zweite Aspekt, der hinsichtlich der Durchführung des Wortassoziationsversuchs empirisch abgeklärt werden sollte, ist die Anordnung von Stimuli, welche unterschiedlich konkret sind und damit einen unterschiedlich großen Bedeutungshof aufweisen (Fragestellung 2). In einem Feldexperiment wurden die Stimuli, welche in Ober- bzw. Unterbegriffsrelation zueinander standen, systematisch variiert in Hinblick auf ihre Anordnung innerhalb des Wortassoziationsversuchs. Auf diese Weise wurde den Probanden sowohl eine induktive als auch eine deduktive Anordnung von Stimuli präsentiert, welche der Domäne der Klassifikation psychischer Störungen zugehörten. Es konnten keine bedeutsamen Unterschiede hinsichtlich des Assoziationskennwertes zwischen den beiden Experimentalgruppen ausgemacht werden, die

deduktive Anordnung von Stimuli eines Themenbereichs führte im interindividuellen Vergleich zu vergleichbaren Assoziationsleistungen wie eine induktive Anordnung der gleichen Stimuli.

Zusammenfassend läßt sich also sagen, daß die deduktive vs. induktive Anordnung von Stimuli verschiedener Abstraktionsebenen keinen Einfluß auf die Anzahl assoziierter Konzepte hat. Somit können alle hier vorgestellten empirischen Studien als vergleichbar angesehen werden, unabhängig vom Design der jeweiligen Stimulusliste. Aufgrund von unsystematischen Beobachtungen im Feld ist jedoch ein deduktiver Aufbau des Wortassoziationsversuchs eher zu empfehlen, da es den Empfindungen der Probanden entspricht (Einstieg in die Domäne mit einem globaleren Stimulus, weiteres Vertiefen mit spezialisierteren Stimuli) und ein induktives Vorgehen bisweilen als Wiederholung empfunden wird. Allerdings ist es als eher selten zu beschreiben, daß ein Wortassoziationsversuch komplett aus Stimuli besteht, welche sich in eine klare hierarchische Anordnung bringen lassen – in der Praxis trifft dies i.d.R. für eine Teilmenge der Stimuli eines Wortassoziationsversuchs zu, während die meisten Stimuli in einer Nebenbegriffsrelation zueinander stehen. Ein in dieser Weise gemischter Satz an Stimuli ist nicht nur für den Probanden abwechslungsreicher, sondern geht auch mit einer für die Domäne repräsentativeren Auswahl von Stimuli einher und erhöht so die Inhaltsvalidität des Versuchs.

6. Hauptgütekriterien

Das folgende Kapitel widmet sich der Bestimmung der Hauptgütekriterien. Dabei soll wie folgt vorgegangen werden: nach einem einführenden Überblick über die für das jeweilige Gütekriterium relevanten Studien werden die Ergebnisse der einzelnen Studien separat vorgestellt. Dazu werden an dieser Stelle auch die deskriptiven Statistiken von Kriterien, anderen Verfahren oder ähnlichem berichtet. Die deskriptiven Statistiken der Wortassoziationsversuche wurden bereits bei der Vorstellung der Studien in Kapitel 4 (Seite 55 ff.) dargestellt, in Weiterführung dessen beziehen sich auch in den folgenden Ausführungen alle Berechnungen ausschließlich auf den dort verwendeten Assoziationskennwert, also die einfache Auszählung aller von den Probanden genannten Assoziationen pro Stimulus (auf mögliche Modifikationen des Assoziationskennwert durch Gewichtungen o.ä. wird später in diesem Kapitel eingegangen werden, siehe Kapitel 6.1.4, Seite 55 ff.).

Insgesamt können 15 Studien herangezogen werden, welche einen empirischen Beitrag zur Bestimmung der Hauptgütekriterien leisten, Abbildung 52 gibt einen Überblick darüber. Es wird deutlich, daß die Frage der Validität im Zentrum der empirischen Befunde steht. Viele der hier vorgestellten Studien können dabei die Außenkriterien Klausurleistung bzw. Geschäftsführerurteil einbringen, so daß diese Erhebungen für die Bestimmung der Kriteriumsvalidität nutzbar sind. Für die Bestimmung der konvergenten Konstruktvalidität können Studien mit anderen wissensdiagnostischen Verfahren herangezogen werden, für die diskriminante Konstruktvalidität wurde der Einfluß des Personenmerkmals der generellen Wortflüssigkeit überprüft. Für die Bestimmung der Reliabilität kann nur eine einzige Studie herangezogen werden. Dazu wurde eine Erhebung mit zwei Meßzeitpunkten zur Bestimmung der Retest-Reliabilität durchgeführt, an späterer Stelle wird dies ausführlicher erläutert werden (siehe Kapitel 6.2 auf Seite 55 ff.). Die

Fragen der Objektivität werden nicht auf empirischer Basis beantwortet, vielmehr werden die Aspekte der Durchführungs-, Auswertungs- und Interpretationsobjektivität ausdiskutiert werden.

Nr.	Name der Studie	Kriteriumsvalidität		Konstruktvalidität		Retest-Reliabilität
		Klausurleistung	Geschäftsführerurteil	anderer Wissenstest	Wortflüssigkeit	
1	Modus				x	
2	psychotherapeutische Ambulanz			x (Lückentext)	x	x
3	Sparkassenakademie (A)	x				
4	Sparkassenakademie (B)	x				
5	Biopsychologie	x		x (Strukturlegetechnik; Assoziations-Struktur-Test)	x	
6	Klausur Klinische Psychologie	x				
7	Klausur Arbeitspsychologie	x				
8	Klausur Personalpsychologie	x				
9	Klausur Entwicklungspsychologie	x				
10	Gebäudereiniger	x		x (Assoziations-Struktur-Test)		
11	Bauklempner	x		x (Assoziations-Struktur-Test)		
12	Handelsunternehmen		x	x (Tiefeninterview, Strukturlegetechnik)		
13	Stoffdruckerei		x	x (Strukturlegetechnik)		
14	Seminar Wissensmanagement			x (Multiple-Choice-Test, Assoziations-Struktur-Test)		
15	Wissen über Arbeitsschutz			x (Multiple-Choice-Test)		

Abbildung 52: Übersicht über alle Studien, die einen empirischen Beitrag zur Bestimmung der Hauptgütekriterien leisten

Um die Bedeutsamkeit der gefundenen Unterschiede bzw. Zusammenhänge besser beurteilen zu können, sollen die Effektstärken bestimmt werden, da dieses Maß weitgehend unabhängig von der Stichprobengröße ist (Bühner, 2006; Sedlmeier und Renkewitz, 2007). Dabei ist die Effektstärke sowohl für die Unterschiedstestung in sehr großen Stichproben geeignet, bei denen Signifikanztests die u.U. sehr kleinen Unterschiede überschätzen (Bühner, 2006), aber auch für sehr kleine Stichproben, bei denen Signifikanztests möglicherweise bestehende Unterschiede übersehen (Bortz und Lienert, 2003; Sedlmeier und Renkewitz, 2007). Es geht also darum, wie repräsentativ ein Mittelwert tatsächlich für die Stichprobe ist. Da die vorliegende Arbeit meist auf relativ kleine Stichproben zurückgreifen muß, scheint der Parameter der Effektstärken daher besonders geeignet. Sedlmeier und Renkewitz (2007) unterscheiden die sog. Abstandsmaße d , welche für Gruppenunterschiede bestimmt werden können, sowie die sog. Zusammenhangsmaße r , welche den

korrelativen linearen Zusammenhang zwischen Variablen beschreiben. Dabei beruht diese Korrelation i.S. einer Effektstärke auf den z-transformierten Werten der jeweiligen Variablen. Zur Interpretation der Effektstärke d geben z.B. Bühner (2006) oder Sedlmeier und Renkewitz (2007) folgendes an: mit $d = .20$ liegt ein kleiner Effekt vor, ab $d = .50$ spricht man von einem mittleren Effekt, ab $d = .80$ von einem starken Effekt. Der Korrelationskoeffizient r als Maß für die Effektstärke kann nach Sedlmeier und Renkewitz (2007) ab $r = .1$ als kleiner Effekt, ab $r = .3$ als mittlerer Effekt und ab $r = .5$ als großer Effekt interpretiert werden.

Nach Darstellung der Befunde aus den einzelnen Studien sollen die Daten aus verschiedenen empirischen Erhebungen zusammengeführt werden, da dies erklärtes Ziel der vorliegenden Arbeit ist. Dazu werden die Daten von Studien mit gleichen Ansätzen, beispielsweise wenn sie das gleiche Kriterium zur Bestimmung der Validität nutzen, zusammengefaßt. Im Sinne einer Metaanalyse können die gefundenen Effektstärken der einzelnen Studien nach dem bei Sedlmeier und Renkewitz (2007) beschriebenen Vorgehen zusammengeführt werden zu einem gewichteten Mittelwert aller vorliegenden Effektgrößen. Da von allen Studien die Originaldaten vorlagen (bis auf eine Einschränkung, die sich aber auf die qualitativen Angaben der Assoziationsinhalte bezieht und daher hier unerheblich ist), konnten diese für die weitere Befundintegration herangezogen werden, nach Sedlmeier und Renkewitz (2007) ist dieses Vorgehen am effektivsten. Um dabei die Daten von Personen aus verschiedenen Studien miteinander vergleichen zu können, wurden für den Schritt der Befundintegration alle interessierenden Parameter in z-Werte transformiert, da nach Bortz (1993) dieser Vorgang individuelle Testwerte an der Unterschiedlichkeit innerhalb der jeweiligen Untersuchungsstichprobe relativiert. Der so ermittelte Korrelationskoeffizient r kann somit auch als Effektstärke interpretiert werden (Sedlmeier und Renkewitz, 2007).

6.1. Validität

Zur Beantwortung der Fragestellung 3 betrachtet dieses Kapitel die nach Häcker, Leutner und Amelang (1998) traditionellen Kategorien der Validität: die inhaltsbezogene, kriteriumsbezogene sowie Konstruktvalidität. Wie bereits angedeutet, wird die Frage der Kriteriumsvalidität sowie der Konstruktvalidität empirisch berücksichtigt. Abschließend wird die Frage thematisiert, inwieweit eine Modifikation des individuellen Assoziationskennwertes Auswirkungen auf diese Befunde bzw. auf die Beurteilung des Wissensstandes eines Probanden hat (Fragestellung 6a bzw. 6b). Vorangestellt werden soll zunächst eine kurze Erörterung der Frage der Inhaltsvalidität des Wortassoziationsversuchs.

Da a priori die Annahme bestand, daß ein hoher Assoziationskennwert auch mit einer guten Klausurleistung einhergeht und eine lineare Beziehung zwischen beiden Variablen besteht, wurden zur Signifikanzprüfung einseitige Tests eingesetzt. Gleiches gilt für die Prognose des Geschäftsführerurteils sowie für die Kennwerte der eingesetzten anderen wissensdiagnostischen Verfahren. Zur Validierung wird die Korrelation nach Pearson herangezogen (sofern nicht anders berichtet), zur Illustration sollen jeweils Extremgruppenvergleiche hinsichtlich der jeweiligen Kennwerte diese korrelative Beziehung erhellen. Die Einteilung in diese Extremgruppen basiert auf der Assoziationsleistung der Probanden im Wortassoziationsversuch (wie auch bereits in Kapitel 4 jeweils für die Studien berichtet wurde), zur sprachlichen Vereinfachung sollen – wie bisher

gehandhabt – die Probanden mit hohem Assoziationskennwert als gute Assoziierer bezeichnet werden, diejenigen mit geringem Assoziationskennwert hingegen als schlechte Assoziierer.

6.1.1. Inhaltsvalidität

Der Wortassoziationsversuch will mit Hilfe einer repräsentativen Auswahl an Stimuli das explizite, verbalisierbare Fachwissen eines Beschäftigten erfassen, wobei sich das Fachwissen auf eine ganz bestimmte Wissensdomäne eingrenzen läßt oder vielmehr: eingrenzen lassen muß. Die Erstellung eines Wortassoziationsversuchs für ein umschriebenes Untersuchungsfeld setzt zunächst immer eine eingehende Beschäftigung des Versuchsleiters mit der jeweiligen Wissensdomäne voraus, dies geschieht i.d.R. auf Grundlage von Dokumentenanalysen bzw. Experteninterviews. Durch diesen multiplen Zugang, insbesondere durch die aktive Einbeziehung des oder der Experten, kann von einer Repräsentativität der ausgewählten Stimuli aus einer Gesamtmenge von möglichen Stimuli der Wissensdomäne ausgegangen und somit von einer logischen Validität gesprochen werden. Lienert und Raatz (1998) billigen einem solchen Expertenrating grundsätzlich die Inhaltsvalidität zu, sie sprechen dabei von einem „Konsens der Kundigen“ (Lienert und Raatz, 1998, S. 11), auch Häcker, Leutner und Amelang (1998) beurteilen solche Expertenurteile als geeignet, um die Repräsentativität der Itemstichprobe zu gewährleisten. Gleichsam ist eine differentielle Inhaltsvalidität zu konstatieren, da durch das oben beschriebene Vorgehen ein Wortassoziationsversuch immer eingegrenzt ist auf eine bestimmte Untersuchungskohorte unter einer bestimmten Fragestellung. Insgesamt kann die Inhaltsvalidität des Wortassoziationsversuchs als gegeben angesehen werden.

6.1.2. Kriteriumsvalidität

Für die Überprüfung der Kriteriumsvalidität des Wortassoziationsversuchs wurde in den verschiedenen Studien der Ansatz der Übereinstimmungsvalidität oder auch konkurrenten Validität zugrunde gelegt, d.h. Testergebnis und Kriterium wurden zum gleichen Zeitpunkt erhoben. Dies ist insofern sinnvoll, als daß der Wortassoziationsversuch eine Statusmessung repräsentiert. Der Ansatz einer prognostischen Validität erscheint wenig erfolgversprechend, da man zwar eine gewisse Halbwertszeit beruflichen Wissens annehmen kann (siehe dazu auch die Ausführungen zur Reliabilität in Kapitel 3.7.1.2 auf Seite 44 f.), das Wissen von heute aber sicherlich nur zu geringem Teil eine zukünftige Performanz im Beruf determiniert, v.a. bei sehr langen Zeiträumen zwischen den Erhebungen (z.B. mehrere Jahre). Vielmehr ist zu erwarten, daß – auch mit Blick auf die gesellschaftliche Forderung nach „lebenslangem Lernen“ (z.B. Sonntag und Schaper, 2006) – der einer Leistung zugrundeliegende Wissensbesitz in der Zukunft (bei der Kriteriumsmessung) ein anderer ist als derjenige von heute (Prädiktormessung), da höchstwahrscheinlich Wissenserwerbs- oder mitunter auch Wissensverlustprozesse zwischen beiden Meßzeitpunkten anzunehmen sind.

Zur Klärung der Kriteriumsvalidität lassen sich – wie in Abbildung 52 ersichtlich (Seite 55) – zwei Kategorien von Kriterien benennen, an denen das Ergebnis des Wortassoziationsversuchs validiert wurde: zum einen die Leistungen in Klausuren mit Prüfungscharakter, welche das Wissen der Prüflinge zu einem umschriebenen Wissensgebiet erfassen. Nach der Klassifikation von Lienert und Raatz (1998) stellen sie ein einfaches, subjektives und quantitatives Validitätskriterium dar. Die zweite Kategorie von Validitätskriterien bildet das Geschäftsführerurteil, welches in zwei

Feldstudien in mittelständischen Unternehmen erhoben wurde. Dabei wird auf die Einschätzung des Vorgesetzten bzw. Geschäftsführers zurückgegriffen, welche dieser über die Arbeitsleistung und Fachkompetenz seiner Beschäftigten traf. Damit steht hier also nicht ein Maß für das Wissen selbst als Kriterium zur Verfügung, sondern ein Maß für die Arbeitsleistung, in welche neben dem Wissen auch noch andere Faktoren einfließen. Dieses Kriterium unterscheidet sich auch insofern von beispielsweise einer Klausurleistung, als daß hier nicht der Proband selbst die Variable produziert, sondern der Geschäftsführer, welcher das in Arbeitstätigkeiten realisierte Fachwissen (und andere Merkmale) seiner Mitarbeiter wahrgenommen und bewertet hat. Nach der Klassifikation von Validitätskriterien, wie sie Lienert und Raatz (1998) vorschlagen, wäre dieses Kriterium als quantitativ (es hat numerische Abstufungen), subjektiv (Einschätzung durch den Geschäftsführer) sowie kombiniert (mehrere Teilkriterien setzen sich zu Gesamtkriterium zusammen) zu beschreiben. Diese beiden hier herangezogenen Validierungskriterien (Klausurleistung bzw. Geschäftsführerurteil) zeichnen sich dadurch aus, daß die erhobenen Kriterien eine praktische Relevanz besitzen, sie sind sozusagen ökologisch valide, nach Amelang und Schmidt-Atzert (2006) können sie als „echte Kriterien“ bezeichnet werden.

6.1.2.1. Klausurleistung als Validitätskriterium

Im Folgenden werden neun Studien vorgestellt, in denen als Validitätskriterium die Leistung in einer Klausur herangezogen werden konnte (siehe auch Abbildung 52 auf Seite 55). Die Klausurleistung ist dabei als ein Maß für das Wissen zu verstehen, dessen angemessene Anwendung allerdings keine Rolle spielt. Nicht immer lagen gleichermaßen Daten zur Note als auch zu den erreichten Punkten vor, dennoch soll der jeweils vorhandenen Datenlage Rechnung getragen werden: sofern möglich, sollen daher neben der Zensur auch die in der Klausur erreichten Punkte betrachtet werden, da diese Skala differenzierter ist als die Notenskala. Im Anschluß werden die Befunde dieser Erhebungen zusammengeführt.

6.1.2.1.1. Klausurleistung in Studie 3: Sparkassenakademie (A)

Die Klausurleistung (Punkte) in Studie 3 – Sparkassenakademie (A) – ist normalverteilt, zwischen minimal 26 Punkten und maximal 60 Punkten wurden von den Untersuchungsteilnehmern durchschnittlich $M=39,50$ ($SD=8,48$) Punkte erreicht (siehe Abbildung 53). Für die Bestimmung der Kriteriumsvalidität wird der Assoziationskennwert des dritten Meßzeitpunktes als Prädiktor herangezogen, da zu diesem Zeitpunkt auch die Klausur (Kriterium) geschrieben wurde.

Der in der Klausur erreichte Punktwert läßt sich sehr zuverlässig auf Grundlage des Assoziationskennwertes desselben Meßzeitpunktes vorhersagen: mit $r= .856$ ($p= .000$) findet sich ein hoher und hoch signifikanter Korrelationskoeffizient. Dieser Zusammenhang zwischen Assoziationsleistung und Klausurleistung läßt sich auch durch den Extremgruppenvergleich illustrieren. Auf Grundlage der Assoziationsleistung zum dritten Meßzeitpunkt wurden zwei Extremgruppen gebildet, wie auch bereits in Kapitel 4.3 (Seite 55 ff.) beschrieben. In Abbildung 53 wird deutlich, daß sich die auf Grundlage der Assoziationsleistung gebildeten Extremgruppen auch hinsichtlich ihrer Klausurleistung unterscheiden. Während die guten Assoziierer $M=50,43$ Punkte ($SD=6,29$) erreichten, schnitten die schlechten Assoziierer mit $M=32,86$ Punkten ($SD=7,90$)

signifikant schwächer in der Klausur ab (Mann-Whitney- U: $z= 3,50$; $p= .004$). Dieser Befund erbringt eine Effektstärke von $d=2.46$, also einen außerordentlich bedeutsamen Effekt.

	gesamte Stichprobe (N=38)	Extremstichproben	
		gute Assoziierer (N=7)	schlechte Assoziierer (N=7)
<i>M</i>	39,50	50,43	32,86
<i>SD</i>	8,48	6,29	7,90
Minimum	26	42	26
Maximum	60	60	49

Abbildung 53: Kennwerte der Klausurleistung (Punkte) – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 3 – Sparkassenakademie (A)

6.1.2.1.2. Klausurleistung in Studie 4: Sparkassenakademie (B)

Die Klausurergebnisse in Form von Punkten sind in dieser zweiten Studie zur Sparkassenakademie innerhalb der Stichprobe normalverteilt. Zwischen minimal 52 und maximal 90 Punkten erreichten die Untersuchungsteilnehmer im Durchschnitt $M=73,81$ ($SD=11,00$) Punkte. Es werden auch hier diejenigen Assoziationsdaten für die Bestimmung der Kriteriumsvalidität herangezogen, welche zeitgleich mit der Klausur erhoben wurden (dies ist hier der dritte und letzte Meßzeitpunkt).

Der Zusammenhang zwischen Assoziationskennwert und Klausurleistung ist durch einen schwachen Korrelationskoeffizienten beschreibbar: die Zahl der in der Klausur erreichten Punkte läßt sich mit $r= .338$ ($p= .023$) signifikant durch die Anzahl der Assoziationen zum Zeitpunkt der Klausur vorhersagen. Die Rangkorrelation nach Spearman ergab sogar mit $r= .454$ ($p= .003$) einen Koeffizienten mittlerer Stärke: die Rangreihe der Probanden auf Grundlage des Assoziationskennwertes ist also ein guter Prädiktor für die Rangreihe hinsichtlich der Klausurleistung. Interessanterweise weist die Leistung in der Hausarbeit, welche ebenfalls bepunktet wurde, überhaupt keinen Zusammenhang zur Assoziationsleistung auf. Dies ist insofern nachvollziehbar, als daß die Klausur vorrangig auf die Abfrage und Anwendung von Faktenwissen abzielt (auch der Wortassoziationsversuch fragt Wissen ab), während bei der Hausarbeit die selbständige Bearbeitung eines komplexeren Problems im Vordergrund stand (also eine gänzlich andere Anforderung als im Wortassoziationsversuch), was auch i.S. einer diskriminanten Konstruktvalidierung verstanden werden kann.

Zur Illustration des korrelativen Zusammenhangs soll auch hier der Extremgruppenvergleich herangezogen werden. Auf Grundlage der Assoziationsleistung zum dritten Meßzeitpunkt wurden zwei Extremgruppen mit jeweils sechs Probanden gebildet. Die guten Assoziierer erreichten in der Klausur eine bessere Bewertung ($M=77,25$; $SD=6,85$) als die schlechten Assoziierer ($M=70,67$; $SD=14,88$). Dieser augenscheinliche Unterschied ist allerdings nicht signifikant, was vermutlich auf die relativ große Streuung der Klausurleistung innerhalb der Extremstichproben zurückzuführen ist, die Probanden decken einen ähnlichen Range ab (siehe auch Abbildung 54), die Effektstärke weist mit $d= .57$ aber immerhin auf einen Effekt mittlerer Stärke und somit eine gewisse praktische Relevanz hin.

	gesamte Stichprobe (N=35)	Extremstichproben	
		gute Assoziierer (N=6)	schlechte Assoziierer (N=6)
M	73,81	77,25	70,67
SD	11,00	6,85	14,88
Minimum	52	68	52
Maximum	90	85	88

Abbildung 54: Kennwerte der Klausurleistung (Punkte) – Mittelwerte (M) und Standardabweichungen (SD) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 4 – Sparkassenakademie (B)

6.1.2.1.3. Klausurleistung in Studie 5: Biopsychologie

Im Feldexperiment Studie 5 – Biopsychologie – konnte die Klausurleistung im Fach Biopsychologie herangezogen werden. Die Stichprobe der neun Untersuchungsteilnehmer erreichte im Durchschnitt die Note 3,19 ($SD=0,83$), hinsichtlich der in der Klausur erreichten Punkte $M=123,11$ ($SD=26,53$), siehe Abbildung 55.

	gesamte Stichprobe (N=9)		Extremstichproben			
			gute Assoziierer (N=3)		schlechte Assoziierer (N=3)	
	Note	Punkte	Note	Punkte	Note	Punkte
M	3,19	123,11	2,57	130,83	4,00	113,67
SD	0,83	26,53	0,23	25,15	1,00	34,40
Minimum	2,3	87,0	2,0	113,0	3,0	87,0
Maximum	5,0	159,5	3,0	159,5	5,0	153,0

Abbildung 55: Kennwerte der Klausurleistung (Note und Punkte) – Mittelwerte (M) und Standardabweichungen (SD) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 5 – Biopsychologie

Interessanterweise läßt sich die in der Klausur erreichte Note anhand des Assoziationskennwertes mit $r = -.709$ ($p = .016$) sehr zuverlässig vorhersagen, während der nicht signifikante korrelative Zusammenhang zwischen der Punktzahl in der Klausur und der Assoziationsleistung mit $r = .327$ deutlich geringer ausfällt und aufgrund der hohen Irrtumswahrscheinlichkeit eine zuverlässige Prognose nicht erlaubt. Ähnliches trifft für die Vorhersage der Rangreihen zu: die Rangkorrelation nach Spearman erbringt für die Note einen hohen und signifikanten Koeffizienten von mit $r = -.783$ ($p = .013$), die Rangreihe der Klausurpunkte jedoch läßt sich nicht vorhersagen ($r = .450$; $p > .05$).

Um diese Zusammenhänge zu illustrieren, wurden auch hier Extremgruppen auf Grundlage der Assoziationsleistung gebildet, dazu wurden die jeweils drei Probanden mit dem höchsten bzw. geringsten Assoziationskennwert zur Gruppe der guten bzw. schlechten Assoziierer zusammengefaßt (siehe auch Abbildung 55). Es deutet sich an, daß die guten Assoziierer im Mittel mehr Punkte ($M=130,83$; $SD=25,15$) und auch eine bessere Klausurnote ($M=2,57$; $SD=0,23$) erreichen als die schlechten Assoziierer (Punkte: $M=113,67$; $SD=34,40$; Note: $M=4,00$; $SD=1,00$). Diese Unterschiede werden allerdings nicht signifikant, was vermutlich auf den sehr geringen Stichprobenumfang zurückzuführen ist. Die Effektstärken jedoch weisen auf eine praktische

Bedeutsamkeit hin: während hinsichtlich der Klausurpunkte mit $d = .57$ ein Effekt mittlerer Stärke vorliegt, ist bzgl. der Klausurnote mit $d = 1.97$ sogar ein starker Effekt zu vermerken. Zu bedenken ist zudem, daß in dieser feldexperimentellen Studie der Wortassoziationsversuch nicht im unmittelbaren Anschluß an die Klausur durchgeführt wurde, sondern zu einem separaten Zeitpunkt im Labor stattfand, wenn auch sehr zeitnah zur Klausur (wenige Tage vor bzw. nach der Klausur). Eine zwischenzeitliche Veränderung des individuellen Wissensbesitzes kann also nicht gänzlich ausgeschlossen werden.

6.1.2.1.4. Klausurleistung in Studien 6 bis 9: Klausuren in der Psychologie

Die Studien 6 bis 9 wurden primär zur Validierung der Assoziationsleistung am Außenkriterium Klausurleistung durchgeführt. Die Benotung der Klausuren erfolgte anhand von Schulnoten mit jeweils drei Zwischennoten (1,0; 1,3; 1,7; 2,0; 2,3 usw.). In Abbildung 56 sind die statistischen Kennwerte für Klausurleistungen in den einzelnen Fächern aufgelistet (sowohl Note als auch die erreichte Punktzahl). Da die Klausur in der Klinischen Psychologie aus zwei separat benoteten Teilklausuren bestand, wird hier ein Gesamtwert für dieses Fach dargestellt, welcher sich aus dem Mittelwert der beiden Teilklausuren errechnet (Zensur bzw. Punkte). In allen Klausuren sind die jeweiligen Punkte sowohl über alle Klausurteilnehmer als auch über die Substichprobe der Versuchsteilnehmer normalverteilt. Es zeigt sich in allen Klausuren die Tendenz, daß die Teilnehmer am Wortassoziationsversuch eher zu denen gehörten, die in der Klausur besser abschnitten, dennoch weisen die Leistungskennwerte in den Substichproben eine hinreichend große Streuung auf.

	Klinische Psychologie (N=38)		Arbeitspsychologie (N=13)		Personalpsychologie (N=8)		Entwicklungs- psychologie (N=17)	
	Note	Punkte	Note	Punkte	Note	Punkte	Note	Punkte
<i>M</i>	1,82	36,22	2,33	37,50	1,79	36,25	2,33	38,38
<i>SD</i>	0,69	3,13	1,00	7,38	0,59	3,57	0,82	5,45
Minimum	1,0	17,5	1,0	18,0	1,3	29,0	1,0	29,0
Maximum	3,9	30,0	5,0	49,5	3,0	39,5	3,7	50,5

Abbildung 56: Kennwerte der Klausurleistung (Note und Punkte) – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in den Studie 6 bis 9 – Klausuren in Fächern der Psychologie

Abbildung 57 gibt einen Überblick über die Korrelationskoeffizienten zwischen Assoziationskennwerten und Klausurleistungen. In der Erhebung zur Klinischen Psychologie läßt sich die Vorhersagegüte des Wortassoziationsversuchs für die Punkte in der Klausur ($r = .351$; $p = .015$) sowie die Note ($r = -.357$; $p = .014$) mit geringen, aber signifikanten Korrelationskoeffizienten belegen. Im Wortassoziationsversuch zur Arbeitspsychologie beschreiben signifikante Korrelationskoeffizienten mittlerer Stärke den Zusammenhang zwischen Assoziation und Klausur: mit $r = -.529$ ($p = .032$) kann die Klausurnote und mit $r = .533$ ($p = .025$) die erreichte Punktzahl gut anhand des Assoziationskennwertes vorhergesagt werden. Die Leistung in der Klausur zur Personalpsychologie kann hingegen nicht zuverlässig durch den Assoziationskennwert vorhergesagt werden: zwar ergeben sich Korrelationskoeffizienten mittlerer Stärke (siehe Abbildung 57), die sind jedoch nicht signifikant. Zu bedenken ist dabei allerdings auch der sehr

geringe Stichprobenumfang von acht Probanden. Legt man jedoch die Rangkorrelation nach Spearman zugrunde, finden sich hohe signifikante Korrelationen zwischen Assoziationskennwert und Klausurleistung: während die Rangreihe des Assoziationskennwertes die Rangreihe der Noten mit $r = -.768$ ($p = .013$) sehr zuverlässig vorhersagen kann, findet sich zur Vorhersage der Rangreihe der Punktzahl sogar ein Korrelationskoeffizient von $r = .777$ ($p = .012$).

	Klinische Psychologie	Arbeitspsychologie	Personalpsychologie	Entwicklungspsychologie
erreichte Note	$r = -.357$ * ($p = .014$)	$r = -.529$ * ($p = .032$)	$r = -.541$ ($p > .05$)	$r = -.125$ ($p > .05$)
erreichte Punktzahl	$r = .351$ * ($p = .015$)	$r = .533$ * ($p = .025$)	$r = .585$ ($p > .05$)	$r = .099$ ($p > .05$)

Abbildung 57: Korrelationskoeffizienten zwischen mit den Klausurleistungen (Note; erreichte Punktzahl) in den Studien 6 bis 9 – Klausuren in Fächern der Psychologie

Es wurde bereits in Kapitel 4.6.3 (Seite 55) deutlich, daß die Leistungen der Stichprobe Entwicklungspsychologie im Wortassoziationsversuch geringer und auch homogener ausfielen als bei den anderen Klausurstichproben. Aufgrund dieser geringen Streuung der individuellen Leistungen im Wortassoziationsversuch wundert es auch nicht, daß sich kein korrelativer Zusammenhang zwischen Assoziationskennwert und Klausurleistung finden läßt (siehe Abbildung 57). In dieser Teiluntersuchung läßt sich also die Klausurleistung nicht anhand des Assoziationskennwertes vorhersagen.

Zur Illustration der korrelativen Zusammenhänge zwischen Assoziationsleistung und Klausurleistung sollen wieder Extremgruppenvergleiche herangezogen werden. Zugunsten der Übersichtlichkeit sollen hier nur die Kennwerte für die Klausurpunkte berichtet werden (siehe Abbildung 58 und Abbildung 59), da sie eine höhere Streuung aufweisen als die Note.

	Klinische Psychologie		Arbeitspsychologie	
	gute Assoziierer (N=6)	schlechte Assoziierer (N=6)	gute Assoziierer (N=5)	schlechte Assoziierer (N=5)
<i>M</i>	27,33	22,67	41,90	33,40
<i>SD</i>	1,86	2,89	5,14	9,16
Minimum	24,5	19,0	37,0	18,0
Maximum	29,5	27,0	49,5	41,5

Abbildung 58: Mittelwerte (*M*) und Standardabweichungen (*SD*) der Klausurleistung (Punkte) in den Extremgruppen (gem. Assoziationskennwert) in Studie 6 (Klausur Klinische Psychologie) und Studie 7 (Klausur Arbeitspsychologie)

Es wird deutlich, daß in der Klinischen Psychologie die Gruppe der guten Assoziierer im Mittel einen höheren Punktwert in der Klausur erreicht als die Gruppe der schlechten Assoziierer (siehe auch Abbildung 58), dieser Unterschied ist signifikant (Mann-Whitney-U: $z = 3,000$; $p = .015$) und erbringt mit $d = 1.92$ einen starken Effekt. Auch in der Klausur der Arbeitspsychologie erreichen die guten Assoziierer einen deutlich höheren Punktwert als die schlechten Assoziierer (siehe Abbildung 58), dieser augenscheinliche Unterschied ist allerdings nicht signifikant, was vermutlich auch auf die großen Standardabweichungen innerhalb der Extremgruppen zurückzuführen sein dürfte, die Effektstärke deutet aber mit $d = 1.14$ auf einen starken Effekt hin.

	Personalpsychologie		Entwicklungspsychologie	
	gute Assoziierer (N=3)	schlechte Assoziierer (N=3)	gute Assoziierer (N=4)	schlechte Assoziierer (N=4)
<i>M</i>	38,33	32,67	39,50	37,50
<i>SD</i>	1,26	3,33	2,74	4,20
Minimum	37,0	29,0	35,5	32,0
Maximum	39,5	35,5	41,5	42,0

Abbildung 59: Mittelwerte (*M*) und Standardabweichungen (*SD*) der Klausurleistung (Punkte) in den Extremgruppen (gem. Assoziationskennwert) in Studie 8 (Klausur Personalpsychologie) und Studie 9 (Klausur Entwicklungspsychologie)

In der Klausur zur Personalpsychologie erreichte die Gruppe der guten Assoziierer eine höhere Punktzahl als die Gruppe der schlechten Assoziierer (siehe Abbildung 59), dieser augenscheinliche Unterschied ist allerdings nicht signifikant, was vermutlich auch auf den geringen Stichprobenumfang zurückzuführen ist, mit $d=2.25$ jedoch liegt ein starker Effekt vor. Wie zu erwarten unterscheiden sich die beiden Extremgruppen in der Studie zur Entwicklungspsychologie kaum hinsichtlich ihrer Klausurleistungen, dennoch ist immerhin die Tendenz zu erkennen, daß die guten Assoziierer mehr Punkte erreichen, was auch durch einen Effekt mittlerer Stärke mit $d= .57$ gestützt werden kann.

6.1.2.1.5. Klausurleistung in Studie 10: Gebäudereiniger

Studie 10 – Gebäudereiniger – bietet als Kriterium eine Klausur an, welche die Berufsschüler im Rahmen ihrer regulären Ausbildung zu absolvieren hatten. Dabei konnten für diese Stichprobe der Gebäudereiniger sowohl Klausurnote als auch die jeweils erreichte Punktzahl herangezogen werden. Zu einigen Probanden lagen keine Angaben zur Klausurleistung vor, insgesamt aber konnten 28 Probanden einbezogen werden, welche sowohl die Klausur als auch den Wortassoziationsversuch bearbeiteten.

	gesamte Stichprobe (N=28)		Extremstichproben			
			gute Assoziierer (N=7)		schlechte Assoziierer (N=7)	
	Note	Punkte	Note	Punkte	Note	Punkte
<i>M</i>	4,54	65,98	4,00	76,57	5,14	49,64
<i>SD</i>	1,07	25,94	1,16	27,61	0,69	19,61
Minimum	2,0	19,0	3,0	38,5	4,0	19,0
Maximum	6,0	122,5	6,0	122,0	6,0	81,0

Abbildung 60: Kennwerte der Klausur – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 10 – Gebäudereiniger

Die in der Klausur erbrachte Leistung ist insgesamt als relativ schlecht zu beschreiben (siehe auch Abbildung 60): mit einer mittleren Note von $M=4,54$ ($SD= 1,07$) wird ein Leistungsspektrum zwischen der Note 2 und der Note 6 abgedeckt. Auch der erreichte Punktwert der gesamten Klausur spiegelt diese enormen Unterschiede in der Leistung der Probanden wider: zwischen minimal 19

und maximal 122,5 Punkten beträgt die mittlere Leistung in der Stichprobe 65,98 Punkte ($SD=25,94$).

Insgesamt ist dem Assoziationskennwert eine geringe bis mittlere Vorhersagekraft der Klausurleistung zu bescheinigen. Die Klausurnote läßt sich mit einem signifikanten Korrelationskoeffizienten von $r = -.411$ ($p = .030$) einigermaßen sicher vorhersagen. Interessanterweise ist die Vorhersagequalität für die Note damit günstiger ausgeprägt als im Falle der Klausurpunkte: deren schwächerer Koeffizient von $r = .373$ ist auch nur in der Tendenz signifikant. Die hier beschriebene Klausur bestand aus mehreren Teilen mit unterschiedlicher Anforderungscharakteristik (Multiple-Choice-Test, offene Fragen, technisches Zeichnen, Kalkulation, Essay), zu denen auch separate Punktvergaben vorlagen, Abbildung 61 listet die Vorhersage aufgrund des Assoziationskennwertes für die einzelnen Klausurteile auf. Es wird deutlich, daß der Multiple-Choice-Test der Klausur eine ähnlich gute Korrelation zur Assoziationsleistung aufweist wie die Klausurnote im Ganzen: mit $r = .417$ ($p = .027$) kann der Punktwert des Multiple-Choice-Tests gut anhand des Assoziationskennwertes vorhergesagt werden. Die beste Prognose ermöglicht der Wortassoziationsversuch für die Vorhersage der Leistung im Klausurbestandteil Essay: mit $r = .461$ ($p = .014$) findet sich hier fast schon eine mittlere Korrelationsstärke. Dies ist insofern interessant, als daß Schreiben eines Essays der Anforderung bei Assoziieren am nächsten kommt: beide Verfahren können als Form der freien Reproduktion verstanden werden, bei denen vergleichsweise wenig Lenkung oder Eingrenzungen durch die Aufgabenstellung zu finden sind. Ebenso soll auf den in keiner Weise vorhandenen (inhaltlichen) Zusammenhang zwischen Assoziationskennwert und der Leistung im technischen Zeichnen hingewiesen werden, welcher sich auch im Korrelationskoeffizienten widerspiegelt ($r = .048$; $p > .80$): während die Wissenserfassung mit dem Wortassoziationsversuch einen semantischen Bezug zum Lehrstoff aufweist, stellt das technische Zeichnen eher visuell-abstrakte bzw. prozedurale Anforderungen. Dieser Kontrast schlägt sich auch in der korrelativen Beziehung nieder, welche vermittelt, daß es praktisch keine Beziehung zwischen diesen beiden Variablen gibt.

	Korrelationskoeffizient	
Klausurnote	$r = -.411 *$	$(p = .030)$
Klausurpunkte (gesamt)	$r = .373$	$(p = .051)$
Klausurpunkte: Multiple-Choice-Test	$r = .417 *$	$(p = .027)$
Klausurpunkte: offene Fragen	$r = .323$	$(p = .094)$
Klausurpunkte: technisches Zeichnen	$r = .048$	$(p = .808)$
Klausurpunkte: Kalkulationen	$r = .205$	$(p = .296)$
Klausurpunkte: Essay	$r = .461 *$	$(p = .014)$

Abbildung 61: Korrelationskoeffizienten zwischen mit der Klausurleistung; differenziert nach deren Bestandteilen in Studie 10 – Gebäudereiniger

Zur Illustration sollen die Kennwerte des Kriteriums in den beiden Extremgruppen berichtet werden. Wie bereits erwähnt, lagen nicht von allen Probanden Klausurdaten vor – dennoch können die in Kapitel 4.7.3 (Seite 55 f.) vorgestellten Extremgruppen herangezogen werden, da sich die „Ausfälle“ gleichmäßig auf die beiden Extremgruppen verteilen (die Extremgruppen verringerten

sich jeweils um einen Probanden). Abbildung 60 stellt für diese beiden Gruppen die jeweiligen statistischen Kennwerte der Klausurnote dar. Es zeigt sich in der Tendenz, daß die gemäß ihrer Assoziationsleistung als gute Assoziierer benannten Probanden erwartungsgemäß auch etwas besser in der Klausur abschneiden als diejenigen, welche wenig assoziiert haben, diese Unterschiede sind aber nur relativ gering und nicht signifikant. Zu bedenken ist dabei die insgesamt sehr geringe Streuung der Klausurleistungen, so daß man hier das Kriterium nur wenig differenzieren kann. Allerdings belegt die Effektstärke mit $d=1.12$ einen starken Effekt für den Extremgruppenvergleich. Ein ähnliches Bild ergibt sich bei Betrachtung der in der Klausur erreichten Punkte: auch hier erreichten die guten Assoziierer im Mittel eine höhere Punktzahl als die schlechten Assoziierer (siehe auch Abbildung 60), auch hier ist der augenscheinliche Unterschied statistisch nicht signifikant, weist aber mit $d=1.13$ auf einen starken Effekt hin. Ergänzend soll an dieser Stelle auf die einzelnen Klausurbestandteile hingewiesen werden: abgesehen vom technischen Zeichnen weisen die guten Assoziierer in den einzelnen Teilen der Klausur (auf Basis der erreichten Punkte) tendenziell eine höhere Leistung auf als die schlechten Assoziierer. Dieser Unterschied wird allerdings nur im Bereich des Multiple-Choice-Tests signifikant (Mann-Whitney-U: $z=-2,166$; $p=.038$), hier erreichen die guten Assoziierer mit $M=11,14$ Punkten ($SD=2,54$) eine bessere Klausurleistung als die schlechten Assoziierer mit $M=7,86$ Punkten ($SD=2,34$), mit $d=1.34$ ein starker Effekt.

6.1.2.1.6. Klausurleistung in Studie 11: Bauklempler

Auch in Studie 11 – Bauklempler – konnte als Kriterium eine Klausur herangezogen, welche die Berufsschüler im Rahmen ihrer regulären Ausbildung zu absolvieren hatten, für die jedoch lediglich die Klausurnote vorlag. Zwischen der Bestnote 3 und der schlechtesten Note 6 wurde von den Berufsschülern durchschnittlich die Note 4,14 erreicht ($SD=1,17$), siehe auch Abbildung 62. Die korrelative Beziehung zwischen Assoziationskennwert und Klausurnote ist mit $r=-.281$ ($p>.05$) schwach ausgeprägt und v.a. nicht signifikant. Dieses Ergebnis ist vermutlich auf die kleine Stichprobe und v.a. auf die geringe Varianz innerhalb der Variable Klausurnote zurückzuführen.

	gesamte Stichprobe (N=14)	Extremstichproben	
		gute Assoziierer (N=3)	schlechte Assoziierer (N=3)
<i>M</i>	4,14	3,00	4,33
<i>SD</i>	1,17	0,00	1,53
Minimum	3,0	3,0	3,0
Maximum	6,0	3,0	6,0

Abbildung 62: Kennwerte Klausurleistung (Note) – Mittelwerte (M) und Standardabweichungen (SD) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 11 – Bauklempler

Dennoch sollen zur Illustration die Extremgruppen (gemäß der Assoziationsleistung) hinsichtlich ihrer Klausurleistung verglichen werden. Abbildung 62 stellt die jeweiligen statistischen Kennwerte der Klausurnote gegenüber. Es zeigt sich erwartungsgemäß die Tendenz, daß die gemäß ihres Assoziationskennwertes als gute Assoziierer benannten Probanden auch etwas besser in der Klausur abschneiden als diejenigen, welche wenig assoziiert haben, dieser Unterschied ist aber nur relativ gering und nicht signifikant, die Effektstärke von $d=1.23$ belegt jedoch dennoch einen starken

Effekt. Zu bedenken ist dabei wiederum die insgesamt sehr geringe Streuung der Klausurleistungen und der geringe Stichprobenumfang in den Extremgruppen.

6.1.2.1.7. Zusammenfassung und Befundintegration der Studien zum Validitätskriterium Klausurleistung

Anhand der Ergebnisse der hier vorgestellten neun Studien, welche einen Beitrag zur Kriteriumsvalidierung anhand einer Klausurleistung erbringen, läßt sich eine erfreuliche Bilanz ziehen hinsichtlich der Vorhersagekraft des Assoziationskennwertes (siehe auch Abbildung 63). Lediglich drei Studien erbringen keine signifikante Korrelation (was vermutlich auch auf die sehr geringen Stichprobenumfänge zurückzuführen ist), wobei in nur einer Studie davon (Studie 9 – Entwicklungspsychologie) die Höhe des Korrelationskoeffizienten andeutet, daß es überhaupt keinen Zusammenhang zwischen den beiden Maßen gibt. Die signifikanten Zusammenhänge von Assoziationskennwert und Klausurnote bzw. Klausurpunkte erbringen überwiegend geringe (minimal $r = .338$ in Studie 4) bis mittlere, aber teilweise auch sehr hohe (maximal $r = .856$ in Studie 3) Korrelationskoeffizienten. Die mittlere gewichtete Effektstärke über alle einbezogenen Studien erbringt insgesamt für das Kriterium der Klausurnote ($d = .38$) sowie der in der Klausur erreichten Punkte ($d = .22$) einen geringen Effekt.

	Klausurnote	Klausurpunkte
<i>Einzelstudien</i>		
3 - Sparkassenakademie (A)	---	$r = .856^{**}$ ($p = .000$)
4 - Sparkassenakademie (B)	---	$r = .338^*$ ($p = .023$)
5 - Biopsychologie	$r = -.709^*$ ($p = .016$)	$r = .327$ ($p > .05$)
6 - Klausur Klinische Psychologie	$r = -.357^*$ ($p = .014$)	$r = .351^*$ ($p = .015$)
7 - Klausur Arbeitspsychologie	$r = -.529^*$ ($p = .032$)	$r = .533^*$ ($p = .025$)
8 - Klausur Personalpsychologie	$r = -.541$ ($p > .05$)	$r = .585$ ($p > .05$)
9 - Klausur Entwicklungspsychologie	$r = -.125$ ($p > .05$)	$r = .099$ ($p > .05$)
10 - Gebäudereiniger	$r = -.411^*$ ($p = .030$)	$r = .373$ ($p = .051$)
11 - Bauklempner	$r = -.281$ ($p > .05$)	---
<i>Befundintegration</i>		
gesamte Stichprobe	$r = -.353^{**}$ ($p = .000$)	$r = .436^{**}$ ($p = .000$)
Extremgruppen	$r = -.371^{**}$ ($p = .000$)	$r = .458^{**}$ ($p = .000$)

Abbildung 63: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerte und Klausurleistung) sowie Befundintegration als Korrelationen zwischen den standardisierten Klausurleistungen (Note bzw. Punkte, z-Werte) und Assoziationskennwerten (z-Werte)

Faßt man die Ergebnisse aller Studien zusammen anhand der z-transformierten Daten, so konnten insgesamt 127 Probanden in die Auswertung hinsichtlich der Klausurnote berücksichtigt werden, in Hinblick auf die in der Klausur erreichten Punkte lagen Daten von insgesamt 186 Probanden vor. Abbildung 63 führt die Korrelationskoeffizienten der Befundintegration auf. Danach läßt sich die Klausurnote mit $r = -.353$ ($p = .000$) zuverlässig durch den Assoziationskennwert voraussagen, die

Punktzahl in der Klausur sogar noch zuverlässiger mit $r = .436$ ($p = .000$). Da in die Korrelationen ausschließlich z-transformierte Werte eingingen, können diese Koeffizienten gleichzeitig als Effektstärke aufgefaßt werden, so daß diese als ein Effekt mittlerer Stärke interpretierbar sind. Werden nur diejenigen Probanden herangezogen, die in ihrer jeweiligen Substichprobe zu den Extremgruppen zusammengefaßt wurden ($N_{\text{Note}}=88$; $N_{\text{Punkte}}=124$), so wird der korrelative Zusammenhang erwartungsgemäß noch etwas akzentuiert (siehe Abbildung 63).

Zusammenfassend ist also festzuhalten, daß der Wortassoziationsversuch in der Lage ist, die Leistung von Probanden in einer Klausur vorherzusagen, signifikante geringe bis mittlere Korrelationskoeffizienten lassen eine solche Prognose zu. Lediglich in einer Studie (Studie 9 – Entwicklungspsychologie) gab die Höhe des Korrelationskoeffizienten keinen Hinweis auf einen Zusammenhang. Warum sich hier ausgerechnet die Klausur der Entwicklungspsychologie durch das Nichtvorhandensein jeglicher Beziehung zwischen Assoziations- und Klausurkennwerten abhebt, ist nicht eindeutig erklärbar. Zu bedenken ist, daß diese Klausur dem Grundstudium zuzuordnen ist, d.h. die Probanden waren Studenten des ersten und dritten Semesters. Eine gewisse Ungeübtheit im Umgang mit dem Lehrstoff und Klausuren dieser Art kann angenommen werden. Einen nicht unwesentlichen Einfluß könnte auch der enorme Schreibaufwand in dieser Klausur gehabt haben (es sollten max. 64 Punkte erreicht werden durch ausführliche Erläuterungen – im Gegensatz zur Klausur in Klinischer Psychologie beispielsweise, in der zwar ebenso viele Punkte zu erlangen waren, aber nur mit einzelnen Worten oder Phrasen geantwortet werden sollte). Der insgesamt geringe Umfang der Assoziationsleistungen sowie die geringe Standardabweichung in dieser Stichprobe der Studie 9 könnten sich also aus der vorherigen massiven Beanspruchung der Schreibfertigkeiten und daraus resultierendem Motivationsverlust erklären.

Es sei daran erinnert, daß in den meisten der hier vorgestellten Studien der Wortassoziationsversuch erst im Anschluß an die Klausur bearbeitet wurde (ausgenommen in den beiden Studien 10 und 11 der Berufsschüler). Unter Experimentalbedingungen wäre es sicherlich vorzuziehen, den Wortassoziationsversuch an vorderster Stelle im Design zu plazieren, um möglicherweise Priming-Effekte zu vermeiden. So wäre sichergestellt, daß die Aktivierung von Assoziationen tatsächlich auf den Stimulus zurückzuführen ist und nicht auf unmittelbar zuvor bearbeitete Inhalte. Da die hier beschriebenen Erhebungen aber als Feldversuch einzuordnen sind, war darauf Rücksicht zu nehmen, daß es sich um Klausuren mit Prüfungscharakter handelte, deren Bewertung auf dem Zeugnis erscheint und somit als konsequenzenreich und justiziabel beschrieben werden kann. Somit war den Probanden (auch unter dem rechtlichen Aspekt) nicht zuzumuten, zuerst den freiwilligen Part (Wortassoziationsversuch) zu bearbeiten und anschließend die Prüfung abzulegen, zumal die Zeit zur Bearbeitung des Wortassoziationsversuchs individuell verschieden ist und daher nur schwer kalkuliert werden kann. Die Psychologie-Klausuren selbst waren sehr knapp formuliert, so daß ein Erinnern von Fachbegriffen aus den Klausurfragen eher unwahrscheinlich ist (trifft für die Studien 6 bis 9 zu, für alle anderen Studien ist darüber nichts bekannt). Vielmehr läßt sich vermuten, daß die Probanden sich während der Klausur bereits aktiv mit einem oder mehreren der Stimulusbegriffe gedanklich auseinandergesetzt haben, und bestimmte Verbindungen zwischen Knoten im semantischen Netzwerk sozusagen schon „vorgewärmt“ wurden durch deren intensiven Gebrauch während der Klausur. Neben der persönlichen Relevanz des Prüfungsereignisses ist der Aspekt der individuellen Beanspruchung durch dieses Ereignis zu beachten. Im Zuge der

Versuchspersonenrekrutierung unmittelbar im Anschluß an die Psychologie-Klausuren in den Studien 6 bis 9 berichteten viele Probanden darüber, nun völlig erschöpft zu sein und sich nicht mehr konzentrieren zu können. Nicht zuletzt die Anforderung, in der Klausur innerhalb von 1,5 Stunden möglichst viel handschriftlich zu produzieren, dämpfte doch erheblich die Bereitschaft, im Wortassoziationsversuch nun wieder schreiben zu müssen.

Es bleibt festzuhalten, daß unter Beachtung all dieser geschilderten Aspekte die Vorhersagequalität von Klausurleistungen aufgrund des Assoziationskennwertes mit einem geringen bis mittleren Korrelationskoeffizienten erfreulich hoch ausfällt.

6.1.2.2. Geschäftsführerurteil als Validitätskriterium

Im Folgenden werden zwei Feldstudien aus mittelständischen Unternehmen vorgestellt, in denen der Geschäftsführer die Arbeitsleistung seiner Mitarbeiter beurteilte. Anschließend werden die Befunde beider Erhebungen zusammengeführt. Das hier zugrunde gelegte Kriterium ist als ein Maß für die Arbeitsleistung zu verstehen, zu beachten ist, daß das fachspezifische Wissen dabei nur einen Teilaspekt neben einigen anderen darstellt.

6.1.2.2.1. Geschäftsführerurteil in Studie 12: Handelsunternehmen

Wie in Kapitel 4.9 (Seite 55 ff.) ausgeführt, wurde in Studie 12 – Handelsunternehmen – das Urteil des Geschäftsführers als Kriterium herangezogen, indem dieser verschiedene Einzelaspekte von Arbeitsleistungen (darunter auch die Fachkompetenz) bewertete (als Schulnote), und diese schließlich zu einer Gesamtleistung zusammengefaßt wurden. Abbildung 64 bzw. Abbildung 65 listen die statistischen Kennwerte zur Gesamtbeurteilung sowie zur Fachkompetenz als Einzelkategorie auf. Danach wurde die Gesamtleistung im Mittel mit der Note 2,90 ($SD=0,83$) beurteilt, bei der Fachkompetenz alleine fiel die Beurteilung im Mittel etwas schlechter aus ($M=3,35$; $SD=1,27$). Zu beachten ist hierbei die sehr geringe Standardabweichung, v.a. bei der Gesamtleistung.

	gesamte Stichprobe (N=17)	Extremstichproben	
		gute Assoziierer (N=4)	schlechte Assoziierer (N=4)
M	2,90	2,08	3,03
SD	0,83	0,30	0,86
Minimum	1,70	1,70	2,20
Maximum	4,20	2,40	4,20

Abbildung 64: Kennwerte des Geschäftsführerurteils zur Gesamtkompetenz (Schulnoten) – Mittelwerte (M) und Standardabweichungen (SD) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 12 – Handelsunternehmen

Für die Vorhersage der Gesamtbeurteilung des Geschäftsführers durch den Assoziationskennwert ergibt sich, wie auch in Abbildung 66 zu sehen, ein signifikanter mittlerer Korrelationskoeffizient $r = -.466$ ($p = .036$). Für die Vorhersage der Beurteilung der Fachkompetenz alleine hingegen ergibt sich ein deutlich schwächerer Korrelationskoeffizient, welcher auch nicht signifikant ist ($r = -.230$; $p > .05$). Betrachtet man die korrelativen Zusammenhänge lediglich in der Extremstichprobe, so fällt die Vorhersagegüte durch die Assoziationsleistung erwartungsgemäß deutlich prägnanter aus, auch

hier ist der Zusammenhang von Assoziationskennwert und Gesamtleistung mit $r = -.792$ ($p = .019$) der Vorhersage des Einzelkriteriums Fachkompetenz mit $r = -.310$ ($p > .05$) deutlich überlegen.

	gesamte Stichprobe (N=17)	Extremstichproben	
		gute Assoziierer (N=4)	schlechte Assoziierer (N=4)
<i>M</i>	3,35	2,75	3,50
<i>SD</i>	1,27	1,50	1,29
Minimum	1,00	1,00	2,00
Maximum	5,00	4,00	5,00

Abbildung 65: Kennwerte des Geschäftsführerurteils zur Fachkompetenz (Schulnoten) – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 12 – Handelsunternehmen

	gesamte Stichprobe (N=17)
Gesamtleistung	$r = -.446^*$ ($p = .036$)
Fachkompetenz	$r = -.230$ ($p > .05$)

Abbildung 66: Korrelationen zwischen den Geschäftsführerurteilen (Schulnoten) und Assoziationskennwerten in Studie 12 – Handelsunternehmen

Zur Illustration der korrelativen Zusammenhänge sei für die beiden Extremgruppen auf die statistischen Kennwerte der Geschäftsführerbeurteilungen verwiesen, welche in Abbildung 64 bzw. Abbildung 65 dargestellt sind. Danach zeigen die guten Assoziierer sowohl für die Gesamtbeurteilung als auch für die Fachkompetenz separat in der Tendenz eine bessere Beurteilung durch den Geschäftsführer als die schlechten Assoziierer. Die aufgeführten Unterschiede sind allerdings statistisch nicht signifikant, dennoch zeigen sich hinsichtlich Beurteilung der Fachkompetenz mit $d = .54$ eine mittlere Effektstärke, bzgl. der Gesamtbeurteilung sogar mit $d = 1.50$ sogar ein starker Effekt.

6.1.2.2.2. Geschäftsführerurteil in Studie 13: Stoffdruckerei

Auch in Studie 13 – Stoffdruckerei – wurde das Urteil des Geschäftsführers über die Arbeitsleistung und Fachkompetenz seiner Mitarbeiter als Außenkriterium herangezogen. Die statistischen Kennwerte für die Gesamtbeurteilung des Geschäftsführers sind in Abbildung 67, die für die separate Beurteilung der Fachkompetenz in Abbildung 68 aufgelistet. Im Mittel wird die Fachkompetenz der Beschäftigten mit 2,52 ($SD = 0,93$) benotet, die Gesamteinschätzung beträgt über die Stichprobe $M = 2,76$ ($SD = 0,65$). Insgesamt ist auf die geringen Standardabweichungen hinzuweisen.

Durch die Anzahl von Assoziationen im Wortassoziationsversuch kann in dieser Studie die Geschäftsführerbeurteilung kaum vorhergesagt werden: es ergeben sich keine signifikanten Korrelationen zwischen den Kennwerten, lediglich die Fachkompetenz erbringt mit $r = -.350$ ($p = .060$) einen in der Tendenz signifikanten Korrelationskoeffizienten (siehe Abbildung 69). Aus diesem Grunde zeigen auch die beiden Extremstichproben, die auf Grundlage des individuellen Assoziationskennwertes gebildet wurden, nur geringe und nicht signifikante Unterschiede in ihrer

mittleren Geschäftsführerbeurteilung (sowohl in Fachkompetenz als auch Gesamtbeurteilung), aber immerhin weisen die angedeuteten Unterschiede in die erwartete Richtung (gute Assoziierer haben eine bessere Benotung), die Effektstärken für die Extremgruppenvergleiche deuten mit $d = .36$ für die Gesamtbeurteilung bzw. $d = .39$ für die Fachkompetenz zumindest einen geringen Effekt an.

	gesamte Stichprobe (N=21)	Extremstichproben	
		gute Assoziierer (N=7)	schlechte Assoziierer (N=7)
<i>M</i>	2,76	2,55	2,80
<i>SD</i>	0,65	0,59	0,77
Minimum	1,81	1,81	1,91
Maximum	4,28	3,19	4,28

Abbildung 67: Kennwerte des Geschäftsführerurteils zur Gesamtkompetenz (Schulnoten) – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 13 – Stoffdruckerei

	gesamte Stichprobe (N=21)	Extremstichproben	
		gute Assoziierer (N=7)	schlechte Assoziierer (N=7)
<i>M</i>	2,52	2,29	2,71
<i>SD</i>	0,93	1,25	0,95
Minimum	1,00	1,00	1,00
Maximum	4,00	4,00	4,00

Abbildung 68: Kennwerte des Geschäftsführerurteils zur Fachkompetenz (Schulnoten) – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe sowie in den Extremgruppen gem. des Assoziationskennwertes in Studie 13 – Stoffdruckerei

gesamte Stichprobe (N=21)	
Gesamtleistung	$r = -.272$ ($p > .10$)
Fachkompetenz	$r = -.350$ ($p = .060$)

Abbildung 69: Korrelationen zwischen den Geschäftsführerurteilen (Schulnoten) und Assoziationskennwerten in Studie 13 – Stoffdruckerei

6.1.2.2.3. Zusammenfassung und Befundintegration der Studien zum Validitätskriterium Geschäftsführerurteil

Zur Beurteilung der Vorhersagekraft des Assoziationskennwertes hinsichtlich der Leistungsbeurteilungen durch den Geschäftsführer konnten zwei Studien herangezogen werden, denen Daten von insgesamt 38 Probanden zugrunde lagen. Die Ergebnisse hinsichtlich der Validitätskoeffizienten sind uneindeutig: nur in Studie 12 (Handelsunternehmen) konnte eine signifikante Korrelation mittlerer Stärke zwischen dem Assoziationskennwert und der Gesamtbeurteilung des Geschäftsführers ermittelt werden, die Fachkompetenz ließ sich in dieser Studie nicht zuverlässig vorhersagen. In der zweiten Studie hingegen verhielt es sich in der Tendenz anders herum: für die Gesamtbeurteilung zeigten sich geringere korrelative Zusammenhänge als für die Fachkompetenz, allerdings erwiesen sich hier keine der Korrelationskoeffizienten als signifikant

(siehe auch Abbildung 70). Die mittlere gewichtete Effektstärke über beide Studien erbringt für die Gesamtbeurteilung mit $d = .35$ ebenso wie für die Beurteilung der Fachkompetenz mit $d = .30$ jeweils einen kleinen Effekt.

	Gesamtleistung	Fachkompetenz
<i>Einzelstudien</i>		
12 - Handelsunternehmen	$r = -.446^*$ ($p = .036$)	$r = -.230$ ($p > .05$)
13 - Stoffdruckerei	$r = -.272$ ($p > .10$)	$r = -.350$ ($p > .05$)
<i>Befundintegration</i>		
gesamte Stichprobe (N=38)	$r = -.349^*$ ($p = .016$)	$r = -.297^*$ ($p = .035$)
Extremgruppen (N=22)	$r = -.467^*$ ($p = .014$)	$r = -.343$ ($p = .059$)

Abbildung 70: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerte und Geschäftsführerurteil) sowie Befundintegration als Korrelationen zwischen den standardisierten Geschäftsführerurteilen (Schulnoten, z-Werte) und Assoziationskennwerten (z-Werte)

Faßt man die Daten beider Studien zusammen, indem die z-transformierten Werte der Variablen herangezogen werden, so ergeben sich schwache, aber signifikante Korrelationskoeffizienten zwischen Assoziationskennwert und Geschäftsführerurteil, welche auch i.S. einer Effektstärke als Effekt mittlerer Stärke interpretiert werden können. Mit $r = .349$ ($p = .016$) erbringt der Wortassoziationsversuch einen geringen und signifikanten Korrelationskoeffizienten zur Vorhersage der Gesamtbeurteilung der Arbeitsleistung. Betrachtet man nur die Gruppe der Probanden mit extremen Assoziationsleistungen (insgesamt 22 Probanden), so wird dieser Effekt erwartungsgemäß noch deutlicher ($r = -.467$; $p = .014$). Die Vorhersage der Fachkompetenz allein ist ebenfalls durch einen geringen und signifikanten Koeffizienten belegt, der mit $r = .297$ ($p = .035$) allerdings etwas geringer ausfällt als zur Vorhersage der Gesamtbeurteilung; legt man nur die Gruppe der extremen Assoziierer zugrunde, so erhöht sich zwar auch hier der Korrelationskoeffizient, ist aber nur noch in der Tendenz signifikant (siehe Abbildung 70). Dieser Vorsprung der Gesamtbeurteilung gegenüber der Fachkompetenzbeurteilung ist insofern bemerkenswert, als daß man zwischen Fachwissen und der wahrgenommenen Fachkompetenz einen unmittelbareren und eindeutigeren Bezug vermutet, während in die Gesamtbewertung der Arbeitsleistung noch viele andere Aspekte neben dem Fachwissen einfließen. Letztlich aber bleibt festzuhalten, daß auf Grundlage des Assoziationskennwertes im Wortassoziationsversuch eine Vorhersage der Beurteilung der Arbeitsleistung durch den Geschäftsführer möglich ist.

6.1.3. Konstruktvalidität

Zur Abklärung der Konstruktvalidität können sowohl konvergente als auch diskriminante Belege herangezogen werden. Häcker, Leutner und Amelang (1998) verstehen unter einem Konstrukt eine theoretische Konstruktion über Zusammenhänge und Bedingungen menschlichen Verhaltens. Der Nachweis der konvergenten Konstruktvalidität ist laut Häcker, Leutner und Amelang (1998) v.a. über Korrelationen möglich, welche das zu validierende Verfahren mit anderen Tests, welche nachweislich dasselbe Konstrukt erfassen, aufweisen sollte. In der vorliegenden Arbeit wird als Konstrukt das fach- oder berufsbezogene Wissen verstanden, somit soll nun der

Wortassoziationsversuch in Beziehung gesetzt werden zu Verfahren, welche ebenfalls für sich beanspruchen, das Konstrukt des Fachwissens zu erheben (siehe auch im Überblick Abbildung 52, Seite 55).

Für den Nachweis der diskriminanten Konstruktvalidität wird in drei Studien ein Verfahren als Kontrollvariable herangezogen, welches das Konstrukt der generellen Wortflüssigkeit erfaßt. Es wird also erwartet, daß zwischen dem Assoziationskennwert und dem Kennwert der Wortflüssigkeit keine Beziehungen existieren, da Wortflüssigkeit nicht die Erfassung des gedächtnismäßig verankerten Wissens, welche der Wortassoziationsversuch für sich beansprucht, beeinflussen sollte.

6.1.3.1. Konvergente Konstruktvalidität

Wie in Abbildung 52 (Seite 55) dargestellt, wurden in acht der hier vorgestellten Studien neben dem Wortassoziationsversuch auch andere wissensdiagnostische Verfahren eingesetzt. Dabei handelt es sich in der Mehrzahl um Strukturlegetechniken sowie den Assoziations-Struktur-Test nach Meyer (2007), aber auch Multiple-Choice-Tests, ein Lückentext sowie ein Tiefeninterview wurden eingesetzt. Zu bedenken ist, daß diese anderen Wissenstests selbst (noch) nicht hinreichend auf Validität und Reliabilität überprüft sind, Amelang und Schmidt-Atzert (2006) bezeichnen sie als daher „Quasikriterien“. Dies ist aber auch darauf zurückzuführen, daß sie – ebenso wie der Wortassoziationsversuch selbst – keinen festen Satz an Items bieten, sondern daß eine grundsätzliche Methode auf wechselnde Wissensdomänen angepaßt wird.

In den nachfolgenden Ausführungen soll betrachtet werden, wie gut der Wortassoziationsversuch mit der Leistung in einem der anderen hier vorgestellten Diagnoseverfahren korreliert. Darüber hinaus soll – soweit in den einzelnen Studien realisierbar – ein Ansatz verfolgt werden, welcher der inkrementellen Validität entspricht (siehe auch Bühner, 2006): die Vorhersagekraft des Wortassoziationsversuchs für ein Außenkriterium (Klausurleistung oder Geschäftsführerurteil) wird verglichen mit der Vorhersagekraft des alternativen Diagnoseverfahrens (Strukturlegetechnik, Tiefeninterview usw.) für das gleiche Außenkriterium. Es ist dabei die Frage zu beantworten, durch welches Verfahren (Wortassoziationsversuch vs. alternatives Verfahren) die bessere Vorhersage des Außenkriteriums möglich ist. Im Folgenden werden jeweils die eingesetzten Methoden vorgestellt und ggf. kurz deren Kennwerte erläutert, im Anschluß daran werden die Befunde in den einzelnen Studien berichtet und schließlich zu einer Befundintegration zusammengeführt.

6.1.3.1.1. Lückentext als konstruktverwandter Test in Studie 2: psychotherapeutische Ambulanz

Im Rahmen von Studie 2 – psychotherapeutische Ambulanz – wurde zur Validierung des Wortassoziationsversuchs ein Lückentext entwickelt, welcher am Ende der zweiten Sitzung (Wiederholungsmessung) bearbeitet wurde. Das freie Antwortformat des Lückentextes wurde gewählt, um Zufallslösungen auszuschließen (siehe z.B. Bühner, 2006). Wie bereits in Kapitel 4.2 (Seite 55 f.) ausgeführt, entsprachen die im Lückentext behandelten Inhalte den Stimuli im Wortassoziationsversuch, so daß hier eine direkte Vergleichbarkeit beider Verfahren erreicht werden konnte – beide Instrumente greifen also im freien Antwortformat auf denselben Ausschnitt der Wissensdomäne zurück.

Es waren maximal 48 Punkte im Lückentext zu vergeben. Die 20 Probanden erreichten im Mittel 33,20 Punkte ($SD=5,28$), die Leistung im Lückentext ist über die Stichprobe normalverteilt (siehe

Abbildung 71). Die Leistung eines Probanden im Lückentext läßt sich aufgrund signifikanter Korrelationen mittlerer Stärke gut anhand seiner Leistung in den Wortassoziationsversuchen vorhersagen (siehe auch Abbildung 72). Legt man den Assoziationskennwert in der Wiederholungsmessung zugrunde (zum gleichen Meßzeitpunkt wurde ja auch der Lückentext bearbeitet), so ergibt sich ein signifikanter Korrelationskoeffizient nach Pearson von $r = .449$ ($p = .024$), die Rangkorrelation nach Spearman zeigt mit $r = .518$ ($p = .010$) ebenfalls aussagekräftige Zusammenhänge zwischen beiden Variablen. Interessanterweise ergeben sich bei Zugrundelegung des Assoziationskennwertes der ersten Erhebung mit $r = .475$ ($p = .017$) nach Pearson und $r = .527$ ($p = .008$) nach Spearman sogar noch deutlichere Zusammenhänge (siehe Abbildung 72), obwohl zwei Wochen Zeit zwischen der Erhebung beider Variablen lag. Da aber die Grundannahme bestand, daß beide Testzeitpunkte zu vergleichbaren Ergebnissen im Wortassoziationsversuch führen (was auch bestätigt werden wird, siehe Kapitel 6.2 auf Seite 55), sollen diese geringen Unterschiede in den Korrelationskoeffizienten nicht weiter interpretiert werden, zumal sie zu den gleichen Aussagen führen.

	gesamte Stichprobe (N=20)	Extremstichproben	
		gute Assoziierer (N=3)	schlechte Assoziierer (N=3)
<i>M</i>	33,20	36,17	30,33
<i>SD</i>	5,28	3,82	5,80
<i>Minimum</i>	21	31	25
<i>Maximum</i>	41	39	37

Abbildung 71: Mittelwerte (M) und Standardabweichungen (SD) des Lückentextes in Studie 2 – psychotherapeutische Ambulanz (die Einteilung der Extremgruppen erfolgt anhand der Assoziationsleistungen in T2)

Lückentext	
Assoziationskennwert in T1	$r = .475^*$ ($p = .017$)
Assoziationskennwert in T2	$r = .449^*$ ($p = .024$)

Abbildung 72: Korrelationen des Lückentextes (Punkte) mit dem jeweiligen Assoziationskennwert (zu beiden Meßzeitpunkten) in Studie 2 – psychotherapeutische Ambulanz

Wie Abbildung 71 außerdem zeigt, erreichen die Probanden, welche anhand ihres Assoziationskennwertes in der zweiten Erhebung als gute vs. schlechte Assoziierer zu Extremgruppen zusammengefaßt wurden, auch tendenziell unterschiedliche Punktwerte im Lückentext. Während die guten Assoziierer im Mittel mit 36,17 Punkte ($SD=3,82$) abschnitten, waren es bei den schlechten Assoziierern durchschnittlich 30,33 Punkte ($SD=5,80$). Die große Nähe der Maximalwerte innerhalb der beiden Extremgruppen (siehe Abbildung 71) macht allerdings schon deutlich, daß dieser augenscheinliche Mittelwertsunterschied zwischen den Gruppen nicht signifikant ist (Mann-Whitney-U: $z = -1,091$; $p > .05$), die Effektstärke weist mit $d = 1.19$ weist allerdings auf einen großen Effekt hin. Es sei erwähnt, daß auch hier der Verweis auf die erste Erhebung (T1) etwas prägnantere Ergebnisse erbringt: legt man deren Assoziationskennwert für die Bildung der Extremgruppen zugrunde, so deckt die Gruppe der guten Assoziierer mit $M=35,83$ ($SD=4,20$) genau den gleichen Wertebereich ab wie ihr Pendant in T2, die Gruppe der schlechten Assoziierer jedoch setzt sich mit ihrer Leistung im Lückentext deutlicher ab ($M=24,67$; $SD=3,51$)

mit einem Maximalwert von 28 Punkten. Allerdings ist auch dieser Mittelwertsunterschied nur in der Tendenz signifikant (Mann-Whitney-U: $z=-1,964$; $p=.050$), mit $d=2.88$ liegt allerdings auch hier ein starker Effekt vor.

Zusammenfassend bleibt festzuhalten, daß der Wortassoziationsversuch in dieser Studie zuverlässig in der Lage ist, auf Grundlage der Leistung eines Probanden im Wortassoziationsversuch dessen Leistung im Lückentext vorherzusagen.

6.1.3.1.2. Tiefeninterview als konstruktverwandter Test in Studie 12: Handelsunternehmen

Rothe und Hinnerichs (2005) bestimmten in Studie 12 – Handelsunternehmen – durch eine inhaltsanalytische Auswertung der Tiefeninterviews die Differenz zum Soll-Wissen, welches auf Basis des hypothetischen Wissensmodells definiert worden war. Jede Übereinstimmung mit den definierten Soll-Wissensinhalten wurde mit einem Punkt bewertet, so daß das Sollwissen einem „Basiswissen“ über die Domäne entspricht, darüber hinausgehende Nennungen wurden ebenfalls bepunktet und als sog. „Spezialwissen“ verstanden (so ist auch der Maximalwert von 147 % des Soll-Wissens in der Gruppe der Produktmanager usw. zu verstehen, siehe Abbildung 73). Die von Hinnerichs (Rothe und Hinnerichs, 2005) durchgeführten und bewerteten Tiefeninterviews mit zwölf Mitarbeitern des Unternehmens sind in Abbildung 73 anhand ihrer quantitativen Kennwerte beschrieben. Da für die Beschäftigten im Bereich Service ein anderer (in einigen Teilen differenzierterer) Interviewleitfaden vorgelegt wurde, werden diese hier auch separat aufgeführt. Das in Prozent ausgedrückte Verhältnis zum jeweiligen Soll-Wissen stellt bereits eine Relativierung dar und macht beide Substichproben miteinander vergleichbar, daher kann dieser Kennwert auch für die weiteren Berechnungen herangezogen werden.

	Produktmanagement, Außendienst, Verkauf (N=10)		Service (N=2)	
	Punktzahl	Verhältnis zum Soll- Wissen in %	Punktzahl	Verhältnis zum Soll-Wissen in %
maximal erreichbarer Wert	86	100	77	100
<i>M</i>	64,10	89,50	63,00	81,50
<i>SD</i>	18,00	28,72	8,48	10,61
Minimum	42,0	50,0	57,0	74,0
Maximum	96,0	147,0	69,0	89,0

Abbildung 73: Kennwerte der Tiefeninterviews – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die Stichprobe (differenziert nach Tätigkeitsbereichen) in Studie 12 – Handelsunternehmen

Der Assoziationskennwert erbringt keine befriedigende Vorhersage des Kriteriums Tiefeninterview (ausgedrückt als prozentuales Verhältnis zum Soll-Wissen): mit $r=.273$ zeigt sich zwar ein schwacher Korrelationskoeffizient, der aber aufgrund der hohen Irrtumswahrscheinlichkeit keine zuverlässige Prognose zuläßt. Hingegen erweisen sich die von Rothe und Hinnerichs (2005) erhobenen quantitativen Kennwerte der Tiefeninterviews selbst als ein sehr guter Prädiktor für die Vorhersage sowohl der Fachkompetenz als auch der Gesamtleistung in der Beurteilung durch den Geschäftsführer (siehe dazu auch Abbildung 74). Mit $r=-.843$ ($p=.000$) beschreibt der Interview-Kennwert einen hohen (und sehr signifikanten) korrelativen Zusammenhang des Tiefeninterviews

mit der Gesamtbeurteilung des Geschäftsführers, die Vorhersagekraft für die Beurteilung der Fachkompetenz alleine ist nur geringfügig schwächer. Damit besitzt das Tiefeninterview in dieser Studie eine deutlich höhere Vorhersagekraft für die Beurteilung durch den Geschäftsführer als der Wortassoziationsversuch selbst, welcher mit $r = -.446$ „nur“ einen signifikanten korrelativen Zusammenhang mittlerer Stärke erbringt (vgl. auch Abbildung 66 auf Seite 55).

	Tiefeninterview (N=12)	
Assoziationskennwert	$r = .273$	$(p > .05)$
Geschäftsführerbeurteilung:		
Gesamtleistung	$r = -.843^{**}$	$(p = .000)$
Fachkompetenz	$r = -.757^{**}$	$(p = .002)$

Abbildung 74: Korrelationen des Tiefeninterviews (Verhältnis zum Soll-Wissen in %) mit Assoziationskennwert und Geschäftsführerurteil (Schulnoten) in Studie 12 – Handelsunternehmen

Im Sinne der inkrementellen Validität ist also festzuhalten, daß Tiefeninterview und Wortassoziationsversuch offenbar nicht genau das Gleiche erfassen. Obwohl in den Tiefeninterviews und in den Wortassoziationsversuchen die gleichen Teildomänen bearbeitet wurden, führen beide Verfahren trotzdem zu unterschiedlicher Vorhersagekraft für das Außenkriterium Geschäftsführerurteil und erbringen außerdem nur (i.S. der konvergenten Validität) geringe korrelative Zusammenhänge untereinander. Zum einen ist dieser Befund nachvollziehbar aufgrund der Verfahren selbst: das Tiefeninterview bietet eine natürlichere Gesprächssituation, in der sowohl Fragen als auch Antworten im ganzen Satz gegeben werden, während der Wortassoziationsversuch sowohl seine Frage- als auch Antwortformate auf das Nennen von Schlagwörtern reduziert. Zum anderen erfordert der Wortassoziationsversuch die freie Reproduktion auf einen Stimulus hin, ohne inhaltliche Einschränkungen. Im Tiefeninterview hingegen wird der Suchraum über die Startfrage hinaus gelenkt (z.B. „Beschreiben Sie bitte den Ablauf bei der Vorbereitung und Nachbereitung zum Elektrodenschweißen! Berücksichtigen Sie bei Ihrer Beschreibung auch, wovon die jeweiligen Verfahrensschritte abhängig sind!“; siehe Hinnerichs und Rothe, 2002), es können dabei also auch möglicherweise inaktive Wissensinhalte reaktiviert werden.

Nicht zuletzt ist die unterschiedliche Vorhersagekraft der beiden Verfahren auch vor dem Hintergrund des Leistungskriteriums zu sehen: beim Wortassoziationsversuch wird (lediglich) die Quantität der Nennungen erhoben, beim Tiefeninterview wird ein inhaltlicher Abgleich vorgenommen mit einem Soll-Wissen, welches der Geschäftsführer definiert hat. Das Kriterium, an dem die Vorhersagegüte gemessen wird (Fachkompetenz bzw. Gesamtbeurteilung der Arbeitsleistung), ist ebenfalls durch den Geschäftsführer definiert worden, wobei sicherlich gleiche Bezugsnormen des Geschäftsführers (für Interview vs. Leistungsbeurteilung) anzunehmen sind.

6.1.3.1.3. Strukturlegetechnik als konstruktverwandter Test

Die in einigen der hier vorgestellten empirischen Studien eingesetzte Strukturlegetechnik basiert auf der von Scheele und Groeben (1984) entwickelten Methodik zur Visualisierung individuellen Wissens (siehe auch Kapitel 2.2.5.3 auf Seite 26). Bei dieser Technik werden zentrale Konzepte

einer Wissensdomäne als sog. Knoten eines Netzwerkes und die relationalen semantischen Beziehungen zwischen zwei Knoten als sog. Kanten begriffen, um so die individuelle Wissensrepräsentation in Form eines semantischen Netzwerkes abzubilden.

In drei Studien wurde die Strukturlegetechnik als weiteres wissensdiagnostisches Instrument eingesetzt, welche im Folgenden näher erläutert werden sollen. Abbildung 75 gibt einen Überblick über diese Studien und ihre Merkmale hinsichtlich der jeweils eingesetzten Strukturlegetechniken. In zwei der drei Studien waren die Knoten der zu legenden Netze durch den Versuchsleiter vorgegeben worden – hier wurde der Wortassoziationsversuch genutzt, um das hypothetische Wissensmodell zu vervollständigen und anschließend aus diesem die Knoten für die zu legenden Netze abzuleiten, welche allen Probanden in gleicher Weise vorgelegt wurden. Nur in einer Studie (Studie 5 – Biopsychologie) wurden die Netze auf Grundlage der zuvor assoziierten Konzepte gelegt, welche einen individuellen Knoten-Pool ergaben – hier diente der Wortassoziationsversuch also gemäß des bei Rothe (2003, 2006) beschriebenen Vorgehens als unmittelbarer Lieferant für das anschließende Strukturlegen.

Nr.	Name der Studie	Knoten	Hinzunahme von zusätzlichen Knoten	Kanten waren vorgegeben nach Klix (1992)
5	Biopsychologie	selbst assoziiert	möglich	ja
12	Handelsunternehmen	vorgegeben	möglich	ja
13	Stoffdruckerei	vorgegeben	--	ja

Abbildung 75: Übersicht über die Merkmale der Strukturlegetechnik in den einzelnen Studien

In zwei der vorgestellten Studien war es den Probanden möglich, während des Strukturlegens weitere (eigene) Knoten zu ergänzen (siehe Abbildung 75). Es sei an dieser Stelle daran erinnert, daß in Studie 5 – Biopsychologie – verschiedene Regeln des Strukturlegens systematisch untersucht wurden. Hier soll lediglich auf diejenige Variante zurückgegriffen werden, welche sich hinsichtlich der zu berücksichtigenden Regeln am besten mit den anderen hier vorgestellten Studien vergleichen läßt. Allen Studien ist gemein, daß die Probanden die Kanten mit semantischen Relationen entsprechend der Kategorien von Klix (1992) zu benennen hatten. Durchführungsvarianten in der Art, daß die Benennung der semantischen Relationen völlig frei durch den Probanden getätigt werden konnte oder gänzlich darauf verzichtet wurde, spielen in der vorliegenden Arbeit keine Rolle. Alle Strukturlegetests wurden im Paper-Pencil-Format durchgeführt.

Die Auswertung der so gelegten Netze erfolgte anhand folgender quantitativen Kennwerte nach Eckert (1998, 2000): Anzahl der Knoten im Netz (N), Anzahl der Verknüpfungen (U ; synonym: Umfang), Verknüpfungsdichte (V), Durchmesser (D) sowie – soweit verfügbar – der gewichtete Korrespondenzkoeffizient (C_w). Die Anzahl der Knoten beschreibt die absolute Anzahl der vom Probanden in das Netz aufgenommenen Begriffe oder Konzepte, der Umfang die absolute Anzahl der gelegten Verknüpfungen zwischen diesen Knoten. Die Verknüpfungsdichte beschreibt die Anzahl der vom Probanden vorgenommenen Verknüpfungen im Verhältnis zu den theoretisch möglichen (also in einem Netz, in dem jeder Knoten mit jedem verknüpft ist; damit ist dieser Wert auch von der Anzahl der Knoten beeinflusst), dieser Parameter kann einen Wert zwischen 0 und +1 annehmen. Der Durchmesser eines Netzwerkes wird oftmals als Maß für die Elaboriertheit des

dahinterstehenden Wissensbesitzes interpretiert und definiert sich als die längste der kürzesten Kantenfolgen im Netz (Eckert, 1998). Dazu wird für alle Dyaden von Knoten die kürzeste Verbindung zwischen ihnen bestimmt, also der Weg über möglichst wenige Kanten. In einer Matrix, in der jeder Knoten mit jedem in Beziehung gebracht wird, läßt sich unter all diesen jeweils kürzesten Verbindungen die längste Kantenfolge ausmachen, welche dann den Wert des Durchmessers bestimmt.

Der gewichtete Korrespondenzkoeffizient ist in Abgrenzung zu den bisher vorgestellten Kennwerten ein Maß für die Übereinstimmung mit einem Außenkriterium, i.d.R. dient dazu ein von einem Experten erstelltes Referenznetz. Dabei wird geprüft, inwieweit die vom Probanden vorgenommenen Verknüpfungen bzw. Nicht-Verknüpfungen zwischen zwei Knoten mit denen des Experten übereinstimmen. Die Gewichtung berücksichtigt dabei das Verhältnis des Umfangs des Probandennetztes zum Korrespondenznetz, um so Verzerrungen aufgrund einer unterschiedlichen Verknüpfungstendenz entgegenzuwirken. Der gewichtete Korrespondenzkoeffizient kann einen Wert zwischen +1 und -1 annehmen, wobei +1 die vollständige Übereinstimmung beider Netze ausdrückt, während -1 ein komplett gegenläufiges Verknüpfungsverhalten des Probanden beschreibt (Eckert, 1998).

6.1.3.1.3.1. Strukturlegetechnik in Studie 5: Biopsychologie

Im Feldexperiment der Studie 5 wurde den Probanden auch die Strukturlegetechnik zur Bearbeitung vorgelegt. Es stand kein Referenznetzwerk eines Experten zur Verfügung, mit dem die Probandennetzwerke hätten verglichen werden können, so daß hier folglich der Kennwert des gewichteten Korrespondenzkoeffizienten nicht berichtet werden kann. Mit durchschnittlich $M=24,89$ Knoten ($SD=13,71$) und $M=27,56$ Kanten ($SD=13,83$) stellen sich die Netze einigermaßen komplex dar, wobei große interindividuelle Unterschiede zu beobachten sind; die statistischen Kennwerte der Netze sind in Abbildung 76 dargestellt.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)
<i>M</i>	24,89	27,56	0,14	7,56
<i>SD</i>	13,71	13,83	0,11	3,58
Minimum	10	10	0,04	3
Maximum	52	57	0,42	14

Abbildung 76: Statistische Kennwerte (Mittelwert und Standardabweichung) des Strukturlegens in Studie 5 – Biopsychologie

Der Kennwert des Wortassoziationsversuchs kann die Anzahl der Knoten im Netz sehr gut vorhersagen: die mit $r = .912$ ($p = .000$) sehr hohe und hoch signifikante Korrelation ist allerdings auch zu erwarten, da ja in dieser Studie die Netzwerke auf Grundlage der zuvor individuell von den Probanden assoziierten Konzepten erstellt wurden. Die jeweils zur Verfügung stehenden Knoten des Netzwerkes stellen also eine Teilmenge der gesamten Assoziationsmenge des Wortassoziationsversuchs dar. Der Durchmesser des Netzwerkes, welcher als ein Maß für die Elaboriertheit des Wissens interpretiert werden kann, weist sogar einen noch stärkeren Zusammenhang zum Assoziationskennwert auf ($r = .933$; $p = .000$). Auch die anderen Parameter lassen sich sehr gut durch den Wortassoziationsversuch vorhersagen, es zeigen sich hohe und sehr

hohe hochsignifikante korrelative Zusammenhänge, welche in Abbildung 77 zusammengetragen sind.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)
Assoziationskennwert	$r = .912^{**}$ ($p = .000$)	$r = .883^{**}$ ($p = .001$)	$r = -.850^{**}$ ($p = .002$)	$r = .933^{**}$ ($p = .000$)
Klausurnote	$r = -.630^*$ ($p = .034$)	$r = -.703^*$ ($p = .017$)	$r = .266$ ($p > .05$)	$r = -.633^*$ ($p = .034$)

Abbildung 77: Korrelationen zwischen den Strukturlegekennwerten und Assoziationskennwerten bzw. Klausurnote in Studie 5 – Biopsychologie

Der Kennwert im Wortassoziationsversuch erweist sich also in dieser Studie als ein hervorragender Prädiktor, um die Leistungsparameter der Probandennetzwerke im Strukturlegeversuch vorherzusagen zu können, wobei dies wie gesagt auch zu erwarten war, da beide Verfahren eng miteinander zusammenhängen. Abbildung 78 illustriert diesen Zusammenhang anschaulich anhand des Extremgruppenvergleichs: die Probanden mit hohem Assoziationskennwert weisen auch deutlich umfangreichere und komplexere Netze in der Strukturlegetechnik auf (mehr Knoten und Kanten, höherer Durchmesser) als diejenigen mit geringer Assoziationsleistung. Aufgrund der extrem kleinen Stichprobe (in jeder Extremgruppe befanden sich nur drei Probanden) wurden diese augenscheinlichen Unterschiede zwar nicht signifikant, jedoch weisen die Effektstärken auf bedeutsame Effekte hin ($d_N=3.33$, $d_U=2.65$, $d_V=1.79$, $d_D=2.85$).

	gute Assoziierer (N=3)				schlechte Assoziierer (N=3)			
	N	U	V	D	N	U	V	D
<i>M</i>	39,67	42,00	0,06	5,00	12,33	15,33	0,25	2,67
<i>SD</i>	11,24	13,45	0,01	1,00	3,21	4,73	0,15	0,58
<i>Minimum</i>	30	31	0,04	4	10	10	0,14	2
<i>Maximum</i>	52	57	0,07	6	16	19	0,42	3

Abbildung 78: Mittelwerte (M) und Standardabweichungen (SD) des Netzes im Strukturlegen in den Extremgruppen (gem. Assoziationskennwert) in Studie 5 – Biopsychologie

Die Parameter der Strukturlegetechnik ihrerseits sind ebenfalls gut geeignet, um das Außenkriterium Klausurnote vorherzusagen zu können (siehe Abbildung 77). Dabei erweist sich v.a. die Anzahl der gelegten Kanten (Umfang) mit einem signifikanten Korrelationskoeffizienten von $r = -.703$ ($p = .017$) als besonders vorhersagekräftig. Dieser korrelative Zusammenhang ist zwar hoch, aber übertrifft nicht die Vorhersagekraft, welche der Assoziationskennwert in Bezug auf die Klausurnote erzielen konnte ($r = -.783$; $p = .013$; siehe Kap. 6.1.2.1.3 auf Seite 55). Zur Vorhersage der Klausurpunkte jedoch können die Strukturlege-Parameter – ebenso wie der Assoziationskennwert selbst – in dieser Studie keinen Beitrag leisten.

6.1.3.1.3.2. Strukturlegetechnik in Studie 12: Handelsunternehmen

In Studie 12 – Handelsunternehmen – ermittelte Liebs (2003) für die von den Probanden mit der Strukturlegetechnik erzeugten beiden Wissensnetze zu den Themen „Schweißprinzip“ sowie „Aufbau des Schweißgerätes“ jeweils die quantitativen Kennwerte, welche in Abbildung 79 als

Mittelwert über die Stichprobe sowie im Vergleich zum Referenznetzwerk dargestellt sind. Es wird deutlich, daß die Stichprobe eine verhältnismäßig geringe Streuung der Kennwerte aufweist.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)	gewichteter Korrespondenzkoeffizient (C_w)
<i>Teilnetz „Schweißprinzip“</i>					
Referenznetzwerk	28	35	0,05	11	--
<i>M</i>	27,3	29,6	0,04	8,2	0,02
<i>SD</i>	2,7	3,8	0,007	2,3	0,3
Minimum	20	23	0,03	4	-0,3
Maximum	30	38	0,06	11	0,5
<i>Teilnetz „Aufbau des Schweißgerätes“</i>					
Referenznetzwerk	58	108	0,03	6	--
<i>M</i>	54,7	59,5	0,020	5,4	-0,25
<i>SD</i>	3,4	3,9	0,002	1,6	0,17
Minimum	47	49	0,017	1	-0,51
Maximum	60	70	0,023	7	0,01

Abbildung 79: Statistische Kennwerte (Mittelwert und Standardabweichung) des Strukturlegens in Studie 12 – Handelsunternehmen – nach Teilnetzen und im Vergleich zum Referenznetzwerk

Die Kennwerte zum Teilnetz „Aufbau des Gerätes“ lassen sich so gut wie gar nicht durch die Assoziationsleistungen erklären (siehe auch Abbildung 80). Im Teilnetz „Schweißprinzip“ hingegen erweist sich der Assoziationskennwert als guter Prädiktor für die Anzahl der vom Probanden verwendeten Knoten: mit $r = .568$ ($p = .014$) besteht zwischen diesen Variablen eine signifikante korrelative Beziehung mittlerer Stärke. An dieser Stelle sei daran erinnert, daß allen Probanden die gleiche Auswahl an Begriffen (Knoten) zur Verfügung stand. Für die Anzahl der verwendeten Verknüpfungen dagegen ist der Assoziationskennwert nur ein in der Tendenz signifikanter Prädiktor, die übrigen Parameter dieses Wissensnetzes lassen sich durch den Assoziationskennwert kaum vorhersagen (siehe auch Abbildung 80).

Allerdings muß einschränkend festgestellt werden, daß offenbar der spezifische Inhalt eines zu legenden Netzes eine wichtige Determinante ist: das Thema Schweißprinzip ist eher komplex, beinhaltet Ablaufalgorithmen und Wirkmechanismen und ist somit auf das Verständnis zielend. Das Thema Geräteaufbau hingegen ist vergleichsweise eindimensional, in seiner einfachsten Form ist hier ein sternförmiges Netz mit lediglich Teil-Ganzes-Relationen vorstellbar. Während dies sich in einer geringen Ausprägung der Parameter Kantenanzahl, Verknüpfungsdichte und v.a. Durchmesser auswirken würde, ist ebenso eine Verzerrung des Kennwertes im Wortassoziationsversuch denkbar. So genügt zu diesem Themenkomplex ein Nennen von Geräteteilen, um einen hohen Assoziationskennwert zu erzielen, ohne daß diese Konzepte in ihrer Wirkungsweise tatsächlich verstanden sein müssen – hier könnten also auch Assoziationen produziert worden sein, wenn Fachbegriffe zwar bekannt sind, aber i.S. von tragem Wissen nicht hinreichend eingesetzt werden können. Derartige Einflüsse des Inhalts der Wissensdomäne auf die Kennwerte beider Verfahren lassen also die fehlenden Zusammenhänge zwischen ihnen zumindest nachvollziehbar werden.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)	gewichteter Korrespondenzkoeffizient (C_w)
<i>Assoziationskennwert</i>					
und Teilnetz „Schweißprinzip“	$r = .568^*$ ($p > .014$)	$r = .359$ ($p > .05$)	$r = -.302$ ($p > .05$)	$r = .180$ ($p > .05$)	$r = .199$ ($p > .05$)
und Teilnetz „Aufbau des Schweißgerätes“	$r = -.223$ ($p > .05$)	$r = -.348$ ($p > .05$)	$r = -.046$ ($p > .05$)	$r = .180$ ($p > .05$)	$r = -.186$ ($p > .05$)
<i>Geschäftsführerbeurteilung (und Teilnetz „Schweißprinzip“)</i>					
Gesamtbeurteilung	$r = -.249$ ($p > .05$)	$r = -.326$ ($p > .05$)	$r = .102$ ($p > .05$)	$r = -.635^{**}$ ($p = .006$)	$r = -.191$ ($p > .05$)
Fachkompetenz	$r = -.125$ ($p > .05$)	$r = .018$ ($p > .05$)	$r = .205$ ($p > .05$)	$r = -.313$ ($p > .05$)	$r = .158$ ($p > .05$)

Abbildung 80: Korrelationen zwischen den Strukturlegekennwerten und Assoziationskennwerten (nach Teilnetzen) bzw. Geschäftsführerbeurteilung (in Beziehung zum Teilnetz „Schweißprinzip“) in Studie 12 – Handelsunternehmen

Der gefundene korrelative Zusammenhang für das Teilnetz „Schweißprinzip“ soll auch hier durch die entsprechenden Kennwerte in den Extremgruppen illustriert werden. Liebs (2003) berichtet davon, daß beim Teilnetz „Geräteaufbau“ die schlechten Assoziierer tendenziell mehr Knoten und auch Relationen verwendeten als die guten Assoziierer und vermutet dahinter eine Ratetendenz bei den schlechten Assoziierern. Wie in Abbildung 81 ersichtlich, weisen die Netze der guten Assoziierer in der Teildomäne „Schweißprinzip“ mehr Knoten und Verknüpfungen sowie einen größeren Durchmesser auf als die Netze der schlechten Assoziierer. Besonders bemerkenswert ist hier die Abgrenzung beider Extremgruppen anhand des gewichteten Korrespondenzkoeffizienten (also die Übereinstimmung mit dem Expertennetz): trotz der relativ großen Streuung innerhalb der Extremgruppen erbringen die schlechten Assoziierer im Mittel sogar einen negativen Koeffizienten, ihre Netze waren also sehr unähnlich zu dem des Experten, die Effektstärke $d=0.46$ weist allerdings eher auf einen Effekt mittlerer Stärke hin. Trotz aller augenscheinlichen Unterschiede der Strukturlegeparameter zwischen den Extremgruppen sind diese nicht signifikant, was vermutlich auf die geringe Stichprobengröße zurückzuführen ist (zumal in der Gruppe der schlechten Assoziierer ein Proband nicht am Strukturlegen teilnahm, so daß hier nur noch Daten von drei Probanden herangezogen werden konnten), die Effektstärken weisen jedoch dennoch auf bedeutsame Effekte hin ($d_N=1.14$, $d_U=1.25$, $d_V=1.00$, $d_D=0.83$).

	gute Assoziierer (N=4)					schlechte Assoziierer (N=3)				
	N	U	V	D	C_w	N	U	V	D	C_w
<i>M</i>	29,00	32,25	0,04	9,25	0,11	25,33	27,00	0,05	7,33	-0,05
<i>SD</i>	1,41	4,57	0,01	1,50	0,29	4,62	3,61	0,01	3,06	0,41
<i>Minimum</i>	27	27	0,03	7	-0,20	20	23	0,04	4	-0,31
<i>Maximum</i>	30	38	0,04	10	0,48	28	30	0,06	10	0,42

Abbildung 81: Mittelwerte (M) und Standardabweichungen (SD) der Strukturlegekennwerte (für das Teilnetz „Schweißprinzip“) in den Extremgruppen (gem. Assoziationskennwert) in Studie 12 – Handelsunternehmen

Mit Blick auf die inkrementelle Validität lassen sich keine systematischen Korrelationen zwischen den Strukturlegekennwerten und den Geschäftsführerurteilen berichten (Abbildung 80). Lediglich der Durchmesser als Maß der Elaboriertheit der gelegten Netze liefert einen signifikanten mittleren bis hohen korrelativen Zusammenhang: mit $r = -.635$ ($p = .006$) konnte der Durchmesser des Netzes „Schweißprinzip“ die Gesamtbeurteilung des Geschäftsführers vorhersagen, war zur Vorhersage der Fachkompetenz allerdings völlig ungeeignet. Damit übertrifft dieser Strukturlegeparameter dennoch die Vorhersagegüte des Assoziationskennwertes, welcher lediglich eine signifikante Korrelation von $r = -.446$ für die Gesamtbeurteilung des Geschäftsführers erbrachte.

Zusammenfassend ist also zu sagen, daß der Wortassoziationsversuch in dieser Studie nur eingeschränkt in der Lage war, die Kennwerte der Strukturlegetechnik vorherzusagen, lediglich die Anzahl der gelegten Knoten erbrachte einen signifikanten korrelativen Zusammenhang mittlerer Stärke zu der Anzahl von Assoziationen im Wortassoziationsversuch. Für die konvergente Validität lassen sich in dieser Studie also eher weniger Belege finden. In diesem Zusammenhang ist die inhaltliche Validität der für die Strukturlegetechnik gewählten Teildomänen in Frage zu stellen. Zudem bleibt im Sinne der inkrementellen Validierung festzuhalten, daß die Strukturlegetechnik zwar deutlich aufwendiger in Durchführung und Auswertung ist als der Wortassoziationsversuch, in dieser Studie aber kaum in der Lage war, die Geschäftsführerbeurteilung vorherzusagen. Lediglich der Parameter Durchmesser als Maß für die Elaboriertheit des Netzes im Teilnetz „Schweißprinzip“ konnte das Außenkriterium überhaupt vorhersagen, übertraf dabei aber die Vorhersagegüte des Assoziationskennwertes deutlich. Insgesamt verhilft dies der Strukturlegetechnik zu einem im Vergleich zum Wortassoziationsversuch äußerst unbefriedigendem Kosten-Nutzen-Verhältnis.

6.1.3.1.3.3. Strukturlegetechnik in Studie 13: Stoffdruckerei

In Studie 13 wurde ebenfalls die Strukturlegetechnik eingesetzt, bei der die Knoten aus dem hypothetischen Wissensmodell abgeleitet und allen Probanden vorgegebenen wurden. Da zwei Probanden aufgrund eines Überforderungsgefühls nicht am Strukturlegen teilnehmen wollten, basieren die folgenden Ausführungen immer auf einer Stichprobe von insgesamt 19 Probanden.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)	gewichteter Korrespondenzkoeffizient (C_w)
Referenznetzwerk	40	57	0,04	15	--
<i>M</i>	32,79	38,37	0,04	12,78	-0,42
<i>SD</i>	8,20	11,71	0,01	4,14	0,20
Minimum	18	18	0,02	5	-0,69
Maximum	53	62	0,06	23	-0,03

Abbildung 82: Statistische Kennwerte (Mittelwert und Standardabweichung) des Strukturlegens im Vergleich zum Referenznetzwerk in Studie 13 – Stoffdruckerei

Abbildung 82 stellt die statistischen Kennwerte für die Strukturlegeparameter über die Stichprobe hinweg dar, ebenso sind dort die Parameter des Referenznetzwerkes berichtet. Es werden große interindividuelle Differenzen deutlich, v.a. bei der Anzahl der gelegten Konzepte ($M=32,79$; $SD=8,20$) und Verknüpfungen ($M=38,37$; $SD=11,71$) sowie beim Durchmesser ($M=12,78$; $SD=4,14$).

In dieser Studie erweist sich der Assoziationskennwert als guter Prädiktor für die Parameter der Strukturlegetechnik. Es ergeben sich signifikante mittlere bis hohe korrelative Zusammenhänge (siehe auch Abbildung 83), insbesondere die Anzahl der gelegten Konzepte ($r = .654$; $p = .001$) sowie der Durchmesser ($r = .655$; $p = .001$) lassen sich sehr gut durch den Assoziationskennwert vorhersagen. Lediglich der gewichtete Korrespondenzkoeffizient weist keinerlei Zusammenhang zum Assoziationskennwert auf. Es soll an dieser Stelle daran erinnert werden, daß die Probanden nicht ihre eigenen Assoziationen im Wissensnetz weiterverarbeiteten, sondern ihnen Knoten aus dem hypothetischen Wissensmodell vorgegeben wurden.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungs-dichte (V)	Durchmesser (D)	gewichteter Korrespondenzkoeffizient (C _w)
Assoziationskennwert	$r = .654^{**}$ ($p = .001$)	$r = .551^{**}$ ($p = .007$)	$r = .524^*$ ($p = .011$)	$r = .655^{**}$ ($p = .001$)	$r = -.072$ ($p > .05$)
<i>Geschäftsführerbeurteilung</i>					
Gesamtbeurteilung	$r = -.170$ ($p > .05$)	$r = -.319$ ($p > .05$)	$r = -.045$ ($p > .05$)	$r = -.453^*$ ($p = .026$)	$r = -.296$ ($p > .05$)
Fachkompetenz	$r = -.148$ ($p > .05$)	$r = -.267$ ($p > .05$)	$r = -.003$ ($p > .05$)	$r = -.460^*$ ($p = .024$)	$r = -.463^*$ ($p = .023$)

Abbildung 83: Korrelationen zwischen den Strukturlegekennwerten (z-transformiert) und Assoziationskennwerten in Studie 13 – Stoffdruckerei

Dieser erfreuliche korrelative Zusammenhang soll nun wieder anhand des Extremgruppenvergleichs illustriert werden, die entsprechenden Kennwerte der beiden Extremgruppen sind in Abbildung 84 dargestellt. Es wird deutlich, daß die Netze der guten Assoziierer auch mehr Knoten und Verknüpfungen aufweisen als die der schlechten Assoziierer sowie einen größeren Durchmesser. Allerdings werden diese augenscheinlichen Unterschiede nicht signifikant, was vermutlich in den großen Standardabweichungen innerhalb der Extremgruppen begründet liegt. Die Effektstärken für die Gruppenvergleiche stützen dies: lediglich hinsichtlich der Knotenanzahl ($d = .57$) sowie der Dichte ($d = .57$) lassen sich Effekte mittlerer Stärke bescheinigen, für den Umfang der Netze ($d = .44$) sowie den gewichteten Korrespondenzkoeffizient ($d = .42$) hingegen liegen kleine Effekte vor. Hinsichtlich der Verknüpfungsdichte ist kein Unterschied zwischen den Assoziierer-Extremgruppen erkennbar ($d = .00$).

	gute Assoziierer (N=7)					schlechte Assoziierer (N=6)				
	N	U	V	D	C _w	N	U	V	D	C _w
<i>M</i>	34,86	41,00	0,04	13,86	-0,47	29,50	35,17	0,04	11,17	-0,38
<i>SD</i>	9,92	13,76	0,01	5,79	0,19	8,62	12,62	0,01	2,93	0,24
<i>Minimum</i>	23	20	0,02	5	-0,62	18	18	0,03	8	-0,69
<i>Maximum</i>	53	62	0,05	23	-0,01	38	48	0,06	15	-0,03

Abbildung 84: Mittelwerte (M) und Standardabweichungen (SD) der Strukturlegekennwerte in den Extremgruppen (gem. Assoziationskennwert) in Studie 13 – Stoffdruckerei

Im Sinne der inkrementellen Validierung wurden die Kennwerte der Strukturlegetechnik zur Leistungsbeurteilung durch den Geschäftsführer in Beziehung gesetzt. Dabei kann lediglich der

Durchmesser (also das Elaboriertheitsmaß) Vorhersagen von geringer bis mittlerer Stärke aufweisen: so erreicht der Korrelationskoeffizient zwischen Durchmesser und Gesamtbeurteilung des Geschäftsführers einen Wert von $r = -.453$ ($p = .026$) und in Bezug auf die Beurteilung der Fachkompetenz sogar den Wert $r = -.460$ ($p = .024$). Die übrigen erhobenen Kennwerte der gelegten Netze (Anzahl der Knoten, Anzahl der Kanten, Verknüpfungsdichte) besitzen keinen Vorhersagewert bzgl. der eingeschätzten Arbeitsleistung (siehe Abbildung 83), lediglich der gewichtete Korrespondenzkoeffizient ist noch zu erwähnen, der mit $r = .463$ ($p = .023$) in der Lage ist, die Beurteilung der Fachkompetenz vorherzusagen. Damit hat in dieser Studie ein Kennwert des Strukturlegens zwar einen etwas höheren Vorhersagewert für die Leistungsbeurteilung durch den Geschäftsführer als es das Assoziieren leisten kann (siehe auch Kap. 6.1.2.2.2), in Anbetracht des enormen Durchführungsaufwandes ist dieser Zuwachs an Vorhersagegüte aber unausgewogen gegenüber der Ökonomie.

6.1.3.1.3.4. Zusammenfassung und Befundintegration zur Vorhersage des Strukturlegens

Die Befunde aus den drei vorgestellten Studien (welche auf insgesamt 44 Probanden beruhen) zur Validierung des Wortassoziationsversuchs anhand der Strukturlegeparameter lassen keine eindeutigen Schlüsse zu. In allen Studien konnte die Anzahl der im Netz gelegten Knoten sehr zuverlässig durch den Assoziationskennwert vorhergesagt werden, mittlere ($r = .568$ in Studie 12) bis sehr hohe ($r = .912$ in Studie 3) signifikante Korrelationskoeffizienten rechtfertigen dies (siehe auch Abbildung 85), unabhängig davon, ob die Knoten vom Probanden jeweils selbst produziert oder aber allen Versuchsteilnehmern die gleichen Knoten vorgegeben wurden.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)	gew. Korrespondenzkoeffizient (C _w)
<i>Einzelstudien</i>					
5 - Biopsychologie	$r = .912^{**}$ ($p = .000$)	$r = .883^{**}$ ($p = .001$)	$r = -.850^{**}$ ($p = .002$)	$r = .933^{**}$ ($p = .000$)	---
12 - Handelsunternehmen	$r = .568^{*}$ ($p = .014$)	$r = .359$ ($p > .05$)	$r = -.302$ ($p > .05$)	$r = .180$ ($p > .05$)	$r = .199$ ($p > .05$)
13 - Stoffdruckerei	$r = .654^{**}$ ($p = .001$)	$r = .551^{**}$ ($p = .007$)	$r = .524^{*}$ ($p = .011$)	$r = .655^{**}$ ($p = .001$)	$r = -.072$ ($p > .05$)
<i>Befundintegration</i>					
gesamte Stichprobe (N=44)	$r = .425^{**}$ ($p = .002$)	$r = .362^{**}$ ($p = .008$)	$r = -.427^{**}$ ($p = .002$)	$r = .530^{**}$ ($p = .000$)	$r = .035$ ($p > .05$)
Extremstichprobe (N=26)	$r = .625^{**}$ ($p = .000$)	$r = .499^{**}$ ($p = .005$)	$r = -.577^{**}$ ($p = .001$)	$r = .652^{**}$ ($p = .000$)	$r = -.018$ ($p > .05$)

Abbildung 85: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerte und Parameter der Strukturlegetechnik) sowie Befundintegration als Korrelationen zwischen den standardisierten Parametern der Strukturlegetechnik (z-Werte) und Assoziationskennwerten (z-Werte)

In Studie 5 und 13 setzt sich diese Vorhersagegüte in analoger Weise für die anderen Strukturlegeparameter fort. Lediglich der gewichtete Korrespondenzkoeffizient, also die Übereinstimmung des Probanden- mit dem Expertennetzwerk, scheint in keinem Verhältnis zur

Assoziationsleistung zu stehen. Faßt man die Effektstärken über die einzelnen Studien hinweg als mittlere gewichtete Effektgröße zusammen, so ergeben sich für die meisten Strukturlegeparameter große Effekte ($d_N = .69$, $d_U = .94$, $d_V = .89$, $d_D = .88$), lediglich die Befunde zur Vorhersage des gewichteten Korrespondenzkoeffizienten lassen insgesamt keinen Effekt erkennen ($d = .056$).

Führt man die Befunde aller drei Studien anhand der z-Transformation zusammen, so ergeben sich durchweg signifikante geringe bis mittlere Korrelationen zwischen dem Assoziationskennwert und den Strukturlegeparametern Anzahl der Knoten, Umfang, Verknüpfungsdichte sowie Durchmesser (siehe Abbildung 85). Die geringste Vorhersagekraft besteht dabei für den Umfang der Netze mit $r = .362$ ($p = .008$), und der stärkste Zusammenhang zeichnet sich zwischen Assoziationskennwert und dem Durchmesser der Netze ab ($r = .530$; $p = .000$), welcher i.S. der Effektstärke bereits als großer Effekt interpretiert werden kann. Zieht man lediglich die Probanden mit extremen Assoziationskennwerten heran, so fallen die korrelativen Beziehungen erwartungsgemäß noch akzentuierter aus (siehe Abbildung 85). Der gewichtete Korrespondenzkoeffizient kann auch in der Befundintegration keine korrelative Beziehung zum Assoziationskennwert ausweisen.

Zusammenfassend läßt sich festhalten, daß der Wortassoziationsversuch insgesamt gut in der Lage ist, Parameter des Strukturlegens zuverlässig vorherzusagen, dies betrifft insbesondere die Parameter Anzahl der Knoten und Kanten im Netz, aber auch die Verknüpfungsdichte sowie den Durchmesser der gelegten Netze. Hierbei ist allerdings auch zu berücksichtigen, daß die Wahrscheinlichkeit für einen hohen Zusammenhang beider Verfahren eng mit der eingesetzten Durchführungsvariante der Strukturlegetechnik verbunden ist: stammen die zu legenden Konzepte aus individuellen Wortassoziationsversuchen, so sind hohe Zusammenhänge zwischen Assoziationskennwert und dem Strukturlegeparameter Anzahl der Knoten zu erwarten. Der Parameter des gewichteten Korrespondenzkoeffizienten (also die Übereinstimmung mit einem Expertennetzwerk) hingegen weist keinerlei Zusammenhänge zum Wortassoziationsversuch auf.

6.1.3.1.4. Assoziations-Struktur-Test als konstruktverwandter Test

Im Folgenden werden vier Studien vorgestellt, in denen der Assoziations-Struktur-Test als wissensdiagnostisches Verfahren eingesetzt wurde, welcher im Folgenden kurz skizziert werden soll. Meyer (2007) entwickelte in seiner Dissertation das computerbasierte Verfahren Assoziations-Struktur-Test zur Erfassung strukturellen impliziten Wissens. Das Verfahren besteht für die Probanden aus zwei Teilen: dem Assoziieren und dem anschließenden Paarvergleich der assoziierten Konzepte. Im ersten Teil des Verfahrens assoziiert der Proband zu einem oder sukzessive mehreren Stimuli, welche zunächst das explizite Wissen zur Domäne erfassen. Dabei soll er nach jeder Assoziation die Enter-Taste drücken, das Programm erfaßt daraufhin die Zeit zwischen dem Enter unmittelbar nach einem Konzept und dem ersten Buchstaben des nächsten Konzeptes, also letztlich die Zeit zwischen zwei Assoziationen. Auf Grundlage dieser Daten markiert das Programm zeitlich eng aufeinanderfolgende Assoziationen als inhaltlich zusammenhängendes Cluster, während längere Pausen zwischen den Eingaben auf die Trennung zwischen zwei Clustern hinweisen (für den zugrundeliegenden Algorithmus siehe Meyer, 2007). Im Ergebnis dieses Versuchsabschnittes erhält der Versuchsleiter eine Auflistung der vom Probanden produzierten Assoziationen (in der Reihenfolge der Eingabe), wobei die Cluster abgetragen sind sowie eine differenzierte Zeitmatrix der Eingabevorgänge verfügbar ist.

Für den Wortassoziationsversuch innerhalb des Assoziations-Struktur-Tests können beliebig viele Stimuli festgelegt werden, aus diesem Pool muß der Versuchsleiter einen Stimulus festlegen, welcher dann die Grundlage bildet für den sich anschließenden Teil des Paarvergleichs. Dabei werden dem Probanden paarweise seine eigenen Assoziationen vorgelegt, wobei der Proband auf einer fünfstufigen Skala die Stärke des Zusammenhangs zwischen diesen beiden Begriffen zu bestimmen hat. Der Versuchsleiter kann zwischen einer impliziten (nur die Stärke des Zusammenhangs bewerten) und einer expliziten (außerdem die Art des Zusammenhangs benennen) Variante wählen. Der Umfang der in den Paarvergleich eingehenden Begriffe kann vom Versuchsleiter begrenzt werden, um den Probanden nicht eine unüberschaubare Zahl an Paarvergleichen zuzumuten. So lassen sich beispielsweise aus 20 assoziierten Konzepten 190 Dyaden für Paarvergleiche ableiten. Sollte ein Proband jedoch mehr Begriffe zum relevanten Stimulus assoziiert haben als die vom Versuchsleiter festgelegte Maximalzahl, so werden zufällig Begriffe für die Paarvergleiche ausgewählt; dabei wird die Größe der Cluster berücksichtigt und aus jedem Cluster mindestens ein Begriff verwendet. Der Assoziations-Struktur-Test ist so programmiert, daß beide Verfahrensteile sowohl unmittelbar aufeinanderfolgend oder aber auch zu zwei getrennten Untersuchungszeitpunkten bearbeitet werden können.

Aus diesen Ähnlichkeitsurteilen in den Paarvergleichen sowie den Angaben über inhaltliche Cluster beim Assoziieren rekonstruiert die Software einen Wissensgraphen (für differenzierte Angaben zum Algorithmus siehe Meyer, 2007), welcher dem Netz aus einem Strukturlegeversuch ähnlich ist, in Abgrenzung zu diesem aber ermöglicht der Wissensgraph des Assoziations-Struktur-Test eine dreidimensionale Darstellung und verzichtet andererseits auf die Beschriftung der Kanten zwischen den Knoten. Es können analog zur Strukturlegetechnik quantitative Kennwerte berechnet werden, Meyer (2007) empfiehlt für die Beschreibung der Wissensgraphen folgende Kennwerte: Anzahl der verwendeten Knoten im Graphen (N), Anzahl der verwendeten Kanten (U), Durchmesser des Netzes (D; die Berechnung dieser Parameter entspricht dem bei der Strukturlegetechnik berichteten Vorgehen, siehe Kapitel 6.1.3.1.3 auf Seite 55) sowie die relative maximale Pfadlänge (RMPL). Letztgenannter Kennwert beschreibt das Verhältnis von Durchmesser des Netzwerkes zum längsten möglichen Pfad und ist als Maß für die Kompaktheit des Netzes zu interpretieren: je kleiner der Wert, desto kompakter ist das Wissensnetz (siehe auch Meyer, 2007). Es sei um Verständnis gebeten, da die Referenz auf verschiedene quantitative Graphenparameter in den einzelnen Studien aufgrund technischer Verfügbarkeit leicht unterschiedlich gehandhabt wurde.

An dieser Stelle soll die Abgrenzung zwischen den Parametern Anzahl der Knoten vs. Assoziationskennwert kurz in Erinnerung gerufen werden. In der vorliegenden Arbeit bezeichnet die Anzahl der Knoten im Assoziations-Struktur-Test die im Graphen verwendeten Konzepte, also basierend auf den Assoziationen des Probanden zu einem *einzigem* ausgewählten Stimulus (egal, wieviele Stimuli in der Assoziationsphase bearbeitet wurden). Dabei ist es durchaus möglich, daß nicht alle assoziierten Konzepte in den Graphen eingingen, z.B. wenn ein Begriff im Paarvergleich als sehr unähnlich zu allen anderen Begriffen beurteilt wurde. In Abgrenzung dazu soll der sog. Assoziationskennwert die Menge *aller* Assoziationen auf *alle* Stimuli hin beschreiben, und auch unabhängig davon, ob ein Konzept in den Wissensgraphen einging – dabei wird also die Assoziationsphase des Assoziations-Struktur-Tests als eigenständiges und in sich abgeschlossenes Verfahren begriffen.

6.1.3.1.4.1. Assoziations-Struktur-Test in Studie 5: Biopsychologie

In Studie 5 – Biopsychologie – wurde neben der Strukturlegetechnik auch der Assoziations-Struktur-Test von Meyer (2007) eingesetzt. Die statistischen Kennwerte der Parameter des Assoziations-Struktur-Tests sind in Abbildung 86 dargestellt; über die Stichprobe hinweg wurden im Mittel 24,56 Knoten ($SD=15,35$) bzw. $M=60,22$ Kanten ($SD=33,64$) verwendet, es zeigen sich dabei allerdings große interindividuelle Unterschiede.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)
<i>M</i>	24,56	60,22	0,34	3,44
<i>SD</i>	15,35	33,64	0,24	1,33
Minimum	9	20	0	2
Maximum	54	129	1	6

Abbildung 86: Statistische Kennwerte (Mittelwert und Standardabweichung) des Wissensgraphen im Assoziations-Struktur-Tests in Studie 5 – Biopsychologie

Diese Parameter der Wissensgraphen lassen sich sehr zuverlässig durch den Kennwert des Wortassoziationsversuchs vorhersagen: v.a. die Anzahl der in den Graphen eingegangenen Knoten weist eine hohe signifikante Korrelation mit dem Assoziationskennwert auf ($r = .833$; $p = .003$). Dies war auch zu erwarten, da der Assoziations-Struktur-Test in hohem Maße auf den im Wortassoziationsversuch assoziierten Konzepten aufbaut. Aber auch übrigen Graphenkennwerte Umfang, Verknüpfungsdichte und Durchmesser erbringen hohe signifikante Korrelationskoeffizienten (siehe Abbildung 87), so daß also die Kennwerte des Assoziations-Struktur-Tests in dieser Studie zuverlässig durch die Kennwerte des Wortassoziationsversuchs vorhergesagt werden können.

	Anzahl der Knoten (N)	Umfang (U)	Verknüpfungsdichte (V)	Durchmesser (D)
Assoziationskennwert	$r = .833^{**}$ ($p = .003$)	$r = .699^{*}$ ($p = .018$)	$r = -.675^{*}$ ($p = .023$)	$r = .809^{**}$ ($p = .004$)
Klausurnote	$r = -.644^{*}$ ($p = .031$)	$r = -.702^{*}$ ($p = .017$)	$r = -.310$ ($p > .05$)	$r = -.401$ ($p > .05$)

Abbildung 87: Korrelationen (einseitige Testung) zwischen den Kennwerten des Assoziations-Struktur-Tests und Assoziationskennwerten bzw. Klausurnote in Studie 5 – Biopsychologie

Der Extremgruppenvergleich (siehe Abbildung 88) macht anschaulich, daß Probanden mit hohem Assoziationskennwert im Assoziations-Struktur-Test auch deutlich kompaktere und umfangreichere Wissensgraphen aufweisen als diejenigen Probanden mit geringer Assoziationsleistung. Die augenscheinlich hohen Unterschiede zwischen den beiden Extremgruppen sind aber – vermutlich aufgrund der sehr geringen Stichprobengröße – nicht signifikant, die Effektstärken verweisen aber durchweg auf starke Effekte hin ($d_N=3.90$, $d_U=2.38$, $d_V=4.17$, $d_D=2.02$).

Der Assoziations-Struktur-Test selbst ist ebenfalls in der Lage, die Leistung der Probanden in der Klausur vorherzusagen: wie in Abbildung 87 ersichtlich, weisen sowohl die Anzahl der Knoten im Wissensgraphen ($r = -.644$; $p = .031$) wie auch die Anzahl von Verknüpfungen ($r = -.702$; $p = .017$) hohe signifikante Korrelationen zur Klausurnote aus. Die anderen Graphenkennwerte hingegen

können keinen Vorhersagebeitrag leisten, ebenso ergeben sich keine korrelativen Beziehungen zu den in der Klausur erreichten Punkten. Die Kennwerte der Wissensgraphen im Assoziations-Struktur-Test sind also in dieser Studie i.S. der inkrementellen Validität in der Lage, die Klausurleistung vorherzusagen, können dabei aber nicht die bereits hohe Vorhersagegüte aufgrund des Assoziationskennwertes ($r = -.783$; $p = .013$; siehe Kap. 6.1.2.1.3 auf Seite 55) übertreffen.

	gute Assoziierer (N=3)				schlechte Assoziierer (N=3)			
	N	U	V	D	N	U	V	D
<i>M</i>	42,00	93,00	0,00	4,67	10,67	31,00	0,59	2,33
<i>SD</i>	11,14	33,41	0,00	1,53	1,53	15,71	0,20	0,58
<i>Minimum</i>	32	63	0	3	9	20	0	2
<i>Maximum</i>	54	129	0	6	12	49	1	3

Abbildung 88: Mittelwerte (*M*) und Standardabweichungen (*SD*) des Wissensgraphen im Assoziations-Struktur-Tests in den Extremgruppen (gem. Assoziationskennwert) in Studie 5 – Biopsychologie

6.1.3.1.4.2. Assoziations-Struktur-Test in Studie 10: Gebäudereiniger

In Studie 10 setzte Meyer (2007) den von ihm entwickelten Assoziations-Struktur-Test ein, bei dem Berufsschüler zunächst assoziierten und anschließend Paarvergleiche zwischen ihren Assoziationen vornahmen. Aus den so gewonnenen Daten konnten individuelle Wissensgraphen erstellt werden, deren statistische Kennwerte in Abbildung 89 dargestellt sind. Es zeigt sich hinsichtlich dieser Parameter eine große Varianz innerhalb der Stichprobe der 33 Gebäudereinigungs-Berufsschüler, insgesamt aber sind die Netze als wenig umfangreich zu beschreiben.

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
<i>M</i>	5,94	11,64	1,64	0,44
<i>SD</i>	3,20	10,94	0,60	0,24
Minimum	2	1	1,0	0,17
Maximum	14	39	3,0	1,00

Abbildung 89: Statistische Kennwerte (Mittelwert und Standardabweichung) des Wissensgraphen im Assoziations-Struktur-Tests in Studie 10 – Gebäudereiniger

Die Parameter der Wissensgraphen lassen sich sehr gut durch den Assoziationskennwert vorhersagen, zwischen den Kennwerten konnten hochsignifikante mittlere bis hohe korrelative Zusammenhänge gefunden werden (siehe Abbildung 90). Die besten Vorhersagen aufgrund des Assoziationskennwertes lassen sich dabei für die Anzahl der Knoten ($r = .688$; $p = .000$) bzw. für die Anzahl der Verknüpfungen oder auch Umfang treffen ($r = .690$; $p = .000$). Die relative maximale Pfadlänge als Maß für die Kompaktheit des Graphen läßt sich immerhin durch eine Korrelation mittlerer Stärke vorhersagen ($r = -.538$; $p = .001$). Zu bedenken ist dabei, daß die Knoten des Wissensgraphen aus dem Pool derjenigen Assoziationen stammen, die vom Probanden im Wortassoziationsversuch zu einem der zehn Stimuli produziert wurden. Daß die Korrelation dennoch nicht extrem hoch ausfällt, läßt sich wie folgt begründen: der Assoziationskennwert des Wortassoziationsversuchs basiert auf der Assoziationsleistung bei zehn Stimuli, im Assoziations-

Struktur-Test bezieht er sich hingegen nur auf einen Stimulus aus dieser Menge; außerdem müssen nicht alle assoziierten Konzepte notwendigerweise auch in den Wissensgraphen eingehen.

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
Assoziationskennwert	$r = .688^{**}$ ($p = .000$)	$r = .690^{**}$ ($p = .000$)	$r = .637^{**}$ ($p = .000$)	$r = -.538^{**}$ ($p = .001$)
Klausurnote	$r = -.156$ ($p > .05$)	$r = -.166$ ($p > .05$)	$r = .046$ ($p > .05$)	$r = .426^*$ ($p = .008$)
Klausurpunkte	$r = .159$ ($p > .05$)	$r = .134$ ($p > .05$)	$r = -.030$ ($p > .05$)	$r = -.454^{**}$ ($p = .008$)

Abbildung 90: Korrelationen zwischen den Kennwerten des Assoziations-Struktur-Tests und Wortassoziationsversuch bzw. Klausurleistung in Studie 10 – Gebäudereiniger

Diese gute Vorhersagegüte der Graphenkennwerte aufgrund des Assoziationskennwertes läßt sich auch anhand des Extremgruppenvergleiches illustrieren. Abbildung 91 stellt die statistischen Kennwerte der Wissensgraphen dar im Vergleich für die Gruppe mit extrem hoher vs. extrem geringer Assoziationsleistung. Daß die guten Assoziierer signifikant mehr Knoten in ihren Netzen aufweisen als die schlechten Assoziierer (Mann-Whitney-U: $z = -2,655$; $p = .007$), scheint angesichts der bereits beschriebenen Überlappungen zwischen Wortassoziationsversuch und Assoziations-Struktur-Test wenig überraschend (mit $d = 1.71$ ein starker Effekt). Es zeigt sich aber auch, daß ihre Netze signifikant mehr Verknüpfungen (Mann-Whitney-U: $z = -2,649$, $p = .007$; $d = 1.39$), einen größeren Durchmesser (Mann-Whitney-U: $z = -2,893$, $p = .007$; $d = 1.94$) und ebenso eine größere Kompaktheit haben (Mann-Whitney-U: $z = -2,551$, $p = .010$; $d = 1.82$) als die Netze der schlechten Assoziierer. Diese Strukturparameter lassen sich im Sinne eines elaborierteren Wissensbestandes interpretieren. Der Wortassoziationsversuch ist also imstande, zwischen den Leistungsstarken vs. Leistungsschwachen im Assoziations-Struktur-Test zu differenzieren.

	gute Assoziierer (N=8)				schlechte Assoziierer (N=8)			
	N	U	D	RMPL	N	U	D	RMPL
<i>M</i>	9,13	22,6	2,13	0,28	3,75	6,1	1,13	0,71
<i>SD</i>	3,14	11,2	0,64	0,05	3,14	12,5	0,35	0,33
<i>Minimum</i>	4	6	1	0,2	2	1	1	0,17
<i>Maximum</i>	14	39	3	0,3	13	37	2	1,0

Abbildung 91: Mittelwerte (M) und Standardabweichungen (SD) des Wissensgraphen im Assoziations-Struktur-Tests in den Extremgruppen (gem. Assoziationskennwert) in Studie 10 – Gebäudereiniger

Die Kennwerte der Wissensgraphen des Assoziations-Struktur-Tests ihrerseits sind im Sinne der inkrementellen Validität nur eingeschränkt in der Lage, die Leistung in der Klausur vorherzusagen. In der Stichprobe der künftigen Gebäudereiniger ist der Parameter der relativen maximalen Pfadlänge der beste Prädiktor für die Leistung der Berufsschüler in der Klausur: die Note läßt sich durch den genannten Parameter mit $r = .426$ ($p = .012$) und die erreichte Punktzahl mit $r = -.454$ ($p = .008$) signifikant vorhersagen. Die andern drei Kennwerte des Wissensgraphen erlauben

hingegen keine Vorhersage der Klausurleistung (siehe Abbildung 90). Es sein daran erinnert, daß der Wortassoziationsversuch mit $r = -.411$ ($p = .030$) ähnlich gut in der Lage war, die Klausurnote vorherzusagen (siehe Kapitel 6.1.2.1.5 auf Seite 55) – der Einsatz des Assoziations-Struktur-Tests erbringt also gegenüber dem Wortassoziationsversuch keinen Gewinn (i.S. der inkrementellen Validität) für die Güte der Vorhersage.

6.1.3.1.4.3. Assoziations-Struktur-Test in Studie 11: Bauklempner

In Studie 11 überprüfte Meyer (2007) ebenfalls den von ihm entwickelten Assoziations-Struktur-Test in einer weiteren Stichprobe von Berufsschülern. Die Kennwerte der Wissensgraphen sind in Abbildung 92 dargestellt. Es zeigen sich große interindividuelle Differenzen in den einzelnen Parametern, insbesondere bei der Anzahl der Verknüpfungen (Umfang).

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
<i>M</i>	17,71	28,64	2,50	0,30
<i>SD</i>	14,65	22,14	1,01	0,26
Minimum	2	1	1,0	0,08
Maximum	54	74	4,0	1,00

Abbildung 92: Statistische Kennwerte (Mittelwert und Standardabweichung) des Wissensgraphen im Assoziations-Struktur-Tests in Studie 11 - Bauklempner

Auch in dieser Stichprobe der 14 Berufsschüler lassen sich die Kennwerte des Wissensgraphen sehr gut durch den Kennwert im Wortassoziationsversuch vorhersagen: insbesondere die Anzahl der Knoten ($r = .712$; $p = .002$) sowie der Durchmesser des Netzes ($r = .710$; $p = .002$) erbringen hohe Korrelationskoeffizienten. Der Umfang sowie die relative maximale Pfadlänge lassen sich immerhin durch signifikante mittlere korrelative Zusammenhänge vorhersagen (siehe auch Abbildung 93).

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
Assoziationskennwert	$r = .712^{**}$ ($p = .002$)	$r = .507^*$ ($p = .032$)	$r = .710^{**}$ ($p = .002$)	$r = -.527^*$ ($p = .027$)
Klausurnote	$r = -.479^*$ ($p = .042$)	$r = -.406$ ($p > .05$)	$r = -.323$ ($p > .05$)	$r = .121$ ($p > .05$)

Abbildung 93: Korrelationen zwischen den Kennwerten des Assoziations-Struktur-Tests und Assoziationskennwerten bzw. Klausurleistung in Studie 11 – Bauklempner

Der Wortassoziationsversuch ist also sehr gut in der Lage, die Kennwerte des Assoziations-Struktur-Tests vorherzusagen, was auch zu erwarten war. Ebenso lassen sich die Leistungsunterschiede hinsichtlich der Parameter des Assoziations-Struktur-Tests auch im Extremgruppenvergleich abbilden. Erwartungsgemäß zeigen sich Unterschiede in den statistischen Kennwerten beider Gruppen, welche in Abbildung 94 zusammengetragen sind: die Wissensgraphen derjenigen Probanden, die einen extrem hohen Assoziationskennwert aufwiesen, hatten nicht nur eine deutlich höhere Anzahl Knoten und Verknüpfungen im Netz, sondern auch eine höhere Dichte

und Kompaktheit als diejenigen, die im Wortassoziationsversuch einen extrem geringen Assoziationskennwert erreicht haben. Diese augenscheinlichen Unterschiede werden allerdings nicht statistisch signifikant, was wohl dem geringen Stichprobenumfang zuzuschreiben ist, dennoch läßt sich die Unterscheidung zwischen den Extremgruppen stützen anhand der durchweg großen Effektstärken ($d_N=2.43$, $d_U=1.28$, $d_D=2.20$, $d_{RMPL}=1.55$).

	gute Assoziierer (N=3)				schlechte Assoziierer (N=3)			
	N	U	D	RMPL	N	U	D	RMPL
<i>M</i>	38,0	52,0	3,67	0,11	7,33	19,33	1,67	0,56
<i>SD</i>	15,52	19,9	0,58	0,03	8,84	30,04	1,15	0,41
<i>Minimum</i>	23	35	3,0	0,08	2	1	1,0	0,19
<i>Maximum</i>	54	74	4,0	0,14	17	54	3,0	1,00

Abbildung 94: Mittelwerte (M) und Standardabweichungen (SD) des Wissensgraphen im Assoziations-Struktur-Tests in den Extremgruppen (gem. Assoziationskennwert) in Studie 11 – Bauklempner

Die Kennwerte der Wissensgraphen des Assoziations-Struktur-Tests ihrerseits sind eingeschränkt in der Lage, die Leistung in der Klausur vorherzusagen (siehe Abbildung 93). Die Anzahl der Knoten im Wissensnetz bietet mit $r = -.479$ ($p = .042$) eine signifikante mittlere Vorhersagegüte für die Klausurnote an. Die Parameter Umfang, Dichte des Graphen sowie die relative maximale Pfadlänge lassen keine Vorhersage der Klausurleistung zu. Es sei daran erinnert, daß der Wortassoziationsversuch alleine nicht in der Lage war, die Klausurnote vorherzusagen (siehe Kap. 6.1.2.1.6 auf Seite 55). Somit ergibt sich hier im Sinne der inkrementellen Validierung durch den Einsatz des Assoziations-Struktur-Tests ein Erkenntnisgewinn gegenüber dem der alleinigen Verwendung des Wortassoziationsversuchs.

6.1.3.1.4.4. Assoziations-Struktur-Test in Studie 14: Seminar Wissensmanagement

Zentrales Anliegen der Studie 14 – Seminar Wissensmanagement – war ein Beitrag zur Verfahrensvalidierung des Assoziations-Struktur-Tests. Die in Abbildung 95 dargestellten statistischen Kennwerte für die Wissensgraphen zeigen zu beiden Erhebungszeitpunkten große interindividuelle Unterschiede, v.a. in den Parametern Knoten (N) und Umfang (U) der Graphen.

	T1 (vor Seminarbeginn)				T2 (Seminarende)			
	N	U	D	RMPL	N	U	D	RMPL
<i>M</i>	10,00	27,00	2,47	0,34	15,30	54,26	2,79	0,23
<i>SD</i>	4,22	23,42	1,02	0,22	6,09	34,55	0,98	0,10
<i>Minimum</i>	3	2	1	0,14	7	4	2	0,10
<i>Maximum</i>	18	91	5	1,00	30	133	5	0,50

Abbildung 95: Mittelwerte (M) und Standardabweichungen (SD) des Wissensgraphen im Assoziations-Struktur-Tests für die gesamte Stichprobe zu den Meßzeitpunkten in Studie 14 – Seminar Wissensmanagement

Wie bereits in Kapitel 4.11 (Seite 55 ff.) dargestellt, steigt die Anzahl der Assoziationen im Wortassoziationsversuch am Ende des Seminars signifikant an. Da der Assoziations-Struktur-Test auf den individuellen Assoziationen der Probanden aufbaut, zeichnet sich diese Entwicklung hin zum zweiten Erhebungszeitpunkt auch im Kennwert der Knoten des Wissensgraphen ab. Aus dem

gleichen Grunde ist die korrelative Beziehung zwischen diesen beiden Parametern – erwartungsgemäß – extrem hoch (siehe Abbildung 96). Interessanterweise fallen die Korrelationskoeffizienten zwischen dem Assoziationskennwert und den Parametern des jeweiligen Wissensgraphen zum ersten Erhebungszeitpunkt sogar noch höher aus als bei der zweiten Erhebung (siehe Abbildung 96), obwohl zu letzterem Zeitpunkt die größere Expertise bei den Probanden vorlag. Eine mögliche Erklärung für diesen Effekt mag in der absoluten Anzahl der Assoziationen liegen. Es sei daran erinnert, daß Konzepte beim Assoziations-Struktur-Test nicht direkt in den Graphen aufgenommen werden, sondern vermittelt werden durch die Paarvergleiche: bei wenigen assoziierten Konzepten ist es wahrscheinlicher, daß alle stark miteinander zusammenhängen und auf diese Weise in den Wissensgraphen eingehen, während bei einer umfangreichen Assoziationsliste in größerem Maße Differenzierungen in der Beziehungsstärke zwischen Konzepten wahrscheinlich sind. Da aber sämtliche Koeffizienten signifikante mittlere bis sehr hohe korrelative Zusammenhänge abbilden (siehe Abbildung 96), soll an dieser Stelle nicht weiter darauf eingegangen werden.

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
Assoziationskennwert zu Seminarbeginn T1 (mit Wissensgraph aus T1)	$r= 1.000^{**}$ ($p= .000$)	$r= .756^{**}$ ($p= .000$)	$r= .456^*$ ($p= .025$)	$r= -.664^{**}$ ($p= .001$)
Assoziationskennwert zu Seminarende T2 (mit Wissensgraph aus T2)	$r= .926^{**}$ ($p= .000$)	$r= .707^{**}$ ($p= .000$)	$r= .425^*$ ($p= .035$)	$r= -.622^{**}$ ($p= .002$)
Punkte im Multiple-Choice-Test T2 (mit Wissensgraph aus T2)	$r= -.269$ ($p> .10$)	$r= -.096$ ($p> .10$)	$r= .054$ ($p> .10$)	$r= .078$ ($p> .10$)

Abbildung 96: Korrelationen zwischen den Kennwerten des Assoziations-Struktur-Tests, Assoziationskennwerten und Leistung im Multiple-Choice-Test in Studie 14 – Seminar Wissensmanagement

Diese sehr hohen Korrelationen zwischen den Verfahren Wortassoziationsversuch und Assoziations-Struktur-Test lassen sich eindrucksvoll veranschaulichen, wenn Probanden mit geringem vs. hohem Assoziationskennwert hinsichtlich ihrer Parameter im Assoziations-Struktur-Test verglichen werden. Zu Gunsten der Übersichtlichkeit soll an dieser Stelle der Extremgruppenvergleich nur für den zweiten Erhebungszeitpunkt dargestellt werden. Die Mittelwerte der Graphen-Kennwerte in den Extremgruppen sind in Abbildung 97 dargestellt. Dabei zeigt sich, daß die guten Assoziierer deutlich bessere Ergebnisse beim Assoziations-Struktur-Test aufweisen als die schlechten Assoziierer. Dies spiegelt sich sowohl in den Mittelwerten wider als auch im Range der Parameter: Probanden mit einem hohen Assoziationskennwert weisen im Vergleich zu denen mit geringem Assoziationskennwert einen Wissensgraphen mit signifikant mehr Knoten auf (Mann-Whitney-U: $z=-2,929$; $p= .002$; $d=4.59$), signifikant mehr Verknüpfungen zwischen den Knoten (Mann-Whitney-U: $z=-2,739$; $p= .002$; $d=2.50$) sowie einen signifikant geringe maximale Pfadlänge, also einen kompakteren Graphen (Mann-Whitney-U: $z=-2,384$; $p= .017$; $d=1.85$). Lediglich der augenscheinliche höhere Wert des mittleren Durchmessers der guten Assoziierer konnte nicht statistisch abgesichert werden, weisen aber mit $d=1.21$ eine bedeutsame Effektstärke auf.

	gute Assoziierer (N=6)				schlechte Assoziierer (N=6)			
	N	U	D	RMPL	N	U	D	RMPL
<i>M</i>	22,50	86,40	3,40	0,16	8,67	22,00	2,33	0,32
<i>SD</i>	3,99	34,86	1,14	0,07	1,51	10,84	0,52	0,10
<i>Minimum</i>	21	43	2	0,10	7	4	2	0,22
<i>Maximum</i>	32	133	5	0,26	10	35	3	0,50

Abbildung 97: Mittelwerte (M) und Standardabweichungen (SD) des Wissensgraphen im Assoziations-Struktur-Tests in den Extremgruppen (gem. Assoziationskennwert) in Studie 14 – Seminar Wissensmanagement (zu T2)

Die Parameter des Assoziations-Struktur-Tests selbst sind in dieser Studie nicht geeignet, um i.S. der inkrementellen Validität die Leistung eines Probanden im Multiple-Choice-Test vorherzusagen. Exemplarisch sei hier auf den zweiten Erhebungszeitpunkt verwiesen, die entsprechenden korrelativen Beziehungen zwischen den Assoziations-Struktur-Test-Parametern und Multiple-Choice-Test-Leistung sind in Abbildung 96 aufgeführt. Die Korrelationskoeffizienten können keine Beziehung abbilden, allerdings sei darauf verwiesen, daß auch der Assoziationskennwert in dieser Studie die Leistung im Multiple-Choice-Test nicht vorhersagen konnte (siehe Kap. 6.1.3.1.5.1 auf Seite 55).

6.1.3.1.4.5. Zusammenfassung und Befundintegration zur Vorhersage des Assoziations-Struktur-Tests

Die Parameter des Assoziations-Struktur-Tests lassen sich sehr gut durch den Kennwert des Wortassoziationsversuchs vorhersagen. Dies ist erwartungsgemäß, da der Wortassoziationsversuch als ein Teilverfahren des Assoziations-Struktur-Tests angesehen werden kann, seine Ergebnisse sind die Grundlage für das weitere Vorgehen im Assoziations-Struktur-Test. Somit sind die Assoziationen und die Knoten im Graphen in hohem Maße deckungsgleich. Dies trifft insbesondere zu, wenn der Assoziationsteil des Assoziations-Struktur-Tests nur aus einem Stimulus besteht. Umfaßt der Assoziationsteil hingegen mehrere Stimuli, so stellen die Knoten des Wissensgraphen immerhin eine Teilmenge der Assoziationen des Wortassoziationsversuchs dar, welche die ebenso hohen Korrelationen erklären. Eng gekoppelt an die Anzahl von Knoten ist die Anzahl der vorgenommenen Verknüpfungen zwischen den Knoten. Dieser Umstand findet sich auch in den korrelativen Beziehungen beider Verfahren (Wortassoziationsversuch und Assoziations-Struktur-Test) wieder, welche sich durchgehend durch mittlere bis sehr hohe hochsignifikante Korrelationskoeffizienten auszeichnen (siehe auch im oberen Bereich von Abbildung 98). Die schwächste Vorhersage war mit $r = -.527$ ($p = .027$) in Studie 11 durch die relative maximale Pfadlänge gegeben, die stärkste Beziehung in Studie 14 durch die Anzahl der Knoten mit $r = .926$ ($p = .000$). Faßt man die Effektstärken der einzelnen Studien als mittlere gewichtete Effektstärke zusammen, so ergeben sich durchweg starke Effekte: für die Anzahl der Knoten $d = .77$, für den Umfang der Netze $d = .66$, für den Durchmesser $d = .62$ sowie für die relative maximale Pfadlänge $d = .56$.

Führt man die Befunde der vier hier vorgestellten Studien zusammen, so lassen sich die z-transformierten Daten von insgesamt 76 Probanden heranziehen (siehe Abbildung 98, unterer Bereich). Im Ergebnis der Befundintegration läßt sich die Anzahl der Knoten im Wissensgraphen

am sichersten durch den Assoziationskennwert vorherzusagen, mit $r = .711$ ($p = .000$) steht ein hoher und hochsignifikanter Korrelationskoeffizient zur Verfügung. Die Anzahl der Kanten im Wissensgraphen ($r = .636$; $p = .000$) sowie dessen Durchmesser ($r = .620$; $p = .000$) lassen ebenso gut durch den Assoziationskennwert vorherzusagen. Die relative maximale Pfadlänge als Maß für die Kompaktheit des Graphen kann mit einer mittleren Korrelation von $r = -.556$ ($p = .000$) ebenfalls gut prognostiziert werden. Interpretiert man diese Korrelationskoeffizienten als Effektstärken, so lassen sich hinsichtlich aller Parameter große Effekte beschreiben. Bei alleiniger Betrachtung der Extremstichproben aus allen Einzelstudien ($N=42$) werden die bestehenden Zusammenhänge etwas mehr akzentuiert, indem zu allen Parametern höhere und durchweg hoch signifikante Korrelationskoeffizienten gefunden wurden. Zusammenfassend ist also festzuhalten, daß der Assoziationskennwert sehr gut in der Lage ist, die Parameter des Assoziations-Struktur-Tests vorherzusagen.

	Anzahl der Knoten (N)	Umfang (U)	Durchmesser (D)	relative maximum path length (RMPL)
<i>Einzelstudien</i>				
5 - Biopsychologie	$r = .833^{**}$ ($p = .003$)	$r = .699^{*}$ ($p = .018$)	$r = .809^{**}$ ($p = .004$)	---
10- Gebäudereiniger	$r = .688^{**}$ ($p = .000$)	$r = .690^{**}$ ($p = .000$)	$r = .637^{**}$ ($p = .000$)	$r = -.538^{**}$ ($p = .001$)
11 - Bauklempner	$r = .712^{**}$ ($p = .002$)	$r = .507^{*}$ ($p = .032$)	$r = .710^{**}$ ($p = .002$)	$r = -.527^{*}$ ($p = .027$)
14 - Seminar Wissensmanagement (T2)	$r = .926^{**}$ ($p = .000$)	$r = .707^{**}$ ($p = .000$)	$r = .425^{*}$ ($p = .035$)	$r = -.622^{**}$ ($p = .002$)
<i>Befundintegration</i>				
gesamte Stichprobe (N=76)	$r = .711^{**}$ ($p = .000$)	$r = .636^{**}$ ($p = .000$)	$r = .620^{**}$ ($p = .000$)	$r = -.556^{***}$ ($p = .000$)
Extremstichproben (N=42)	$r = .776^{**}$ ($p = .000$)	$r = .660^{*}$ ($p = .000$)	$r = .774^{**}$ ($p = .000$)	$r = -.637^{**}$ ($p = .000$)

Abbildung 98: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerte und Parameter des Assoziations-Struktur-Test) sowie Befundintegration als Korrelationen zwischen den standardisierten Parametern des Assoziations-Struktur-Test (z-Werte) und Assoziationskennwerten (z-Werte)

6.1.3.1.5. Multiple-Choice-Tests als konstruktverwandter Test

In zwei Studien wurde ein Multiple-Choice-Test entwickelt und eingesetzt, der als Validitätskriterium für die Assoziationsleistung dienen kann. Im Folgenden sollen Befunde dieser beiden Studien vorgestellt und anschließend zusammengeführt werden.

6.1.3.1.5.1. Multiple-Choice-Test in Studie 14: Seminar Wissensmanagement

In Studie 14 – Seminar Wissensmanagement – wurde zu zwei Meßzeitpunkten assoziiert sowie jeweils ein kurzer Multiple-Choice-Test vorgelegt. Abbildung 99 stellt die statistischen Kennwerte des Multiple-Choice-Tests dar, beide Test-Kennwerte sind normalverteilt. Es wird deutlich, daß die

durchschnittliche Leistung im Multiple-Choice-Test am Seminarende mit $M=6,25$ ($SD=1,33$) erreichten Punkten deutlich höher liegt als vor dem Seminar ($M=3,90$; $SD=1,97$). Dieser Unterschied ist sehr signifikant ($t(19)=-7,378$; $p=.000$). Anhand des Multiple-Choice-Tests kann also belegt werden, daß die Teilnahme am Seminar zu einer signifikanten Erhöhung des Wissens über das behandelte Thema geführt hat. Wie bereits beschrieben (siehe Kapitel 4.11, Seite 55 ff.) erhöhte sich auch die Assoziationsleistung im Laufe des Seminars signifikant: von $M=10,00$ ($SD=4,22$) Assoziationen zu Seminarbeginn auf $M=16,35$ ($SD=7,50$) Nennungen bei Seminarende.

	T1 (vor Seminarbeginn)	T2 (Seminarende)
<i>M</i>	3,90	6,25
<i>SD</i>	1,97	1,33
Minimum	1	4
Maximum	9	9

Abbildung 99: Kennwerte des Multiple-Choice-Tests – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 14 – Seminar Wissensmanagement

Zu Seminarbeginn besitzt der Assoziationskennwert keine Vorhersagekraft in Hinblick auf die Leistung im Multiple-Choice-Test (siehe Abbildung 100). Dies ist insofern nachvollziehbar, als daß der Wissensstand bei den Probanden zu diesem Zeitpunkt noch sehr eingeschränkt ist, außerdem ist der mit zehn Fragen sehr kurze Multiple-Choice-Test auch unter Anwendung von Heuristiken oder schlicht durch Raten zu bewältigen. Am Seminarende erweist sich der Assoziationskennwert jedoch als ebenso ungeeignet für die Vorhersage der Leistung im Multiple-Choice-Test. Mit einem nicht-signifikanten Korrelationskoeffizienten von $r=-.357$ ($p=.061$) zeigt sich sogar eine gegenläufige Tendenz: eine hohe Leistung im Wortassoziationsversuch geht scheinbar eher mit einem geringeren Punktwert im Multiple-Choice-Test einher. Aufgrund der unklaren und in dieser Weise auch nicht erklärbaren Beziehungen zwischen Multiple-Choice-Test und Wortassoziationsversuch soll an dieser Stelle auf die Illustration anhand von Extremgruppen verzichtet werden.

	T1 (vor Seminarbeginn)	T2 (Seminarende)
Assoziationskennwert	$r=.165$ ($p>.20$)	$r=-.357$ ($p>.05$)

Abbildung 100: Korrelationen des Multiple-Choice-Test (Punkte) mit dem Assoziationskennwert in Studie 14 – Seminar Wissensmanagement

6.1.3.1.5.2. Multiple-Choice-Test in Studie 15: Wissen über Arbeitsschutz

Studie 15 – Wissen über Arbeitsschutz – stellt in der vorliegenden Arbeit mit Abstand die größte Stichprobe dar: mit $N=472$ ist der Datenpool für arbeitspsychologische Studien und noch dazu bei wissensdiagnostischen Fragestellungen außerordentlich groß. Auf der andere Seite ist anzumerken, daß die Probanden insgesamt (im Gegensatz zu den anderen Studien) am wenigsten über Expertise in der untersuchten Wissensdomäne verfügten, es bestand sogar a priori die Hypothese, daß sie nichts oder relativ wenig über die interessierende Wissensdomäne Arbeits- und Gesundheitsschutz wissen (siehe auch Kapitel 6.1.4.3, Seite 55 ff).

Im Multiple-Choice-Test dieser Studie waren maximal 58 Punkte zu erreichen. Über die gesamte Stichprobe wurden im Mittel $M=27,69$ Punkte ($SD=8,85$) erreicht, also etwas weniger als die Hälfte

des definierten Soll-Wissens. Allerdings wird dabei auch ein sehr großer Range abgedeckt, welcher von null bis 50 richtige Antworten reicht (siehe Abbildung 101).

	gesamte Stichprobe (N=472)	Extremstichproben	
		gute Assoziierer (N=57)	schlechte Assoziierer (N=67)
<i>M</i>	27,69	30,28	24,88
<i>SD</i>	8,85	5,96	10,82
<i>Minimum</i>	0	19	0
<i>Maximum</i>	50	44	44

Abbildung 101: Mittelwerte (M) und Standardabweichungen (SD) des Multiple-Choice-Test in Studie 15 – Wissen über Arbeitsschutz

Die Vorhersage dieses Punktwertes im Multiple-Choice-Test ist auf Grundlage des Assoziationskennwertes nur sehr ungenau möglich. Mit $r = .160$ ($p = .000$) ist der Korrelationskoeffizient zwar signifikant, aber schwach ausgeprägt. Allerdings ist auch daran zu erinnern, daß der Assoziationskennwert über die gesamte Stichprobe mit $SD = 1,56$ ($M = 2,76$) eine sehr geringe Streuung aufweist. Angesichts dessen ist auch diese schwache Korrelation beachtlich. Diese geringe Streuung ist vermutlich darauf zurückzuführen, daß – wie bereits erwähnt – diese Stichprobe größtenteils nicht über das Fachwissen verfügt, wie man es üblicherweise in arbeitspsychologischen Wissensdiagnosen annimmt. Während man beim Multiple-Choice-Test noch durch Wiedererkennen, Abwägen von Wahrscheinlichkeiten oder schlichtes Raten der richtigen Antworten Punkte erreichen kann, bleiben dem Probanden beim Wortassoziationsversuch diese Strategien verwehrt. Das Assoziieren als Verfahren der freien Reproduktion kann eben nur dort etwas produzieren, wo Wissen (oder zumindest Halbwissen) dahinter steht. Zur Illustration dieses Umstandes sei auf Abbildung 102 verwiesen; aufgrund der großen Stichprobe in dieser Studie wird im Scatterplot recht deutlich, wie die individuellen Kennwerte der Probanden im Wortassoziationsversuch bzw. im Multiple-Choice-Test zusammenhängen: ein geringer Assoziationskennwert kann ebenso mit einem geringen wie auch einem mittleren oder hohen Punktwert im Multiple-Choice-Test einhergehen. Ein hoher Assoziationskennwert hingegen geht einher mit einer hohen Leistung im Multiple-Choice-Test, so daß beide Testleistungen nur durch das Vorhandensein von Wissen erklärt werden können (statt durch Raten). Die Form der Punktwolke der individuellen Kennwerte in Abbildung 102 macht ebenso den geringen Korrelationskoeffizienten (s.o.) zwischen beiden Parametern nachvollziehbar.

Berücksichtigt man lediglich die Probanden der Extremstichproben ($N = 124$), so wird die schwache Beziehung zwischen Assoziationskennwert und Leistung im Multiple-Choice-Test etwas deutlicher: mit $r = .307$ ($p = .000$) ist der Korrelationskoeffizient nach Spearman zwar immer noch als gering zu interpretieren, aber dennoch deutlich höher als bei Betrachtung der Gesamtstichprobe. Diese insgesamt schwache Beziehung zwischen Wortassoziationsversuch und Multiple-Choice-Test bildet sich auch im Extremgruppenvergleich ab. Die in Abbildung 101 dargestellten statistischen Kennwerte der beiden Extremgruppen zeigen, daß diejenigen Probanden mit einem höheren Assoziationskennwert im Mittel auch einen höheren Punktwert im Multiple-Choice-Test erreichten, während diejenigen mit geringer Assoziationsleistung auch im Multiple-Choice-Test schlechter abschnitten. Dieser Unterschied ist signifikant ($t(122) = 3,358$; $p = .001$), mit $d = .61$ liegt auch ein

großer Effekt vor, jedoch differenziert er beide Extremgruppen nicht so eindeutig wie erwartet. V.a. der große Range bzw. die Standardabweichung in der Gruppe der schlechten Assoziierer weisen auf eine sehr heterogene Substichprobe hin. Es gibt auch schlechte Assoziierer, die einen hohen Punktwert im Multiple-Choice-Test erreichen, während bei den guten Assoziierern keiner der Probanden eine extrem niedrige Multiple-Choice-Test-Leistung aufweist. Zusammenfassend ist zu sagen, daß der Assoziationskennwert in dieser Studie eine schwache Vorhersage des Leistungskennwertes im Multiple-Choice-Test erlaubt.

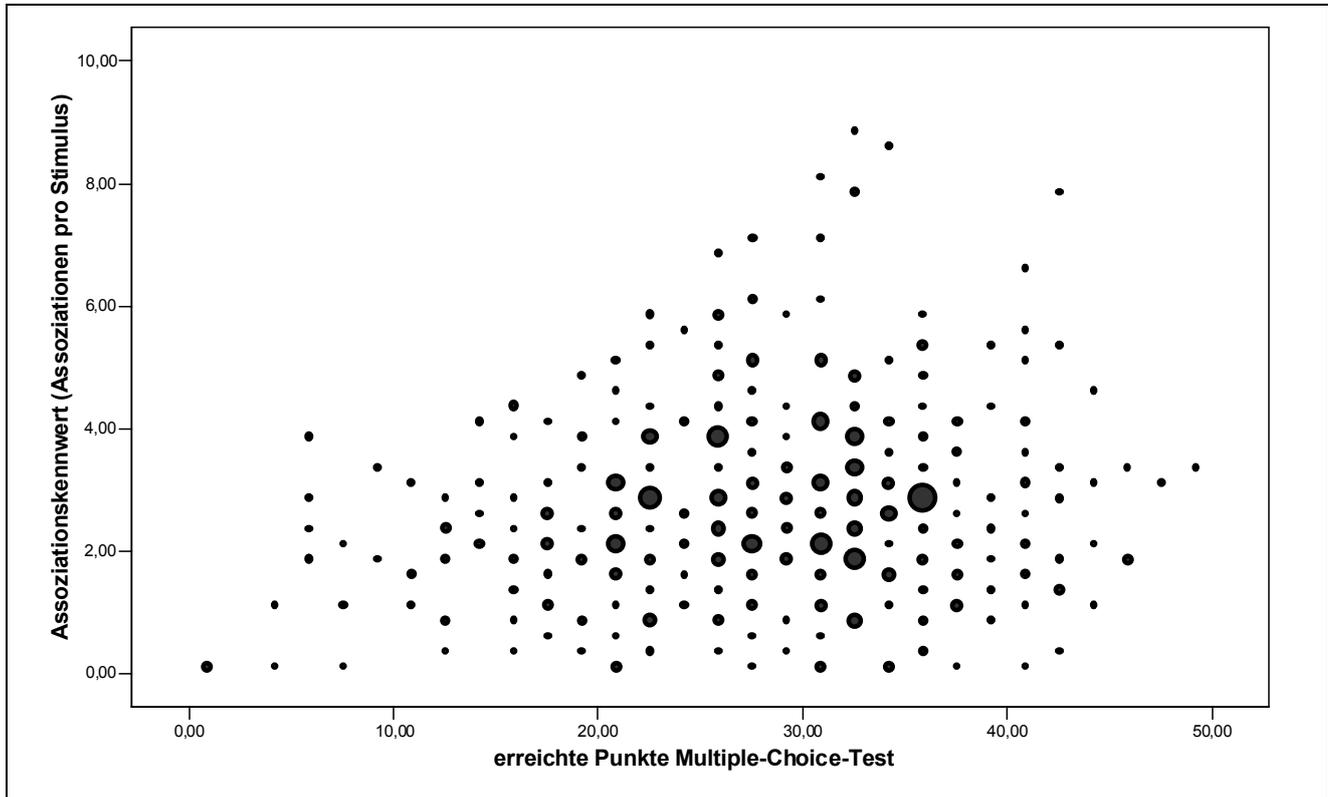


Abbildung 102: Scatterplot zum Verhältnis der individuellen Assoziationskennwerte zum jeweiligen Multiple-Choice-Test-Ergebnis in Studie 15 – Wissen über Arbeitsschutz (die Größe eines Datenpunktes ist ein Indikator für die Anzahl der dahinterstehenden Probanden)

6.1.3.1.5.3. Zusammenfassung und Befundintegration zur Vorhersage des Multiple-Choice-Tests

Die Befunde zur Validierung des Wortassoziationsversuchs an der Leistung in einem Multiple-Choice-Test sind wenig ermutigend (siehe Abbildung 103). In den zwei hier vorgestellten Studien konnte zum einen eine schwache signifikante Vorhersagekraft des Assoziationskennwertes ($r = .160$; $p = .000$) belegt werden, zum anderen war keine bedeutsame Beziehung zwischen den Kennwerten nachzuweisen, es deutete sich sogar ein erwartungswidriger Zusammenhang an. Die mittlere gewichtete Effektstärke über beide Studien belegt dann auch mit $d = .14$, daß offenbar kein bedeutsamer Zusammenhang zwischen Leistung im Wortassoziationsversuch und im Multiple-Choice-Test besteht.

Führt man die z-transformierten Ergebnisse beider Studien zusammen, so findet sich immer noch ein schwacher, signifikanter Korrelationskoeffizient von $r = .140$ ($p = .001$). Vermutlich profitiert dieser Koeffizient von der sehr großen Stichprobe der Studie 15 – Wissen über Arbeitsschutz, welche die zusammengeführte Stichprobe ($N = 491$) dominiert. Interpretiert man diesen

Korrelationskoeffizienten i.S. der Effektstärke, so zeichnet sich hier immerhin noch ein schwacher Effekt ab. Legt man lediglich die Probanden mit den geringsten bzw. höchsten Assoziationskennwerten aus beiden Studien zusammen (Extremstichprobe mit $N=130$), so erhöht sich die Prognosequalität dieses Kennwertes deutlich, mit $r = .263$ ($p = .001$) ist dieser Korrelationskoeffizient allerdings immer noch als schwach zu interpretieren.

Multiple-Choice-Test	
<i>Einzelstudien</i>	
14 - Seminar Wissensmanagement (T2)	$r = -.357$ ($p > .05$)
15 - Wissen über Arbeitsschutz	$r = .160$ *** ($p = .000$)
<i>Befundintegration</i>	
gesamte Stichprobe ($N=492$)	$r = .140$ ** ($p = .001$)
Extremgruppen ($N=130$)	$r = .263$ ** ($p = .001$)

Abbildung 103: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerte und Punkte im Multiple-Choice-Test) sowie Befundintegration als Korrelationen zwischen den standardisierten Multiple-Choice-Test-Leistungen (z-Werte) und Assoziationskennwerten (z-Werte)

Zusammenfassend läßt sich also festhalten, daß der Kennwert des Wortassoziationsversuchs eingeschränkt in der Lage ist, ungefähr die Leistung in einem Multiple-Choice-Test vorherzusagen. Grundsätzlich ist daran zu erinnern, daß Multiple-Choice-Tests sehr anfällig sind gegenüber Ratenendenzen der Probanden bzw. dem Einsatz von Heuristiken, ohne daß diesen Strategien tatsächlich fundiertes Wissen zugrunde liegt. Ein hoher Kennwert im Wortassoziationsversuch hingegen kann nur auf entsprechendes Wissen (oder zumindest Halbwissen) zurückzuführen sein – Raten ist bei der freien Reproduktion eine wenig erfolgreiche Strategie. Aufgrund der aufgeführten Einwände sind Multiple-Choice-Tests offenbar weniger gut geeignet als Kriterium für die Verfahrensvalidierung und die hier berichteten diesbezüglichen Befunde eher als nachrangig zu bewerten.

6.1.3.2. Diskriminante Konstruktvalidität

Liebs (2003) bemerkt, daß in ihrer Studie (Studie 12 – Handelsunternehmen) diejenigen Mitarbeiter mit der größten Schweiß erfahrung (besitzen Schweißpaß) zwar über das umfangreichste Wissen über die Domäne Schweißen verfügen sollten, sich aber nicht in der oberen Extremgruppe gemäß ihrer Assoziationsleistung befanden. Neben motivationalen Faktoren vermutet sie als Ursache für diesen Umstand, daß die Probanden mit der Anforderung des Wortassoziationsversuchs, nur in Begriffen zu antworten, überfordert waren, oder aber das Wissen über das Schweißen aufgrund der Geübtheit des Vorgangs nicht mehr in vollem Umfang verbalisierbar war. Auch andere Versuchsleiter berichten immer wieder über Schwierigkeiten der Probanden, flüssig in Stichworten zu antworten.

Zur Klärung der diskriminanten Konstruktvalidität soll daher der Einfluß des Personenmerkmals Wortflüssigkeit thematisiert werden. Um abzusichern, daß die Assoziationsleistung eines Probanden das Konstrukt Fachwissen erfaßt und nicht durch die allgemeine Fähigkeit zur flüssigen Wortproduktion determiniert ist, wurde in drei Studien ein standardisierter Test zur Wortflüssigkeit

eingesetzt. Dazu wurde der Untertest „Wortflüssigkeit“ aus dem WILDE-Intelligenztest (Jäger und Althoff, 1994) herangezogen, welcher gemäß der Autoren die quantitative Seite i.S. der Flüssigkeit assoziativer Worteinfälle erfassen soll.

Der Untertest zur Wortflüssigkeit aus dem WILDE-Intelligenztest (Jäger und Althoff, 1994) ist ein Speed-Test. Es wurde für die in der vorliegenden Arbeit vorgestellten Untersuchungen die Kurzform gewählt, um den zeitlichen Aufwand für die Probanden möglichst gering zu halten. Die Aufgabe in diesem Test läßt sich wie folgt beschreiben: den Probanden werden ein Anfangs- und ein Endbuchstabe vorgegeben, sie sollen innerhalb einer begrenzten Zeit möglichst viele Worte finden, die mit den jeweiligen Buchstaben beginnen bzw. aufhören, wobei jegliche Abwandlungen, Konjugationen oder sonstige in der deutschen Sprache üblichen Erscheinungsformen dieser Worte erlaubt sind. Der Proband hat zwei Übungsbeispiele (mit limitierter Bearbeitungszeit) zu absolvieren und anschließend fünf solcher Aufgabenstellungen mit jeweils 60 Sekunden Bearbeitungszeit. Die beiden Beispielaufgaben gehen nicht in die Wertung ein. Es werden alle gültigen Worte der fünf Testaufgaben aufsummiert, diesem Rohwert wird ein Normwert entsprechend der Normtabelle (siehe Jäger und Althoff, 1994) zugewiesen, wobei ein Standardwert von 100 dem Durchschnitt entspricht. Die bevölkerungsrepräsentativen Normtabellen unterscheiden zwischen verschiedenen Altersbereichen bis zu einem maximalen Alter von 38 Jahren. Für Probanden höheren Lebensalters wurde ebenfalls der letztgenannte Altersbereich zugrunde gelegt, dabei kann davon ausgegangen werden, daß die Standardwerte diese Probanden eher zu streng bewerten.

6.1.3.2.1. Wortflüssigkeit in Studie 1: Modus

In Studie 1 – Modus der Erhebung – wurde die mündliche der schriftlichen Durchführung des Wortassoziationsversuchs miteinander verglichen sowie Stimuli des Allgemeinwissensbereich einer spezialisierten Wissensdomäne gegenübergestellt (siehe auch zum Versuchsaufbau Kapitel 4.1.2 auf Seite 55). Als Kontrollvariable wurde dabei auch das Merkmal der Wortflüssigkeit erhoben. In dieser Stichprobe von 30 Probanden ist der Standardwert des WIT-Testes normalverteilt, zwischen einem Minimum von 87 und einem Maximum von 120 wurde im Mittel ein Standardwert von 101,80 erreicht ($SD=8,92$).

Abbildung 104 listet die Korrelationskoeffizienten auf, welche der WIT-Standardwert mit dem Assoziationskennwert über den gesamten Wortassoziationsversuch hinweg bzw. zu seinen Teilbereichen erbringt. Im Wesentlichen werden schwache, aber nicht signifikante Beziehungen zwischen der Wortflüssigkeit und der Assoziationsleistung sichtbar: sowohl hinsichtlich des Assoziationskennwertes über den ganzen Versuch hinweg (insgesamt 40 Stimuli) als auch in Bezug auf den schriftlichen Teil, den Allgemeinwissensbereich sowie den Spezialwissensbereich (jeweils 20 Stimuli). Lediglich der mündliche Teil des Wortassoziationsversuchs erbringt einen schwachen, aber signifikanten korrelativen Zusammenhang zum WIT-Standardwert ($r = .348$; $p = .030$). Differenziert man diesen mündlichen Teil weiter in seinen Allgemeinwissensbestandteil sowie seinen Spezialwissensbereich (jeweils 10 Stimuli), so zeigt sich eine korrelative Beziehung zur Wortflüssigkeit ausschließlich im Teil des Allgemeinwissens: mit $r = .336$ ($p = .035$) kann der WIT-Standardwert den Assoziationskennwert im mündlichen Teil des Allgemeinwissens signifikant vorhersagen. Das mündliche Spezialwissen hingegen ist nicht durch die Wortflüssigkeit

determiniert ($r = .190$; $p > .05$). Da in dieser Studie kein Außenkriterium zur Verfügung stand, können hier keine Partialkorrelationen berichtet werden.

	WIT-Standardwert
Assoziationskennwert über gesamten Wortassoziationsversuch	$r = .273$ ($p > .05$)
Wortassoziationsversuch mündlicher Teil	$r = .348^*$ ($p = .030$)
Wortassoziationsversuch schriftlicher Teil	$r = .212$ ($p > .05$)
Wortassoziationsversuch Allgemeinwissen	$r = .263$ ($p > .05$)
Wortassoziationsversuch Spezialwissen	$r = .218$ ($p > .05$)

Abbildung 104: Korrelationen zwischen Wortflüssigkeit und Assoziationskennwert (differenziert nach verschiedenen Teilbereichen) in Studie 1 – Modus der Erhebung

Im Extremgruppenvergleich, bei dem die Probanden mit hohem den mit geringem Assoziationskennwert gegenüber gestellt werden, zeigen die guten Assoziierer mit $M=103,71$ ($SD=8,81$) einen signifikant höheren WIT-Standardwert als die schlechten Assoziierer, welche im Mittel einen Standardwert von $M=93,38$ ($SD=5,48$) erreichen (Mann-Whitney-U: $z=-2,435$; $p = .014$). Zusammenfassend ist zu sagen, daß die Befunde zum Einfluß der Wortflüssigkeit in dieser Studie uneindeutig sind. Es deutet sich ein schwacher, aber nicht statistisch bedeutsamer Zusammenhang zwischen der Assoziationsleistung und dem Merkmal Wortflüssigkeit an, bei genauerer Betrachtung besteht dieser Zusammenhang jedoch vor allem bei Stimuli aus dem Allgemeinwissensbereich, welche mündlich bearbeitet werden, nicht jedoch für Stimuli aus einer spezialisierten Wissensdomäne. Für den Einsatz des Wortassoziationsversuchs im Bereich des Fachwissens lassen sich aus diesen Ergebnissen somit noch keine Einschränkungen ableiten hinsichtlich der Überlegenheit eines bestimmten Durchführungsformates.

6.1.3.2.2. Wortflüssigkeit in Studie 2: psychotherapeutische Ambulanz

Im Retest-Design von Studie 2 – psychotherapeutische Ambulanz – kam ebenfalls der Test zur Wortflüssigkeit nach Jäger und Althoff (1994) in seiner Kurzform zum Einsatz, er war zu Beginn der ersten Sitzung plaziert. Die Stichprobe erreichte dabei im Mittel einen Standardwert von $M=112,70$ ($SD=7,64$), wobei sich alle Probanden zwischen einem minimalen Standardwert von 101 und einem maximalen von 129 als recht wortflüssig darstellen. Die Extremgruppe der guten Assoziierer (gemäß ihres Assoziationskennwertes in der ersten Erhebung) unterscheidet sich mit einem Standardwert von $M=108,67$ ($SD=8,15$) nicht signifikant von der Gruppe der schlechten Assoziierer ($M=111,67$; $SD=7,02$).

Der Standardwert des Wortflüssigkeitstests weist keine signifikanten korrelativen Beziehungen auf zu den Assoziationskennwerten zu beiden Meßzeitpunkten sowie zur Leistung im Lückentext. Kontrolliert man in einer Partialkorrelation den Einfluß des Personenmerkmals Wortflüssigkeit, so ergibt sich mit $r = .856$ ($p = .000$) eine hohe und hochsignifikante korrelative Beziehung zwischen den Assoziationskennwerten zu beiden Meßzeitpunkten (um es vorwegzunehmen: bei der bivariaten Korrelation zur Reliabilitätsbestimmung war der Zusammenhang vergleichbar mit $r = .849$ und $p = .000$; vgl. Kapitel 6.2, Seite 55 ff.). Auch für die Vorhersage des Leistungskennwertes im Lückentext auf Basis der Assoziationsleistung (erste Messung) erbringt die Partialkorrelation bei

Kontrolle der Variable Wortflüssigkeit mit $r = .485$ ($p = .035$) eine signifikante Korrelation mittlerer Stärke (im Vergleich zur unkontrollierten Korrelation mit $r = .475$; $p = .017$; vgl. Kapitel 6.1.3.1.1, Seite 55 ff.). Der Assoziationskennwert sowie das Leistungsmaß im Lückentext sind also in dieser Studie offenbar unbeeinflusst vom Merkmal der generellen Wortflüssigkeit.

6.1.3.2.3. Wortflüssigkeit in Studie 5: Biopsychologie

Im Feldexperiment der Studie 5 – Biopsychologie – wurde u.a. der Test zur Wortflüssigkeit aus dem WILDE-Intelligenztest (Jäger & Althoff, 1994) eingesetzt. Dabei zeigte die Stichprobe überdurchschnittliche Leistungen: mit $M = 114,22$ ($SD = 6,28$) sowie dem minimal erreichten Normwert 105 bzw. maximalen Normwert 124 stellen sich die Probanden durchweg als sehr wortflüssig dar, wobei dennoch eine ausreichende Streuung der Normwerte vorhanden ist. Die korrelativen Beziehungen zwischen diesem Normwert und den Parametern Assoziationskennwert sowie Klausurleistung bzw. –note erbringen zwar geringe Korrelationskoeffizienten zwischen den Parametern, jedoch sind sie weit von einer Signifikanz entfernt. Im Extremgruppenvergleich, bei dem die Probanden mit sehr hoher denjenigen mit sehr geringer Assoziationsleistung gegenüber gestellt werden, finden sich ebenfalls keine Hinweise auf systematische Zusammenhänge zwischen den Variablen. Die Gruppe der guten Assoziierer zeichnete sich durch einen mittleren WIT-Normwert von 120,33 ($SD = 4,01$) aus, während die schlechten Assoziierer mit $M = 111,67$ ($SD = 1,53$) einen etwas geringeren Normwert aufwiesen. Dieser Unterschied ist aber nicht statistisch signifikant (Mann-Whitney-U: $z = -1,964$; $p > .05$).

Um den etwaigen Einfluß, den das Merkmal Wortflüssigkeit auf die Assoziationsleistung haben könnte, aufzudecken, wurde außerdem zur Vorhersage der Klausurleistung durch den Assoziationskennwert die Variable Wortflüssigkeit in einer Partialkorrelation kontrolliert. Dadurch konnte ein mittlerer signifikanter Zusammenhang zwischen Assoziationskennwert und Klausurnote von $r = -.628$ ($p = .048$) ermittelt werden, der völlig frei vom Einfluß der Wortflüssigkeit ist. Dieser Koeffizient fällt zwar etwas geringer aus als in einer Korrelation, in der diese Variable nicht kontrolliert wurde ($r = -.783$; $p = .013$; siehe Kapitel 6.1.2.1.3, Seite 55 ff.), dennoch weist diese signifikante mittlere Korrelation darauf hin, daß der Wortassoziationsversuch eher das Wissen als die generelle Fähigkeit zur Wortgewandtheit erfaßt. Sowohl die Assoziationsleistung als auch die Klausurleistung hängen also nicht systematisch mit dem Merkmal Wortflüssigkeit zusammen.

6.1.3.2.4. Zusammenfassung und Befundintegration der Studien zum Einfluß der Wortflüssigkeit

Die empirischen Befunde zur Rolle der allgemeinen Wortflüssigkeit weisen überwiegend darauf hin, daß dieses Personenmerkmal keinen Einfluß auf die Menge der im Wortassoziationsversuch produzierten Konzepte hat. Die letzten beiden hier vorgestellten Studien, in denen ein solches Nichtvorhandensein einer solchen Beziehung anschaulich nachgewiesen werden konnte, greifen jeweils auf eine Stichprobe eher höheren Intelligenzniveaus zurück, die Probanden sind alle dem akademischen Bereich zuzuordnen. Die Verteilungen des WIT-Standardwertes beschreiben eine deutlich überdurchschnittliche Leistung in den Stichproben.

Die Normwerte des Wortflüssigkeitstests in der Stichprobe in Studie 1 hingegen sind deutlich breiter gestreut und auch im unterdurchschnittlichen Leistungsbereich angesiedelt. Global betrachtet zeigt sich eine Tendenz, daß ein hoher Assoziationskennwert auch mit einem hohen Grad an

Wortflüssigkeit einherzugehen scheint. Betrachtet man dies unter Berücksichtigung des spezifischen Designs dieser Studie, so scheint diese Beziehung v.a. durch diejenigen Stimuli determiniert, welche im mündlichen Modus auf Allgemeinwissen abzielten. Für diejenigen Stimuli dieser Studie, welche das Spezialwissen thematisierten, konnte hingegen keine Beziehung zwischen Assoziationsleistung und Wortflüssigkeit nachgewiesen werden. Es sei an dieser Stelle daran erinnert, daß diese experimentelle Studie nicht auf eine spezifische Kohorte von Arbeitspersonen o.ä. zurückgreift. Somit läßt sich aus dem in dieser Studie angedeuteten Zusammenhang noch keine zuverlässige Schlußfolgerung für den Einsatz des Wortassoziationsversuchs in Wissensdomänen ableiten, welche das berufsspezifische Wissen i.e.S. beinhalten. Es ist aber anzunehmen, daß das Assoziieren von Fachwissen von der Fähigkeit zur Wortflüssigkeit unbeeinflusst bleibt, da es bereits auf die spezialisierte Wissensdomäne zutraf. Insgesamt ist anhand der empirischen Befunde also davon auszugehen, daß das Personenmerkmal der generellen Wortflüssigkeit keinen systematischen Einfluß auf die Assoziationsleistung von Probanden bei der Erhebung berufs- oder fachspezifischen Wissens hat.

6.1.4. Modifikationen des Assoziationskennwertes

Die bisherigen Ausführungen zur Validität basieren auf einem Assoziationskennwert, bei dem alle sinnvollen Assoziationen eines Probanden ausgezählt und über die Anzahl der dargebotenen Stimuli gemittelt wurden. Es stellt sich nun die Frage, ob dieser Kennwert, welcher relativ einfach durch den Versuchsleiter zu erheben ist, die Probanden hinreichend gut repräsentiert, oder ob nicht andere Auswerteprozeduren zu ergiebigeren numerischen Kennwerten führen. Dazu sollen zwei Ansätze der Modifikation vorgestellt werden: zum einen wird der Beobachtung vieler Versuchsleiter Rechnung getragen, daß die Qualität der Äußerungen über die Probanden hinweg sehr unterschiedlich sein kann. Aus diesem Grunde wurden alle Assoziationen hinsichtlich ihrer Qualität bewertet und die so entstandenen Kategorien von Assoziationen mit unterschiedlichem Gewicht belegt. Exemplarisch sollen hier vier Studien vorgestellt werden, welche empirische Daten zu dieser Frage beitragen (siehe Abbildung 105).

Ein anderer Zugang zur Modifizierung des individuellen Assoziationskennwertes besteht darin, gemäß des bei Rothe (2003) beschriebenen Vorgehens auf den Konsens innerhalb der Untersuchungsgruppe Bezug zu nehmen. Dabei werden nur innerhalb der Stichprobe hochfrequente Assoziationen berücksichtigt und gehen in den individuellen Assoziationskennwert ein. Hierzu werden fünf Studien vorgestellt, Abbildung 105 gibt einen Überblick darüber. Es soll geklärt werden, inwieweit diese modifizierten Kennwerte zu einer anderen Beurteilung der Probanden hinsichtlich ihres Wissensbesitzes führen als bei Referenz auf den „einfachen“ Assoziationskennwert. Im Sinne einer inkrementellen Validierung wird geprüft, ob der modifizierte Kennwert eine Verbesserung der Vorhersage hinsichtlich eines Außenkriteriums (soweit vorhanden) über den „einfachen“ Assoziationskennwert hinaus erbringen kann. Im Anschluß an diese quantitativ ausgerichteten Zugänge soll anhand einer weiteren Studie abschließend demonstriert werden, welchen Zugewinn eine stärker qualitativ ausgerichtete Analyse der Assoziationen für die Beantwortung einer praktischen Fragestellung erbringen kann.

Nr.	Name der Studie	Gewichtung	hochfrequente Nennungen
3	Sparkassenakademie (A)		x
4	Sparkassenakademie (B)		x
12	Handelsunternehmen		x
13	Stoffdruckerei	x	x
14	Seminar Wissensmanagement		x
15	Wissen über Arbeitsschutz	x	
16	Call Center (A)	x	
17	Call Center (B)	x	

Abbildung 105: Übersicht über alle Studien, die einen empirischen Beitrag hinsichtlich der Modifikation des Assoziationskennwertes leisten

6.1.4.1. Gewichtung unterschiedlicher Qualitäten von Assoziationsinhalten

In verschiedenen Studien, v.a. wenn die Probanden Nicht-Akademiker waren, fiel auf, daß die Qualität der Assoziationen erhebliche inter- und intraindividuelle Unterschiede aufwiesen. Der quantitative Ansatz der Auswertung des Wortassoziationsversuchs sieht vor, für jede (sinnvolle) Assoziationen einen Punkt zu vergeben, ohne Rücksicht auf die Qualität dieser Äußerung. So geht beispielsweise in einem Wortassoziationsversuch zur Störungsdiagnose von technischen Anlagen eine Assoziation wie „kommt selten vor“ mit dem gleichen Gewicht ein wie eine Äußerung zu einem ganz konkreten Störungstool (z.B. die Adresse einer diesbezüglichen Internetseite). Erstgenannte Assoziation ist dabei keineswegs als unsinnig abzutun, denn immerhin ist diese sehr ungenaue Einschätzung der Prävalenz eines Störungsbildes durchaus praxisrelevant und eine auf Basis von Erfahrungen bzw. Wissen getroffene Feststellung. Dennoch ist ein erheblicher Niveauunterschied zwischen beiden Äußerungen zu konstatieren. Diesem Umstand wurde in einigen Studien versucht, Rechnung zu tragen, indem a posteriori alle von den Probanden produzierten Assoziationen hinsichtlich ihrer Qualität klassifiziert wurden, diese Kategorien dann unterschiedliche Gewichte erhielten und so unterschiedlich gewichtig in den individuellen Assoziationskennwert eingingen. So kann durch wenige sehr präzise Assoziationen ein ebenso hoher Assoziationskennwert erzielt werden wie durch das Nennen vieler, aber sehr allgemeiner Assoziationen. Im Folgenden werden vier Studien vorgestellt, in denen der Assoziationskennwert auf diese Weise modifiziert wurde.

6.1.4.1.1. Gewichtung in Studie 13: Stoffdruckerei

Neben dem reinen Auszählen der Assoziationen wurde von Pokrandt (2004) in Studie 13 – Stoffdruckerei – außerdem eine Gewichtung der Assoziationen i.S. einer qualitativen Auswertung vorgenommen. Aufgrund der großen Bandbreite und unterschiedlicher inhaltlicher Niveaus der

Nennungen bewertete sie die einzelnen Aussagen anhand der in Abbildung 106 dargestellten Kategorien. Als inhaltlich erwartungsgemäß kann das Basiswissen verstanden werden, welches bei allen Mitarbeitern des Unternehmens vorausgesetzt werden kann – Nennungen dieser Art gingen mit einem Gewicht von 1 in die Auswertung ein. Davon hoben sich sehr differenzierte oder umfassende Antworten ab, diese wurden als Spezialwissen klassifiziert und gingen doppelt in die Auswertung ein. Davon abzugrenzen waren sehr undifferenzierte Meinungen und ähnliche Äußerungen allgemeiner Natur, welche als sog. Individualkenntnisse mit einem Gewicht von 0,5 in die Auswertung eingingen. Grundlage für die Klassifizierungen waren die Ergebnisse der Wortassoziationsversuche mit dem internen Experten sowie den jeweiligen Abteilungsleitern (Pokrandt, 2004).

Kategorie	Erläuterung	Beispiele	Gewicht
Individualkenntnisse	individuelle Erfahrungen, v.a. emotionale Äußerungen sowie sehr weit entfernte Generierungen auf den Stimulus	wichtig; ist kein Problem	0,5
Basiswissen	Assoziationen, die von allen Mitarbeitern vorausgesetzt werden konnten bzw. allgemeiner Natur waren, aber über das niedrigere Niveau der Individualkenntnisse hinausgingen	Kontrollzettel; Kunden- bestätigungen	1
Spezialwissen	sehr umfassende und qualitativ anspruchsvolle Antworten, die über das allgemeine Basiswissen hinausgingen	Audit; Quali-Review	2

Abbildung 106: Erläuterung der inhaltlichen Gewichtungen in Studie 13 – Stoffdruckerei (Beispiele für den Stimulus Qualitätsablaufplan; vgl. Pokrandt, 2004)

Pokrandt (2004) berichtet, daß durch die Gewichtung der Assoziationen ein maximaler Summenwert (über alle Stimuli) von 785,5 Nennungen erzielt wurde, im Mittel 243,89 Nennungen ($SD=168,19$), das entspricht einer Steigerung des Scores auf 126 % für den Maximalwert, auf 122 % für den Mittelwert. Dabei habe sich jedoch das Ranking der Probanden kaum verändert: die unteren Ränge blieben unverändert, lediglich im oberen Bereich gab es geringfügige Wechsel der Rangplätze (Pokrandt, 2004).

Abbildung 107 stellt die statistischen Kennwerte dar, bezogen auf Assoziationen pro Stimulus. Danach werden bei Auszählung aller Assoziationen im Mittel 6,85 Äußerungen pro Stimulus produziert ($SD=4,01$). Gewichtet man die Äußerungen nach inhaltlichem Gehalt, so erhöht sich der Punktwert bei Einbezug aller Nennungen auf 8,14 Äußerungen pro Stimulus ($SD=5,46$). Pokrandt (2004) legte außerdem noch eine weitere Strategie an, indem sie bei der Bildung des gewichteten Wissensscores die Kategorie der Individualkenntnisse nicht einbezog. In diesem Falle reduzierte sich der mittlere Assoziationskennwert auf $M=7,86$ ($SD=5,29$).

Durch Rangkorrelationen nach Spearman konnte die Beobachtung von Pokrandt (2004) ebenfalls belegt werden: der ungewichtete Assoziationskennwert korreliert sehr hoch mit dem gewichteten Score ($r=.970$; $p=.000$) und hoch mit dem gewichteten Score, in dem die Individualkenntnisse vernachlässigt worden sind ($r=.864$; $p=.000$). Die Gewichtung der Assoziationsinhalte führt also lediglich zu einer Betonung der bereits bestehenden Tendenz – sie führt nicht dazu, daß man die Assoziationsleistung von Probanden grundsätzlich anders bewerten würde.

	Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus (ohne Individualkenntnisse)
<i>M</i>	6,85	8,14	7,86
<i>SD</i>	4,01	5,46	5,29
Minimum	2,83	3,11	3,04
Maximum	21,54	28,07	26,71

Abbildung 107: Kennwerte des Wortassoziationsversuchs mit Gewichtung – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 13 – Stoffdruckerei

Die Gewichtung der Assoziationsinhalte beeinflusst die Vorhersagekraft des Assoziationskennwertes hinsichtlich der Geschäftsführerbeurteilung. Zur Vorhersage der Beurteilung der Fachkompetenz erweist sich das Gewichtungsmodell (Individualkenntnisse*0,5 + Basiswissen + Spezialwissen*2) als brauchbarer Prädiktor: mit $r = -.371$ ($p = .049$) zeigt sich eine etwas höhere und signifikante Zuverlässigkeit der Vorhersage dieses Außenkriteriums als der ungewichtete Assoziationskennwert mit $r = -.350$ ($p = .060$). Zur Vorhersage der Gesamtbeurteilung allerdings konnte auch dieser modifizierte Kennwert keinen Mehrertrag erbringen: mit einem schwachen und nicht signifikanten Korrelationskoeffizienten von $r = -.287$ ($p > .10$) ist er ähnlich ungeeignet wie der ungewichtete Assoziationskennwert, welcher (wie bereits berichtet) einen Koeffizienten von $r = -.272$ ($p > .10$) aufwies. Das Gewichtungsmodell, in dem die Individualkenntnisse unberücksichtigt blieben (Basiswissen + Spezialwissen*2), erwies sich hingegen als weniger geeignet zur Vorhersage der Geschäftsführerbeurteilungen, weder die Gesamtbeurteilung ($r = -.289$; $p > .10$) noch die Fachkompetenz ($r = -.359$; $p > .05$) konnten zuverlässig durch den so modifizierten Assoziationskennwert vorhergesagt werden.

6.1.4.1.2. Gewichtung in Studie 15: Wissen über Arbeitsschutz

Da in der Online-Erhebung in Studie 15 – Wissen über Arbeitsschutz – ebenfalls ein breites Spektrum an unterschiedlichen Qualitäten der Assoziationen zu beobachten war, wurden auch hier alle geäußerten Assoziationen einer Inhaltsanalyse unterzogen, so daß sie in Kategorien gemäß ihrer Wissensqualität zusammengefaßt werden konnten. Diese Kategorien sowie deren Gewichte sind in Abbildung 108 veranschaulicht. Grundsätzlich wurden alle Nennungen der mittleren Kategorie „Basiswissen“ zugeordnet. Lediglich deutliche qualitative Abweichungen nach oben bzw. nach unten wurden als „Spezialwissen“ oder aber als „Allgemeinwissen“ kodiert. Letztgenanntes beschreibt eher auf umgangssprachlichem Niveau angesiedelte unspezifische Äußerungen oder persönliche Meinungen, während das Basiswissen in Abgrenzung dazu Fakten auf fachlicher Ebene und Grundlagenwissen zur Thematik beschreibt. Davon läßt sich das Spezialwissen abgrenzen, welches sich durch sehr differenzierte und spezifische Sachverhalte auszeichnet.

Durch die Gewichtung wurde das Feld der Assoziationskennwerte auseinander gestreckt: mit $M=6,05$ und $SD=3,63$ besitzt der Kennwert nun eine deutlich größere Streubreite als ohne Gewichtung (siehe Abbildung 109). Analog zum Vorgehen von Pokrandt (2004) wurde auch hier ein weiterer Score gebildet, in dem die niedrigste Kategorie der Wissensqualitäten (hier: sog. Allgemeinwissen) unberücksichtigt blieb. Auch hier ergibt die Rangkorrelation nach Spearman mit $r = .896$ ($p = .000$) eine hohe Korrelation zwischen diesem in dieser Weise gewichteten und dem ungewichteten Kennwert. Werden hingegen alle Nennungen in die Gewichtung einbezogen, so

zeigt sich eine sehr hohe Korrelation nach Spearman ($r = .965$; $p = .000$). Die Rangplätze der Probanden in der Verteilung der Kennwerte bleiben also durch die Gewichtung im Wesentlichen erhalten, und bestehende Tendenzen erfahren eine Akzentuierung.

Kategorie	Erläuterung	Beispiele	Gewicht
Allgemeinwissen	individuelle Bewertungen zum Stimulus oder unspezifische Sachverhalte, die noch nicht als Fachwissen angesehen werden können	oft vernachlässigt; nicht betrunken Autofahren	1
Basiswissen	alle Assoziationen, die weder die Bedingung für die Kategorie Allgemeinwissen noch für die Kategorie Spezialwissen erfüllen	Gefahrenquellen; Schutz vor Unfällen	2
Spezialwissen	sehr konkrete Sachverhalte werden beschrieben oder sehr spezifische Fachtermini benutzt, die über das Basiswissen hinausgingen	keine Nahrungsmittel im Labor; S-Sätze	3

Abbildung 108: Erläuterung der inhaltlichen Gewichtungen in Studie 15 – Wissen über Arbeitsschutz (Beispiele für den Stimulus Arbeitsschutz)

	Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus (ohne Allgemeinwissen)
<i>M</i>	2,76	6,05	3,29
<i>SD</i>	1,56	3,63	2,15
Minimum	0,0	0,0	0,0
Maximum	9,0	20,0	11,7

Abbildung 109: Kennwerte des Wortassoziationsversuchs mit Gewichtung – Mittelwerte (*M*) und Standardabweichungen (*SD*) in Studie 15 – Wissen über Arbeitsschutz

Im Sinne der inkrementellen Validierung wurde auch in dieser Studie, in der als alternatives Verfahren der Wissenserfassung ein Multiple-Choice-Test zur Verfügung stand, der gewichtete Assoziationskennwert hinsichtlich seiner Vorhersagekraft für dieses Kriterium geprüft. Danach führt die Gewichtung der Assoziationen nach inhaltlichen Erwägungen zu einer leichten Erhöhung der Vorhersagekraft des Wortassoziationsversuchs in Bezug auf die Wissensleistung im Multiple-Choice-Test. Während der korrelative Zusammenhang zwischen Multiple-Choice-Test und ungewichteter Assoziationscore $r = .160$ ($p = .000$) beträgt, kann die Referenz auf den gewichteten Assoziationscore den Koeffizienten auf $r = .177$ ($p = .000$) erhöhen. Vernachlässigt man bei der Aggregation des gewichteten Assoziationskennwertes das Allgemeinwissen, so erbringt dies einen Korrelationskoeffizienten von $r = .176$ ($p = .000$). Allerdings ist auch anhand der so modifizierten Assoziationskennwerte der Zusammenhang zwischen beiden Verfahren immer noch nur als sehr schwach zu bezeichnen.

6.1.4.1.3. Gewichtung in Studie 16: Call Center (A)

Die Fragestellung in Studie 16 – Call Center (A), inwieweit sich die organisatorische Unterscheidung zwischen Mitarbeitern nach First und Second Level auch in einem unterschiedlichen Ausmaß an Wissensbesitz niederschlägt, ließ sich anhand der einfachen

Auszählung von assoziierten Begriffen nicht beantworten. Aus diesem Grund und aufgrund der Beobachtung, daß die Qualität der Assoziationen eine große Streubreite aufwies, wurde eine Gewichtung der Assoziationsinhalte vorgenommen. Zu prüfen war die Annahme, das Wissen der Second-Level-Agents sei spezieller als das des First-Levels. Es sei an dieser Stelle berichtet, daß durch die Gewichtung qualitativ unterschiedlicher Assoziationen keine Unterschiede im Wissensbesitz von First- vs. Second-Level-Agents nachvollzogen werden konnten – die formale Zugehörigkeit zu einem Level war offenbar kein geeignetes Kriterium, um Wissensunterschiede zu erklären (siehe auch Ceglarek und Rothe, 2003b).

Der Inhalt der Assoziationen wurde bewertet und den drei Kategorien Basiswissen, sog. erweitertes Basiswissen oder Spezialwissen zugeordnet (Abbildung 110 erläutert diese Kategorien). Es wurde ein Wissensscore gebildet, in den Basiswissen einfach einging, erweitertes Basiswissen zweifach und Spezialwissen dreifach. Als Referenz für diese Klassifizierung wurden die Schulungsunterlagen des Unternehmens sowie Experteninterviews herangezogen.

Kategorie	Erläuterung	Beispiele	Gewicht
Basiswissen	allgemeine Erklärungen und Grundlagen zum ausgewählten Produktbereich (z.B. Umschreibungen des Stimulusbegriffs, Synonyme auf nicht-fachlicher Ebene, Aufzählung von konkreten Beispielen in diesem Bereich)	aufwendig; braucht Vorkenntnisse	1
erweitertes Basiswissen	spiegelt Niveau der Schulungsunterlagen wider (z.B. Fachbegriffe, vergleichende Betrachtungen, Funktionsbeschreibungen, zu beachtende Bedingungen, Beziehungen zwischen Komponenten u.ä.)	serverseitig; verschiedene Skriptsprachen	2
Spezialwissen	über das eigentliche Produkt hinausgehend, dessen Beherrschung ist nicht zwingend für die Bewältigung der Arbeitsaufgaben (z.B. Algorithmen im Computer, Besonderheiten der Hardware- und Softwarearchitektur, Spezialbegriffe)	Upload-Script; PHP- Skripte	3

Abbildung 110: Erläuterung der inhaltlichen Gewichtungen in Studie 16 und 17 (Beispiele für den Stimulus Skripte) (vgl. Ruprecht, 2004)

Abbildung 111 listet die statistischen Kennwerte der Assoziationskennwerte auf, welche durch die verschiedenen Gewichtungsprozeduren entstanden, im Vergleich zum ungewichteten Assoziationskennwert. Durch die vorgenommene Gewichtung (Basiswissen*1+erweitertes Basiswissen*2 + Spezialwissen*3) erhöht sich der Assoziationskennwert auf $M=5,30$, wobei sich gleichsam die Streuung erhöht ($SD=2,28$). Die Rangkorrelation nach Spearman erbringt einen sehr hohen signifikanten Korrelationskoeffizienten von $r= .970$ ($P= .000$) zwischen dem ungewichteten und dem gewichteten Assoziationskennwert: Probanden haben also bei beiden Vorgehensweisen einen annähernd identischen Rangplatz innerhalb der Stichprobe inne. Die Gewichtung von Assoziationsinhalten führt demnach nicht zu einer grundsätzlich anderen Beurteilung der Assoziationsleistungen von Probanden.

Analog zum Vorgehen bei Pokrandt (2004) wurde auch hier ein weiterer Score gebildet, bei dem die niedrigste Kategorie der Wissensqualitäten unberücksichtigt blieb, während die anderen beiden Kategorien gewichtet in den Kennwert einfließen. Durch dieses Vorgehen reduzierte sich der Assoziationskennwert auf $M=2,64$ ($SD=1,61$) und fiel damit sogar noch geringer aus als der

ungewichtete Assoziationskennwert (siehe Abbildung 111). Die Rangkorrelation nach Spearman ergibt einen hoch signifikanten Korrelationskoeffizienten von $r = .903$ ($p = .000$) zwischen diesen beiden Assoziationskennwerten, so daß man bei beiden Vorgehensweisen zu einer vergleichbaren Rangreihe der Probanden auf Grundlage ihres Assoziationskennwertes gelangen würde. Da in dieser Studie kein Außenkriterium oder ein weiteres Verfahren zur Wissensdiagnose zur Verfügung stand, kann hier auch keine inkrementelle Validität geprüft werden.

	Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus (ohne Basiswissen)
<i>M</i>	3,95	5,30	2,64
<i>SD</i>	1,43	2,28	1,61
Minimum	1,9	2,3	08
Maximum	7,6	11,9	8,0

Abbildung 111: Kennwerte des Wortassoziationsversuchs mit Gewichtung – Mittelwerte (*M*) und Standardabweichungen (*SD*) in Studie 16 – Call Center (A)

6.1.4.1.4. Gewichtung in Studie 17: Call Center (B)

Im Wortassoziationsversuch in Studie 17 – Call Center (B) – merkte Ruprecht (2004) eine große Varianz bzgl. des Niveaus der Assoziationen an. Daher bewertete sie alle Äußerungen hinsichtlich ihres fachspezifischen Niveaus (in Anlehnung an das Vorgehen in Studie 16) entsprechend der in Abbildung 110 dargestellten drei Kategorien Basiswissen, erweitertes Basiswissen sowie Spezialwissen. Die Zuordnungen von Wissensinhalten zu diesen Kategorien geschahen zunächst auf Basis der Schulungsunterlagen und wurde abschließend mit Hilfe eines Experten der untersuchten Arbeitsgruppe abgesichert. Diese drei Kategorien hatten unterschiedliches Gewicht bei der Bildung eines Assoziationscores (siehe auch Abbildung 110): danach ging das Basiswissen aufgrund seines unspezifischen Charakters einfach in die Berechnung ein, das sog. erweiterte Basiswissen als das zur Aufgabenerfüllung erforderliche Fachwissen wurde doppelt gewichtet, Spezialwissen wurde als besonders bedeutsam und qualitativ hochwertiger interpretiert und daher mit einem Gewicht von 3 belegt.

	Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus	gewichtete Anzahl Assoziationen pro Stimulus (ohne Basiswissen)
<i>M</i>	5,26	8,68	6,06
<i>SD</i>	2,30	4,39	3,66
Minimum	2,23	3,63	2,00
Maximum	13,30	24,07	18,77

Abbildung 112: Kennwerte des Wortassoziationsversuchs mit Gewichtung – Mittelwerte (*M*) und Standardabweichungen (*SD*) in Studie 17 – Call Center (B)

Gemäß dieser Definitionen wurden alle Äußerungen der Probanden als zu einer der drei Kategorien gehörig eingestuft. Durch die Gewichtung hat sich der Assoziationskennwert gegenüber dem ungewichteten Kennwert ($M=5,24$; $SD=2,29$) auf $M=8,68$ ($SD=4,39$) erhöht, ebenso hat sich erwartungsgemäß die Streuung der individuellen Kennwerte vergrößert (siehe Abbildung 112). Dennoch sind die Auswirkungen auf die Veränderung der Rangfolge bei Hinzunahme der

Gewichtung gering, der Korrelationskoeffizient der Rangkorrelation nach Spearman zwischen dem ungewichteten und dem gewichteten Assoziationscore ist mit $r = .971$ ($p = .000$) sehr hoch und signifikant. Ruprecht (2004) weist darauf hin, daß das Heranziehen des gewichteten Scores zu einer annähernd gleichen Zuordnung der Probanden zu den Extremgruppen geführt hat wie beim Heranziehen des ungewichteten Kennwertes.

In Anlehnung an das bisherige Vorgehen wurde auch hier ein weiterer Score gebildet, welcher auf dem gewichteten Kennwert beruht, bei dem die Assoziationen der untersten Kategorie (Basiswissen) unberücksichtigt blieben. Durch dieses Vorgehen wird im Mittel ein Assoziationskennwert von 6,06 ($SD = 3,66$) erreicht (siehe Abbildung 112). Auch dieser Kennwert korreliert bei der Rangkorrelation nach Spearman sehr hoch und hoch signifikant mit dem ungewichteten Assoziationskennwert ($r = .923$; $p = .000$). Somit würde auch die Zugrundelegung dieses Gewichtungsmodells zu keiner grundlegend anderen Einordnung der Leistung von Probanden führen. Auch in dieser Studie also führt die Gewichtung von Assoziationsinhalten nicht zu einer gänzlich anderen Beurteilung von Probanden im Vergleich zur ungewichteten Auswertung. Da in dieser Studie kein Außenkriterium oder ein weiteres Verfahren zur Wissensdiagnose zur Verfügung stand, kann hier auch keine inkrementelle Validität geprüft werden.

6.1.4.1.5. Zusammenfassung und Befundintegration der Studien zur Gewichtung von Assoziationsinhalten

In den vier soeben vorgestellten Studien, in denen die Gewichtung von Assoziationsinhalten vorgenommen wurde, ergeben sich durchweg extrem hohe Übereinstimmungen in der Rangreihe der Probanden über die verschiedenen Auswerteprozeduren hinweg. Abbildung 113 gibt einen Überblick über die Koeffizienten der Rangkorrelation nach Spearman in den einzelnen Studien, wobei jeweils der ungewichtete Assoziationskennwert den zwei Varianten der gewichteten Kennwerte gegenübergestellt wurde (Gewichtung aller drei Inhaltskategorien vs. Gewichtung ohne Berücksichtigung der niedrigsten Kategorie).

Korrelation des ungewichteten Assoziationskennwertes mit gewichtetem Assoziationskennwert	... gewichtetem Assoziationskennwert (ohne Individual- bzw. Basiswissen)
13 - Stoffdruckerei	$r = .970$ ** ($p = .000$)	$r = .864$ ** ($p = .000$)
15 - Wissen über Arbeitsschutz	$r = .965$ ** ($p = .000$)	$r = .896$ ** ($p = .000$)
16 - Call Center (A)	$r = .970$ ** ($p = .000$)	$r = .903$ ** ($p = .000$)
17 - Call Center (B)	$r = .971$ ** ($p = .000$)	$r = .923$ ** ($p = .000$)

Abbildung 113: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (ungewichteter mit verschiedenen gewichteten Assoziationskennwerten) als Rangkorrelation nach Spearman

Es wird dabei deutlich, daß die Berücksichtigung *aller* Kategorien von Assoziationen bei der Gewichtung zu einem jeweils höheren Koeffizienten führt als das Vernachlässigen von Assoziationen der untersten Kategorie (sog. Individual-, Allgemein- oder Basiswissen). Eine Zusammenführung dieser Befunde als mittlere gewichtete Effektstärke erbringt mit $d = .97$ einen großen Effekt. Durch Gewichtung der Assoziationsinhalte werden Probanden also nicht grundsätzlich anders hinsichtlich ihres Wissensbesitzes beurteilt als bei ungewichteter Auszählung aller Assoziationen.

Da in den Studien teilweise verschiedene Gewichtungformeln verwendet wurden, soll das Vorgehen für alle vier Studien vereinheitlicht werden, indem jeweils ein gewichteter Kennwert nach der Formel ($x*0,5+y*1+z*2$) sowie nach der Formel ($x*1+y*2+z*3$) berechnet wird. Der Unterschied zwischen beiden Formeln liegt in der Akzentuierung der qualitativsten Kategorie: in der ersten Formel steigen die Gewichte exponentiell (Spezialwissen wird besonders betont), in der zweiten linear. Außerdem wurde ein Score gebildet ($x*0+y*1+z*2$), bei dem die jeweils niedrigste Kategorie (Basis- bzw. Individualwissen) vernachlässigt wurde, da unterstellt wird, diese sei u.U. zu trivial (siehe Pokrandt, 2004). Auf diese Weise kamen drei verschiedene Gewichtungskennwerte zustande, die als z-transformierte Werte auf eine Gesamtstichprobe von $N=545$ angewendet werden konnten. Die Ergebnisse der Befundintegration sind in Abbildung 114 dargestellt. So läßt sich bei einer Rangkorrelation nach Spearman anhand der Gewichtungformel ($x*1+y*2+z*3$) am wahrscheinlichsten die Rangreihe der ungewichteten Assoziationskennwerte replizieren ($r= .965$; $p= .000$), auch die Gewichtungformel ($x*0,5+y*1+z*2$) liefert eine sehr hohe signifikante Rangkorrelation ($r= .940$, $p= .000$). Im Vergleich der verschiedenen Gewichtungsmodelle am schlechtesten, aber dennoch sehr gut, stellt sich mit $r= .896$ ($p= .000$) jenes Modell dar, in dem Assoziationen der untersten Qualitätsebene nicht in den individuellen Kennwert eingingen. Die ungewichteten Kennwerte selbst (z-transformiert) lassen sich durch die unterschiedlichen gewichteten Kennwerte sogar mit noch größerer Präzision vorhersagen, in allen Gewichtungsmoellen ist hier ein sehr hoher und hoch signifikanter Koeffizient bei der Korrelation nach Pearson zu vermerken (siehe Abbildung 114). Das ist insofern auch zu erwarten, als daß bei allen Rechenvorgängen ja immer auf den gleichen Grundstock an Daten zurückgegriffen wurde (ein Proband mit wenigen Assoziationen beispielsweise kann durch Gewichtungen auch nur einen begrenzten Zugewinn für seinen Assoziationskennwert erlangen). Der Vollständigkeit halber sei erwähnt, daß die beschriebenen Zusammenhänge noch sehr viel prägnanter ausfallen, wenn nur die Extremstichprobe zugrunde gelegt wird – da aber auch bei der Gesamtstichprobe bereits extrem hohe signifikante Korrelationen zu finden waren, soll auf das Berichten der Werte der Extremstichprobe verzichtet werden.

Gewichtungsmodell	Korrelation nach Pearson	Rangkorrelation nach Spearman
$x*0,5+y*1+z*2$	$r= .953 **$ ($p= .000$)	$r= .940 **$ ($p= .000$)
$x*1+y*2+z*3$	$r= .973 **$ ($p= .000$)	$r= .965 **$ ($p= .000$)
$x*0+y*1+z*2$	$r= .918 **$ ($p= .000$)	$r= .896 **$ ($p= .000$)

Abbildung 114: Korrelationen zwischen dem ungewichteten Assoziationskennwert mit verschiedenen gewichteten Assoziationskennwerten (z-transformiert)

Es läßt sich also zusammenfassen, daß die Gewichtung von Assoziationsinhalten – auch nach verschiedenen Formeln – kaum Auswirkungen auf die Beurteilung von Assoziationsleistungen von Probanden hat. In den zwei vorgestellten Studien, in denen ein Außenkriterium zur Verfügung stand, erhöhte sich durch die Gewichtungen der Korrelationskoeffizient zur Vorhersage dieses Kriteriums, allerdings nur gering. Das ist auch zu erwarten, da die ungewichteten und die gewichteten Assoziationskennwerte untereinander hoch korrelieren. Die Rangreihen der Probanden veränderten sich durch das Prozedere ebenfalls nur geringfügig, v.a. bei einer groben Einteilung der Stichprobe in Extremgruppen ändern sich die Gruppenzusammensetzungen kaum. Dennoch bleibt

festzuhalten, daß ein in dieser Weise modifizierter Assoziationskennwert geringfügig höhere Zuverlässigkeit in der Vorhersage eines Außenkriteriums (hier: Geschäftsführerbeurteilung der Fachkompetenz, Leistung im Multiple-Choice-Test) erbringt als der ungewichtete Kennwert.

Somit ist der Vorgang der Gewichtung von Assoziationsinhalten unterschiedlicher Qualität zwar ein bisweilen nicht unaufwendiges Verfahren, erbringt aber keine völlig anderen Erkenntnisse hinsichtlich der Beurteilung der Probanden auf Grundlage der numerischen Kennwerte. Das Feld der Assoziationskennwerte innerhalb der Stichprobe wird zwar durch die Gewichtung auseinandergestreckt, akzentuiert damit aber lediglich bereits bestehende Unterschiede zwischen Probanden. Dennoch kann ein solches Vorgehen wertvoll sein in Hinblick auf die Deskription der Assoziationsdaten: es können Aussagen darüber getroffen werden, inwieweit der individuelle Wissensbesitz oder aber auch als Mittel über die Stichprobe verschiedene Qualitäten von Fachwissen abbildet, als Beispiel für einen solchen illustrierenden Ansatz ist die Arbeit von Ruprecht (2004) zu nennen.

6.1.4.2. Assoziationskennwert auf Basis der hochfrequenten Assoziationen

Wie bereits erwähnt waren die Inhalte der Assoziationen innerhalb einer Stichprobe sehr unterschiedlich. Ein weiterer Ansatz, diesem Umstand zu begegnen, ist neben der soeben vorgestellten Gewichtung der Qualität individueller Assoziationen, die Relevanz dieser Äußerungen auf Basis der Gruppenleistung zu beurteilen. Unter bestimmten Fragestellungen, z.B. wenn der Wortassoziationsversuch zur Modifikation eines hypothetischen Wissensmodells herangezogen werden soll, ist es sinnvoll, so etwas wie den Konsens über die Domäne zu erfassen. Rothe (2003) schlägt dazu eine Datenaggregation auf Basis der Häufigkeit einzelner Assoziationen innerhalb der Stichprobe vor: diejenigen Assoziationen, die von mindestens zwei Probanden der Stichprobe benannt wurden (bei größeren Stichproben kann dieses Kriterium auch ausgeweitet werden), scheinen auch eine gewisse Relevanz für die Wissensdomäne zu besitzen. In der Konsequenz gehen in den individuellen Assoziationskennwert dann auch nur alle in dieser Weise relevanten oder hochfrequenten Assoziationen ein, während diejenigen Äußerungen, die nur von einem einzelnen Probanden gemacht werden, bei der Auszählung unberücksichtigt bleiben. Allerdings ist anzumerken, daß keine Befunde hinsichtlich der Bestimmung des Kriteriums bekannt sind, wie angemessen also diese kritische Mindestmenge ist. Es unterliegt also einer gewissen Willkür, ob eine Assoziation von mindestens zwei Probanden übereinstimmend benannt werden muß, um gezählt zu werden, oder ob ein bestimmter Prozentsatz der Stichprobe einen bestimmten Inhalt assoziieren sollte.

Zur sprachlichen Abgrenzung sei dieser Wertung hochfrequenter oder relevanter Assoziationen die Auszählung der sog. absoluten Assoziationen gegenüber gestellt, bei der sämtliche Äußerungen eines Probanden berücksichtigt werden (wie es auch in der vorliegenden Arbeit grundsätzlich gehandhabt wurde). Im Folgenden wird der Erkenntnisgewinn, den ein solches Vorgehen erbringen kann, exemplarisch anhand von fünf Studien illustriert.

6.1.4.2.1. Auszählung hochfrequenter Assoziationen in Studie 3: Sparkassenakademie (A)

Wie bereits beschrieben, erhob Wegwarth (2002) in der Studie 3 – Sparkassenakademie (A) – das Fachwissen von Weiterbildungsteilnehmern zu vier Zeitpunkten: zu Beginn der Weiterbildung, am

Ende der Präsenzlernphase, am Ende der Selbstlernphase zeitgleich zur Prüfung sowie vier Monate danach. Aus Gründen der Datenreduzierung bezog Wegwarth (2002) nur hochfrequente Assoziationen in diese Analyse ein, d.h. Assoziationen, die von mindestens acht der insgesamt 38 Probanden genannt wurden (dies entspricht etwa 20 % der Stichprobe). Da die Rohdaten dazu nicht mehr zur Verfügung standen und keine individuellen Kennwerte der hochfrequenten Auszählung für die einzelnen Probanden existieren, ist es an dieser Stelle nicht möglich, korrelative Zusammenhänge von hochfrequenter zu absoluter Auszählung der Assoziationen sowie zur Klausurleistung zu berichten. Ebenso konnten keine diesbezüglichen Signifikanzprüfungen vorgenommen werden. Aus diesem Grunde basieren die in Abbildung 115 bis Abbildung 118 dargestellten Befunde auf den in der Arbeit von Wegwarth (2002) berichteten Ergebnissen. Da Wegwarth (2002) ihre Berechnungen immer auf einzelne Stimuli bezog (und nicht, wie in der vorliegenden Arbeit gehandhabt, als mittlere Assoziationsanzahl über alle Stimuli hinweg), können hier auch nur diese zitiert und eben diese Analyseebene beibehalten werden.

	hochfrequente Assoziationen zum Stimulus				absolute Assoziationen zum Stimulus			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>M</i>	4,13	5,97	8,15	4,48	12,2	11,2	14,6	11,1
<i>SD</i>	1,83	2,30	2,78	2,13	5,38	4,35	4,97	5,32

Abbildung 115: Assoziationskennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – für Stimulus „Strategische Unternehmensführung“ bei Auszählung hochfrequenter vs. aller Assoziationen in Studie 3 – Sparkassenakademie (A); (vgl. Wegwarth, 2002)

	hochfrequente Assoziationen zum Stimulus				absolute Assoziationen zum Stimulus			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>M</i>	2,03	5,07	7,86	4,86	9,7	10,6	13,5	10,1
<i>SD</i>	0,74	2,14	2,63	2,29	3,54	4,46	4,54	4,77

Abbildung 116: Assoziationskennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – für Stimulus „Führen von Gruppen“ bei Auszählung hochfrequenter vs. aller Assoziationen in Studie 3 – Sparkassenakademie (A); (vgl. Wegwarth, 2002)

	hochfrequente Assoziationen zum Stimulus				absolute Assoziationen zum Stimulus			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>M</i>	3,63	3,93	8,41	4,83	9,6	8,7	14,5	10,3
<i>SD</i>	1,42	1,71	2,62	2,50	3,74	3,81	4,51	5,31

Abbildung 117: Assoziationskennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – für Stimulus „Personalentwicklung“ bei Auszählung hochfrequenter vs. aller Assoziationen in Studie 3 – Sparkassenakademie (A); (vgl. Wegwarth, 2002)

	hochfrequente Assoziationen zum Stimulus				absolute Assoziationen zum Stimulus			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>M</i>	1,13	3,37	4,89	3,39	6,3	7,0	9,8	6,8
<i>SD</i>	0,60	1,92	2,09	2,03	3,33	3,99	4,18	4,05

Abbildung 118: Assoziationskennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – für Stimulus „Motivationstheorien“ bei Auszählung hochfrequenter vs. aller Assoziationen in Studie 3 – Sparkassenakademie (A); (vgl. Wegwarth, 2002)

In Abbildung 115 bis Abbildung 118 sind die statistischen Kennwerte der Assoziationsleistungen zu den einzelnen Stimuli aufgelistet, im Vergleich zwischen der Auszählung sämtlicher vs. nur der hochfrequenten Stimuli. Es wird deutlich, daß sich v.a. zum ersten Meßzeitpunkt der Assoziationskennwert auf diese Weise drastisch reduziert, während das Verhältnis von hochfrequenten zu absoluten Kennwerten beim dritten Meßzeitpunkt schon deutlich ausgeglichener ist. Dies weist darauf hin, daß insbesondere vor Beginn der Weiterbildung der Anteil eines geteilten Wissens in der Gruppe gering ausgeprägt ist (d.h. relativ wenige Assoziationen wurden von mindestens acht Probanden benannt). Die Probanden assoziieren zwar etwas, aber ihre Assoziationsinhalte sind nicht identisch. Durch die Weiterbildung jedoch gleichen sich die individuellen Wissensbestände einander an, was ja durchaus als Ziel einer solchen Veranstaltung verstanden werden kann.

Zusammenfassend ist zu sagen, daß die Referenz auf die innerhalb der Stichprobe hochfrequenten Assoziationen in dieser Studie zu einem interessanten Erkenntnisgewinn geführt hat, der sich v.a. auf den Verlauf über die Erhebungszeitpunkte bezieht. Im Vergleich zum absoluten Assoziationskennwert wird deutlich, daß vor der Weiterbildung noch relativ wenige Konzepte das Kriterium für die hochfrequenten Nennungen erfüllten, während dieser Anteil nach Abschluß der Präsenzlernphase und schließlich bei der Prüfung zunahm. Es konnte also im Verlauf der Weiterbildung zunehmend so etwas wie geteiltes Wissen in der Gruppe entwickelt werden. In der Weiterbildung vermittelte Konzepte fanden erfolgreich Eingang in den individuellen Wissensbestand vieler Probanden, so daß diese Maßnahme als erfolgreich angesehen werden kann.

6.1.4.2.2. Auszählung hochfrequenter Assoziationen in Studie 4: Sparkassenakademie (B)

Auch in der zweiten Studie an der Sparkassenakademie (B) konnte eine Auswertung hinsichtlich der Berücksichtigung der hochfrequenten Assoziationen vorgenommen werden. Dabei wurde als Kriterium zugrunde gelegt, daß eine Assoziation von mindestens zwei Probanden der Stichprobe genannt worden sein muß, um in die Auszählung einzugehen. Auf diese Weise wurde erwartungsgemäß ein kleinerer Assoziationskennwert erreicht als durch das Zugrundelegen sämtlicher Assoziationen, Abbildung 119 gibt einen Überblick über die Entwicklung der Kennwerte über die Meßzeitpunkte im Vergleich zwischen absoluter vs. hochfrequenter Auszählung, wobei diese beiden unterschiedlich erhobenen Kennwerte sich zu jedem Meßzeitpunkt jeweils signifikant voneinander abgrenzen lassen ($t_{T1}(34)=12,724; p=.000$; $t_{T2}(34)=13,543; p=.000$; $t_{T3}(34)=12,115; p=.000$). Bei Auszählung der relevanten Assoziationen wurde mit $M=6,90$ ($SD=2,03$) zur ersten Erhebung ein signifikant geringerer Kennwert erreicht als zum zweiten Meßzeitpunkt mit $M=9,69$ und $SD=1,91$ ($t(34)=-7,602; p=.000$), der Kennwert der zweiten Erhebung unterscheidet sich aber nicht signifikant vom letzten Meßzeitpunkt ($M=10,03; SD=2,89$).

	hochfrequente Assoziationen pro Stimulus			absolute Assoziationen pro Stimulus		
	T1	T2	T3	T1	T2	T3
<i>M</i>	6,90	9,69	10,03	11,29	13,59	13,39
<i>SD</i>	2,03	1,91	2,89	3,14	3,03	3,94
Minimum	3,67	5,67	1,00	5,00	7,50	2,33
Maximum	11,33	12,83	15,83	18,50	21,33	20,33

Abbildung 119: Kennwerte des Wortassoziationsversuchs bei Auszählung hochfrequenter Assoziationen – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 4 – Sparkassenakademie (B)

Die korrelativen Zusammenhänge nach Pearson zwischen dem absoluten und dem hochfrequenten Assoziationskennwert eines jeden Meßzeitpunktes erweisen sich als signifikant hoch ($r_{T1} = .745$; $p = .000$; $r_{T2} = .692$; $p = .000$; $r_{T3} = .871$; $p = .000$). Auch die Rangkorrelationen nach Spearman erbringen hohe signifikante Beziehungen zwischen den Auszählmethoden zu jedem der Meßzeitpunkte: $r_{T1} = .776$ ($p = .000$); $r_{T2} = .652$ ($p = .000$); $r_{T3} = .677$ ($p = .000$). Die Beurteilung der Assoziationsleistung eines Probanden führt also nicht zu grundsätzlich anderen Ergebnissen in Abhängigkeit davon, ob die Auszählmethode sämtliche Assoziationen eines Probanden berücksichtigt oder nur die innerhalb der Stichprobe hochfrequenten Assoziationen.

Während der Assoziationskennwert der absoluten Auszählung wie bereits beschrieben die in der Klausur erreichte Punktzahl mit $r = .338$ ($p = .023$) signifikant vorhersagen kann, ist dies anhand des durch Auszählen der hochfrequenten Assoziationen ermittelten Kennwertes kaum möglich ($r = .215$; $p > .100$). Allerdings läßt sich anhand der Rangfolge der Probanden hinsichtlich des letztgenannten Kennwertes eine gute Vorhersage der Rangfolge der Probanden in der Klausurleistung treffen: die Rangkorrelation nach Spearman erbringt mit $r = .403$ ($p = .008$) eine signifikante mittlere Korrelation. Damit kann der Kennwert der hochfrequenten Assoziationen die Rangfolge der Probanden bzgl. der Klausurleistung ähnlich gut vorhersagen wie der Kennwert der absoluten Assoziationen ($r = .454$; $p = .003$).

Die Zugrundelegung der hochfrequenten Assoziationen bei der Bildung individueller Kennwerte erbringt also in dieser Studie keinen zusätzlichen Gewinn hinsichtlich der Vorhersagequalität des Kriteriums. Zudem führen beide hier eingesetzten Auszählmethoden zu einer annähernd gleichen Beurteilung der Assoziationsleistung der Probanden innerhalb der Stichprobe.

6.1.4.2.3. Auszählung hochfrequenter Assoziationen in Studie 12: Handelsunternehmen

Liebs (2003) wertete den Wortassoziationsversuch der Studie 12 unter quantitativem Aspekt gemäß zweier Herangehensweisen aus: der Auszählung absoluter vs. hochfrequenter Assoziationen. Als hochfrequent wurde eine Nennung dann bewertet, wenn sie von mindestens zwei Probanden assoziiert wurde. Bei der Auszählung absoluter und hochfrequenter Assoziationen berichtet Liebs (2003) von großen interindividuellen Unterschieden. Abbildung 120 stellt die statistischen Kennwerte für die unterschiedlichen Auswertemöglichkeiten der Assoziationsergebnisse dar. Bei Auszählung nur der in der Gruppe hochfrequenten Assoziationen beträgt der Assoziationskennwert im Mittel 3,14 ($SD = 0,90$), während der Kennwert bei Auszählung sämtlicher Nennungen sich mit $M = 4,54$ ($SD = 1,66$) signifikant davon unterscheidet (Wilcoxon: $z = -3,621$; $p = .000$).

	hochfrequente Assoziationen pro Stimulus	absolute Assoziationen pro Stimulus
<i>M</i>	3,14	4,54
<i>SD</i>	0,90	1,66
Minimum	1,42	1,77
Maximum	5,27	8,42

Abbildung 120: Kennwerte des Wortassoziationsversuchs bei Auszählung hochfrequenter Assoziationen – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 12 – Handelsunternehmen

Ein auf Grundlage der hochfrequenten Assoziationen modifizierter Assoziationskennwert führt jedoch nicht zu einer dramatisch anderen Beurteilung der Probanden hinsichtlich ihrer Assoziationsleistung. Der Kennwert der hochfrequenten Auszählung läßt sich mit $r = .895$ ($p = .000$) sehr sicher durch die Gesamtmenge aller Assoziationen vorhersagen. Auch die Rangplätze korrelieren sehr hoch und signifikant zwischen den beiden Auswertestrategien ($r = .878$; $p = .000$). Der Bezug auf die hochfrequenten Assoziationen führt also nicht zu grundsätzlich verschiedenen Aussagen in Hinblick auf Leistungsgruppen innerhalb der Stichprobe – man würde auf Grundlage der verschiedenen Assoziationskennwerte einen Probanden nicht gänzlich unterschiedlich beurteilen. Liebs (2003) weist darauf hin, daß v.a. die Probanden mit extremen Assoziationsleistungen hinsichtlich ihres Rangplatzes relativ stabil bleiben.

Da in dieser Studie die Beurteilung der Probanden durch ihren Geschäftsführer zur Verfügung stand, sollen die modifizierten Assoziationskennwerte i.S. einer inkrementellen Validierung zu diesen Außenkriterien in Beziehung gesetzt werden. Wie bereits berichtet wurde, kann der Assoziationskennwert auf Basis sämtlicher Assoziationen (absolute Auszählung) die Geschäftsführerbeurteilung der Fachkompetenz kaum voraussagen ($r = -.230$; $p > .05$), die Gesamtbeurteilung der Arbeitsleistung jedoch mit $r = -.446$ ($p = .036$) recht zuverlässig. Der durch das Auszählen der hochfrequenten Assoziationen gewonnene Kennwert kann diese Vorhersagegüte nicht überbieten – es ergeben sich im Vergleich schwächere und zudem nicht signifikante Korrelationskoeffizienten ($r_{\text{Gesamturteil}} = -.315$; $p > .05$; $r_{\text{Fachkompetenz}} = -.169$; $p > .05$). Zusammenfassend läßt sich sagen, daß in dieser Studie durch die vorgenommene Modifizierung des Assoziationskennwertes die Beurteilung der Assoziationsleistung der Probanden im Vergleich zur Stichprobe kaum verändert. Eine Erhöhung der Vorhersagegüte des Außenkriteriums konnte durch die Modifizierung des Assoziationskennwertes nicht erreicht werden.

6.1.4.2.4. Auszählung hochfrequenter Assoziationen in Studie 13: Stoffdruckerei

In Ergänzung zu den von Pokrandt (2004) vorgenommenen Auswertungen wurden die Nennungen im Wortassoziationsversuch auch unter dem Aspekt der Frequenz der geäußerten Inhalte innerhalb der Stichprobe ausgezählt. Demnach wurden also die Assoziationen eines Probanden dahin gehend beurteilt, ob dieser Inhalt von mindestens zwei Probanden der Stichprobe geäußert wurde. War dies der Fall, ging diese Assoziation in die Berechnung des individuellen Assoziationskennwertes ein (Auszählung hochfrequenter Assoziationen), andernfalls blieb diese Assoziation unberücksichtigt. Dem wird die Auszählung sämtlicher Assoziationen gegenübergestellt, die entsprechenden Kennwerte finden sich in Abbildung 121. Danach werden bei Auszählung aller Assoziationen im Mittel 6,85 Äußerungen pro Stimulus produziert ($SD = 4,01$), bei Berücksichtigung derjenigen

Äußerungen, die von mindestens zwei Probanden der Stichprobe genannt wurden, konnten pro Stimulus durchschnittlich 4,49 Äußerungen gewertet werden ($SD=1,73$), dieser Unterschied ist signifikant (Wilcoxon: $z=-4,015$, $p=.000$).

	hochfrequente Assoziationen pro Stimulus	absolute Assoziationen pro Stimulus
<i>M</i>	4,49	6,85
<i>SD</i>	1,73	4,01
Minimum	2,21	2,83
Maximum	9,39	21,54

Abbildung 121: Kennwerte des Wortassoziationsversuchs bei Auszählung hochfrequenter Assoziationen – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 13 – Stoffdruckerei

Beide Assoziationskennwerte hängen stark miteinander zusammen ($r=.853$; $p=.000$); die Rangreihen, welche aufgrund der verschiedenen Auszählvarianten entstehen, ergeben ebenfalls einen hohen und signifikanten Korrelationskoeffizienten von $r=.789$ ($p=.000$) nach Spearman. Die Vorhersage des Außenkriteriums der Geschäftsführerbeurteilung (Fachkompetenz und Gesamtbeurteilung) durch den Assoziationskennwert auf Basis sämtlicher Nennungen erbrachte lediglich schwache und nicht-signifikante Korrelationskoeffizienten (siehe Kapitel 6.1.2.2.2, Seite 55 ff.), der individuelle Kennwert auf Grundlage der hochfrequenten Nennungen kann die Vorhersagequalität jedoch nicht erhöhen ($r_{Gesamturteil}=-.271$; $p>.05$; $r_{Fachkompetenz}=-.303$; $p>.05$).

Zusammenfassend ist zu sagen, daß die Auszählung der in der Stichprobe hochfrequenten Assoziationen in dieser Studie zwar eine bedeutsame Modifikation des Assoziationskennwertes bewirkt, dies jedoch keinen Einfluß auf die Beurteilung der Probanden hat. Ebenso erlaubt ein solcher Kennwert hier keine bessere Prognose des Außenkriteriums der Geschäftsführerbeurteilung als es der Assoziationskennwert auf Basis sämtlicher Assoziationen vermag.

6.1.4.2.5. Auszählung hochfrequenter Assoziationen in Studie 14: Seminar Wissensmanagement

In Studie 14 – Seminar Wissensmanagement – wurden ebenfalls Assoziationskennwerte auf Grundlage der hochfrequenten Nennungen innerhalb der Stichprobe gebildet. Dazu wurden alle Assoziationen herangezogen, die von mindestens zwei Probanden der Stichprobe benannt wurden. In Abbildung 122 wird deutlich, daß sich der Wissensbesitz in der Gruppe über die Individuen hinweg angleicht, homogener wird (da die Auszählung hochfrequenter Assoziationen die individuelle Assoziationsleistung in Relation zu der der restlichen Gruppe setzt). Vor Beginn des Seminars halbiert sich der mittlere Assoziationskennwert ($M=4,65$; $SD=2,72$), wenn statt sämtlicher nur noch die hochfrequenten Assoziationen gezählt werden. Bei Ende des Seminars hingegen beträgt der „Verlust“ aufgrund der strengeren Auszählmethode (hochfrequente Assoziationen; $M=10,26$; $SD=4,12$) nur noch etwa ein Drittel des absoluten Assoziationskennwertes. Die so modifizierten Assoziationskennwerte, deren statistische Kennwerte in Abbildung 122 dargestellt sind, unterscheiden sich sowohl bei der ersten Erhebung ($t(19)=-9,058$; $p=.000$) als auch zur zweiten Messung ($t(19)=-5,563$; $p=.000$) signifikant vom jeweiligen sog. absoluten Assoziationskennwert. Die mittlere Anzahl hochfrequenter Assoziationen des ersten Meßzeitpunktes unterscheidet sich dabei signifikant von der der zweiten Erhebung ($t(19)=-5,706$;

$p = .000$). Legt man also den Gruppenkonsens als Kriterium an, so gelingt es nach dem Treatment relativ häufiger, über diese Hürde zu springen, als vor dem Treatment.

	T1 (Seminarbeginn)		T2 (Seminarende)	
	hochfrequent	absolut	hochfrequent	absolut
<i>M</i>	4,65	10,00	10,26	16,35
<i>SD</i>	2,72	4,22	4,12	7,50
Minimum	0	3	5	7
Maximum	10	18	19	32

Abbildung 122: Kennwerte des Wortassoziationsversuchs bei Auszählung hochfrequenter Assoziationen – Mittelwerte (*M*) und Standardabweichungen (*SD*) über die gesamte Stichprobe in Studie 14 – Seminar Wissensmanagement

Der Assoziationskennwert der hochfrequenten Auszählung bei T1 läßt sich sehr gut anhand des absoluten Assoziationskennwertes vorhersagen, mit $r = .794$ ($p = .000$) wurde ein hoher und signifikanter korrelativer Zusammenhang (nach Pearson) deutlich, die Rangreihe der Probanden (nach Spearman) erbrachte ebenfalls einen signifikanten hohen Koeffizienten von $r = .774$ ($p = .000$). Für den zweiten Erhebungszeitpunkt wurde diese Beziehung sogar noch deutlicher: mit $r = .857$ ($p = .000$) zeigt sich eine hohe signifikante Korrelation nach Pearson, mit $r = .899$ ($p = .000$) läßt sich die Rangreihe der Probanden sehr zuverlässig vorhersagen. Beide Kennwerte führen also zu einer annähernd gleichen Beurteilung der Probanden hinsichtlich ihrer Assoziationsleistung.

Im Sinne der inkrementellen Validierung soll auch hier die Vorhersage des modifizierten Kennwertes hinsichtlich der Leistung im Multiple-Choice-Test geprüft werden. Der Kennwert der absoluten Auszählung erbrachte sowohl mittels der Korrelation nach Pearson als auch anhand der Rangkorrelation nach Spearman zu beiden Meßzeitpunkten lediglich geringe, nicht signifikante Korrelationskoeffizienten (in der zweiten Erhebung sogar mit negativem Vorzeichen, siehe Kapitel 6.1.3.1.5.1, Seite 55 ff.). Auch der Assoziationskennwert auf Grundlage der hochfrequenten Assoziationen kann die Vorhersagegüte nicht verbessern: zur ersten Erhebung zeigt sich eine nicht signifikante Korrelation zur Leistung im Multiple-Choice-Test ($r = .258$; $p > .05$), ebenso wie bei der zweiten Messung ($r = -.323$; $p > .05$). Weder der absolute noch der hochfrequente Assoziationskennwert kann in dieser Studie also die Testleistung im Multiple-Choice-Test angemessen vorhersagen.

Neben der rein quantitativen Dominanz der während der Lehrveranstaltung erworbenen Wissensinhalte ist eine weitergehende qualitative Betrachtung aufschlußreich, die als quantitative Inhaltsanalyse aufgefaßt werden kann. Dies soll nun hier exemplarisch für Studie 14 vorgenommen werden. Dabei wurde für jeden der beiden Erhebungszeitpunkte ausgezählt, wie oft eine Assoziation innerhalb der Stichprobe benannt wurde. Abbildung 123 und Abbildung 124 stellen diese hochfrequenten Assoziationen für den ersten bzw. zweiten Erhebungszeitpunkt dar. Dabei wurden aus Gründen der Anschaulichkeit lediglich diejenigen Assoziationen berücksichtigt, die von mindestens vier Probanden geäußert wurden, das entspricht 20 % der Stichprobe.

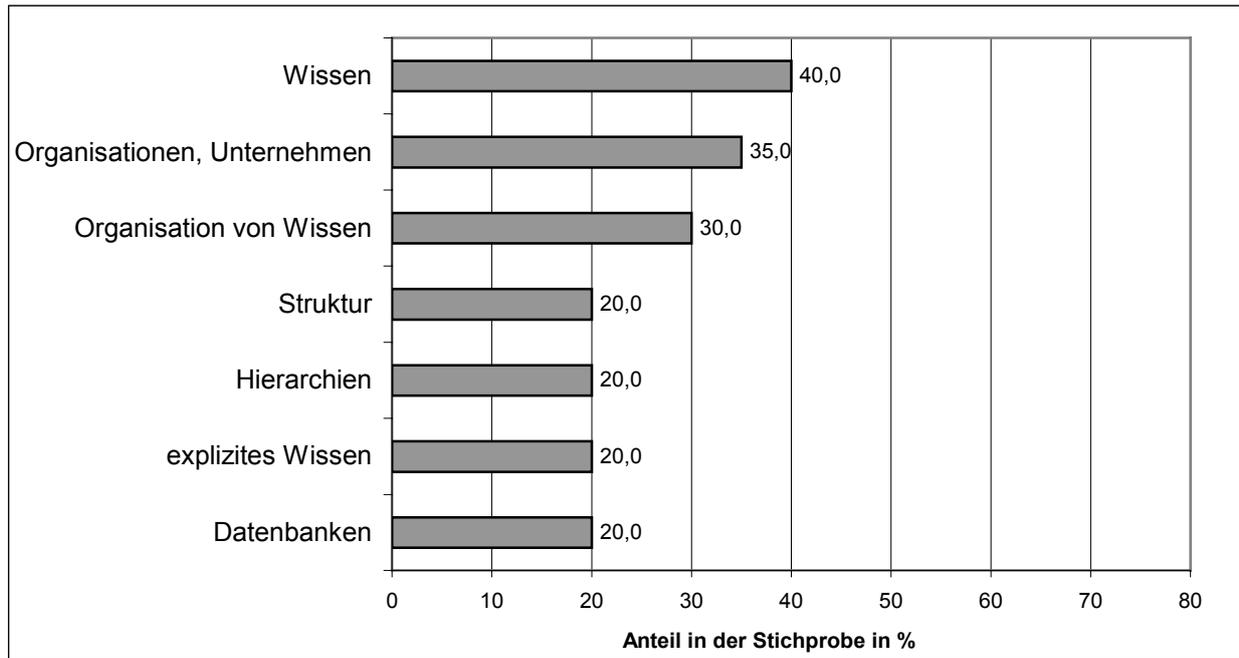


Abbildung 123: Hochfrequente Assoziationen vor Seminarbeginn (T1) in Studie 14 – Seminar Wissensmanagement

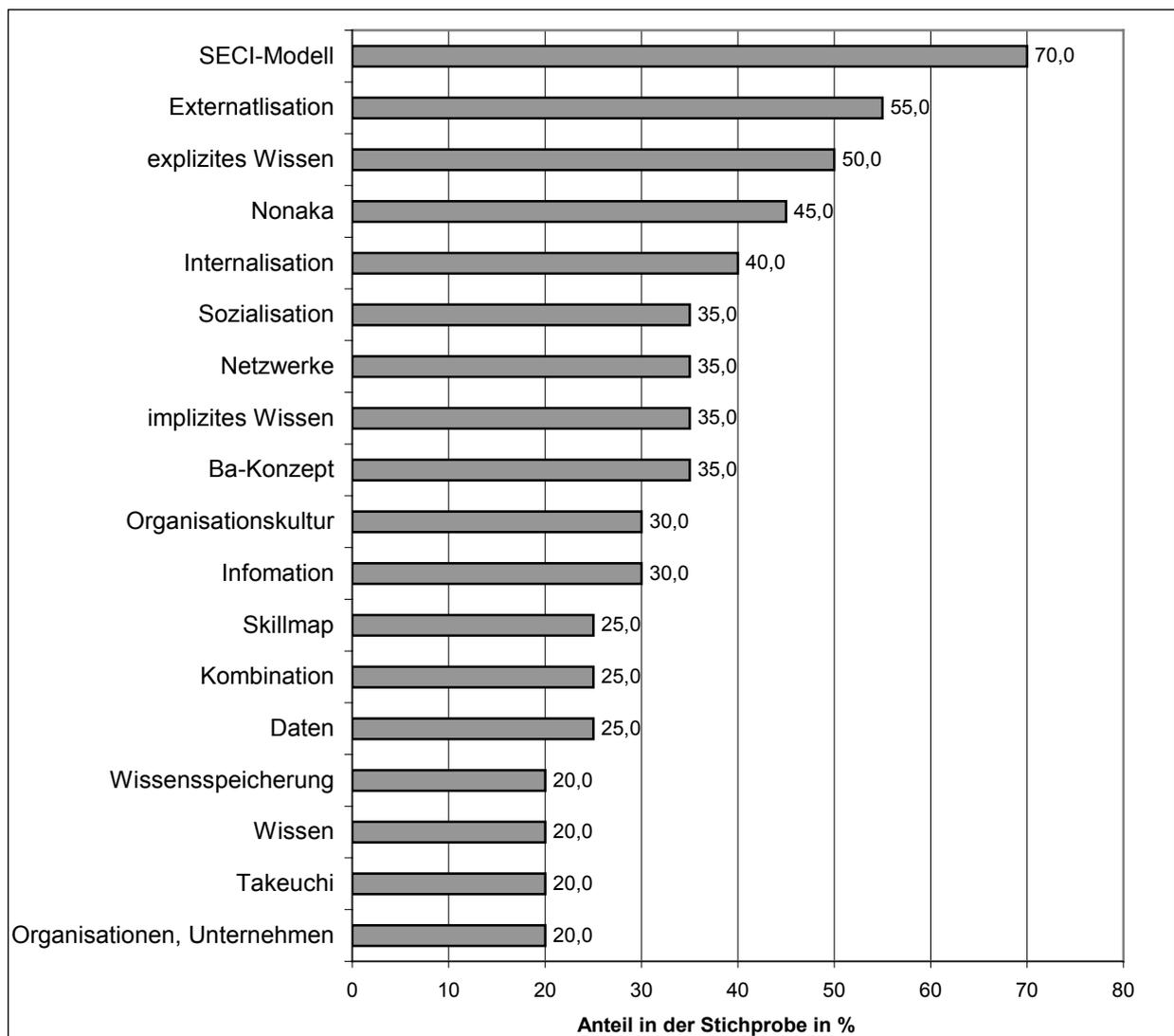


Abbildung 124: Hochfrequente Assoziationen bei Seminarende (T2) in Studie 14 – Seminar Wissensmanagement

Auf den ersten Blick wird deutlich, daß der Wissensbesitz innerhalb der Stichprobe in der Post-Erhebung einheitlicher wird. Während vor dem Seminar nur sieben Nennungen das Kriterium für die hochfrequenten Assoziationen erfüllten, waren es hingegen am Ende des Seminars 18 Konzepte. Zudem wurden die hochfrequenten Assoziationen der zweiten Erhebung von einem größeren Anteil der Stichproben benannt als zum ersten Meßzeitpunkt. So benannten zu Seminarbeginn 40 % der Stichprobe (entspricht acht Probanden) die Assoziation mit der höchsten Frequenz („Wissen“), hingegen wurde die häufigste Nennung der Post-Erhebung („SECI-Modell“) von 70 % der Stichprobe assoziiert (entspricht 14 Probanden). Ebenso wird deutlich, daß die zu Seminarbeginn hochfrequenten Assoziationen eher allgemeinerer Natur sind (z.B. Wissen, Organisation, Struktur), während bei Seminarende überwiegend hochspezifische Konzepte assoziiert wurden (z.B. SECI-Modell, Nonaka, Ba-Konzept, Internalisation).

Zusammenfassend ist also zu sagen, daß die Bildung eines Assoziationskennwertes auf Grundlage der hochfrequenten Assoziationen innerhalb der Stichprobe in dieser Studie zu keiner grundlegend anderen Beurteilung der Probanden führt als bei Einbeziehung sämtlicher Assoziationen. Auch eine genauere Vorhersage der individuellen Leistungen im Multiple-Choice-Test kann der so modifizierte Kennwert nicht leisten. Die Orientierung an den hochfrequenten Nennungen machte allerdings eindrucksvoll deutlich, inwiefern das Treatment die Bildung eines geteilten Wissens innerhalb der Stichprobe vorantrieb. Es konnte gezeigt werden, daß die im Seminar vermittelten Inhalte Eingang in die individuellen Wissensbestände fanden.

6.1.4.2.6. Zusammenfassung und Befundintegration der Studien zur Auszählung hochfrequenter Assoziationen

In den fünf soeben vorgestellten Studien, in denen ein modifizierter Assoziationskennwert auf Grundlage der in der Stichprobe hochfrequenten Assoziationen gebildet wurde, ergeben sich durchweg hohe Übereinstimmungen in der Rangreihe der Probanden verglichen mit jenem Assoziationskennwert, in den sämtliche Assoziationen eines Probanden eingingen. Abbildung 125 gibt einen Überblick über die Koeffizienten der Rangkorrelation nach Spearman zwischen absolutem und hochfrequentem Assoziationskennwert in den einzelnen Studien (für Studie 3 – Sparkassenakademie A – waren aufgrund der Datenlage keine solchen Angaben mehr verfügbar). Führt man diese Koeffizienten i.S. einer mittleren gewichteten Effektstärke zusammen (bei Studien mit mehreren Meßzeitpunkten wurde jeweils der letzte berücksichtigt), so zeigt sich mit $d = .79$ ein großer Effekt.

Führt man die Befunde aller Studien anhand der z-transformierten Daten zusammen, so kann man dabei auf eine Gesamtstichprobe von $N=92$ zurückgreifen (bei den Studien 4 und 14 wurde jeweils der letzte Erhebungszeitpunkt zugrunde gelegt). In Abbildung 126 sind die Korrelationskoeffizienten dargestellt, welche die Beziehung zwischen dem absoluten und dem hochfrequenten Assoziationskennwert darstellen. So läßt sich der eine Kennwert sehr gut aus dem anderen vorhersagen ($r = .891$; $p = .000$), ebenso weisen die Rangreihen beider Auszählvarianten signifikante hohe korrelative Beziehungen auf ($r = .873$; $p = .000$). Interpretiert man diese Koeffizienten i.S. von Effektstärken, so zeichnen sich hier starke Effekte ab. Erwartungsgemäß wird dieser Befund noch ausgeprägter, wenn man nur die Probanden mit extremer Assoziationsleistung zugrunde legt (siehe Abbildung 126).

Rangkorrelation zwischen absoluten vs. hochfrequentem Assoziationskennwert	
4 - Sparkassenakademie (B)	T1: $r = .776^{**}$ ($p = .000$)
	T2: $r = .652^{**}$ ($p = .000$)
	T3: $r = .677^{**}$ ($p = .000$)
12 - Handelsunternehmen	$r = .878^{**}$ ($p = .000$)
13 - Stoffdruckerei	$r = .789^{**}$ ($p = .000$)
14 - Seminar Wissensmanagement	T1: $r = .774^{**}$ ($p = .000$)
	T2: $r = .899^{**}$ ($p = .000$)

Abbildung 125: Übersicht über die Korrelationskoeffizienten der einzelnen Studien (Assoziationskennwerten nach absoluter vs. hochfrequenter Auszählung) als Rangkorrelation nach Spearman

	Korrelation nach Pearson	Rangkorrelation nach Spearman
Gesamtstichprobe (N=92)	$r = .891^{**}$ ($p = .000$)	$r = .873^{**}$ ($p = .000$)
Extremstichprobe (N=45)	$r = .929^{**}$ ($p = .000$)	$r = .900^{**}$ ($p = .000$)

Abbildung 126: Korrelationen zwischen dem absoluten und dem hochfrequenten Assoziationskennwert (z-transformiert)

In Studie 4 – Sparkassenakademie (B) – konnte der modifizierte Assoziationskennwert eine ähnlich gute Vorhersage des Außenkriteriums (Klausurleistung) erbringen wie der absolute Kennwert; in Studie 14 – Seminar Wissensmanagement – können beide Kennwerte das Außenkriterium (Multiple-Choice-Test) ähnlich schlecht vorhersagen. In den Studien 12 – Handelsunternehmen – und 13 – Stoffdruckerei – konnten Geschäftsführerbeurteilungen der Fachkompetenz sowie Gesamtbeurteilungen als Außenkriterium herangezogen werden, die z-transformierten Daten beider Studien sollen nun zusammengeführt werden (N=38). Während in den beiden einzelnen Studien eine Vorhersage des Außenkriteriums durch den modifizierten Assoziationskennwert kaum möglich war, treten bei der Befundintegration deutliche Zusammenhänge zutage (siehe Abbildung 127). So kann durch den hochfrequenten Assoziationskennwert die Fachkompetenz mit $r = -.290$ ($p = .039$) einigermaßen zuverlässig vorhergesagt werden, zur Gesamtbeurteilung besteht mit $r = -.318$ ($p = .026$) ebenfalls eine schwache signifikante korrelative Beziehung. Somit erbringt der modifizierte Assoziationskennwert eine ähnlich gute Vorhersagegüte für das Außenkriterium wie der absolute Kennwert. Bei Betrachtung der Extremstichprobe (N=22) verstärkt sich wie erwartet der korrelative Zusammenhang, wenngleich er zur Vorhersage der Fachkompetenz nicht signifikant wird (siehe Abbildung 127).

	Fachkompetenz		Gesamtbeurteilung	
	absolut	hochfrequent	absolut	hochfrequent
Gesamtstichprobe (N=38)	$r = -.297^*$ ($p = .035$)	$r = -.290^*$ ($p = .039$)	$r = -.349^*$ ($p = .016$)	$r = -.318^*$ ($p = .026$)
Extremstichprobe (N=22)	$r = -.343$ ($p = .059$)	$r = -.305$ ($p > .05$)	$r = -.467^*$ ($p = .014$)	$r = -.486^*$ ($p = .011$)

Abbildung 127: Korrelationen zwischen der Geschäftsführerbeurteilung (Fach- bzw. Gesamtkompetenz; z-transformiert) mit dem absoluten bzw. hochfrequenten Assoziationskennwert (z-transformiert) für die Gesamtstichprobe sowie die Extremstichproben

Zusammenfassend bleibt festzuhalten: bei Auszählung der hochfrequenten Assoziationen ergeben sich keine grundsätzlich anderen Beurteilungen der Probanden als man es aufgrund der Auszählung sämtlicher Assoziationen tun würde, beide Parameter weisen einen hohen korrelativen Zusammenhang auf. Ebenso erbringt der modifizierte Assoziationskennwert im Vergleich mit dem absoluten Kennwert keinen bedeutsamen Zugewinn zur Vorhersage von Außenkriterien.

Grundsätzlich stellt sich die Frage, inwieweit man die Beurteilung der Assoziationsleistung eines Probanden davon abhängig machen sollte, wie „gut“ oder „schlecht“ die anderen Mitglieder der Stichprobe sind. Assoziiert beispielsweise ein Spezialist einer Wissensdomäne vorrangig hoch differenzierte Inhalte, welche von seinen weniger spezialisierten Kollegen nicht genannt werden (vielleicht benennen diese eher Konzepte allgemeinerer Natur), so würde dieser Proband bei Auszählung der hochfrequenten Assoziationen relativ schlecht abschneiden – er hätte zwar viel (für die Wissensdomäne Relevantes) assoziiert, aber leider waren es andere Inhalte als bei seinen Kollegen. Der so gebildete Assoziationskennwert könnte solche Spezialisten oder Experten nicht abbilden und sie – im Vergleich zur Stichprobe – eher schlechter erscheinen lassen als sie tatsächlich sind. Als Grundlage für eine Leistungsbeurteilung einzelner Probanden ist dieses Vorgehen also zumindest mit Vorsicht zu betrachten.

Dennoch ist das Erheben von hochfrequenten Assoziationen ein durchaus lohnendes Vorgehen, wenn die Gruppe als Ganzes beschrieben werden soll. Im Ergebnis einer quantitativen Inhaltsanalyse lassen sich Aussagen darüber treffen, wie relevant ein Wissenselement in der Untersuchungsgruppe ist. So kann beispielsweise im Anschluß an eine Bildungsmaßnahme geprüft werden, ob die von der Lehrperson als zentral erachteten und entsprechend vermittelten Begriffe einer Thematik auch tatsächlich in den individuellen Wissensbesitz der einzelnen Lernenden integriert wurden. Auch zur Konstruktion eines hypothetischen Wissensmodells hat sich die Referenz auf die hochfrequenten Assoziationsinhalte bewährt. Die Analyse der hochfrequenten Assoziationen innerhalb einer Stichprobe gibt Hinweise auf so etwas wie geteiltes Wissen in der Gruppe. Als Beurteilungsgrundlage von Probanden hingegen scheint ein in dieser Weise modifizierter Assoziationskennwert wenig geeignet und wenig hilfreich.

6.1.4.3. Beispiel für einen qualitativen Zugang bei der Auswertung von Assoziationsdaten

Bereits bei der Erörterung der Frage der Gewichtung unterschiedlicher Assoziationensqualitäten (Kapitel 6.1.4.1 auf Seite 55) sowie bei der Auszählung hochfrequenter Assoziationen innerhalb der Stichprobe (Kapitel 6.1.4.2 auf Seite 55) wurde den Inhalten von Assoziationen Beachtung geschenkt und gezeigt, daß nicht alleine das *Wieviel* an Assoziationen Aufschluß gibt, sondern auch das *Was* analysiert werden kann. Im Folgenden soll anhand des Beispiels der Studie 15 – Wissen über Arbeitsschutz – demonstriert werden, welchen Erkenntnisgewinn eine noch stärker qualitativ geprägte Herangehensweise für die Beantwortung einer praktischen Fragestellung erbringen kann. In diesem Beispiel gelang es nicht, erwartungskonforme Unterschiede zwischen definierten Subgruppen der Stichprobe anhand rein quantitativer Kennwerte befriedigend zu beschreiben. Durch eine qualitative Betrachtung der Assoziationen jedoch wurde diese Differenzierung zwischen den Gruppen sichtbar. Der methodische Zugang läßt sich dabei als quantitative Inhaltsanalyse beschreiben, da einerseits die Inhalte der Assoziationen im Fokus der Auswertungen stehen, und

andererseits dabei versucht wird, diese in verschiedenen Kategorien zusammenzufassen und so zu quantifizieren.

6.1.4.3.1. Einführung in die praxisrelevante Fragestellung von Studie 15: Wissen über Arbeitsschutz

Wie bereits beschrieben, richtete sich die zentrale Zielstellung in Studie 15 auf die Wissensbestände von Studierenden in Bezug auf die Domäne Arbeits- und Gesundheitsschutz. Aufgrund von Voruntersuchungen (Interviews mit universitären Experten des Sicherheitswesens, Vorlesungsverzeichnisse, Curricula) bestanden bzgl. des arbeitsschutzrelevanten Wissens für verschiedene Fächergruppen bestimmte Erwartungen bzw. Vermutungen, welche im Folgenden kurz beschrieben werden sollen (siehe auch Ceglarek, 2007). So war bekannt, daß in einigen Studienfächern Themen des Arbeits- und Gesundheitsschutz regulärer Bestandteil des Curriculums sind. Dies betrifft die Fachbereiche Biologie, Chemie, Biochemie, Ernährungswissenschaften sowie Arbeitslehre. Aufgrund der Arbeit in Laboren bzw. an Werkmaschinen sind entsprechende Unterweisungen für die Studierenden unumgänglich – dem erhöhten Unfallrisiko wird durch entsprechende, curricular verankerte Lehrveranstaltungen, Testate, Unterweisungen u.ä. systematisch Rechnung getragen. Aufgrund dessen wird in Studiengängen dieser Art ein höherer Kenntnisstand zum Arbeits- und Gesundheitsschutz erwartet. Auch bei Studenten im Studiengang Sportwissenschaften wurde ein hoher Wissensstand angenommen, da diese aufgrund ihrer sportlicher Betätigung bzw. beim Umgang mit diversen Sportgeräten eine erhöhte Wahrscheinlichkeit haben, einen Sportunfall zu erleiden (was auch durch die höchste Unfallquote innerhalb der Universität belegt ist). Durch Experteninterviews war bekannt, daß Unfallprävention immer wieder während der praktischen Übungen thematisiert und dabei sehr konkret auf z.B. Unfallgefahren bei einer bestimmten Sportart eingegangen wird, aber nicht in dieser systematischen Form wie in den oben beschriebenen naturwissenschaftlichen Fächern. Zu allen anderen Studienfächern bestand die Annahme, daß die Studierenden geringe bis gar keine Kenntnisse im Arbeits- und Gesundheitsschutz besitzen, da solche Inhalte nicht Gegenstand von Lehrveranstaltungen oder Curricula waren.

6.1.4.3.2. Quantitativer Zugang in Studie 15: Wissens über Arbeitsschutz

Die Grundannahme in dieser Studie läßt sich also wie folgt zusammenfassen: einige Studenten sollten im Rahmen ihrer Curricula bereits Wissen über Arbeitsschutz erworben haben, während andere während ihres Studiums keinen Kontakt zu solchen Lehrinhalten haben und daher vermutlich über vergleichsweise wenig Wissen darüber verfügen. Davon ausgehend wurde die Stichprobe auf Grundlage der Studienfächer in zwei Gruppen eingeteilt: Studenten der einschlägigen Fächer (Biologie, Chemie, Biochemie, Ernährungswissenschaften, Arbeitslehre, Sportwissenschaften) als sog. Experten vs. Studenten aller anderen Fächer als sog. Laien. So war etwa ein Viertel der gesamten Stichprobe den Experten zuzuordnen (N=112), die Mehrheit der Probanden (N=360) wurde aufgrund des Studienfaches als Laie klassifiziert.

Die Experten hoben sich hinsichtlich der erhobenen Wissensvariablen nicht besonders deutlich von den Laien ab. Es gab zwar einen (quantitativen) Wissensvorsprung der Experten, aber der Abstand zur Gruppe der Laien war relativ gering und im Ganzen war der Umfang des Wissens geringer

ausgefallen als erwartet. Dies trifft sowohl für die Menge richtiger Antworten im Multiple-Choice-Test zu als auch für die Leistungen im Wortassoziationsversuch. Abbildung 128 stellt die mittleren Assoziationskennwerte der beiden Gruppen sowie deren Leistung im Multiple-Choice-Test gegenüber. Danach assoziieren die Experten mit $M=3,10$ ($SD=1,48$) signifikant mehr ($t(469)=-2,635$, $p=.009$) als die Laien mit $M=2,66$ ($SD=1,57$), der Unterschied ist aber marginal, zumal in beiden Substichproben ein ähnlicher Range abgedeckt wird. Diese Signifikanz dürfte wohl v.a. auf die Stichprobengröße zurückzuführen sein, die Effektstärke ist mit $d=.28$ als gering zu bezeichnen. Ebenso wird in Abbildung 128 deutlich, daß auch das Ergebnis des Multiple-Choice-Test nicht besonders gut zwischen Experten und Laien unterscheiden kann. Während die Gruppe der Experten von 58 maximal erreichbaren Punkten im Mittel 30,88 ($SD=7,89$) Punkte erreicht, schafft es die Gruppe der Laien auf immerhin durchschnittlich 26,70 Punkte ($SD=8,90$). Auch dieser Unterschied ist statistisch bedeutsam ($t(470)=-4,445$, $p=.000$), aber grenzt die beiden Gruppen eher unscharf voneinander ab, auch hier ist die Effektstärke mit $d=.48$ als eher gering zu interpretieren.

	Assoziationskennwert		Punkte im Multiple-Choice-Test	
	Experten (N=112)	Laien (N=360)	Experten (N=112)	Laien (N=360)
<i>M</i>	3,10	2,66	30,88	26,70
<i>SD</i>	1,48	1,57	7,89	8,90
<i>Minimum</i>	0,0	0,0	11	0
<i>Maximum</i>	9,0	8,7	46	50

Abbildung 128: Kennwerte des Wortassoziationsversuchs sowie der erreichten Punkte im Multiple-Choice-Test – Mittelwerte (M) und Standardabweichungen (SD) im Experten-Laien-Vergleich in Studie 15 – Wissen über Arbeitsschutz

Auf der Ebene der quantitativen Betrachtung der beiden Wissensdiagnoseverfahren wurden also große interindividuelle Unterschiede zwischen den Probanden beobachtet (beide Parameter sind normalverteilt mit einer großen Streubreite). Dennoch lassen sich diese Unterschiede nicht befriedigend allein durch das Studienfach erklären (Vergleich Experten vs. Laien). Aus diesem Grunde sollte regressionsanalytisch geklärt werden, ob andere Merkmale der Person die erheblichen interindividuellen Wissensunterschiede womöglich besser erklären könnten als das Studienfach. Als Variablen wurden dabei berücksichtigt: Alter der Probanden, Geschlecht, Erwerbstätigkeit oder Ausbildung vor dem Studium, Anzahl der Semester, eigene Erfahrungen mit (Arbeits-) Unfällen sowie die Angaben, ob die Probanden die Thematik Arbeitsschutz in der Lehre behandelt oder aber sich selbständig dazu informiert haben. Sowohl für den Assoziationskennwert als auch für den Punktwert im Multiple-Choice-Test konnte keine befriedigende Varianzaufklärung auf Basis der Personenmerkmale erbracht werden (siehe auch Ceglarek, 2007). Keine der in dieser Studie erhobenen Variablen war also in der Lage, die großen quantitativen interindividuellen Wissensunterschiede hinreichend zu erklären.

Als mögliche Erklärung für die geringen Wissenskennwerte der Experten werden Transferprobleme zwischen Erwerb und Abruf angenommen: das z.B. im Rahmen der Laborpraktika erworbene Wissen ist kontextual eingebettet in das Setting Labor und an die ganz konkreten Bedingungen der zu absolvierenden Experimente gebunden. Demgegenüber steht eine vergleichsweise abstrakte Ebene, auf der sich Wortassoziationsversuch sowie Multiple-Choice-Test bewegen. Es ist denkbar,

daß bei den Probanden das (konkrete) Wissen aus den Praktika kaum unter der (globalen) Kategorie Arbeitsschutz abgelegt ist. Gleichsam ist zu bedenken, daß das im Labor erworbene Wissen in starkem Maße handlungsorientiert ist – pädagogisches Ziel der Praktika ist das sichere Handeln und Verhalten im Labor (im Ergebnis der Experteninterviews sowie aufgrund der Unfallstatistik wurde deutlich, daß dies gelungen ist). Für die Erfassung von prozeduralem Wissen sind Wortassoziationsversuch und auch Multiple-Choice-Test weniger geeignete Instrumente, wenngleich die Studierenden aufgrund der praktikumsbegleitenden Vorlesungen und Testate auch über deklaratives Wissen verfügen müßten.

6.1.4.3.3. Qualitativer Zugang in Studie 15: Wissens über Arbeitsschutz

Die Aussage, daß ein Studium mit arbeitsschutzrelevanten Inhalten nicht zu bedeutend mehr Wissen über Arbeitsschutz führt als ein Studium ohne solche Thematiken, ließe den Schluß zu, die akademische Lehre wäre wenig wirksam und somit verzichtbar. Wie soeben diskutiert, gibt es durchaus methodische Bedenken hinsichtlich der Angemessenheit der Wissensdiagnoseinstrumente mit Blick auf die Art bzw. Repräsentationsform des in der Lehre erworbenen Wissens. Aber auch auf Basis der vorhandenen Assoziationsdaten läßt sich die o.g. Schlußfolgerung bereits widerlegen. Denn die Nachhaltigkeit der akademischen Lehre kann in der vorliegenden Studie weniger durch das *Wieviel* an Wissen belegt werden, statt dessen zeigt sich, daß *etwas anderes* gewußt wird.

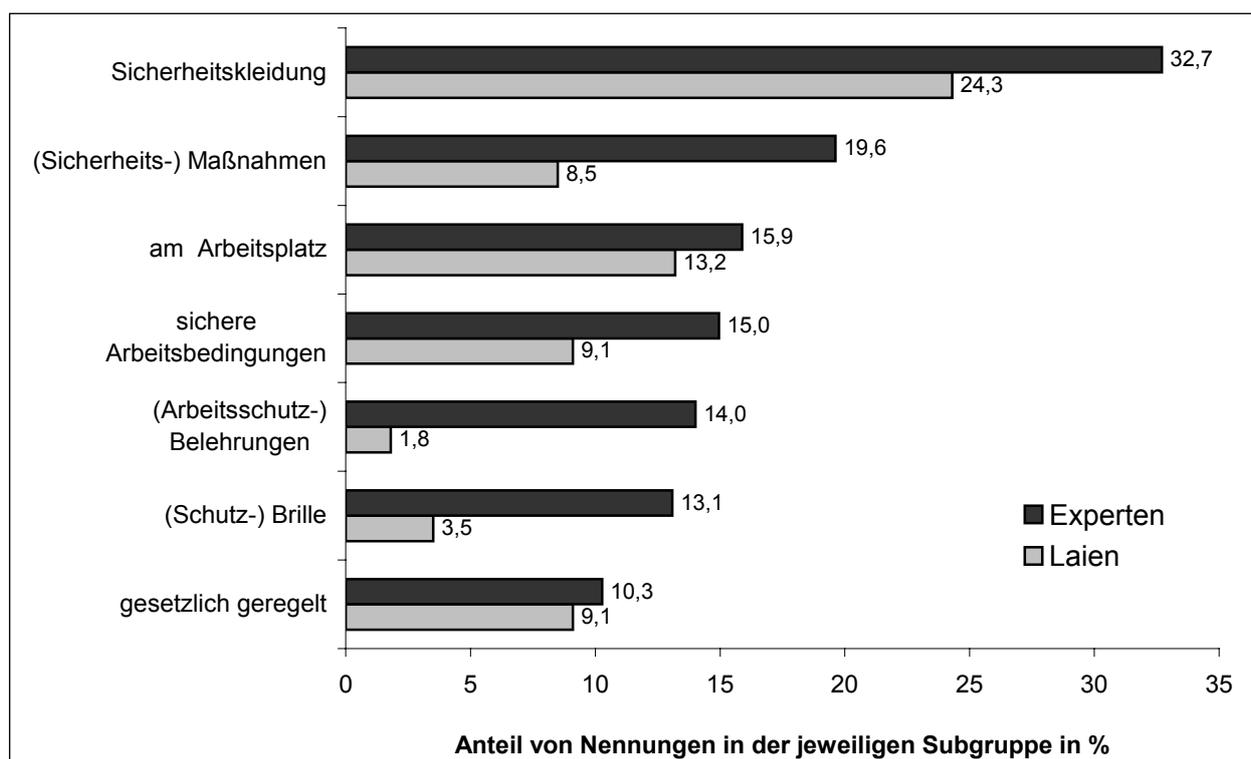


Abbildung 129: Hochfrequente Assoziationen zum Stimulus „Arbeitsschutz“ im Experten-Laien-Vergleich in Studie 15 – Wissen über Arbeitsschutz

Da nun die großen interindividuellen Unterschiede in beiden Wissensparametern (Multiple-Choice-Test und Wortassoziationsversuch) weder durch das Merkmal Studienfach noch andere Personenmerkmale erklärt werden konnten, wurden die Assoziationen unter inhaltlichem Aspekt analysiert. Ziel dabei war, möglicherweise Unterschiede im Wissensbesitz von Experten vs. Laien

unter qualitativem Aspekt beleuchten zu können, da sich diese wie gesagt rein quantitativ nicht befriedigend abbilden ließen. Abbildung 129 stellt beispielhaft die hochfrequenten Assoziationen zum Stimulus „Arbeitsschutz“ im Vergleich der beiden Subgruppen Experten vs. Laien dar. Dabei wurden nur diejenigen Äußerungen einer jeden Gruppe berücksichtigt, die von mindestens 10 % der Probanden der jeweiligen Gruppe genannt wurden. Erfüllte eine Assoziation in einer der Gruppen dieses Kriterium, so wurde zum Vergleich auch deren Häufigkeit in der anderen Gruppe gegenübergestellt. Es wird deutlich, daß es keine (hochfrequenten) Assoziation gibt, welche in höherem Maße von den Laien benannt wurde – innerhalb der Gruppe der Experten wurden alle Assoziationen ähnlich häufig oder häufiger benannt als in der Gruppe der Laien. Somit kann für die Gruppe der Experten in stärkerem Maß von einem gemeinsamen Wissen oder auch geteilten Wissensbesitz gesprochen werden. Die Assoziation „Sicherheitskleidung“ ist die insgesamt häufigste Äußerung, in der Gruppe der Experten benannte dies etwa jeder dritte Probanden (32,7 %), in der Gruppe der Laien etwa jeder Vierte (24,3 %), dieser Unterschied ist tendenziell signifikant ($\chi^2(1, N=472)=3,06; p > .05$).

Betrachtet man im Weiteren die Inhalte der Assoziationen, so fällt auf, daß sich die Gruppe der Experten durch eine größere Spezifität und Fachausdrücke auszeichnet. Bei den in Abbildung 129 dargestellten hochfrequenten Assoziationen betrifft dies insbesondere die Begriffe (Sicherheits-) Maßnahmen ($\chi^2(1, N=472)=10,32; p = .002$), (Arbeitsschutz-) Belehrungen ($\chi^2(1, N=472)=27,63; p = .000$) sowie (Schutz-) Brille ($\chi^2(1, N=472)=13,79; p = .001$), welche nicht nur einen augenscheinlichen, sondern auch einen statistisch signifikanten Häufigkeitsunterschied der Assoziationen im Experten-Laien-Vergleich markieren. Betrachtet man den gesamten Pool an geäußerten Assoziationen in den beiden Stichproben (also nicht nur die hochfrequenten), so zeichnet sich die Experten-Gruppe durch solche hochspezifischen Begriffe wie „Gefahrstoffsymbole“, „MAK-Werte“ oder „Abzug“ aus, während Äußerungen dieser Art bei der Gruppe der Laien kaum anzutreffen sind.

6.1.4.3.4. Zusammenfassung der Befunde zur qualitativen Auswertung in Studie 15: Wissens über Arbeitsschutz

Ausgangsbasis in Studie 15 war die Frage, inwieweit Studierende über ein allgemeines Wissen über Arbeits- und Gesundheitsschutz verfügen. Dabei bestand die Annahme, daß Studierende einschlägiger Fachrichtungen (z.B. Biochemie), in denen aufgrund der Arbeit in Laboren oder an Werkmaschinen o.ä. Arbeitsschutzmaßnahmen in das Studium integriert sind, besser abschneiden sollten als Studierende anderer Fachrichtungen. Die erwarteten Wissensunterschiede zwischen diesen zwei Gruppen ließen sich nicht zufriedenstellend anhand der rein quantitativen Auszählung assoziierter Konzepte abbilden. Die Betrachtung i.S. einer quantitativen Inhaltsanalyse machte jedoch die Wissensunterschiede auf einer qualitativen Ebene deutlich: die Gruppe der sog. Experten assoziierte zwar im Mittel nicht sehr viel mehr als die Gruppe der sog. Laien, dennoch zeigten die von den Experten benannten Assoziationen eine höhere Spezifität und wurden außerdem von einem größeren Teil der Gruppe geäußert, so daß hier von einem gemeinsamen Wissen innerhalb der Gruppe gesprochen werden kann. Somit kann gezeigt werden, daß Inhalte der Lehrveranstaltungen einschlägiger Fächer in die individuellen Wissensbestände vieler Studierender, welche der Gruppe der Experten zuzuordnen waren, übergegangen sind und erinnert werden konnten.

6.1.4.4. Zusammenfassung der Befunde zu den Modifikationen des Assoziationskennwertes

In Ergänzung zu den Befunden die Validität betreffend sollte geprüft werden, inwieweit eine Modifikation des numerischen Assoziationskennwertes die Beurteilung eines Probanden beeinflussen. Dazu wurden zwei Ansätze verfolgt: zum einen die Gewichtung unterschiedlicher Qualitäten von Assoziationsinhalten, zum anderen die Bildung eines Assoziationskennwertes in Referenz zu den innerhalb der Stichprobe hochfrequenten Assoziationen. In vier Studien wurde eine quantitative Inhaltsanalyse vorgenommen, indem die Inhalte der Assoziationen hinsichtlich ihrer Qualität beurteilt wurden und dann gewichtet in den Assoziationskennwert eingingen (Fragestellung 6a). Es fanden sich durchweg hohe Übereinstimmungen zwischen den Rangreihen des gewichteten vs. ungewichteten Kennwertes, so daß dieses Vorgehen keine Auswirkungen auf die Beurteilung der Probanden hat. Auch die Vorhersagegüte für ein Außenkriterium konnte durch die Gewichtung nicht erhöht werden gegenüber dem sonst üblichen Assoziationskennwert. Allerdings läßt sich durch eine solche qualitativ orientierte Auseinandersetzung mit den Wissensinhalten der Wissensbesitz in der Stichprobe sehr anschaulich beschreiben.

Ähnliches ist zu sagen bei einem zweiten Ansatz i.S. einer quantitativen Inhaltsanalyse (Fragestellung 6b): in fünf Studien wurde der Assoziationskennwert modifiziert auf Grundlage der in der Stichprobe hochfrequenten Nennungen (alleinig von einem Probanden gemachte Äußerungen gingen nicht in den Kennwert ein). Auch hier weisen der modifizierte und der ursprüngliche Assoziationskennwert hohe Rangkorrelationen auf, zudem kann auch ein Außenkriterium nicht besser durch den modifizierten Kennwert vorhergesagt werden als es bereits durch den bisher üblichen Assoziationskennwert möglich war. Dennoch ist auch dieser Analyseansatz aufschlußreich, um die Stichprobe zu beschreiben, da dieser modifizierte Kennwert auf so etwas wie das geteilte Wissen in der Gruppe zurückgreift. Insbesondere bei Mehrfacherhebung lassen sich z.B. Effekte von Bildungsmaßnahmen auf diese Weise anschaulich abbilden.

In Ergänzung zu diesen an numerischen Werten orientierten Ansätzen wurde demonstriert, welchen Gewinn eine stärker inhaltlich ausgerichtete Auseinandersetzung mit den Assoziationen der Probanden erbringen kann. Im vorgestellten Beispiel konnten vermutete Leistungsunterschiede zwischen Subgruppen der Stichprobe nicht besonders deutlich anhand numerischer Kennwerte abgebildet werden. Durch die (qualitativ-quantitative) Analyse der jeweiligen Assoziationsinhalte jedoch ließen sich beide Gruppen voneinander abgrenzen.

Zusammenfassend ist also zu sagen, daß der in der vorliegenden Arbeit herangezogene Assoziationskennwert, welcher die mittlere Anzahl von Assoziationen pro Stimulus beschreibt, ein brauchbarer Parameter zu sein scheint. Die soeben vorgestellten Möglichkeiten der Modifikation dieses Kennwertes beeinflussen die Rangreihe der Probanden innerhalb der Stichprobe nicht wesentlich und können daher auch keinen zusätzlichen Erkenntnisgewinn im Sinne der inkrementellen Validität erbringen. Allerdings können Modifikationen dieser Art sehr nützlich sein, um den Wissensbesitz innerhalb der Stichprobe anschaulich zu beschreiben. Unter diesem Aspekt besitzen sie also durchaus ihre Berechtigung.

6.1.5. Zusammenfassung der Befunde zur Validität

Der Wortassoziationsversuch kann sich hinsichtlich seiner Validität gut behaupten (Fragestellung 3). Die Inhaltsvalidität kann als gegeben hingenommen werden – die Relevanz oder

Repräsentativität potentieller Stimuli für die interessierende Wissensdomäne ist abgesichert durch die intensive Auseinandersetzung des Versuchsleiter mit der Wissensdomäne im Vorfeld der Erhebungen sowie die Einbeziehung betrieblicher Experten.

Hinsichtlich der Kriteriumsvalidität konnten empirische Befunde aus insgesamt 11 verschiedenen Studien herangezogen werden, wobei zwei verschiedene Arten von Außenkriterien zur Verfügung standen. In neun Studien im Kontext akademischer sowie betrieblicher Aus- und Weiterbildung konnte der Wortassoziationsversuch eingesetzt und das in einer Klausur abgeprüfte Wissen als Außenkriterium herangezogen werden. Lediglich in drei dieser Studien fanden sich keine (signifikanten) Zusammenhänge zwischen Assoziationskennwert und Klausurleistung, in der Mehrzahl der Studien jedoch zeigten sich signifikante Korrelationskoeffizienten, welche schwach, mittel oder hoch ausgeprägt waren. Integriert man die Befunde aller Studien, so läßt sich die Klausurnote aus dem Assoziationskennwert vorhersagen (schwache signifikante Korrelation); die in der Klausur erreichten Punkte lassen sich ebenfalls gut aus der Assoziationsleistung vorhersagen, was sich in einer signifikanten Korrelation mittlerer Stärke ausdrückt.

In zwei Studien wurde das Urteil des Geschäftsführers über die Arbeitsleistung seiner Mitarbeiter (als Gesamturteil sowie separates Urteil über die Fachkompetenz) herangezogen, die Befunde dazu sind uneindeutig. Während in einer Studie signifikante Korrelationen mittlerer Stärke zwischen Assoziationskennwert und Gesamtbeurteilung gefunden werden konnten, zeichnete sich in der zweiten Studie nur eine schwache Tendenz ab. Integriert man die Befunde beider Studien anhand ihrer z-transformierten Werte, so ergeben sich schwache, aber signifikante Korrelationen – der Assoziationskennwert ist also in der Lage, sowohl die eingeschätzte Fachkompetenz als auch die Gesamtbeurteilung des Geschäftsführers vorherzusagen. Es sei an dieser Stelle daran erinnert, daß das Geschäftsführerurteil die gesamte Arbeitsleistung umfaßt, woran das fachspezifische Wissen nur einen Teilaspekt ausmacht.

Für die Bestimmung der (konvergenten) Konstruktvalidität standen in mehreren Studien verschiedene andere wissensdiagnostische Instrumente zur Verfügung. So wurde in einer Studie ein Lückentext eingesetzt, welcher dieselben Inhalte aufgriff wie die Stimuli des Wortassoziationsversuchs. Dabei erweist sich der Assoziationskennwert als ein guter Prädiktor für die Leistung im Lückentext, belegt durch eine signifikante Korrelation mittlerer Stärke. In einer weiteren Studie kam ein Tiefeninterview zum Einsatz, welches die gleiche Wissensdomäne thematisierte wie der Wortassoziationsversuch. Es konnten keine signifikanten Zusammenhänge zwischen den Kennwerten beider Verfahren gefunden werden. Zudem erwies sich das Tiefeninterview selbst als ein sehr guter Prädiktor für das Außenkriterium Geschäftsführerurteil, dem Wortassoziationsversuch dabei um ein Vielfaches überlegen. Allerdings ist anzumerken, das ein Tiefeninterview in weitaus stärkerem Maße als der Wortassoziationsversuch die Wissensreproduktion lenkt und darüber hinaus in seiner Auswertung die inhaltliche Passung zum Soll-Wissen berücksichtigt, während der Assoziationskennwert sich „lediglich“ aus der Anzahl assoziierter Konzepte zusammensetzt. Allerdings ist ein Tiefeninterview nur im Einzelversuch durchführbar und dadurch wenig ökonomisch.

In drei Studien wurde neben dem Wortassoziationsversuch die Strukturlegetechnik eingesetzt, wobei allerdings geringfügige Unterschiede hinsichtlich der konkreten Durchführungsbedingungen zu berücksichtigen sind. Grundsätzlich ist festzustellen, daß das Zurückgreifen auf zuvor

selbstassoziierte Konzepte zu einem höheren Zusammenhang zwischen den Parametern des Strukturlegens und dem Assoziationskennwert führt als wenn für alle Probanden ein einheitlicher Pool an zu legenden Knoten vorgegeben wurde. Der gewichtete Korrespondenzkoeffizient (also die Übereinstimmung eines Probandennetzwerkes mit einem Expertennetzwerk) erweist sich als völlig unbrauchbarer Kennwert, in keiner Studie konnten bedeutsame Zusammenhänge zum Assoziationskennwert gefunden werden (im Übrigen auch nicht zu einem anderen Außenkriterium). Insbesondere die Anzahl der Knoten im Netz läßt sich gut durch den Assoziationskennwert vorhersagen (mittlere bis sehr hohe Korrelationen), und das, obwohl in den Studien überwiegend die Knoten den Probanden vorgegeben wurden (alle Probanden hatten die gleiche Menge Knoten zur Verfügung). Integriert man die Befunde aller Studien, so weist der Strukturlegeparameter des Durchmessers den stärksten Zusammenhang zum Assoziationskennwert auf (signifikante Korrelation mittlerer Stärke), die Parameter Anzahl der Knoten, Anzahl der Kanten sowie Verknüpfungsdichte können ebenfalls gut aus der Assoziationsleistung vorhergesagt werden. In den Studien, in denen außer der Strukturlegetechnik noch ein weiteres Außenkriterium zur Verfügung stand, konnten die Parameter der Strukturlegetechnik keine bessere Vorhersage dieses Kriteriums leisten als es bereits der Assoziationskennwert kann.

In vier Studien wurde der Assoziations-Struktur-Test als weitere wissensdiagnostische Methode eingesetzt. Da dieses Verfahren unmittelbar an die Ergebnisse des Wortassoziationsversuchs anknüpft (ein Teil oder alle der assoziierten Konzepte werden im weiteren Verlauf des Assoziations-Struktur-Tests bearbeitet), finden sich durchweg mittlere bis sehr hohe signifikante Korrelationen zwischen den Parametern des Assoziations-Struktur-Tests und dem Assoziationskennwert. In der Befundintegration aller Studien erweist sich der Assoziationskennwert insbesondere für die Vorhersage der im Wissensgraphen verwendeten Knoten als ein sehr guter Prädiktor (signifikante sehr hohe Korrelation). Die Parameter des Assoziations-Struktur-Tests selbst können die etwaigen anderen Außenkriterien gut vorhersagen, gehen dabei aber überwiegend über die Vorhersagegüte des Wortassoziationsversuchs kaum hinaus.

Zwei der Studien griffen auf Multiple-Choice-Tests zurück. Während die eine dieser Studien eine signifikante, aber schwache Korrelation zwischen Assoziationskennwert und der im Multiple-Choice-Test erreichten Punktzahl erbringt, läßt sich in der anderen Studie kein Zusammenhang zwischen den Kennwerten nachweisen, es deutet sich sogar eine negative Tendenz an. Zwar erbringt die Integration der Befunde eine signifikante schwache Korrelation, dies ist aber vermutlich v.a. auf die Größe der Stichprobe zurückzuführen. Insgesamt betrachtet erweist sich der Multiple-Choice-Test also am wenigsten geeignet, um einen Zusammenhang mit dem Assoziationskennwert nachzuweisen, allerdings kann die Leistung in diesem Verfahren – im Vergleich zu den anderen hier vorgestellten Methoden – am ehesten durch Raten verzerrt werden.

Insgesamt ist auffällig, daß sich beim Extremgruppenvergleich die Gruppe derjenigen Probanden, die beim Wortassoziationsversuch sehr viel assoziiert haben, immer etwas homogener erscheinen hinsichtlich ihrer Ausprägung im Kriterium (geringere Standardabweichung) als die Gruppe der Probanden mit sehr geringer Assoziationsleistung. Eine hohe Assoziationsleistung geht also in der Regel mit einer hohen Leistung im Kriterium einher, während dies nicht analog auf die untere Extremgruppe übertragen werden kann: Probanden mit geringem Assoziationskennwert haben nicht notwendigerweise auch eine geringe Ausprägung im Kriterium, sondern decken oftmals einen

großen Wertebereich hinsichtlich des Kriteriums ab. Wenn ein Probanden im Sinne der freien Reproduktion viel assoziiert hat, so kann dies nur durch einen entsprechend umfangreichen Wissensbesitz erklärt werden (Raten o.ä. ist kaum möglich), während bei einer geringen Menge assoziierter Begriffe nicht eindeutig geklärt werden kann, ob dies auf einen geringen Wissensumfang oder aber auf mangelnde Teilnahmebereitschaft zurückzuführen ist.

Ferner wurde zur Bestimmung der diskriminanten Konstruktvalidität in drei Studien der Einfluß des Personenmerkmals der allgemeinen Wortflüssigkeit auf die Assoziationsleistung thematisiert. In zwei dieser Feldstudien konnte ein solcher Einfluß ausgeschlossen werden, in der dritten Studie (mit experimentellem Ansatz) deutet sich ein Zusammenhang an, der allerdings nur für die mündlich dargebotenen Stimuli aus dem Allgemeinwissensbereich nachgewiesen werden konnte, für das spezialisierte Stimulusmaterial dieser Erhebung war ein solcher Effekt nicht erkennbar. Aufgrund der empirischen Befunde kann insgesamt davon ausgegangen werden, daß die Assoziationsleistung von Arbeitspersonen in einer berufsspezifischen Wissensdomäne unbeeinflusst ist von der generellen Wortflüssigkeit eines Probanden. Im Sinne der diskriminanten Konstruktvalidität ist also festzuhalten, daß der Wortassoziationsversuch in der Lage ist, das Konstrukt Fachwissen zu erfassen, ohne dabei von einem anderen Faktor (generelle Wortflüssigkeit) beeinflusst zu sein.

Zusammenfassend kann die Validität des Wortassoziationsversuchs als belegt angesehen werden. In vielen Einzelstudien sowie in den Befundintegrationen konnten bedeutsame Zusammenhänge zu verschiedenen Außenkriterien bzw. alternativen Wissensdiagnoseinstrumenten nachgewiesen werden.

6.2. Reliabilität

Das folgende Kapitel setzt sich mit der Reliabilität des Wortassoziationsversuchs auseinander (Fragestellung 4). Methodische Ansätze zur Reliabilitätsbestimmung wie Paralleltestverfahren oder interne Konsistenz sind für den Wortassoziationsversuch unangemessen. Alle diese Verfahren setzen eine Vergleichbarkeit von Items voraus, wodurch sich zwei äquivalente Itemgruppen mit identischer Schwierigkeit entwickeln lassen. Diese Forderung ist beim Wortassoziationsversuch nicht erfüllbar aufgrund der Spezifität der Stimuli und der dahinterstehenden Inhalte. Somit steht einzig die Testwiederholung als Methode der Reliabilitätsbestimmung zur Verfügung. Da dabei die Annahme besteht, daß die Assoziationsleistung zu beiden Meßzeitpunkten durch den Rückgriff auf den gleichen Wissensbesitz bestimmt wird, ist also zwischen den beiden wissensdiagnostischen Erhebungen ein angemessenes Zeitintervall zu plazieren, in dem keine Veränderung des Wissensbesitzes (z.B. durch Fortbildungsveranstaltungen) zu erwarten sind. Dazu wurde eine separate Erhebung konzipiert, die als Feldexperiment genau diesem Anspruch Rechnung trägt. Dabei besteht die Annahme, daß Fachwissen zwar grundsätzlich veränderbar ist, aber durchaus über einen gewissen Zeitraum stabil sein kann. Im Unterschied zu den anderen hier vorgestellten Studien mit mehreren Meßzeitpunkten, bei denen zwischen den Erhebungen ein Treatment stattfand und daher das Ziel darin bestand, durch den Wortassoziationsversuch Veränderungen im Wissensbesitz abzubilden, ist das zentrale Merkmal dieses Designs, daß der individuelle Wissensbestand zwischen beiden Erhebungen nicht gezielt beeinflusst wurde.

Zur Bestimmung der Retest-Reliabilität soll nun also Studie 2 – psychotherapeutische Ambulanz – herangezogen werden (für Details zum Versuchsaufbau sei auf Kapitel 4.2.2, Seite 55 f. verwiesen). Lienert und Raatz (1998) raten vom Einsatz der Testwiederholung zur Reliabilitätsbestimmung ab mit dem Argument, die Reliabilität würde aufgrund von Erinnerungseffekten überschätzt werden. Diesem Einwand kann wie folgt begegnet werden: zum einen war die Stimulusliste des Wortassoziationsversuchs mit insgesamt 14 Stimuli verhältnismäßig lang, noch dazu umfaßte sie wenig spezifische Stimuli (Diagnosegruppen für psychische Störungen im Kindes- und Jugendalter), so daß die Probanden im Durchschnitt eine bis anderthalb Stunden für die Bearbeitung der Stimuli benötigten. Es wurde bewußt dieser relativ große Umfang an Stimulusbegriffen gewählt, da auf diese Weise ein Erinnern von Assoziationen aus der erstmaligen Testsituation unwahrscheinlicher ist. Als Zeitintervall zwischen den Erhebungen waren zwei Wochen gewählt worden, da dieser Zeitraum lang genug schien, um die Erinnerung an das eigene Testverhalten zu minimieren, und andererseits kurz genug, um den Einfluß von Lernereignissen auf den Wissensbesitz möglichst gering zu halten. Nach Wirtz und Caspra (2002) wird durch die Testwiederholung v.a. die zeitliche Stabilität der Testwerte erfaßt.

6.2.1. Retest-Reliabilität des Assoziationskennwertes in Studie 2: psychotherapeutische Ambulanz

Wie bereits in Kapitel 4.2 (Seite 55 ff.) besprochen, unterscheiden sich die mittleren Assoziationskennwerte in der Stichprobe nicht signifikant zwischen den beiden Meßzeitpunkten (Wilcoxon: $z = -.149$; $p > .05$). Während zum ersten Erhebungszeitpunkt im Mittel 21,70 Konzepte ($SD = 16,67$) assoziiert wurden, waren es bei der Wiederholungsmessung $M = 19,14$ ($SD = 9,50$). Die mittleren Assoziationsleistungen sind also zu beiden Erhebungen vergleichbar. Dieser Befund bildet sich auch durch einen hohen und hoch signifikanten Korrelationskennwert ab: mit $r = .849$ ($p = .000$) läßt sich der Assoziationskennwert eines Probanden in der einen Erhebung sehr sicher anhand seines Assoziationskennwertes in der zweiten Erhebung vorhersagen. Auch die Rangkorrelation nach Spearman erbringt eine signifikante mittlere bis hohe Korrelation ($r = .665$; $p = .001$). Somit kann die Reliabilität des Assoziationskennwertes als gegeben hingenommen werden.

In Ergänzung zur Reliabilitätsbestimmung anhand des Assoziationskennwertes, welcher die Assoziationsleistung eines Probanden über alle Stimuli zusammenfaßt, soll nun kurz auf die Aspekte der Testwiederholung auf der nächst differenzierteren Ebene eingegangen werden. Dazu wurde für jeden Probanden separat ermittelt, inwiefern die Assoziationsmenge zu einem bestimmten Stimulus bei der Wiederholungsmessung repliziert wird. Somit konnte für jeden Probanden ein Korrelationskoeffizient nach Pearson ermittelt werden, welcher (gemäß der 14 Stimuli) auf 14 Wertepaaren basiert, die Ergebnisse sind in Abbildung 130 dargestellt. Für die Mehrheit der Probanden fanden sich mittlere bis hohe signifikante korrelative Beziehungen. Somit erlaubt nicht nur der Assoziationskennwert als über alle Stimuli aggregierter Parameter eine Vorhersage der Assoziationsleistung bei Testwiederholung, sondern auch auf Ebene einzelner Stimuli (eine hohe Assoziationsleistung zu Stimulus X geht auch bei wiederholter Testung mit ähnlich hoher Assoziationsleistung beim selben Stimulus einher). Lediglich bei fünf Probanden konnte keine eindeutige Beziehung der stimulusbezogenen Assoziationsleistungen zu den beiden Erhebungszeitpunkten belegt werden. Es ist unklar, wodurch diese Befunde verursacht wurden.

Proband	Korrelationskoeffizient (nach Pearson)	Proband	Korrelationskoeffizient (nach Pearson)
A	$r = .893^{**}$ ($p = .000$)	K	$r = .665^{**}$ ($p = .005$)
B	$r = .870^{**}$ ($p = .000$)	L	$r = .651^{**}$ ($p = .006$)
C	$r = .859^{**}$ ($p = .000$)	M	$r = .586^*$ ($p = .014$)
D	$r = .809^{**}$ ($p = .000$)	N	$r = .519^*$ ($p = .029$)
E	$r = .727^{**}$ ($p = .002$)	O	$r = .492^*$ ($p = .037$)
F	$r = .713^{**}$ ($p = .002$)	P	$r = .329$ ($p > .05$)
G	$r = .706^{**}$ ($p = .002$)	Q	$r = .261$ ($p > .05$)
H	$r = .705^{**}$ ($p = .002$)	R	$r = .253$ ($p > .05$)
I	$r = .670^{**}$ ($p = .004$)	S	$r = .203$ ($p > .05$)
J	$r = .667^{**}$ ($p = .005$)	T	$r = .016$ ($p > .05$)

Abbildung 130: Intraindividuelle Reliabilität auf Stimulusebene in Studie 2 – psychotherapeutische Ambulanz

6.2.2. Qualitative Aspekte der Testwiederholung in Studie 2: psychotherapeutische Ambulanz

Neben der Berücksichtigung des quantitativen Assoziationskennwertes läßt sich die Testwiederholung auch unter qualitativem Aspekt betrachten. Dabei stellt sich die Frage, inwieweit über die Replikation der reinen Anzahl assoziierter Konzepte hinaus die Assoziationen auch inhaltlich übereinstimmen. Dazu wurden für jeden Probanden für jeden Stimulus die inhaltlich übereinstimmenden Assoziationen bestimmt und deren Anteil an der jeweiligen individuellen Gesamtmenge von Assoziationen gemäß folgender Gleichung berechnet:

$$\text{Anteil inhaltlicher Übereinstimmung} = \frac{100 * (\text{Anzahl identischer Assoziationen} * 2)}{(\text{Anzahl Assoziationen T1} + \text{Anzahl Assoziationen T2})}$$

Im Ergebnis dieser Gleichung ergibt sich für jeden Probanden der prozentuale Anteil inhaltlich übereinstimmender Konzepte an der Gesamtmenge der zu beiden Meßzeitpunkten assoziierten Konzepten zu einem Stimulus. Zur Veranschaulichung sei an dieser Stelle angemerkt, daß eine Übereinstimmung von 100 % dann erreicht werden kann, wenn zu einem Stimulus in der zweiten Erhebung genauso viele und außerdem inhaltlich identische Assoziationen produziert wurden wie bei der ersten Erhebung. Würden hingegen beispielsweise bei T2 genau doppelt so viel assoziiert wie bei T1, wobei sämtliche Konzepte aus T1 in der zweiten Erhebung wiederholt würden, so betrüge der Anteil an inhaltlich übereinstimmenden Assoziationen 66 %.

Anteil inhaltlicher Übereinstimmung in %	
<i>M</i>	45,78
<i>SD</i>	9,56
Minimum	31,03
Maximum	60,48

Abbildung 131: Statistische Kennwerte (Mittelwert und Standardabweichung) der Anteile inhaltlich übereinstimmender Assoziationen zwischen T1 und T2 (in %) in Studie 2 – psychotherapeutische Ambulanz

Auf diese Weise wurden für jeden Probanden 14 Werte gemäß der 14 Stimuli bestimmt, woraus wiederum für jeden Probanden ein Mittelwert über die Anteile inhaltlicher Übereinstimmungen gebildet werden konnte. Wie dieser Wert sich über die gesamte Stichprobe verteilt, ist in Abbildung 131 dargestellt. Danach beträgt der mittlere Anteil inhaltsgleicher Assoziationen $M=45,78\%$ ($SD=9,56$). Es wurde also etwa knapp die Hälfte derjenigen Assoziationen, welche in der ersten Erhebung benannt wurden, auch in der Wiederholungsmessung produziert. Dieser Mittelwert bewegt sich dabei zwischen minimal etwa einem Drittel und maximal etwa zwei Dritteln inhaltlicher Übereinstimmungen (siehe Abbildung 131).

An dieser Stelle sollen zwei Beobachtungen zur Illustration berichtet werden: zum einen war es verblüffend, wie manche Probanden bei der Testwiederholung stellenweise die exakt gleichen Formulierung in der exakt gleichen Reihenfolge analog zur ersten Erhebung produzierten. Andererseits wurde mitunter sehr deutlich, daß ein Proband unterschiedliche inhaltliche Schwerpunkte setzte, indem er z.B. bei Ersterhebung v.a. auf die therapeutischen Aspekte der einzelnen Störungsbilder fokussierte, während bei Testwiederholung offenbar die Diagnostik dieser Störungsbilder im Zentrum seiner Aufmerksamkeit stand. Da es sich bei den dargebotenen 14 Stimuli durchweg um sehr globale Reizworte handelte (die Bezeichnungen für psychische Störungsbilder; siehe Kapitel 4.2.3 auf Seite 55), welche theoretisch einen sehr großen Pool möglicher Wissensinhalte umfassen können, ist der gefundene Anteil inhaltlich übereinstimmender Assoziationen bei Testwiederholung bemerkenswert hoch.

6.2.3. Stimuli verschiedener Abstraktionsebenen bei Testwiederholung in Studie 2: psychotherapeutische Ambulanz

Die Frage der inhaltlichen Reliabilität soll nun abschließend in Beziehung zur Hierarchieebene der Stimuli gesetzt werden, welche bereits in Kapitel 5.2 (Seite 55 ff.) thematisiert wurde. Wie im vorangegangenen Kapitel dargestellt und auch in Abbildung 132 ersichtlich, betrug die mittlere Übereinstimmung der Assoziationsinhalte beider Erhebungszeitpunkte über alle Stimuli hinweg $45,78\%$ ($SD=9,56$) – betrachtet man dies differenziert nach den einzelnen Abstraktionsebenen der Stimuli (Ober-, Zwischen-, Unterbegriffe), so ändert sich dieser Befund kaum.

	über alle Stimuli	Oberbegriffe	Zwischenbegriffe	Unterbegriffe
<i>M</i>	45,78	46,89	45,57	45,58
<i>SD</i>	9,56	14,15	10,56	10,16
Minimum	31,03	26,92	28,77	29,45
Maximum	60,48	71,71	64,70	59,65

Abbildung 132: Statistische Kennwerte (Mittelwert und Standardabweichung) der Anteile (in %) inhaltlicher Übereinstimmungen zwischen T1 und T2 in Studie 2 – psychotherapeutische Ambulanz

Selbst, wenn man weiter differenziert nach den Abstraktionsebenen in den Themen F8 bzw. F9, lassen sich keine signifikanten Unterschiede zwischen den Stimuluskategorien finden. Unterscheidet man dabei auch noch danach, ob die jeweiligen Themen in induktiver oder deduktiver Anordnung präsentiert wurden, so lassen sich wiederum keine Unterschiede zwischen den Gruppen belegen. Es spielt also für die inhaltliche Übereinstimmung der Assoziationen über die

beiden Meßzeitpunkte hinweg keine Rolle, auf welchem Abstraktionsniveau sich ein Stimulus bewegt bzw. in welcher Reihenfolge Stimuli verschiedener Abstraktionsniveaus angeordnet sind.

6.2.4. Zusammenfassung der Befunde zur Reliabilität

Aufgrund der durch die Stimuli eines Wortassoziationsversuchs bedingten inhaltlichen Spezifität steht zur Bestimmung der Reliabilität dieses Verfahrens nur die Methode der Testwiederholung zur Verfügung. Durch die hier vorgestellten empirischen Befunde eines Feldexperiments kann die Retest-Reliabilität des Wortassoziationsversuchs belegt werden (Fragestellung 4): mit $r = .849$ ($p = .000$) weisen die Assoziationskennwerte zweier Erhebungszeitpunkte mit identischem Wortassoziationsversuch einen hohen korrelativen Zusammenhang auf. Die Anzahl assoziierter Konzepte eines Probanden bleibt also bei Testwiederholung einigermassen stabil, wenn zwischen den Erhebungszeitpunkten keine gezielten Maßnahmen den Wissensbesitz betreffend stattfanden. Lienert und Raatz (1998) bewerten einen Reliabilitätskoeffizienten, der kleiner als .60 ist, als niedrig, für die Retest-Reliabilität von standardisierten Tests (was für den Wortassoziationsversuch nicht zutrifft) fordern sie einen Koeffizienten von mindestens .70. Daran orientiert kann der hier gefundene Koeffizient als ausreichend betrachtet werden.

Selbst, wenn man über die bloße Anzahl der assoziierten Konzepte hinaus auch deren Inhalte betrachtet, so ergibt sich eine hohe Übereinstimmung zwischen den Erhebungszeitpunkten: im Mittel war knapp die Hälfte der Assoziationen bei beiden Testdurchführungen zu finden (auf Ebene einzelner Stimuli bei einzelnen Probanden). Dieser hohe Anteil inhaltlich übereinstimmender Assoziationen zwischen beiden Erhebungszeitpunkten ist um so bemerkenswerter, als daß die im Feldexperiment dargebotenen Stimuli sehr allgemeiner Natur waren. Daher wäre bei deutlich spezifischeren Stimuli (für die vorliegende Studie z.B. „Diagnostik von Störung X“) eine noch größere inhaltliche Übereinstimmung zwischen den Meßzeitpunkten zu erwarten, dies sollte ggf. in weiteren Feldexperimenten geprüft werden. Differenziert man die Assoziationsleistungen hinsichtlich der Eigenschaften des Stimulus (Oberbegriff zur Domäne, Unterbegriffe auf niedrigeren Abstraktionsebenen), so erbringen die Probanden über die beiden Erhebungszeitpunkte hinweg vergleichbare Anteile inhaltlicher Übereinstimmungen in diesen Klassen von Stimuli, unabhängig davon, in welcher Weise diese Stimuli verschiedener Abstraktionsebenen angeordnet waren. Die zeitliche Stabilität (Wirtz und Caspar, 2002) sowohl der Assoziationskennwerte als auch der dahinterstehenden semantischen Inhalte können aufgrund der vorliegenden empirischen Befunde als gegeben hingenommen werden.

6.3. Objektivität

Zur Objektivität (Fragestellung 5) wurden keine empirischen Daten erhoben, vielmehr sollen die verschiedenen Aspekte der Objektivität ausdiskutiert werden. Diese betreffen die Durchführung und Auswertung des Wortassoziationsversuchs sowie die Interpretation der erhobenen Daten.

6.3.1. Durchführungsobjektivität

Die Durchführungsobjektivität des Wortassoziationsversuchs kann grundsätzlich als gegeben hingenommen werden (siehe dazu auch Rothe, 2003), es bestehen aber dennoch Zusammenhänge zu den Durchführungsbedingungen, welche den Probanden unterschiedlich große Freiräume lassen.

Die größere Gefahr für die Objektivität besteht bei einer mündlichen Durchführung des Wortassoziationsversuchs: die Interaktion mit dem Probanden erweist sich als kritisch, wenn der Proband so etwas wie eine natürliche Gesprächssituation herzustellen versucht, z.B. durch Rückfragen an den Versuchsleiter. Produziert der Proband statt schlagwortartiger Antworten komplexe Ausführungen in ganzen Sätzen, so kann sich der Versuchsleiter genötigt sehen, ihn an die Instruktion zu erinnern. Dem kann entgegnet werden durch eine schriftliche Instruktion, die vom Versuchsleiter vorgelesen oder dem Probanden vorgelegt wird (so wie es in allen hier vorliegenden Studien gehandhabt wurde), sowie das Vorbereiten entsprechender Floskeln, um an die Instruktion zu erinnern (z.B. „Bitte versuchen Sie, nur in Schlagworten oder kurzen Phrasen zu antworten!“). Ein „Helfen“ oder „Vorsagen möglicher Lösungen“ durch den Versuchsleiter ist aufgrund des Umstandes, daß der Wortassoziationsversuch ein Verfahren der freien Reproduktion ist, ohnehin auszuschließen.

Bei einer schriftlichen Durchführung des Wortassoziationsversuchs ist diese Interaktion zwischen Probanden und Versuchsleiter grundsätzlich auf ein Minimum reduziert, insbesondere wenn der Wortassoziationsversuch als Gruppenversuch durchgeführt wird. Auch hier mindert das Vorliegen einer Instruktion den Einfluß des Versuchsleiters. Durch den Einsatz computerbasierter Tools wie z.B. dem Assoziations-Struktur-Test von Meyer (2007) läßt ein Höchstmaß an Standardisierung realisieren. Zusammenfassend kann die Durchführungsobjektivität des Wortassoziationsversuchs als grundsätzlich gegeben angenommen werden, was durch das Zurückgreifen auf eine standardisierte, niedergeschriebene Instruktion untermauert werden kann.

6.3.2. Auswerteobjektivität

Die Auswerteobjektivität ist sicherlich der empfindlichste Aspekt der Objektivität des Wortassoziationsversuchs. Hier sind bei Verfahren mit freien Antworten grundsätzlich Einschränkungen zu machen (siehe z.B. Lienert und Raatz, 1998). Im Falle des Wortassoziationsversuchs hängt die Auswerteobjektivität eng mit den Durchführungsbedingungen zusammen: je mehr Freiheiten in Bezug auf das Antwortformat ein Durchführungsmodus dem Probanden läßt, desto mehr Probleme ergeben sich daraus u.U. für die Auswerteobjektivität. Die Instruktion für die Auswertung des Wortassoziationsversuchs verlangt vom Versuchsleiter ein Auszählen der (sinnvollen) Nennungen auf einen Stimulus hin, nicht berücksichtigt werden sollten Äußerungen, welche nicht Wissen darstellen (z.B. „da fällt mir nichts ein“) bzw. Wiederholungen der identischen Assoziation auf einen Stimulus. Eine exakte a priori Definition derjenigen Antworten, die als richtig gelten, so wie sie z.B. Bühner (2006) für Verfahren mit offenen Antworten fordert, ist unrealistisch (siehe dazu auch die Ausführungen im folgenden Kapitel 6.3.3). Statt dessen sollte a priori ein hypothetisches Wissensmodell konstruiert werden, um den möglichen Wissensraum zu beschreiben, dies kann auch für den Versuchsleiter ein gutes Exzerpt seiner Analysen der Wissensdomäne darstellen. Damit dürfte ein großer Teil möglicher Nennungen antizipiert sein. Da aber individuelle Repräsentationen eines Wissensgebietes durchaus über dieses Modell hinausgehen oder eine andere Richtung einschlagen können, scheint es sinnvoll, a posteriori festzulegen, welche Nennungen (sofern sie über das hypothetische Wissensmodell hinaus gehen) als sinnvoll vs. unsinnig im Kontext dieser Wissensdomäne anzusehen sind.

Dieses Auszählen von Assoziationen ist unproblematisch und wird verschiedene Versuchsleiter zu einem identischen Ergebnis führen, wenn der Proband – wie in der Instruktion gefordert – nur in Stichworten oder kurzen Phrasen antwortet. Bei einem mündlichen Wortassoziationsversuch fällt genau dies manchen Probanden schwer, vielmehr bieten sie im Sinne einer natürlichen Gesprächssituation ganze Sätze an. In diesem Falle ist im Zuge der Auswertung eine Zerlegung in kleinere Sinneinheiten unumgänglich, i.d.R. bieten sich dazu als formales Kriterium die Grenzen von Teilsätzen oder längere Pausen im Gedanken- bzw. Redefluß an. An dieser Stelle erlebt die Auswerteobjektivität erste Erschütterungen. Als Maxime sollte daher gelten, so wenig wie möglich von einer solchen Zerlegung Gebrauch zu machen.

Bei der schriftlichen Darbietungsform des Wortassoziationsversuchs ist diese Problematik weniger relevant. Probanden ist es i.d.R. nicht fremd, schriftlich Stichworte aufzulisten. Dies kann in der Papier-Bleistift-Form auch optisch unterstützt werden durch vorgegebene Anstriche auf dem Assoziationsbogen und / oder die Andeutung eines Spaltenformates, so daß kurze Antworten suggeriert werden. Noch günstiger als die handschriftliche Durchführung wirkt sich die computergestützte Durchführung des Assoziierens auf die Auswerteobjektivität aus. Die technische Realisation führt zu eindeutigen Grenzen zwischen Assoziationen, die über die Programmierung gesteuert werden können. Hier ist v.a. die Umsetzung im Assoziations-Struktur-Test von Meyer (2007) zu nennen: die Eingabe einer identischen Nennung auf einen Stimulus hin wird vom System abgelehnt; die für den Probanden permanente listenartige Ansicht der bereits von ihm assoziierten Begriffe suggeriert das Antworten in Stichworten oder kurzen Phrasen. Das Setzen der Enter-Taste am Computer nach jeder Assoziation definiert eindeutig das Ende einer Nennung, welches auch in der Auswerteansicht des Assoziations-Struktur-Tests eindeutig definiert und klar erkennbar ist. Das Auszählen der Assoziationen ist in Anbetracht dieser technischen Regelmäßigkeit als objektiv zu betrachten. Lediglich das Ausstreichen ggf. sinnloser Nennungen kann nicht durch das System erfolgen, sondern nur durch den Auswerter. Eine klare Festlegung, welcher Art solche Nennungen sein könnten, ermöglicht aber auch hier eine hohe Auswerteobjektivität.

Zusammenfassend ist die Auswerteobjektivität des Wortassoziationsversuchs als kritisch zu betrachten. Je mehr jedoch die Durchführung des Assoziierens technisch unterstützt wird und so auch die Probanden zu einer eindeutigen Abgrenzung ihrer Assoziationen zwingt, desto höher kann auch die Auswerteobjektivität ausfallen.

6.3.3. Interpretationsobjektivität

Im Gegensatz zu vielen anderen Wissenstests (z.B. Multiple-Choice-Test, Lückentext) gibt es beim Wortassoziationsversuch keine richtigen oder falschen Antworten. Wenn die Probanden instruktionsgemäß nur in Stichworten antworten, ist eine Entscheidung des Versuchsleiters darüber, ob diese Assoziation nun richtig oder falsch ist, schlicht nicht möglich. Dies ist darin begründet, daß – um im Bild des semantischen Netzwerkes zu bleiben – nur das Abrufen der Knoten gefordert ist, die Benennung der Kanten hingegen, also die Beziehung zwischen zwei Knoten, wird nicht gefordert. Eine Beurteilung in richtiges vs. falsches Wissen aber setzt die Kenntnis eben dieser Relationen voraus. So werden zwar meist positive Beziehungen zwischen Assoziationen erwartet, der Proband kann aber beispielsweise das Antonym eines Stimulus geäußert haben, welches vom Auswerter fälschlicherweise als Synonym und damit als falscher Wissenszusammenhang

interpretiert werden könnte. Ähnliche Effekte sind außerdem denkbar z.B. hinsichtlich Ober-, Unter- und Nebenbegriffskategorien. Die Aufdeckung der Relationen zwischen Begriffen und damit deren Interpretation und Bewertung als richtig oder falsch ist nur durch differenziertere Verfahren wie die Strukturlegetechnik möglich, welche genau diese Zusammenhänge explizieren. Eine Interpretation der Korrektheit des individuellen Wissensbesitzes auf Grundlage des Wortassoziationsversuchs ist also völlig ausgeschlossen.

Die interpretatorische Aussage, die man letztlich anhand des Wortassoziationsversuchs treffen kann, ist: wie war jemand ad hoc in der Lage, viele oder wenige Wissens Elemente auf einen Stimulus hin zu aktivieren. Eine hohe Assoziationsmenge legt nahe, daß viele Knoten mit dem Stimulus und untereinander verbunden sein müssen und ihre Verbindungen untereinander schnell abrufbar und verbalisierbar sind. Ob und wie sinnvoll der Proband tatsächlich in der Lage ist, mit diesem Wissen umzugehen, bleibt zunächst Spekulation und ist auch nicht Aufgabe des Wortassoziationsversuchs (s.o.). Gemäß der theoretischen Annahmen des Modells eines semantischen Netzwerkes haben Kanten dann eine erhöhte Wahrscheinlichkeit, abgerufen zu werden, wenn sie oft genutzt und daher gut trainiert sind, über eine besondere Bedeutungshaltigkeit verfügen oder die Knoten sehr nah am Stimulus liegen. Einer hohen Assoziationsmenge wird also unterstellt, daß das dahinterstehende semantische Netzwerk sehr viele Verknüpfungen aufweist, was sich als umfangreicher oder elaborierter Wissensbesitz interpretieren läßt. Darüber aber, was eine hohe oder aber eine geringe Assoziationsmenge ist, läßt sich keine generelle Aussage treffen, da dies von den hinter den Stimuli stehenden spezifischen Inhalten maßgeblich beeinflusst wird.

An dieser Stelle sei an Mazur (2006) erinnert, der davor warnte, von einem Nicht-Reden auf ein Nicht-Wissen zu schließen. Ebenso sei auf die Beobachtungen im Rahmen der Validitätsprüfungen verwiesen (siehe Kapitel 6.1.5, Seite 55), wonach die guten Assoziierer i.d.R. auch einen hohen Wert in einem Außenkriterium oder alternativen Wissensdiagnoseverfahren aufweisen, während Probanden mit geringer Assoziationsmenge nicht notwendigerweise auch eine geringe Ausprägung im Kriterium haben müssen. In bestimmten Einsatzgebieten der Arbeitspsychologie allerdings läßt sich das Reden als Teil der Arbeitsaufgabe verstehen, wenn wie im Dienstleistungsbereich (z.B. Call Center mit Beratungsleistungen oder Verkaufsgesprächen) permanent Fachwissen verbalisiert werden muß. In diesem Falle wäre beim Nicht-Reden im Wortassoziationsversuch der Aspekt des Motivationsmangels in Erwägung zu ziehen. Dafür mag ein solch offenes Antwortformat wie das des Wortassoziationsversuchs zwar eher anfällig sein als andere Formate, aber die Gefahr von Motivationsdefiziten bei den Probanden besteht generell bei jeder Form arbeitspsychologischer Untersuchungen.

So ergeben sich für die Interpretation von Assoziationsdaten folgende zwei Ansätze: die Interpretation a) in Relation zur jeweiligen Untersuchungsstichprobe (X aktiviert mehr Wissen als Y; oder: X befindet sich in der oberen Extremgruppe bzgl. des Assoziationskennwertes der Untersuchungsgruppe) sowie b) in Relation zum hypothetischen Wissensmodell (auf Grundlage von betrieblichen Unterlagen, Experteninterviews usw. wird ein Sollwissen definiert: X aktiviert alles für die Erfüllung der Arbeitsaufgabe notwendige Basiswissen). Letzterer Ansatz läßt sich v.a. auf qualitativer Ebene gut bewerkstelligen: hat der Proband die zentralen Begriffe des hypothetischen Wissensmodells assoziiert oder eher randständige oder gänzlich andere Wissenskategorien aktiviert? Beide Strategien der Interpretation sind numerisch gut faßbar und somit statistisch

errechenbar, so daß hier kaum Spielraum für abweichende Interpretationen bleibt. Etwas mehr Freiheiten läßt der qualitative Auswertungsansatz, bei dem die Inhalte der Assoziationen berücksichtigt werden (X hat spezifischeres Wissen als Y), dem kann aber durch Expertenratings und / oder der Orientierung am Wissensmodell abgeholfen werden (X nennt die zentralen Elemente der Wissensdomäne), so daß auch hier die Interpretation objektiviert wird.

Zusammenfassend läßt sich also sagen: unter der Einschränkung, daß die Interpretation des Wortassoziationsversuchs sich nicht an einem absoluten Maßstab orientieren kann, sondern immer nur in Relation zum jeweiligen Untersuchungsfeld, kann die Interpretationsobjektivität als gegeben angenommen werden.

6.3.4. Zusammenfassung der Befunde zur Objektivität

Die Objektivität des Wortassoziationsversuchs kann als gegeben hingenommen werden (Fragestellung 5). Die Durchführungsobjektivität kann abgesichert werden durch das Beschränken der Interaktion zwischen Proband und Versuchsleiter auf eine festgelegte Instruktion. Die schriftliche Durchführung (insbesondere computerunterstützt) wirkt sich vorteilhaft auf die Auswerteobjektivität aus, da das Auszählen der Assoziationen dadurch erleichtert wird. Einschränkungen der Auswerteobjektivität können sich ergeben, wenn die Probanden nicht wie in der Instruktion gefordert in Stichworten oder kurzen Phrasen assoziiert haben, sondern in längeren Texten. Dieser Problematik sowie der Frage der inhaltlichen Angemessenheit der Nennungen kann entgegnet werden, indem a priori ein hypothetisches Wissensmodell entwickelt wird, welches seinerseits die Auswerteobjektivität erhöht. Die Interpretationsobjektivität des Wortassoziationsversuchs kann ebenfalls als gegeben hingenommen werden, auch hier unterstützt ein hypothetisches Wissensmodell die Interpretation der Qualität oder Inhalte der Assoziationen.

6.4. Zusammenfassung der Befunde zu den Hauptgütekriterien

Die Hauptgütekriterien Validität, Reliabilität und Objektivität können hinsichtlich des Wortassoziationsversuchs als gegeben hingenommen werden. Die Reliabilität des Verfahrens konnte im Rahmen eines Retest-Designs nachgewiesen werden, der sehr signifikante Korrelationskoeffizient von $r = .849$ ($p = .000$) illustriert den Zusammenhang zwischen den Assoziationskennwerten beider Erhebungszeitpunkte.

Die Inhaltsvalidität des Wortassoziationsversuchs kann aufgrund der intensiven Auseinandersetzung des Versuchsleiters im Vorfeld der Erhebungen als gegeben hingenommen werden. Zur Bestimmung der Kriteriumsvalidität wurden zwei Arten von Außenkriterien herangezogen: zum einen die Leistung in einer Klausur als Maß für das Wissen, zum anderen die Beurteilung der Arbeitspersonen durch den Geschäftsführer als Ausdruck für das in Arbeitshandeln umgesetzte Fachwissen. Dabei kann die Leistung der Klausuren insgesamt recht gut anhand des Assoziationskennwertes vorhergesagt werden, belegt durch geringe bis mittlere signifikante Korrelationskoeffizienten in neun Studien. Deutlich schwächer fällt der Koeffizient hinsichtlich der Geschäftsführerbeurteilung aus, hier zeigt sich über die zwei Feldstudien hinweg eine schwache signifikante Korrelation, wobei sich das Gesamturteil besser vorhersagen läßt als die beurteilte Fachkompetenz alleine.

Die konvergente Konstruktvalidität wurde in vielen Studien mit unterschiedlichen Alternativverfahren empirisch belegt. In den meisten Studien zeigte sich ein signifikanter Zusammenhang zwischen jeweiligem Leistungskriterium und Assoziationskennwert, die Höhe der Korrelationen schwankten – je nach Kriterium – zwischen geringen und sehr hohen Koeffizienten. Im Sinne der konvergenten Konstruktvalidierung also kann sich der Wortassoziationsversuch gut gegenüber den Parametern anderer Wissensdiagnoseverfahren behaupten: der Assoziationskennwert erbringt überwiegend mindestens genauso gute Prognosen eines weiteren Außenkriteriums wie die Parameter dieser anderen Verfahren. Aufgrund der empirischen Befunde in drei Studien kann davon ausgegangen werden, daß das Personenmerkmal der generellen Wortflüssigkeit keinen Einfluß auf die Performanz im Wortassoziationsversuch hat, dieser also tatsächlich das Konstrukt Wissen erfaßt. Somit kann auch eine diskriminante Konstruktvalidität bestätigt werden.

Im Zusammenhang mit der Bestimmung der Validität des Wortassoziationsversuchs wurde auch geprüft, inwieweit der dabei zugrundegelegte individuelle Assoziationskennwert als mittlere Anzahl von Assoziationen pro Stimulus tatsächlich ein geeigneter Parameter zur Quantifizierung des individuellen Wissensbestandes ist, oder ob Modifikationen des Kennwertes angemessener sind. So wurde zum einen die Gewichtung qualitativ unterschiedlicher Assoziationsinhalte thematisiert, zum anderen die Berücksichtigung der innerhalb der Stichprobe hochfrequenten Assoziationen. In beiden Fragen ist festzuhalten, daß die Modifikationen lediglich bereits bestehende Unterschiede zwischen Probanden akzentuierten, aber zu keiner grundsätzlich anderen Beurteilung ihres Wissensbesitzes führten. Zudem konnten die in dieser Weise modifizierten Kennwerte keine höhere Vorhersageleistung für ein Außenkriterium erbringen als bereits der einfache Assoziationskennwert in der Lage war. Nichts desto trotz erbringen solche stärker qualitativ ausgerichteten Auswertestrategien einen interessanten Mehrgewinn hinsichtlich der Beschreibung des Wissensbesitzes innerhalb der Stichprobe.

Die Objektivität des Wortassoziationsversuchs kann ebenfalls als gegeben hingenommen werden. Die Durchführungsobjektivität läßt sich durch bestimmte Durchführungsbedingungen hochhalten (z.B. Instruktionen, Gruppenversuch). Die Interpretation der Assoziationen bietet wenig Spielraum, da ein Richtig oder Falsch der individuellen Assoziationen nicht durch den Versuchsleiter beurteilt werden kann. Somit bleibt nur der Abgleich individueller Assoziationen mit einem hypothetischen Wissensmodell sowie mit anderen Probanden der Stichprobe, was wiederum numerisch gut faßbar und dadurch als objektiv anzusehen ist. Lediglich bei der Auswerteobjektivität sind aufgrund des freien Antwortformats Abstriche zu machen, wobei sie gestärkt werden kann durch beispielsweise den Einsatz computerbasierter Assoziationstools.

Zusammenfassend können die Hauptgütekriterien Validität, Reliabilität und Objektivität als gegeben hingenommen werden. Der Wortassoziationsversuch erweist sich als ein gültiges, zuverlässiges und m.E. objektives Instrument zur Diagnose von Fachwissen.

7. Nebengütekriterien

In diesem Kapitel sollen die wichtigsten Nebengütekriterien erörtert werden. Die Kriterien der Veränderungssensitivität (Fragestellung 7), Akzeptanz (Fragestellung 8) sowie der Ökonomie (Fragestellung 9) können sich auf empirische Befunde stützen, Abbildung 133 gibt einen Überblick über diejenigen Studien, die dazu einen Beitrag leisten. Zunächst wird auf das Nebengütekriterium der Veränderungssensitivität des Wortassoziationsversuchs eingegangen werden, dazu konnten drei Feldstudien herangezogen werden, in denen mehrere Meßzeitpunkte die Veränderung des individuellen Wissensbesitzes in Folge von Lehrveranstaltungen abbilden. In zwei weiteren Studien wurden Daten zur Akzeptanz des Wortassoziationsversuchs durch die Probanden sowie in drei Studien Befunde zur Durchführungsökonomie erhoben. Darüber hinaus können zu diesen beiden Nebengütekriterien auch Erfahrungen anderer Versuchsleiter berichtet werden. Für die Kriterien der Nützlichkeit (Fragestellung 10), Normierung (Fragestellung 11) sowie Vergleichbarkeit (Fragestellung 12) liegen keine empirischen Befunde vor, vielmehr werden diese Kriterien ausdiskutiert.

Nr.	Name der Studie	Veränderungssensitivität	Akzeptanz	Durchführungsökonomie
1	Modus		x	x
2	psychotherapeutische Ambulanz		x	x
3	Sparkassenakademie (A)	x		
4	Sparkassenakademie (B)	x		
13	Stoffdruckerei			x
14	Seminar Wissensmanagement	x		

Abbildung 133: Übersicht über alle Studien, die einen empirischen Beitrag zu den Nebengütekriterien leisten

7.1. Veränderungssensitivität

In Kapitel 6.2 (Seite 55 ff.) konnte die Reliabilität des Wortassoziationsversuchs belegt werden, so daß damit die Grundvoraussetzung gegeben ist, um die Veränderungssensitivität diskutieren zu können. Die hier vorgestellten drei Studien zur Veränderungssensitivität des Wortassoziationsversuchs realisierten im jeweiligen Feld zwei, drei bzw. vier Meßzeitpunkte. Ziel dabei war immer, die intraindividuellen Veränderungen im Wissensbesitz abzubilden, welche aufgrund von Weiterbildungs- oder universitären Lehrveranstaltungen zu erwarten waren. Allen Studien gemein ist die erste Erhebung als Prä-Messung, welche vor Beginn jeglicher Lehre stattfand. Studie 14 (Seminar Wissensmanagement) setzt die Post-Messung an das Ende der Lehrveranstaltung, die beiden Studien zur Sparkassenakademie (Studien 3 und 4) erheben sogar noch an weiteren, für das pädagogische Konzept der Weiterbildungsveranstaltung bedeutsamen Zeitmarken weitere Assoziationsversuche.

In den hier vorgestellten Studien soll die Entwicklung der Assoziationsleistungen über die Erhebungszeitpunkte hinweg betrachtet werden anhand des quantitativen Kennwertes, bei dem die Anzahl der Assoziationen herangezogen wird. Zudem soll ein weiterer Zugang erfolgen, der stärker qualitativ geprägt ist und erstmals von Wegwarth (2002) praktiziert wurde: er beschreibt das Hinzukommen von „neuen“ Assoziationsinhalten eines Probanden im Laufe der Erhebungen. Danach sind sämtliche Äußerungen zum ersten Meßzeitpunkt als „Vorwissen“ zu bewerten (selbst wenn sie nach dem Treatment wiederholt werden), da sich diese bereits vor Beginn der Lehrveranstaltung im individuellen Wissensbesitz befanden. Assoziationen, die nach der Lehrveranstaltung erstmalig geäußert wurden, sind hochwahrscheinlich auf den Einfluß der Lehre zurückzuführen und wurden daher als „fremdgesteuert“ klassifiziert. Da in den beiden Studien zur Sparkassenakademie auch eine Selbstlernphase Bestandteil der Weiterbildung war, konnten Assoziationen, die erstmalig nach dieser Selbstlernphase genannt wurden, als „selbstgesteuert“ klassifiziert werden. Auf diese Weise ergibt sich für jenen Probanden eine individuelle Bilanz, im Ergebnis werden die mittleren Anteile dieser drei Kategorien über den Zeitverlauf dargestellt.

7.1.1. Veränderungssensitivität in Studie 3: Sparkassenakademie (A)

Wie bereits in Kapitel 4.3 (Seite 55 f.) beschrieben, erhob Wegwarth (2002) in der Studie 3 – Sparkassenakademie (A) – das Fachwissen von Weiterbildungsteilnehmern zu vier Zeitpunkten: zu Beginn der Weiterbildung, am Ende der Präsenzlernphase, am Ende der Selbstlernphase zeitgleich zur Prüfung sowie vier Monate danach. Die Entwicklung des Assoziationskennwertes über die Meßzeitpunkte ist in Abbildung 134 graphisch veranschaulicht.

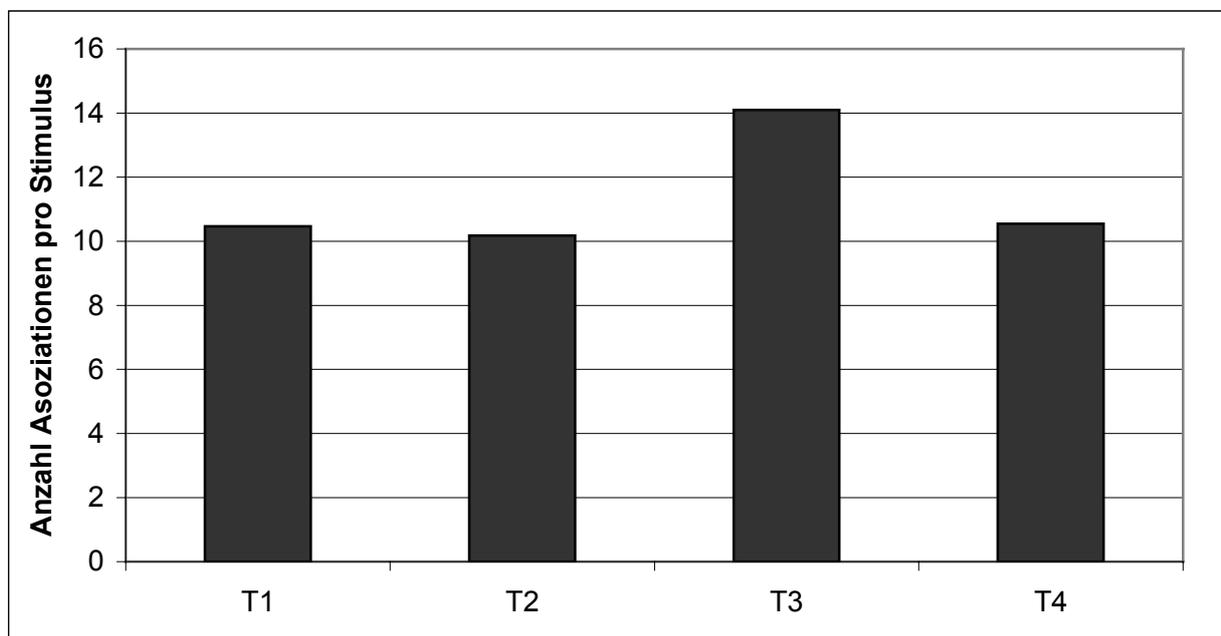


Abbildung 134: Entwicklung des Assoziationskennwertes über die Erhebungszeitpunkte in Studie 3 – Sparkassenakademie (A)

Die Assoziationsleistung zum dritten Meßzeitpunkt, also am Tag der Prüfung, ist mit $M=14,10$ ($SD=4,05$) signifikant höher als zu Beginn der Weiterbildung ($M=10,46$; $SD=3,73$; Wilcoxon: $z=-4,151$, $p=.000$; $d=.94$), zum Ende der Präsenzphase ($M=10,18$; $SD=3,61$; Wilcoxon: $z=-4,338$; $p=.000$; $d=1.00$) sowie vier Monate nach der Prüfung ($M=10,54$; $SD=4,18$; Wilcoxon: $z=-4,259$;

$p = .000$; $d = .87$). Die Weiterbildungsmaßnahme hat also zu einer signifikanten Zunahme des assoziierten Wissens geführt, was anhand des Kennwertes im Wortassoziationsversuch belegt werden konnte.

Abbildung 135 illustriert, wie die Assoziationsleistungen sich auf die verschiedenen Stimuli verteilen. Dabei wird sichtbar, daß sich jeweils der dritte Erhebungszeitpunkt (Prüfung) von den anderen abhebt. Für die Stimuli „Führen von Gruppen“, „Personalentwicklung“ sowie „Motivationstheorien“ konnte Wegwarth (2002) einen signifikanten Anstieg der Assoziationsleistung vom ersten hin zum dritten Meßzeitpunkt nachweisen, alle anderen Unterschiede waren nicht statistisch bedeutsam (siehe Wegwarth, 2002).

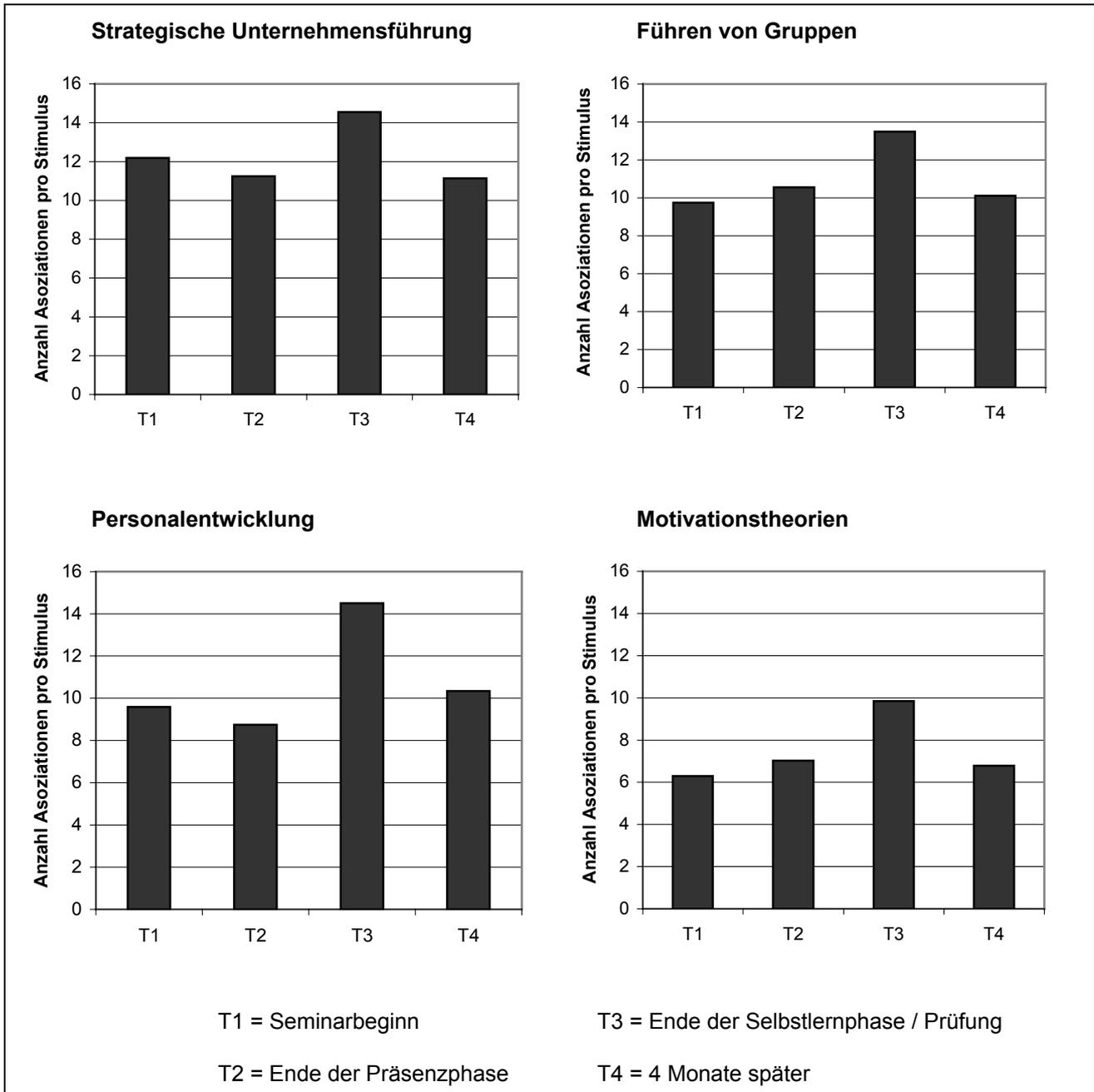


Abbildung 135: Entwicklung des Assoziationskennwertes zu den vier Stimuli über die Erhebungszeitpunkte in Studie 3 – Sparkassenakademie (A); (nach Wegwarth, 2002)

Um den Wissenstransfer im Laufe einer Weiterbildungsveranstaltung abzubilden, differenzierte Wegwarth (2002) separat für jeden Stimulus den Zuwachs von Wissensinhalten, der im Laufe der vier Meßzeitpunkte sichtbar wurde, nach ihrem wahrscheinlichen Ursprung (Vorwissen, Fremdsteuerung, Selbststeuerung), so daß sich für jeden Probanden eine individuelle Bilanz ergab, welche Assoziationsinhalte in welchen Lernphasen erworben wurden. Allerdings wendete Wegwarth (2002) diesen Analysezugang aus Gründen der Datenreduzierung nur auf die sog. hochfrequenten Assoziationen an, d.h. Assoziationen, die von mindestens acht der Probanden genannt wurden (siehe Kapitel 6.1.4.2.1, Seite 55 ff.). An dieser Stelle jedoch soll weiterhin der Kennwert über sämtliche Assoziationen eines Probanden betrachtet werden. Da die Rohdaten dazu nicht mehr verfügbar waren, kann hier nur auf die Arbeit von Wegwarth (2002) verwiesen werden.

Insgesamt beschreibt Wegwarth (2002) den Wortassoziationsversuch als ein geeignetes Verfahren, um die Effizienz einer solchen Weiterbildungsveranstaltung bewerten zu können. Anhand des quantitativen Assoziationskennwertes war der Wortassoziationsversuch in der Lage, Veränderungen im Umfang des Wissensbesitzes abzubilden. Offenbar gelingt diese Abgrenzung zwischen den Erhebungszeitpunkten weitaus besser anhand eines über alle Stimuli aggregierten Kennwertes anstatt bei Betrachtung der Assoziationsleistungen zu einzelnen Stimuli.

7.1.2. Veränderungssensitivität in Studie 4: Sparkassenakademie (B)

Studie 4 – Sparkassenakademie (B) – begleitete ebenfalls mit Hilfe des Wortassoziationsversuchs eine Weiterbildungsveranstaltung über mehrere Meßzeitpunkte. Jordan (2004) konnte in seiner Arbeit für die einzelnen Stimuli eine Zunahme der Assoziationsleistung über die drei Meßzeitpunkte hinweg nachweisen. Anhand des Assoziationskennwertes, welcher sich aus der Summe aller Stimuli ergibt, läßt sich somit ein Wissenszuwachs nachweisen, der v.a. den Unterschied vor vs. nach der Präsenzlernphase markiert (siehe auch Abbildung 136).

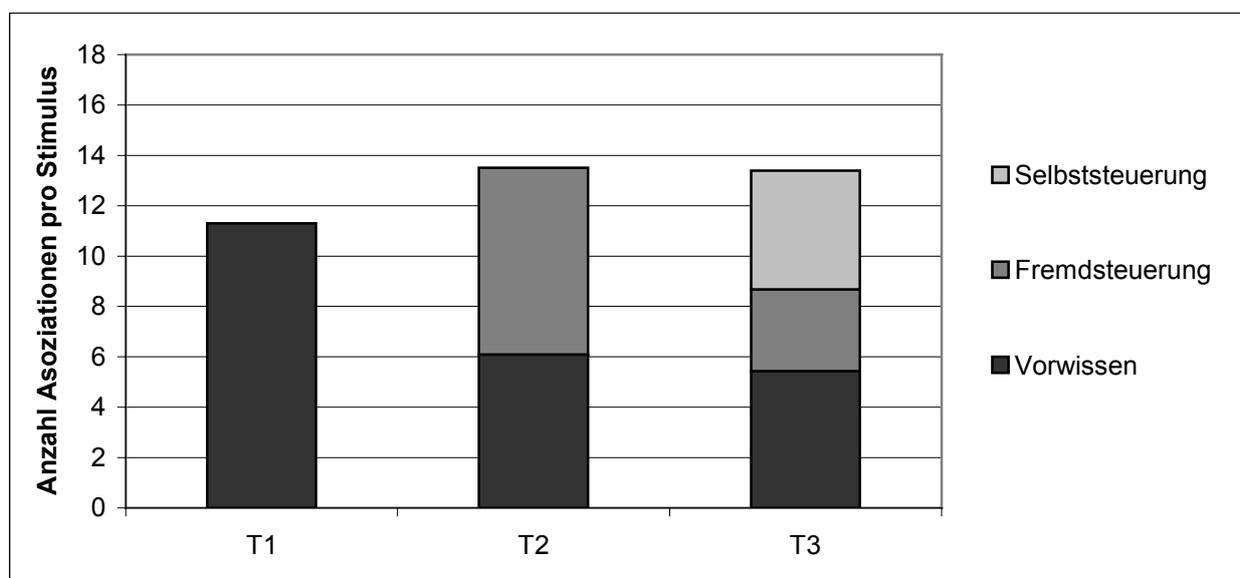


Abbildung 136: Entwicklung des Assoziationskennwertes sowie Umstrukturierungsprozesse der Assoziationen über alle Stimuli in Studie 4 – Sparkassenakademie (B)

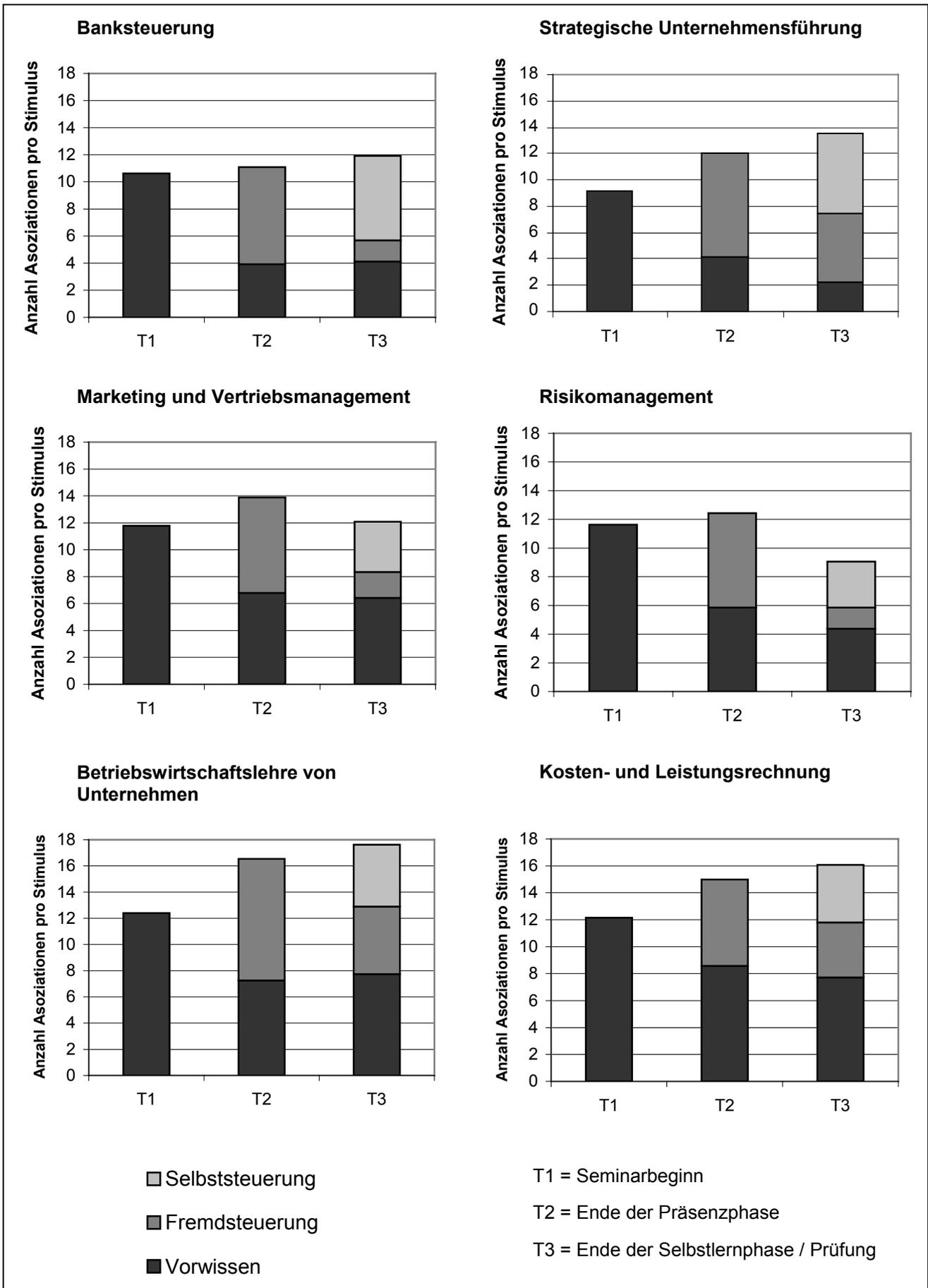


Abbildung 137: Entwicklung des Assoziationskennwertes sowie Umstrukturierungsprozesse der Assoziationen zu den sechs Stimuli in Studie 4 – Sparkassenakademie (B)

Während zu Beginn der Fortbildung durchschnittlich 11,29 Assoziationen ($SD=3,14$) produziert wurden, waren es zum zweiten Meßzeitpunkt mit $M=13,59$ ($SD=3,03$) signifikant mehr ($t(34)=-4,35$; $p=.000$; $d=.75$) und ebenso zum dritten Erhebungszeitpunkt mit $M=13,39$ ($SD=3,94$) ($t(34)=-2,76$; $p=.009$; $d=.59$). Der zweite und dritte Meßzeitpunkt unterschieden sich in der Menge der Assoziationen nicht signifikant voneinander, die Effektstärke von $d=.06$ gibt auch keinen Hinweis auf einen Mittelwertsunterschied. Der quantitative Wissenszuwachs geschieht also in der Phase des Präsenzunterrichts, die Phase des Selbststudiums hingegen erbringt keine Veränderungen in der Wissensmenge.

Um diesen Befund näher zu beleuchten, klassifizierte Jordan (2004) die Assoziationen der einzelnen Probanden in gleicher Weise wie Wegwarth (2002) nach ihrer vermuteten Herkunft: als Vorwissen (erstmalige Nennung zu T1), Fremdsteuerung (erstmalige Nennung zu T2) oder als Selbststeuerung (erstmalige Nennung zu T3). Zusammengefaßt über alle Stimuli ergibt sich das in Abbildung 136 dargestellte Bild. Der Anteil der als Vorwissen klassifizierten Assoziationen verringert sich dabei signifikant vom ersten Meßzeitpunkt ($M=11,29$; $SD=3,14$) zum zweiten hin ($M=6,10$; $SD=1,60$; $t(34)=10,527$; $p=.000$; $d=2.08$) und ebenso im Vergleich zum dritten Meßzeitpunkt ($M=5,43$; $SD=1,64$; $t(34)=10,802$; $p=.000$; $d=2.34$). Diese Assoziationen der Kategorie Vorwissen reduzieren sich zwischen Weiterbildungsbeginn und Ende der Präsenzphase dramatisch (der Kennwert halbierte sich fast) zugunsten der als fremdgesteuert klassifizierten Assoziationen ($M=7,41$; $SD=2,12$). Aber auch diese bilden keinen stabilen Wissensbesitz: nach Abschluß der Selbstlernphase zum dritten Meßzeitpunkt treten auch Assoziationen der Kategorie Fremdsteuerung in den Hintergrund und werden mit $M=3,24$ ($SD=1,23$) signifikant seltener genannt als zum zweiten Meßzeitpunkt ($t(34)=11,937$; $p=.000$; $d=2.41$). Zwischen dem zweiten und dritten Meßzeitpunkt bleibt der Anteil der Vorwissens-Assoziationen in etwa gleich. Am Ende der Selbstlernphase werden statt dessen zu großem Anteil neue, in eben dieser Lernphase erworbene Assoziationen geäußert ($M=4,71$; $SD=2,02$).

Zur Illustration sind diese drei Kategorien in Abbildung 137 im Zeitverlauf über die Meßzeitpunkte separat für jeden Stimulus dargestellt. Es wird deutlich, daß bei Betrachtung der einzelnen Stimuli durchaus unterschiedliche Akzente in der Wirkung einzelner Lernphasen vorhanden sind. Dennoch bildet sich bei allen Stimuli der gleiche Effekt ab: die Bedeutung der als Vorwissen kategorisierten Inhalte nimmt im Laufe der Weiterbildungsveranstaltung ab, ebenso die Relevanz derjenigen Inhalte, die in der Präsenzlernphase erworben wurden.

Offenbar scheint gerade die Kategorie der selbstgesteuerten Assoziationen besondere Vorhersagekraft für das Außenkriterium Klausurleistung zu besitzen: Jordan (2004) modifizierte den Assoziationskennwert dahingehend, daß er lediglich die Assoziationen der sog. Kategorie Selbststeuerung heranzog, also diejenigen Nennungen eines Probanden, welche dieser erstmals zum Zeitpunkt der Klausur äußerte. Außerdem ließ der von Jordan (2004) gebildete Kennwert die Assoziationen zum Stimulus „Risikomanagement“ unberücksichtigt, da diese Thematik nicht Gegenstand der Klausur war. Jordan (2004) berichtet über einen mittleren Korrelationskoeffizienten von $r=.570$ ($p=.000$) zwischen dem so gewonnenen Assoziationskennwert und der Klausurleistung. Zum Vergleich sei an den in Kapitel 6.1.2.1.2 (Seite 55 ff.) berichteten geringen Korrelationskoeffizienten von $r=.338$ ($p=.023$) erinnert, bei dem die Klausurleistung durch einen Assoziationskennwert vorhergesagt wurde, in den sämtliche Assoziationen von allen sechs Stimuli

eingingen. Ein hohes Engagement in der Selbstlernphase, welches sich durch eine große Anzahl an bisher noch nicht geäußerten Assoziationsinhalten (im Vergleich zu den beiden vorangegangenen Meßzeitpunkten) abbilden läßt, geht also auch mit einer hohen Leistung in der Klausur einher.

Zusammenfassend läßt sich festhalten, daß der Wortassoziationsversuch also in dieser Studie in der Lage ist, quantitative wie auch qualitative Veränderungen im Wissensbesitz der Probanden über einen Treatmentzeitraum hinweg abzubilden.

7.1.3. Veränderungssensitivität in Studie 14: Seminar Wissensmanagement

Studie 14 – Seminar Wissensmanagement – wurde ein Prä-Post-Design zur Evaluation des Wissenszuwachses durch die Lehrveranstaltung realisiert. Wie bereits berichtet (siehe Kapitel 4.11, Seite 55 ff.) konnte über die beiden Erhebungszeitpunkte ein quantitativer Anstieg der Assoziationsleistung von $M=10,00$ ($SD=4,22$) Assoziationen vor Seminarbeginn hin zu durchschnittlich $M=16,35$ ($SD=7,50$) Assoziationen signifikant nachgewiesen werden ($t(19)=-4,028$; $p=.001$; $d=1.04$). Auf Grundlage des Assoziationskennwertes zum ersten Meßzeitpunkt läßt sich die Assoziationsleistung eines Probanden zum Ende des Seminars ungefähr vorhersagen: mit einer Korrelation (nach Pearson) von $r=.384$ ($p=.047$) liegt ein geringer signifikanter korrelativer Zusammenhang vor. Eine Rangkorrelation nach Spearman erbringt mit $r=.473$ ($p=.017$) einen mittleren Korrelationskoeffizienten, die Rangreihe der Probanden bzgl. ihrer Assoziationsleistung läßt sich also recht gut vorhersagen. Probanden, die bereits vor dem Seminar in der Lage waren, viel zu assoziieren, konnten auch bei Seminarende viele Assoziationen produzieren.

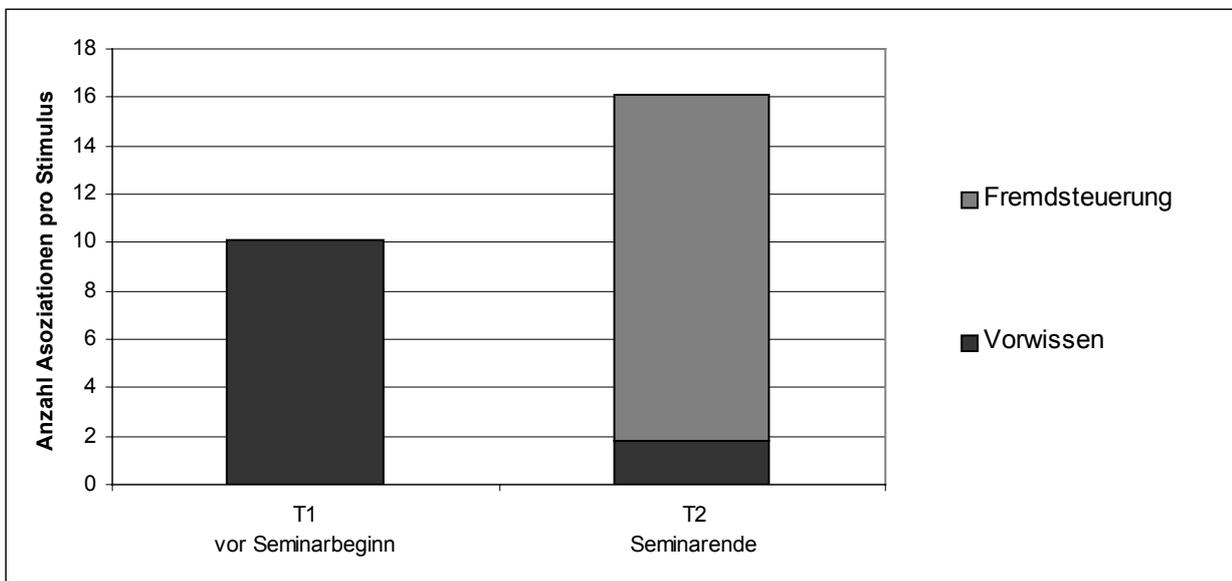


Abbildung 138: Entwicklung des Assoziationskennwertes sowie Umstrukturierungsprozesse der Assoziationen in Studie 14 – Seminar Wissensmanagement

In Anlehnung an das bisherige Vorgehen in den Studien 3 und 4 wurden auch hier die assoziierten Inhalte des zweiten Meßzeitpunktes dahingehend kategorisiert, ob ein Proband diese bereits bei der ersten Erhebung benannte (dies wurde als Vorwissen aufgefaßt) oder aber eine Assoziation erstmals im zweiten Meßzeitpunkt geäußert wurde (dies wurde auf den Einfluß der Lehrveranstaltung zurückgeführt, sog. Fremdsteuerung). Abbildung 138 illustriert das Ergebnis dieser Auswertung: danach sinkt die Bedeutung der als Vorwissen klassifizierten Assoziationen rapide ab, bei

Seminarende werden durchschnittlich nur noch 1,79 ($SD=2,04$) dieser Inhalte assoziiert, dies ist signifikant geringer als die Menge der Assoziationen zur ersten Erhebung ($t(19)=7,651$; $p=.000$; $d=2,48$). Der Wortassoziationsversuch zeigt sich also in der Lage, den Wissenszuwachs während der Lehrveranstaltung auf quantitativer wie auch auf qualitativer Ebene abzubilden.

7.1.4. Zusammenfassung der Befunde zur Veränderungssensitivität

Die drei in diesem Abschnitt vorgestellten Studien demonstrieren eindrucksvoll, wie der Wortassoziationsversuch in der Lage ist, Veränderungen im Wissensbesitz infolge von Lehrveranstaltungen abzubilden (Fragestellung 7). Zum einen ist den Wortassoziationsversuch sensitiv in quantitativer Hinsicht. In allen drei Studien konnte im Prä-Post-Design ein signifikanter Anstieg des Assoziationskennwertes belegt werden, der infolge der Lehrveranstaltungen auch erwartet wurde. Die beiden Studien der Sparkassenakademien illustrieren, wie aufgrund der Weiterbildung der Wissensbesitz zunimmt und seinen maximalen Wert zum Zeitpunkt der Prüfung erreicht (wenn es also für den Probanden besonders relevant ist, viel zu wissen), und wie sich im Follow-up der Studie 3 der Wissensbesitz wieder reduziert (wenn das Wissen nicht mehr notwendigerweise aktiviert werden muß).

Um die Befunde der empirischen Studien zu integrieren, wurde für jeden Probanden der relative Zugewinn des Assoziationskennwertes (in %) von T1 hin zu T2 bestimmt. Da in allen drei Studien die erste Erhebung vor und die zweite Erhebung unmittelbar am Ende der (Präsenz-) Lernphase stattfand, waren diese Meßzeitpunkte miteinander vergleichbar, so daß insgesamt 93 Probanden berücksichtigt werden konnten. Der durchschnittliche Zuwachs zum zweiten Meßzeitpunkt betrug dabei $M=29,90$ % der Assoziationen im Vergleich zu T1 ($SD=61,51$), wobei eine große Spanne zwischen maximal 300,00 % und minimal -65,67 % (also ein Absinken der individuellen Assoziationsleistung) abgedeckt wird. Betrachtet man den Kennwert von T1 als Referenzgröße und setzt ihn mit 100 % gleich, so ist diese mittlere Zunahme von Assoziationen bei T2 statistisch signifikant ($t(92)=-4,688$; $p=.000$).

Analog dazu läßt sich die Veränderung der individuellen Assoziationskennwerte vom zweiten hin zum dritten Meßzeitpunkt beschreiben (dabei entspricht die Leistung von T2 100 %), wobei jedoch nur noch die beiden Studien zur Sparkassenweiterbildung einbezogen werden können ($N=73$), bei denen es einen solchen dritten Erhebungszeitpunkt gab (nach der Selbstlernphase bzw. zur Prüfung). Der mittlere Zuwachs zwischen diesen beiden Zeitpunkten beträgt $M=27,27$ %, wobei ebenfalls eine sehr große Streuung ($SD=48,60$) zwischen den Extremwerten 157,14 % und -77,42 % zu verzeichnen ist. Auch dieser Zuwachs an Assoziationsleistung ist statistisch signifikant ($t(72)=-4,793$; $p=.000$). Insgesamt betrachtet ist der Wortassoziationsversuch also dazu geeignet, die quantitativen Veränderungen im individuellen Wissensbesitz, welche durch die Lehrveranstaltung zu erwarten waren, auch abzubilden.

Zum anderen wurde deutlich, daß der Wortassoziationsversuch auch in der Lage ist, die qualitativen Veränderungen im Wissensbesitz darzustellen. Durch die Analyse, welche Assoziationen eines Probanden während der einzelnen Phasen der Weiterbildung neu in den individuellen Wissensbesitz aufgenommen wurden, konnte nachgewiesen werden, daß die als Vorwissen einzustufenden Assoziationen (wurden vor Beginn der Weiterbildung benannt) im Verlauf der Weiterbildung an Bedeutung verlieren. Sie werden verdrängt zugunsten von Inhalten, welche auf den Einfluß der

Lehrveranstaltung zurückzuführen sind. Somit konnte belegt werden, daß die Phase der Wissensvermittlung auch zu einer Umstrukturierung des individuellen Wissensbesitzes geführt hat. Es wurde am Ende der Weiterbildung nicht nur mehr gewußt, sondern es wurden auch andere Inhalte gewußt als zu Beginn.

7.2. Akzeptanz

In zwei Studien wurden explizit Erhebungen zur Akzeptanz des Verfahrens durch die Probanden durchgeführt. Zum einen wird dies im Vergleich zur Methode des Lückentextes betrachtet, zum anderen im Vergleich zwischen verschiedenen Durchführungsbedingungen des Wortassoziationsversuchs selbst (mündlich vs. schriftlich). Darüber hinaus sollen Erfahrungen der Versuchsleiter der anderen Studien berichtet werden.

7.2.1. Akzeptanz in Studie 1: Modus

In Studie 1 – Modus – wurde der Wortassoziationsversuch im Rahmen eines experimentellen Designs in zwei Durchführungsformen eingesetzt: mündlich vs. schriftlich als Papier-und-Bleistift-Form. In einem kurzen Fragebogen wurden die Probanden gebeten, unmittelbar nach dem mündlichen bzw. schriftlichen Teil zu bewerten, wie leicht oder schwer es ihnen fiel, zu antworten und dabei nur Stichworte zu verwenden. Zudem wurde nach der Verständlichkeit der Instruktion gefragt sowie bei der mündlichen Variante nach der erlebten Unterstützungsfunktion durch die gleichzeitige auditive und schriftliche Präsentation der Stimuli durch den Versuchsleiter. Somit waren vom Probanden insgesamt sechs Items anhand einer fünfstufigen Skala zu bewerten sowie ggf. Kommentare bzw. Begründung für die jeweilige Entscheidung dazu abzugeben (siehe auch Anhang 1). Abschließend wurde den Probanden ein siebentes Item zur Präferenz gegenüber den Verfahrensteilen vorgelegt, welches am Schluß des Arbeitsbogens positioniert war.

	Verständlichkeit der Instruktion	Phrasen (schriftlich)	schriftliches Antworten	Phrasen (mündlich)	mündliches Antworten	visuelle Unterstützung
<i>M</i>	4,69	4,14	3,97	3,53	3,14	4,00
<i>SD</i>	0,60	0,69	0,87	1,11	1,06	0,83
Minimum	3	3	2	2	2	3
Maximum	5	5	5	5	5	5

Abbildung 139: Statistische Kennwerte (Mittelwert, Standardabweichung, Minimal- und Maximalwerte) der Items des Akzeptanz-Fragebogens in Studie 1 – Modus

Die Durchführung des Wortassoziationsversuchs als mündliche bzw. schriftliche Variante stellte für die Probanden insgesamt eine bewältigbare Aufgabe dar. Abbildung 139 listet die statistischen Kennwerte der Items auf: zur Verständlichkeit der Instruktion, das Antworten in Stichworten im schriftlichen Teil sowie das schriftliche Antworten generell, analog dazu das Antworten in Stichworten bzw. das generelle Antworten im mündlichen Teil, außerdem die erlebte Unterstützung durch die visuelle Präsentieren der Stimuli. Die Kennwerte können einen Wert zwischen 1 und 5 annehmen, wobei ein hoher numerischer Wert einer positiven Aussage entspricht („trifft zu“ bzw. „fiel mir leicht“). Es wird deutlich, daß sich die Mittelwerte der Bewertungen immer im positiven

Bereich der Skala bewegen, selbst die Minima weisen selten eine negative Bewertung aus (siehe Abbildung 139).

Die Instruktion für den Wortassoziationsversuch war für alle Probanden gut verständlich ($M=4,69$; $SD= 0,60$), keiner der Versuchsteilnehmer negierte die angebotene Aussage. Es zeigen sich signifikante Unterschiede in der Bewertung der beiden Modi: generell fällt es den Versuchspersonen eher leichter, schriftlich zu antworten ($M=3,97$; $SD= 0,87$) als mündlich ($M=3,14$; $SD=1,06$), (Wilcoxon: $z= -2,96$, $p= .003$; $d= .86$). Diejenigen, denen der mündliche Teil weniger leicht fiel, begründeten dies mit dem Zeit- und Leistungsdruck, schnell antworten zu müssen, den sie in der Erhebungssituation erlebten. Diejenigen, denen der mündliche Teil eher leicht fiel, benannten als Begründung den geringen Zeitaufwand sowie eine generelle Vorliebe für solche Darbietungsform. Das leichtere Bearbeiten des schriftlichen Antwortens wurde in erster Linie durch die Möglichkeit des längeren Nachdenkens und dadurch als geringer erlebten Zeit- und Leistungsdruck erklärt.

Das Antworten in Stichworten oder kurzen Phrasen fiel den Probanden im schriftlichen Teil insgesamt recht leicht ($M=4,14$; $SD=0,69$), im mündlichen Teil war die Zustimmung etwas geringer ($M=3,53$; $SD=1,11$). Auch unter diesem Aspekt war die schriftliche Variante der mündlichen signifikant überlegen (Wilcoxon: $z=-2,35$; $p= .019$; $d= .66$). Bei beiden Modi werden positive Bewertungen von den Probanden begründet mit dem geringeren Aufwand, keine Sätze formulieren zu müssen, und dabei Inhalte auf den Punkt bringen zu können. Diejenigen Probanden, denen ein Antworten in Stichworten eher schwer fiel (v.a. in der mündlichen Form), begründeten dies wiederum mit dem erlebtem Zeit- und Leistungsdruck.

Die visuelle Unterstützung des mündlichen Teils durch die Präsentation des aktuellen Stimulus auf Kärtchen wurde von den Probanden mehrheitlich als erleichternd aufgenommen ($M=4,00$; $SD=0,83$), die Urteile verteilten sich gleichmäßig auf die Kategorien „weder erschwert noch erleichtert“, „etwas erleichtert“ und „erleichtert“. Die Probanden begründeten eine erlebte Erleichterung in erster Linie damit, daß die visuelle Darbietung eine bessere Vorstellung des Begriffes, die Konzentration darauf und außerdem auch eine Entlastung des Gedächtnisses erlaube. Probanden, die angaben, daß die visuelle Präsentation das Assoziieren weder erschwere noch erleichtere, begründeten dies ausschließlich mit der Wahl der Stimuli selbst: die dargebotenen Begriffe waren gut verständlich und dadurch gut zu merken.

Der schriftliche Modus wird signifikant häufiger bevorzugt ($\chi^2(2, N=28)=14,0$; $p= .001$). 18 Probanden bevorzugten die schriftliche Form des Wortassoziationsversuchs, dies wird im Wesentlichen mit der längeren „Denkphase“ oder Bearbeitungszeit begründet, welche das Erleben von Leistungsdruck mindert. Acht Probanden bevorzugten die mündliche Durchführungsform und begründeten dies mit dem geringeren zeitlichen und kognitiven Aufwand, aber auch mit einer angenehmeren Interaktion zwischen Proband und Versuchsleiter. Lediglich zwei Probanden gaben an, beide Formen vergleichbar zu finden, zwei weitere Probanden gaben kein Präferenzurteil ab.

Zusammenfassend bleibt festzuhalten, daß der Wortassoziationsversuch insgesamt von den Probanden akzeptiert wird. Dabei wurde die schriftliche Durchführungsform als angenehmer erlebt, was im Wesentlichen durch den – im Vergleich zur mündlichen Variante – als geringer erlebten Zeit- und Leistungsdruck determiniert wird. Die gleichzeitige visuelle und akustische Präsentation

der Stimuli in der mündlichen Form wurde überwiegend als unterstützend erlebt, zumindest gab keiner der Befragten an, daß dies das Assoziieren erschweren würde.

7.2.2. Akzeptanz in Studie 2: psychotherapeutische Ambulanz

In Studie 2 – psychotherapeutische Ambulanz – wurde den Probanden unmittelbar nach jedem Untersuchungsabschnitt ein Fragebogen zur Akzeptanz des gerade bearbeiteten Verfahrens vorgelegt (siehe Anhang 2 bzw. Anhang 3). So wurde zu beiden Erhebungszeitpunkten nach dem Wortassoziationsversuch ein Urteil hinsichtlich der Verständlichkeit der Instruktion sowie der erlebten Schwierigkeit beim Antworten in Phrasen erbeten. Nach Bearbeiten des Lückentextes zum zweiten Meßzeitpunkt wurden außerdem analoge Items (bezogen auf den Lückentext) vorgelegt. Außerdem wurde bei der zweiten Erhebung erfragt, inwieweit der Wortassoziationsversuch bzw. der Lückentext den Probanden das Gefühl geben, den eigenen Wissensbesitz angemessen abzubilden. Abschließend wurde die Präferenz für eines der beiden bearbeiteten Verfahren erbeten. Ausgenommen der Frage nach der Instruktionsverständlichkeit waren alle Bewertungen nach Möglichkeit zu begründen.

Alle Items sollten anhand einer fünfstufigen Skala bewertet werden, wobei der Wert 5 Zustimmung bedeutet bzw. daß es leicht fiel, in Phrasen zu antworten, während der Wert 1 für Ablehnung des Items steht bzw. daß das Antworten in Phrasen schwer fiel. Die deskriptiven Statistiken für alle Items den Wortassoziationsversuch betreffend sind in Abbildung 140 dargestellt, die Kennwerte bzgl. des Lückentextes sind in Abbildung 141 zu finden.

	T1		T2		
	Verständlichkeit der Instruktion	Antworten in Phrasen	Verständlichkeit der Instruktion	Antwort in Phrasen	Abbild eigenen Wissens
<i>M</i>	4,55	3,20	4,68	3,45	3,30
<i>SD</i>	0,76	1,32	0,58	1,10	0,87
Minimum	3	1	3	2	2
Maximum	5	5	5	5	5

Abbildung 140: Statistische Kennwerte (Mittelwert, Standardabweichung, Minimal- und Maximalwerte) der Items des Akzeptanz-Fragebogens den Wortassoziationsversuch betreffend in Studie 2 – psychotherapeutische Ambulanz

Die Instruktion des Wortassoziationsversuchs wird sowohl bei der ersten ($M=4,55$; $SD=0,76$) als auch bei der zweiten Erhebung ($M=4,68$; $SD=0,58$) als sehr verständlich bewertet, beide Urteile unterscheiden sich dabei nicht voneinander (Wilcoxon: $z=-,378$; $p> .05$; $d= .19$). Das Antworten in Phrasen fiel den Probanden im Mittel weder besonders schwer noch besonders leicht, der Gruppenmittelwert zu T1 ($M=3,20$; $SD=1,32$) unterscheidet sich nicht signifikant (Wilcoxon: $z=-1,387$; $p> .05$; $d= .20$) von dem des zweiten Meßzeitpunktes ($M=3,45$; $SD=1,10$). Diejenigen Probanden, denen es leichter fiel, begründeten dies in erster Linie damit, daß dieses Antwortformat ihrer eigenen Form der Gedächtnisrepräsentation ähnlich sei sowie mit dem Umstand, daß stichwortartiges Antworten mit Zeitersparnis einhergehe. Diejenigen Probanden, denen ein solches Antworten eher Probleme bereitete, begründeten ihr Urteil v.a. damit, daß komplexe Zusammenhänge im Wortassoziationsversuch nicht darstellbar seien. Zudem wurde oftmals auch

die Unsicherheit in der Wissensdomäne angeführt, offenbar wird das Assoziieren von Stichworten in solchen Fällen als erschwert erlebt.

	Verständlichkeit der Instruktion	Antwort in Phrasen	Abbild eigenen Wissens
<i>M</i>	4,25	3,10	3,50
<i>SD</i>	0,97	1,12	1,00
Minimum	2	1	2
Maximum	5	5	5

Abbildung 141: Statistische Kennwerte (Mittelwert, Standardabweichung, Minimal- und Maximalwerte) der Items des Akzeptanz-Fragebogens den Lückentext betreffend (zu T2) in Studie 2 – psychotherapeutische Ambulanz

Der Lückentext wird ähnlich wahrgenommen wie der Wortassoziationsversuch: auch ihm bescheinigen die Probanden mit $M=4,25$ ($SD=0,97$) eine sehr verständliche Instruktion. Die Beurteilung beider Verfahren unterscheidet sich zwar nicht signifikant (Wilcoxon: $z=-1,903$; $p>.05$), allerdings verweist die Effektstärke von $d=.54$ auf einen Effekt mittlerer Stärke zugunsten des Wortassoziationsversuchs. Das stichwortartige Antworten innerhalb des Textgerüsts des Lückentextes wird ähnlich ambivalent erlebt ($M=3,10$; $SD=1,12$) wie beim Wortassoziationsversuch (Wilcoxon: $z=-0,964$; $p>.05$; $d=.32$). Probanden, denen dies eher leicht fiel, begründeten dies damit, daß dieses Format eine klare Zielvorgabe mache, während Probanden, denen es eher schwer fiel, dazu bemerkten, das Ziel der einzelnen Textlücken erschien ihnen oftmals unklar oder uneindeutig.

Beide Verfahren geben den Probanden offenbar nur ein mäßiges Gefühl, das eigene Wissen repräsentativ abbilden zu können, die Beurteilung für den Wortassoziationsversuch ($M=3,30$; $SD=0,87$) unterscheidet sich nicht (Wilcoxon: $z=-0,836$; $p>.05$; $d=.21$) von der des Lückentextes ($M=3,50$; $SD=1,00$). Positive Bewertungen bezüglich des Lückentextes wurden überwiegend damit begründet, daß dieser eine Faktenabfrage realisiere. Die Probanden, welche negative Urteile abgaben, sahen sich v.a. zu sehr durch die Vorgaben des Lückentextformates eingengt. Hinsichtlich des Wortassoziationsversuchs waren die Begründungen der Urteile zu diesem Item breit gestreut. Positiv ausgeprägte Bewertungen wurden begründet mit einer Anregungsfunktion des Wortassoziationsversuchs, der Entsprechung mit der eigenen Gedächtnisrepräsentation oder die Erleichterung dadurch, daß die Stimuli das Thema gut repräsentierten. Hatten die Probanden weniger das Gefühl, der Wortassoziationsversuch könne ein angemessenes Abbild ihres eigenen Wissens sein, dann wurde v.a. auf die Beeinträchtigung bei der Darstellung komplexer Zusammenhänge verwiesen. Zudem wurden Unsicherheiten in der Wissensdomäne benannt, Motivationsschwäche durch den hohen Aufwand des Verfahrens sowie die fehlende Möglichkeit, in der eingesetzten computerbasierten Form des Wortassoziationsversuchs zu einem späteren Zeitpunkt Konzepte zu früheren Stimuli zu ergänzen.

In der abschließenden Frage zur Präferenz hinsichtlich eines der beiden Verfahren wird der Wortassoziationsversuch signifikant häufiger benannt ($\chi^2(2, N=20)=9,7$; $p=.008$), 13 Probanden bevorzugten dieses Verfahren und begründeten dies in erster Linie damit, daß der Wortassoziationsversuch Raum für eigene Formulierungen gebe und somit mehr Freiheiten zulasse. Weiterhin wurde die Einfachheit und Anwendungsfreundlichkeit des Verfahrens hervorgehoben sowie die Unähnlichkeit zu einer Prüfungssituation. Lediglich fünf Probanden bevorzugten den

Lückentext gegenüber dem Wortassoziationsversuch. Sie begründeten dies mit der erlebten Erleichterung durch die Einbettung in einen Kontext, daß es die eigenen Lücken deutlicher mache sowie die Anstrengung durch die Bildschirminteraktion (zur Erinnerung: im vorliegenden Design wurde der Wortassoziationsversuch computergestützt durchgeführt, der Lückentext als Paper-Pencil-Variante). Nur zwei Probanden bekundeten keine Präferenz gegenüber einem der beiden Verfahren, wovon einer den Wortassoziationsversuch zu aufwendig und den Lückentext als zu eingeschränkt erlebte, während der andere Proband beide Verfahren als gleichermaßen langweilig empfand.

Zusammenfassend kann die Akzeptanz des Wortassoziationsversuchs bei den Probanden als gegeben angenommen werden. Die Aufgabenstellung wurde als gut verständlich wahrgenommen, das Antworten in Phrasen wurde als bewältigbar, wenn auch nicht besonders leicht erlebt. Auch das Gefühl, daß der Wortassoziationsversuch ein gutes Abbild des eigenen Wissens sei, war insgesamt mäßig ausgeprägt. Diese Wahrnehmungen des Wortassoziationsversuchs unterscheiden sich nicht von der des Lückentextes. Dennoch wird der Wortassoziationsversuch insgesamt dem Lückentext vorgezogen, was Probanden insbesondere mit den großen Freiheitsgraden der Assoziationstechnik begründen, während der Lückentext als zu einschränkend erlebt wird.

7.2.3. Zusammenfassung der Befunde zur Akzeptanz

Liebs (2003), Pokrandt (2004) sowie Ruprecht (2004) merken an, daß es vielen Probanden schwer fiel, sich an die Instruktion des Wortassoziationsversuchs zu halten und nur in Stichworten oder kurzen Phrasen zu antworten. Während einige Probanden bei Liebs (2003) die Testsituation als Prüfungssituation empfanden, erlebten einzelne Probanden bei Pokrandt (2004) Blockaden beim spontanen Abruf ihres Wissens. Pokrandt (2004) berichtet zudem über Probanden, die wenig motiviert an den Erhebungen teilnahmen, was sie v.a. darauf zurückführt, daß diese Beschäftigten der Gesamtuntersuchung gegenüber mißtrauisch und wenig aufgeschlossen waren. Wegwarth (2002) bewertet den Wortassoziationsversuch als einfach und unkompliziert: da die Probanden ihrer Studie bereits im Seminarkontext Brainstormingverfahren einsetzten, waren sie mit Methoden dieser Art vertraut und nahmen auch den Wortassoziationsversuch gut an. Insbesondere im Vergleich zur Strukturlegetechnik wird dem Wortassoziationsversuch ein Vorsprung in der Akzeptanz durch die Probanden bescheinigt, so resümieren Rothe und Hinnerichs (2005) über die Strukturlegetechnik, sie sei sehr kompliziert und umfangreich, so daß die Motivation der Probanden nachlasse und Akzeptanzprobleme auftreten können.

Diese Erfahrungen aus verschiedenen Feldstudien sowie aus den empirischen Daten, welche explizit die Akzeptanz des Wortassoziationsversuchs durch die Probanden fokussieren, belegen anschaulich, daß dieses Verfahren insgesamt von den Probanden angenommen wird. Dabei ist gemäß der empirischen Befunde die schriftliche Durchführung des Wortassoziationsversuchs aus Sicht der Probanden der mündlichen Variante überlegen. Als Gründe dafür wurden v.a. Zeit- und Leistungsdruck benannt, welche beim mündlichen Versuch in stärkerem Maße von den Probanden erlebt wurden als beim schriftlichen. Führt man den Wortassoziationsversuch mündlich durch, so erweist sich die visuelle Präsentation parallel zum Nennen der Stimuli als sinnvoll, die Probanden gaben überwiegend an, ein solches Vorgehen als unterstützend zu erleben. An dieser Stelle sei auch auf Pokrandt (2004) verwiesen, die berichtet, v.a. bei langen Assoziationsketten vergaßen einige

Probanden bisweilen den ursprünglichen Stimulusbegriff und mußten diesen erneut erfragen (in ihrer Studie wurde der Wortassoziationsversuch mündlich durchgeführt).

In einer zweiten Studie, welche die Akzeptanz des Verfahrens explizit thematisierte, wurde der Wortassoziationsversuch einem Lückentext gegenüber gestellt. Dabei werden beide Verfahren von den Probanden als gleichermaßen gut verständlich hinsichtlich der Aufgabenstellung beurteilt. Die Anforderung, in Phrasen zu antworten, wird bei beiden Verfahren ähnlich wahrgenommen (weder als besonders schwer noch besonders leicht); auch vermitteln Wortassoziationsversuch und Lückentext in gleicher Weise nur mäßig das Gefühl, ein angemessenes Abbild des eigenen Wissens zu sein. Dennoch präferiert die Mehrheit der Probanden den Wortassoziationsversuch als Verfahren der Wissensdiagnose, was v.a. auf das Erleben vergleichsweise großer Freiheitsgrade beim Assoziieren und gleichzeitig großer Einschränkungen beim Lückentext zurückzuführen ist.

Insgesamt kann die Akzeptanz des Wortassoziationsversuchs durch die Probanden als gegeben hingenommen werden (Fragestellung 8). Dabei scheinen die Vorzüge des Assoziierens sich den Probanden v.a. in Kontrast zu anderen Verfahren der Wissensdiagnose abzuheben, welche beispielsweise deutlich zeitaufwendiger sind, größere kognitive Anstrengungen erfordern oder erhebliche Restriktionen beim Zugriff auf den individuellen Wissensbesitz vorgeben.

7.3. Ökonomie

Zur Bewertung der Ökonomie des Wortassoziationsversuchs können empirische Befunde aus drei Studien herangezogen werden, die sich mit dem zeitlichen Aspekt der Durchführungsökonomie auseinandersetzen. Zwei dieser Studien vergleichen den zeitlichen Aufwand des Wortassoziationsversuchs mit dem anderer Verfahren der Wissensdiagnose (Strukturlegetechnik und Lückentext), eine weitere Studie betrachtet die Durchführungszeit von mündlichem vs. schriftlichem Wortassoziationsversuch. Daten zur Vorbereitungs- und Auswerteökonomie wurden nicht erhoben, ohnehin liegt im betrieblichen Kontext aufgrund der betriebswirtschaftlichen Bedeutung der Arbeitszeit der Fokus auf der Durchführungsökonomie. Dennoch sollen auch diese Aspekte hier ausdiskutiert sowie die Erfahrungen verschiedener Versuchsleiter dazu zusammengetragen werden.

7.3.1. Anmerkungen zur Vorbereitungsökonomie

Liebs (2003), Pokrandt (2004) sowie Ruprecht (2004) bewerten die Vorarbeiten zur Erarbeitung der Wissensdomäne als sehr umfangreich. Allerdings ist zu sagen, daß *jedes* Verfahren zur Wissensdiagnose eine intensive Auseinandersetzung mit der zu untersuchenden Wissensdomäne voraussetzt. Die Vorbereitung des Untersuchungsmaterials selbst ist in seinem Aufwand überschaubar und unkompliziert: das Erstellen von Assoziationsbögen ist mit herkömmlichen Textbearbeitungsprogrammen leicht zu realisieren, ebenso ggf. das Fertigen von „Kärtchen“, auf denen beim mündlichen Wortassoziationsversuch die Stimuli visuell dargeboten werden können. Für den mündlichen Versuch ist zudem ein Tonaufzeichnungsgerät vorzubereiten, bei einer computergestützten Versuchsdurchführung wie beim Assoziations-Struktur-Test von Meyer (2007) ist das Erstellen des Tests sowie der Zugangscodes unproblematisch und schnell zu realisieren. Auch Wegwarth (2002) beschreibt das Erstellen des Untersuchungsmaterials als einfach.

7.3.2. Anmerkungen zur Durchführungsökonomie

Liebs (2003) berichtet, daß bei dem von ihr durchgeführten Wortassoziationsversuch (im mündlichen Modus) in einem Handelsunternehmen zwar eine Zeitbegrenzung von zwei Minuten je Stimulus vorgegeben war, diese Obergrenze jedoch von keinem der Probanden erreicht wurde. Die von ihr eingesetzten 17 Stimuli waren also mit einer maximalen Durchführungsdauer von 34 Minuten je Proband angelegt, diese Zeitspanne wurde tatsächlich jedoch nicht benötigt. Ähnliches berichtet Ruprecht (2004): auch bei ihr nahmen die Wortassoziationsversuche weniger Zeit in Anspruch als erwartet. Wegwarth (2002) und Jordan (2004) merken an, daß der Einsatz des Wortassoziationsversuchs sich gut und relativ störungsfrei in den Seminarablauf einer Weiterbildungsveranstaltung integrieren ließ. Als vorteilhaft erwies sich dabei die geringe zeitliche Belastung für die Probanden, ebenso ist die Durchführung des Wortassoziationsversuchs auch für den Versuchsleiter recht einfach (Wegwarth, 2002, Jordan, 2004).

Zwar ist bei der mündlichen Durchführung die Bearbeitungszeit des Wortassoziationsversuchs für den einzelnen Probanden geringer, da dabei jedoch immer nur ein Proband getestet werden kann, ist die Durchführungsökonomie in der Summe für den Versuchsleiter deutlich geschmälert (dieses Argument ist irrelevant, wenn aufgrund der betrieblichen Abläufe ohnehin keine Gruppentestung der Beschäftigten möglich ist). Der schriftliche Modus ist also unter ökonomischem Gesichtspunkt grundsätzlich im Vorteil durch seine Möglichkeit der zeit- und raumparallelen Bearbeitung durch mehrere Probanden sowie durch seine Auswerteökonomie gegenüber dem mündlichen Modus. Wird eine schriftliche Variante computergestützt eingesetzt, so ist das Verfahren zwar auch als Gruppenversuch einsetzbar, setzt aber die Verfügbarkeit mehrerer Computer voraus. Dies ist leicht zu realisieren bei Vorhandensein eines Computer-Pools (z.B. in Weiterbildungseinrichtungen), für den Einsatz des Assoziations-Struktur-Tests in der gegenwärtigen technischen Realisierung von Meyer (2007) ist darüber hinaus ein Internetzugang erforderlich. Sind diese organisatorischen Aspekte berücksichtigt, ist der Einsatz des Verfahrens unproblematisch.

7.3.2.1. Durchführungsökonomie in Studie 1: Modus

In Studie 1 – Modus – wurde die schriftliche Durchführung des Wortassoziationsversuchs der mündlichen gegenübergestellt, wobei auch die Bearbeitungszeiten durch die Probanden erfaßt wurden. Es sei daran erinnert, daß in dieser experimentellen Studie eine sehr heterogene Stichprobe in Hinblick auf Alter oder Berufsstand herangezogen wurde. Das Lebensalter oder auch der Normwert des Wortflüssigkeitstests (nach Jäger und Althoff, 1994) kann als Moderatorvariable der von den Probanden benötigten Bearbeitungszeit ausgeschlossen werden, Partialkorrelationen erbrachten keine Hinweise darauf.

Wie bereits an früherer Stelle beschrieben, wurden sowohl im schriftlichen als auch im mündlichen Wortassoziationsversuch dieser Studie jeweils 20 Stimuli bearbeitet. Im Mittel wurde der mündliche Wortassoziationsversuch in 11,57 Minuten absolviert ($SD=6,71$), für den schriftlichen Teil wurden 46,17 Minuten benötigt ($SD=28,13$). Wie in Abbildung 142 aus den Standardabweichungen sowie Minima und Maxima ersichtlich, gibt es innerhalb eines jeden Verfahrensteils große interindividuelle Unterschiede bei der benötigten Bearbeitungszeit, dennoch ist der Unterschied zwischen beiden Bearbeitungszeiten hoch signifikant (Wilcoxon: $z=-4,783$; $p=.000$; $d=1.69$). Daß die schriftliche Variante länger braucht als die mündliche, ist zu erwarten –

in dieser Studie wurde zur Bearbeitung der gleichen Anzahl Stimuli im schriftlichen Teil etwa viermal soviel Zeit benötigt wie im mündlichen Teil.

	mündlicher Teil	schriftlicher Teil
<i>M</i>	11,57	46,17
<i>SD</i>	6,71	28,13
minimaler Wert	4	15
maximaler Wert	28	115

Abbildung 142: Kennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – der mittleren Bearbeitungszeiten (in Minuten) des Wortassoziationsversuchs im mündlichen vs. schriftlichen Teil in Studie 1 – Modus

7.3.2.2. Durchführungseconomie in Studie 2: psychotherapeutische Ambulanz

In Studie 2 – psychotherapeutische Ambulanz – wurden den Probanden im Wortassoziationsversuch 14 Stimuli dargeboten und anschließend ein Lückentext vorgelegt, welcher zu zwölf dieser Stimulusbegriffe einen kurzen Text mit jeweils vier Lücken beinhaltete. Durch die Deckungsgleichheit der Inhalte ist eine direkte Vergleichbarkeit beider Verfahren gegeben. Die statistischen Kennwerte der Durchführungsdauer sind in Abbildung 143 dargestellt.

	im Wortassoziationsversuch T1	im Wortassoziationsversuch T2	im Lückentext T2
<i>M</i>	61,50	51,05	23,30
<i>SD</i>	30,21	20,68	5,25
Minimum	34	25	16
Maximum	180	120	34

Abbildung 143: Kennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – der mittleren Bearbeitungszeiten (in Minuten) des Wortassoziationsversuchs sowie des Lückentextes in Studie 2 – psychotherapeutische Ambulanz

Da das Design dieser Studie eine Testwiederholung des Wortassoziationsversuchs vorsah, sollen an dieser Stelle auch die Bearbeitungszeiten für beide Meßzeitpunkte berichtet werden. Die enorm großen Standardabweichungen bei den Wortassoziationsversuchen weisen auf erhebliche interindividuelle Differenzen hin, während die Werte des Lückentextes deutlich weniger streuen. Während bei der ersten Durchführung des Wortassoziationsversuchs im Mittel 61,50 Minuten (*SD*=30,21) benötigt wurden, konnte er bei der Wiederholungsmessung in *M*=51,05 Minuten (*SD*=20,68) absolviert werden (Wilcoxon: $z=-2,689$; $p=.007$), dies ist ein Effekt mittlerer Stärke ($d=.40$). Für die Bearbeitung des Lückentextes hingegen wurden im Mittel 23,30 Minuten (*SD*=5,25) benötigt. Verglichen mit dem Wortassoziationsversuch, welcher zum selben Erhebungszeitpunkt stattfand, konnte der Lückentext also in weniger als der Hälfte der Zeit bearbeitet werden, dieser Unterschied ist statistisch signifikant (Wilcoxon: $z=-3,921$; $p=.000$; $d=1.84$). In der Frage der Durchführungseconomie zeigt sich also in dieser Studie der Lückentext dem Wortassoziationsversuch deutlich überlegen, wobei beide Verfahren denselben Wissensgegenstand umfassen.

7.3.2.3. Durchführungsökonomie in Studie 13: Stoffdruckerei

In Studie 13 – Stoffdruckerei – wurde neben dem Wortassoziationsversuch auch die Strukturlegetechnik als wissensdiagnostisches Verfahren eingesetzt. Pokrandt (2004) erhob zu beiden Verfahren die jeweilige Bearbeitungsdauer, da die Erhebungen als Einzelsitzungen durchgeführt wurden. Für den Wortassoziationsversuch benötigten die Probanden der Stichprobe im Mittel etwas mehr als eine Viertelstunde ($M=16,99$ min; $SD=14,42$), wobei der kürzeste Versuch etwa fünf Minuten dauerte, der mit Abstand längste jedoch über eine Stunde (siehe auch Abbildung 144). Es soll dabei in Erinnerung gerufen werden, daß der Wortassoziationsversuch dieser Studie immerhin 28 Stimuli umfaßte.

	Wortassoziationsversuch (N=21)	Strukturlegetechnik (N=19)
<i>M</i>	16,99	53,89
<i>SD</i>	14,42	20,92
Minimum	4,68	17,10
Maximum	71,13	106,51

Abbildung 144: Kennwerte – Mittelwerte (*M*) und Standardabweichungen (*SD*) – der mittleren Bearbeitungszeiten (in Minuten) des Wortassoziationsversuchs sowie der Strukturlegetechnik in Studie 13 – Stoffdruckerei

Für die Strukturlegetechnik, an der zwei der Probanden nicht teilnehmen wollten, wurde im Mittel eine knappe Stunde Zeit benötigt ($M=53,89$; $SD=20,92$), auch hier konnte ein weiter Bereich zwischen minimal etwa einer Viertelstunde und maximal über anderthalb Stunden beobachtet werden (Abbildung 144). Dabei sei daran erinnert, daß die Knoten in den zu legenden Wissensnetzen vom Versuchsleiter vorgegeben waren, allen Probanden standen also die gleiche Menge an Knoten zur Verfügung, die sie in ihr Netz einbauen konnten oder auch nicht. Dennoch läßt sich ein signifikanter korrelativer Zusammenhang finden zwischen der individuellen Bearbeitungszeit für den Wortassoziationsversuch und der Zeit für die Strukturlegetechnik: mit $r= .603$ ($p= .003$) liegt eine signifikante mittlere Korrelation vor – wer bereit ist, viel Zeit in das eine Verfahren zu investieren, tut es auch beim Bearbeiten des anderen Verfahrens.

Um den direkten Vergleich zwischen Wortassoziationsversuch und Strukturlegetechnik zu ermöglichen, seien in Anbetracht der etwas geringeren Stichprobengröße beim Strukturlegen hier die statistischen Kennwerte der Assoziations-Zeiten für diese reduzierte Stichprobe nachgetragen: diese 19 Probanden benötigten im Mittel 17,77 Minuten ($SD=14,98$), wobei der Minimal- bzw. Maximalwert deckungsgleich sind mit den bereits für die gesamte Stichprobe berichteten (siehe Abbildung 144). Damit ist die Bearbeitungszeit für die Strukturlegetechnik dreimal so hoch wie die des Wortassoziationsversuchs. Dieser augenscheinliche Unterschied ist auch statistisch signifikant (Wilcoxon: $z=-3,823$; $p= .000$) und beschreibt mit $d=1.99$ einen sehr großen Effekt.

7.3.3. Anmerkungen zur Auswerteökonomie

In den soeben beschriebenen empirischen Befunden konnten die Nachbearbeitungszeiten der Wortassoziationsversuche durch die Versuchsleiter nicht systematisch erfaßt werden. Wie in allen Verfahren mit freien Antwortformaten ist auch die Auswertung des Wortassoziationsversuchs vergleichsweise aufwendig, Ruprecht (2004) merkt einen hohen zeitlichen Aufwand für die Auswertung der Assoziationsdaten an. Der Aufwand wird dabei von zwei Faktoren determiniert:

zum einen vom Durchführungsmodus, zum anderen von der Herangehensweise der Auswertung. Es liegt auf der Hand, daß ein mündlicher Wortassoziationsversuch nicht auf eine Transkription der Assoziationen verzichten kann. Dies ist allerdings als zeitintensiv und mühsam zu bewerten, und es kann erschwert werden durch die Artikulation der Probanden, Qualität der Tonaufzeichnung, Störgeräusche usw., was oftmals ein mehrfaches Abhören bestimmter Passagen erforderlich macht. Ein Wortassoziationsversuch im Paper-Pencil-Format erfordert i.d.R. ebenfalls eine Übertragung der Daten in ein digitales Format, um die Weiterverarbeitung der Daten zu ermöglichen. Haupterschwerernis dabei ist die Unleserlichkeit mancher Handschrift. Beide Problemfelder können jedoch umgangen werden durch den Einsatz computerbasierter Verfahren, bei denen Probanden per Tastatur ihre Assoziationen selbst in einem digitalen Format ablegen, so daß diese Daten vom Versuchsleiter mit nur wenigen weiteren Handgriffen weiterverarbeitet werden können. Zusammenfassend ist anzumerken, daß unter dem zeitlichen Aspekt das mündliche Format des Wortassoziationsversuchs dem schriftlichen in der Durchführungsökonomie deutlich überlegen ist, sich dieser Effekt aber in der Auswerteökonomie wieder aufhebt.

Der zweite angesprochene Aspekt der Auswerteökonomie betrifft die Auswertestrategie. Will man lediglich einen quantitativen Kennwert ermitteln, welcher sämtliche Assoziationen eines Probanden umfaßt, so reicht u.U. ein einfaches Auszählen der individuellen Assoziationen (unproblematisch, sofern man Daten in irgendeiner schriftlichen Form vorzuliegen hat). Strebt man differenziertere Auswertezugänge an wie z.B. die Gewichtung unterschiedlicher Qualitäten von Assoziationsinhalten oder die Bestimmung von hochfrequenten Assoziationen innerhalb der Stichprobe, ist jedoch ein größerer Aufwand für die Auswertung einzuplanen. Hat man allerdings erst einmal die Assoziationen in geeigneter Weise verarbeitet (z.B. läßt sich in Excel oder anderen Tabellenkalkulationsprogrammen leicht eine Matrix „Versuchsperson x Assoziation“ erstellen), so lassen sich weiterführende Prozeduren relativ effektiv anschließen. Nicht zuletzt ist noch eine weitere Determinante der Auswerteökonomie zu nennen: die Anzahl der Stimuli im Wortassoziationsversuch. Trivialerweise vergrößert sich der Auswerteaufwand mit zunehmender Anzahl von vorgelegten Stimuli. Unter materiellem Aspekt ist anzumerken, daß die Auswertung des Wortassoziationsversuchs keiner speziellen Software oder anderer Materialien bedarf, mit den gängigen Office-Anwendungen lassen sich Auswertungen vornehmen, wobei sich Tabellenkalkulationsprogramme als besonders effizient für weiterführende Prozeduren erwiesen haben.

7.3.4. Zusammenfassung der Befunde zur Ökonomie

Die hier vorgestellten empirischen Befunde zur Durchführungsökonomie des Wortassoziationsversuchs vergleichen dieses Instrument mit dem jeweils parallel eingesetzten Strukturlegeverfahren bzw. Lückentext. Separat betrachtet erscheint der Aufwand für die Durchführung und Auswertung des Wortassoziationsversuchs aufgrund seines freien Antwortformates nicht unerheblich, im Kontrast zur in den vorgestellten Studien oftmals parallel eingesetzten Strukturlegetechnik jedoch kann sich die Ökonomie des Wortassoziationsversuchs gut behaupten. Der Wortassoziationsversuch ist dem Strukturlegeverfahren deutlich überlegen (sowohl in zeitlicher als auch materieller Hinsicht), während er seinerseits hinter dem zeitlichen Aufwand bei Durchführung des Lückentextes zurückstehen muß.

Eine weitere Studie widmete sich der Frage nach den zeitökonomischen Aspekten verschiedener Durchführungsformate des Wortassoziationsversuches selbst. Dabei zeigte sich – wie zu erwarten – der mündliche Wortassoziationsversuch in der Durchführung deutlich zeitsparender als der schriftliche (im intraindividuellen Vergleich als Einzeltestung). Dieser Umstand ist insbesondere im Kontext einer Feldstudie ein gewichtiges Argument. Wird ein Wortassoziationsversuch im Rahmen einer arbeitspsychologischen Fragestellung in einem Unternehmen durchgeführt, so ist ein bedeutsamer Aspekt des Designs die zeitliche Inanspruchnahme des einzelnen Mitarbeiters durch die Erhebung. Unter der Maßgabe einer möglichst geringen Störung bzw. Unterbrechung des Arbeitsablaufs wäre die Durchführung des Wortassoziationsversuchs als mündliche Erhebung die Methode der Wahl. Zwar wäre es für den Versuchsleiter in der Summe deutlich zeitsparender, alle Probanden zeitparallel einen schriftlichen Wortassoziationsversuch in einer Gruppentestung bearbeiten zu lassen, dieser Aspekt ist jedoch nachrangig gegenüber betrieblichen Erfordernissen.

In der Frage Auswerteökonomie liegen keine empirischen Daten vor, jedoch werden entsprechende Erfahrungen v.a. im Kontrast zum Einsatz der Strukturlegetechnik berichtet. Viele Versuchsleiter beklagen einen sehr hohen Auswerteaufwand der Strukturlegetechnik, der mit dem „Speichern“ und Übertragen der als Karten gelegten Netze beginnt (am einfachsten als Fotografie) und in das mühevoll Auszählen der quantitativen Kennwerte mündet. Liebs (2003) beschreibt die Strukturlegetechnik als sehr zeitintensiv (sowohl in Vorbereitung als auch Durchführung), wobei beim Strukturlegen auch nur ein vergleichsweise kleiner Ausschnitt der Wissensdomäne bearbeitet werden kann. Auch Rothe und Hinnerichs (2005) beschreiben die Strukturlegetechnik als sehr aufwendig – sowohl für die Probanden als auch den Versuchsleiter –, während sie den Wortassoziationsversuch als ein sehr rationelles Verfahren bewerten. Gleiches berichten Kermas (2006), Jordan (2004) und auch Pokrandt (2004), welche überdies darauf hinweist, daß die lange Durchführungszeit der Strukturlegetechnik einige Probanden an die Grenzen ihrer Konzentrationsfähigkeit führte. Hingegen bewerten Pokrandt (2004), Ruprecht (2004), Rothe und Hinnerichs (2005) oder auch Liebs (2003) den Wortassoziationsversuch als relativ ökonomisches Verfahren, bei dem Aufwand und Nutzen in einem angemessenen Verhältnis stehen.

Zusammenfassend läßt sich der Wortassoziationsversuch als sehr ökonomisches Verfahren bewerten (Fragestellung 9). Viele der Voraussetzungen dazu, so wie sie z.B. Lienert und Raatz (1998) oder Bühner (2006) definieren, erfüllt das Verfahren: die Durchführungszeit des Wortassoziationsversuchs ist im allgemeinen kurz, v.a. im Vergleich mit der Strukturlegetechnik hebt sich das Assoziieren positiv ab. Für die Durchführung des Wortassoziationsversuchs ist wenig Material erforderlich, dieses ist auch leicht durch den Versuchsleiter selbst zu erstellen und leicht zu handhaben. Da der Wortassoziationsversuch auch als schriftliche Variante einsetzbar ist, kann er als Gruppenversuch durchgeführt werden. Lediglich in der Frage der einfachen und schnellen Auswertung des Versuchs müssen Abstriche gemacht werden, insbesondere wenn eine qualitativ orientierte Auswertung angestrebt ist. Dieses Problem ist allerdings allen Verfahren mit freien Antworten eigen. Zumindest kann die Ökonomie der Datenaufbereitung unterstützt werden durch den Einsatz von computerbasierten Tools wie dem Assoziations-Struktur-Test von Meyer (2007).

7.4. Nützlichkeit

Der Wortassoziationsversuch reiht sich ein in eine Vielzahl wissensdiagnostischer Instrumente, welche in arbeitspsychologischer Forschung und Praxis vielfältige Verwendung finden. V.a. im Zusammenhang mit Weiterbildungsevaluation sowie im Kontext von Wissensmanagement (hier insbesondere die Phase der Wissensidentifizierung) besteht ein praktisches Bedürfnis, Fachwissen von Personen zu erfassen. Die Vorzüge des Wortassoziationsversuchs gegenüber anderen Verfahren liegen dabei in erster Linie in der leichten Handhabbarkeit des Verfahrens, v.a. in der Durchführung vor Ort. In Abhängigkeit von den jeweiligen Rahmenbedingungen kann der Wortassoziationsversuch mündlich als Einzelversuch durchgeführt werden oder als Gruppenversuch schriftlich in der Paper-Pencil-Version oder mit Hilfe einer Computer-Version. Als Verfahren mit einem freien Antwortformat hat er gegenüber Verfahren mit geschlossenem Antwortformat den grundsätzlichen Vorzug, die Wahrscheinlichkeit für das Erraten der „richtigen“ Antwort deutlich zu minimieren. Ein wichtiger, wenn nicht sogar der wichtigste Aspekt für den Praxiseinsatz ist der zeitliche Umfang des Verfahrens für den Probanden: da Unternehmen die arbeitspsychologischen Untersuchungen i.d.R. in den laufenden Betrieb integrieren, wird ein Verfahren benötigt, welches schnell durchführbar ist und den Probanden nicht allzu lange aus dem Arbeitsprozeß fernhält (sollten die Erhebungen außerhalb der Arbeitszeit stattfinden, ist die Teilnahmemotivation und damit die Aussagekraft der Erhebungen ohnehin fraglich). Zudem ist der Wortassoziationsversuch für den Probanden leicht verständlich. Alles in allem kann dem Wortassoziationsversuch eine Nützlichkeit bescheinigt werden (Fragestellung 10).

7.5. Normierung

Eine Normierung mit Normtabellen und Standardwerten, so wie sie z.B. Lienert und Raatz (1998) oder Bühner (2006) fordern, kann der Wortassoziationsversuch nicht leisten (Fragestellung 11). Ein wesentliches Merkmal des Wortassoziationsversuchs ist seine universelle Einsetzbarkeit, da er nicht über einen festen Itemsatz verfügt, sondern vielmehr jedes Einsatzfeld die Konstruktion einer auf die jeweilige Fragestellung abgestimmten Stimulusliste erfordert. Der Wortassoziationsversuch wird also in Hinblick auf die konkrete Fragestellung der Untersuchung bzw. an den konkreten Arbeitsgegenstand angepaßt und „maßgeschneidert“. Dies betrifft sowohl Inhalt als auch Umfang dieser Stimulusliste. Daher ist es nicht möglich (und auch nicht sinnvoll i.S. von ökonomisch), für den Wortassoziationsversuch eine Normierung vorzunehmen. Die von Bühner (2006) geforderte Mindestgröße von 300 Probanden für eine Normstichprobe übersteigt ohnehin um ein Vielfaches die Größe der üblichen Untersuchungsstichproben in der arbeitspsychologischen Praxis. Damit würde die für die Normstichprobe benötigte Zahl von Probanden bei weitem die der Anwendungsstichprobe übertreffen und wäre daher unrentabel.

Die Notwendigkeit, immer wieder einen „neuen“ Wortassoziationsversuch konstruieren zu müssen, beeinflusst allerdings die Interpretierbarkeit der Ergebnisse: durch die Abwesenheit von Normen ist die Einordnung der Assoziationsleistung eines Individuums erschwert. So lassen sich quantitative Assoziationskennwerte immer nur in Relation zur konkreten Stichprobe bzw. zum hypothetischen Wissensmodell interpretieren oder als intraindividuelle Vergleich zwischen verschiedenen Stimuli beispielsweise. Eine grundsätzliche, stichprobenunabhängige Bewertung, was eine hohe oder aber geringe Assoziationsleistung ist, kann es daher nicht geben.

Zu verschiedenen Stimuli kann der Versuchsleiter a priori verschiedene Annahmen über den Umfang der Assoziationen entwickeln. Die Anzahl der geäußerten Assoziationen hängt nicht nur vom individuellen Wissensstand ab, sondern zentral auch vom Stimulus selbst: in welchem Maße er den Suchraum für den Wissensabruf eingrenzt oder nicht, und nicht zuletzt, wie umfangreich oder eingeschränkt der Suchraum auf diesen Stimulus hin überhaupt objektiv ist (z.B. durch ein hypothetisches Wissensmodell darstellbar). So ist es beispielsweise bei einem Stimulus, welcher ein psychopathologisches Störungsbild beschreibt, angemessen zu erwarten, daß 20 bis 30 Assoziationen genannt werden, da der Bedeutungshof dieses Begriffes hier von vornherein sehr breit gefaßt ist, das gesamte Feld von Symptomatik, Ätiologie, Therapie usw. kann unter diesem Störungsbild abgeschritten werden. Lautet der Stimulus hingegen „Symptome von Störung X“, so weist das hypothetische Wissensmodell nur eine eng umschriebene Anzahl von möglichen Inhalten auf, die durchschnittliche Anzahl von Assoziationen wird also erwartungsgemäß deutlich niedriger liegen. Jeder einzelne Wortassoziationsversuch ist konstruiert für eine spezifische Fragestellung und wird einer spezifischen Population vorgelegt.

Die Einordnung eines individuellen Testergebnisses in ein Bezugssystem ist kaum möglich. Lediglich ein Intra-Gruppen-Vergleich innerhalb der abgegrenzten Untersuchungspopulation ist realisierbar. Allerdings ist das individuelle Assoziationsergebnis (sowohl quantitativ als auch qualitativ) nur immer in Relation zu den anderen Gruppenmitgliedern interpretierbar. Die Bestimmung eines Testkonstruktionsparameters wie Trennschärfe oder Schwierigkeitsindex ist beim Wortassoziationsversuch schon allein deshalb nicht möglich, da es keine richtigen oder falschen Antworten gibt, die herangezogen werden könnten. Eine Hilfskonstruktion, um Assoziationen dahingehend zu beurteilen, ob sie in Schlüsselrichtung sind, ist die Formulierung eines hypothetischen Wissensmodells (oder Expertenmodell). Auf dieser Grundlage läßt sich ein Soll-Wissen definieren, anhand dessen auch das individuelle Ergebnis interpretierbar wird, ohne daß dabei die restliche Untersuchungsstichprobe als Vergleichsmaßstab herangezogen werden muß. Die rein qualitative Auswertung und Interpretation der Assoziationen entzieht sich ohnehin der Normierung.

7.6. Vergleichbarkeit

Aus dem gleichen Grund, aus dem es keine Normierung für den Wortassoziationsversuch geben kann (vgl. Kapitel 7.5), ist auch das Kriterium der Vergleichbarkeit zu negieren (Fragestellung 12). Ein Test gilt dann als vergleichbar, wenn mehrere Parallelformen mit dem gleichen Gültigkeitsbereich vorliegen, welche auch zu ähnlichen Ergebnissen führen sollten (z.B. Bühner, 2006). Dies setzt jedoch die Vergleichbarkeit von Items voraus, wodurch sich zwei äquivalente Itemgruppen mit annähernd identischer Schwierigkeit entwickeln lassen. Im Falle des Wortassoziationsversuchs sind die Items die Stimuli, die zwar repräsentativ für die zu untersuchende Wissensdomäne sein sollen, an die aber dennoch immer ein ganz spezifisches Wissensmodell geknüpft ist. Der einzige Zugang, um eine Vergleichbarkeit von Stimuli zu erreichen, wäre der Einsatz von Synonymen. Allerdings gibt es nur in wenigen Bereichen echte Synonyme, und selbst dort wäre fraglich, ob dabei nicht doch ein anderes Wissensnetz aktiviert wird. So wäre beispielsweise im Bereich der Medizin denkbar, die deutsche sowie die lateinische Bezeichnung für ein Krankheitsbild synonym zu verwenden. Gleichsam ist dabei aber vorstellbar,

daß der lateinische Stimulus eher das Fachwissen aktiviert, wie es in der Auseinandersetzung mit der Fachliteratur gebraucht wird, während der deutsche Stimulus eher den Kontext eines Austauschs mit dem Patienten suggeriert. Die Forderung nach Vergleichbarkeit ist also beim Wortassoziationsversuch nicht erfüllbar aufgrund der Spezifität der Stimuli und ihrer konkreten, dahinter stehenden Inhalte.

7.7. Zusammenfassung der Befunde zu den Nebengütekriterien

Auch viele der Nebengütekriterien fallen für den Wortassoziationsversuch positiv aus. Das Verfahren erwies sich als sensitiv gegenüber Veränderungen im individuellen Wissensbesitz. In drei Studien wurden im Prä-Post-Design mehrere Erhebungszeitpunkte realisiert, zwischen denen akademische Lehre bzw. Weiterbildung stattfand. Es konnte ein signifikanter Anstieg der Menge assoziierter Konzepte belegt werden, welcher auf den Einfluß der Lehrphase zurückzuführen ist. Nicht nur der quantitative Wissenszuwachs konnte auf diese Weise nachgewiesen werden, auch qualitative Umstrukturierungen im individuellen Wissensbesitz ließen sich aufzeigen. Der Wortassoziationsversuch war also in der Lage zu belegen, daß nach dem Treatment nicht nur mehr, sondern auch anderes gewußt wird als vor der Weiterbildungsmaßnahme.

Viele Versuchsleiter berichten über eine insgesamt positive Grundhaltung der Probanden gegenüber dem Verfahren der freien Assoziation. Zwar erlebten einige Probanden den Versuch als Prüfungssituation, während andere Schwierigkeiten hatten, sich an die Instruktion zu halten, aber es wurde auch berichtet, daß mit der Technik des Brainstorming vertraute Probanden den Wortassoziationsversuch als sehr ähnlich und daher gut zu bewältigen erlebten. Insbesondere in Abgrenzung zu der in einigen Studien parallel eingesetzten Strukturlegetechnik zeichnete sich der Wortassoziationsversuch als schnell und unkompliziert durchführbar ab. Die empirischen Daten aus zwei Studien belegen ebenfalls eine breite Akzeptanz des Wortassoziationsversuchs durch die Probanden. Dabei wird die schriftliche der mündlichen Durchführungsvariante bevorzugt, da Letztgenannte eher zum Erleben von Zeit- und Prüfungsdruck führt. Im direkten Vergleich von Wortassoziationsversuch und einem Lückentext erleben die Probanden beide Verfahren als ähnlich hinsichtlich der Verständlichkeit der Instruktion oder der Angemessenheit, mit der das eigene Wissen durch das Verfahren abgebildet wird. Dennoch wird der Wortassoziationsversuch gegenüber dem Lückentext von den Probanden präferiert, was in erster Linie mit den beim Assoziieren gegebenen Freiheitsgraden begründet wurde, während der Lückentext zu viele Einschränkungen mache. Zusammenfassend kann der Wortassoziationsversuch als von den Probanden akzeptiert betrachtet werden.

Auch hinsichtlich des Nebengütekriteriums der Ökonomie kann sich der Wortassoziationsversuch gut behaupten. Die Vorbereitungsökonomie fällt günstig aus, da das Versuchsmaterial recht einfach selbst zu erstellen ist. Die Auseinandersetzung mit der Wissensdomäne ist zwar einigermaßen aufwendig, aber unerläßlich und letztlich immer dann erforderlich, wenn man sich mit einem spezifischen Feld mit spezifischen Arbeitsaufgaben und damit verbundenem fachspezifischem Wissen befaßt. Zur Durchführungsökonomie konnten empirische Befunde aus drei Studien herangezogen werden. Danach erweist sich der Wortassoziationsversuch zwar als zeitaufwendiger im Vergleich zu einem Lückentext mit gleichem Gegenstandsbereich, ist aber der Strukturlegetechnik deutlich überlegen. Trivialerweise zeigte sich der mündliche

Wortassoziationsversuch als zeitsparender in der Durchführung als eine schriftliche Variante gleichen Umfangs. Insgesamt jedoch kann der Wortassoziationsversuch als ein zeitökonomisches Verfahren beschrieben werden, welches sich gut in den Arbeits- bzw. Weiterbildungsablauf integrieren ließ, v.a., da er sich sowohl als Einzelversuch wie auch als Gruppentestung (schriftlich als Paper-Pencil oder computerbasiert) einsetzen läßt. Lediglich in der Auswerteökonomie des Wortassoziationsversuchs sind Abstriche zu machen, insbesondere unter zeitlichem Aspekt. Als Verfahren der freien Reproduktion ist die Weiterverarbeitung der produzierten Äußerungen sehr aufwendig. Einfluß auf das Ausmaß des Aufwands haben die Durchführungsmodalität (mündliches Assoziieren erfordert Transkription der Assoziationen, während beim Einsatz von Computertools die Rohdaten unmittelbar zur Verfügung stehen) sowie die Auswertestrategie (für das Erstellen eines einfachen Assoziationskennwertes reicht das Auszählen der Assoziationen, für weitergehende Inhaltsanalysen müssen die Assoziationen weiter aufbereitet werden). Zugunsten des Wortassoziationsversuchs ist anzumerken, daß es für seine Auswertung keine besondere Software braucht, sondern gängige Tabellenkalkulationsprogramme dazu sehr effektiv eingesetzt werden können.

Die Nützlichkeit des Wortassoziationsversuchs kann als gegeben hingenommen werden. Für die Evaluation von Weiterbildungsveranstaltungen oder im Rahmen von Wissensmanagementprozessen besteht ein praktisches Bedürfnis, ein ökonomisches, valides und sensitives Instrument zur Diagnose des individuellen Wissensbesitzes zur Verfügung zu haben. Dies kann der Wortassoziationsversuch leisten durch seine leichte und schnelle Handhabbarkeit, die Möglichkeit, Einzel- oder auch Gruppentestungen vorzunehmen, sowie zur mündlichen, schriftlichen bzw. computerbasierten Durchführung.

Die Nebengütekriterien der Normierung sowie der Vergleichbarkeit jedoch kann der Wortassoziationsversuch nicht erfüllen. Aufgrund der Spezifität der Stimuli für die jeweilige Wissensdomäne ist es nicht möglich, Normen (z.B. für einen „guten“ oder „schlechten“ Assoziationskennwert) zur Verfügung zu stellen, welche die Interpretation der Ergebnisse erleichtern könnten. Aus dem gleichen Grund kann es auch keine Parallelförmigkeiten eines Wortassoziationsversuchs geben, denn eine Vergleichbarkeit der Items respektive der Stimuli ist praktisch ausgeschlossen. Zusammenfassend ist der Wortassoziationsversuch als ein veränderungssensitives, durchführungsökonomisches Verfahren zu beschreiben, das von den Probanden akzeptiert wird.

Teil E - ZUSAMMENFASSUNG UND DISKUSSION DER ERGEBNISSE

Zum Abschluß werden die wesentlichen empirischen Befunde der vorliegenden Arbeit zusammenfassend dargestellt und vor dem Hintergrund der einschlägigen Literatur und v.a. der praktischen Aspekte diskutiert. Dabei findet die Beurteilung der Gütekriterien des Wortassoziationsversuchs besondere Berücksichtigung, ebenso die Implikationen hinsichtlich der Möglichkeiten und Grenzen seines Einsatzes im arbeitspsychologischen Feld. Abschließend werden weiterhin offene Fragen im Zusammenhang mit dem Wortassoziationsversuch skizziert, mit einem Fazit schließt die vorliegende Arbeit ab.

8. Methodische Aspekte des Wortassoziationsversuchs

Die vorliegende Arbeit setzte sich u.a. mit zwei Aspekten der Durchführung des Wortassoziationsversuchs auseinander, da die einschlägige Literatur dazu keine eindeutigen Befunde liefern konnte. So wurde der Durchführungsmodus thematisiert sowie die Anordnung der Stimuli innerhalb des Wortassoziationsversuchs. Dieses Kapitel stellt zunächst die diesbezüglichen Untersuchungsergebnisse kurz vor, anschließend sollen diese Befunde und weitere methodische Aspekte des Wortassoziationsversuchs diskutiert werden, welche sich weitergehend aus den Erfahrungen verschiedener Versuchsleiter in arbeitspsychologischen Wissensanalysen heraus ergaben.

8.1. Zusammenfassung der empirischen Befunde

Zunächst wurde anhand einer experimentellen Studie der **Durchführungsmodus** des Wortassoziationsversuchs thematisiert (Fragestellung 1), wobei mündliche und schriftliche Durchführung (letztere im Papier-und-Bleistift-Format) im intraindividuellen Vergleich gegenübergestellt wurden. Unter Vernachlässigung derjenigen Stimuli, welche in dieser Studie dem Bereich des Allgemeinwissens zuzuordnen waren, führen das mündliche und das schriftliche Assoziieren zu vergleichbaren Assoziationskennwerten. Somit kann von einer Äquivalenz beider Modi ausgegangen werden.

In einer zweiten Vorstudie wurde die **Anordnung von Stimuli** geprüft hinsichtlich des Einflusses auf die Assoziationsleistungen der Probanden (Fragestellung 2). Dazu wurden Stimuli verschiedener Abstraktionsebenen zusammengestellt, welche in Unter- bzw. Oberbegriffsrelation zueinander standen. Auf diese Weise bearbeiten Probanden die Stimuli zum einen gemäß der Strategie „Vom Besonderen zum Allgemeinen“, zum anderen gemäß der Strategie „Vom Allgemeinen zum Besonderen“. Die empirischen Befunde legen nahe, daß die systematische Anordnung gemäß dieser Strategien keinen Einfluß auf die individuelle Assoziationsleistung eines Probanden hat.

8.2. Diskussion der Befunde

Die Studie zum Durchführungsmodus bediente sich als einzige der hier vorgestellten Untersuchungen nicht einer Kohorte aus dem Arbeitsleben, sondern griff auf eine Stichprobe zurück, welche hinsichtlich des Lebensalters, des intellektuellen bzw. beruflichen Hintergrundes sehr heterogen war. Daher konnte auch keine „echte“ fach- oder berufsspezifische Wissensdomäne herangezogen werden, sondern es wurde neben Stimuli aus dem Bereich des Allgemeinwissens ein Lebensbereich gewählt, welcher so etwas wie Spezialwissen erzeugt, in diesem Falle handelte es sich um Bankgeschäfte. Da wie beschrieben zwar geringfügige Unterschiede hinsichtlich des Assoziationskennwertes im mündlichen vs. schriftlichen Modus gefunden wurden, diese aber durch die Stimuli aus dem Bereich des Allgemeinwissens begründet sind, läßt sich folgendes annehmen: wenn bereits bei den alltagsnahen Stimuli des sog. Spezialwissens sich die Unterschiede hinsichtlich des Durchführungsmodus auflösen, so ist anzunehmen, daß bei zunehmender Spezialisierung und Professionalisierung der Wissensdomäne diese Unterschiede noch stärker zurücktreten und somit von einer Äquivalenz beider Modi ausgegangen werden kann.

Zu bedenken ist bei der Wahl des Durchführungsmodus auch der Aspekt der Testfairneß (siehe auch Kubinger, 2006), v.a. wenn man einen computergestützten Wortassoziationsversuch plant. Dabei kann es je nach Untersuchungskohorte einige Probanden unangemessen benachteiligen, wenn sie aufgrund ihrer Arbeitstätigkeit und sonstigen Lebensumstände kaum Erfahrung im Umgang mit Computern haben, der Wortassoziationsversuch aber in einer computerbasierten Variante eingesetzt wird. Mangelnde Fähigkeiten im Tippen könnten so in diesem Beispiel mit Verunsicherung oder einem Motivationsverlust einhergehen, so daß auf diese Weise der Wortassoziationsversuch den Umfang des individuellen Wissensbestandes unterschätzen würde. Auf der anderen Seite sollte bei der Wahl des Untersuchungsdesigns auch in Erwägung gezogen werden, in welcher Form das Fachwissen in der Untersuchungsstichprobe üblicherweise angewendet wird. So bietet sich beispielsweise für Call Center Agents im Front-Office-Bereich an, den Wortassoziationsversuch mündlich durchzuführen, da deren Arbeitstätigkeiten ebenfalls das mündliche Reproduzieren und Anwenden von Fachwissen erfordern.

Die Planung eines Wortassoziationsversuchs erfordert neben der Entscheidung über ein Durchführungsformat aber zunächst eine Auseinandersetzung hinsichtlich der Inhalte, also der Wissensdomäne und der daraus abgeleiteten Stimuli. Wie anhand der empirischen Befunde beschrieben, ist es für den Assoziationskennwert irrelevant, ob dabei zunächst ein Oberbegriff der Wissensdomäne als Stimulus präsentiert wird und anschließend weitere Unterbegriffe folgen, oder ob derart hierarchische Beziehungen zwischen Stimuli in umgekehrter Reihenfolge präsentiert werden. Aus der Praxis heraus hat es sich jedoch als hilfreich erwiesen, die Stimuli (sofern sie überhaupt in hierarchischer Beziehung zueinander stehen) gemäß der Strategie „Vom Allgemeinen zum Besonderen“ anzuordnen, da die Probanden andernfalls so etwas wie Redundanz erlebten und ihren Unwillen bekundeten, sich zu wiederholen (siehe z.B. Pokrandt, 2004). In der vorgestellten empirischen Studie wurde zudem von den Probanden kritisiert, daß die dargebotenen Stimuli (in diesem Falle: die Bezeichnungen psychischer Störungen) insgesamt zu offen gehalten waren, also einen zu großen Bedeutungshof ausfüllten, so daß bisweilen schnell die Lust am umfassenden Bearbeiten der weiteren Stimuli beeinträchtigt wurde.

Bei der Auswahl von Stimuli sollte insgesamt ein mittleres Schwierigkeitsniveau der Stimuli angestrebt werden (hier ist die Antizipation des Bedeutungshofes eines Stimulus relevant), so daß eine gute Trennschärfe der Items zwischen „guten“ und „schlechten“ Probanden zu erwarten ist. Wenn man Wissensdiagnosen zur Individualdiagnose einsetzt, so ist oftmals Ziel, damit die Stellung einer Person im Vergleich zu anderen abzubilden. Gestaltet man die Items eines Leistungstests so, daß die meisten Probanden alle Aufgaben gut lösen können, so fällt die Leistungsdifferenzierung zwischen den Probanden relativ schwer: es ist nicht erkennbar, ob der hohe Leistungskennwert eines Probanden die Grenze seiner Leistungsfähigkeit abbildet oder ob dieser Proband weitere oder schwerere Aufgaben ebenso gut hätte lösen können. Der Wortassoziationsversuch hat nun gegenüber anderen Leistungstests den Vorteil, einen nach oben offenen Leistungskennwert auszugeben, während z.B. ein Multiple-Choice-Test einen maximal erreichbaren Punktwert festlegt. Dennoch bleibt das Problem der zu geringen Streuung von Leistungskennwerten, wenn Stimuli so gestaltet sind, daß zu erwarten ist, daß sie von allen Probanden gleich gut beantwortet werden. Es ist durchaus denkbar, einen Stimulus mit höherem Schwierigkeitsgrad in eine Reihe mittelschwerer Stimuli zu integrieren. Stimuli mit geringem Schwierigkeitsgrad können so etwas wie das Basiswissen erheben. Als leicht wären Items zu beurteilen, von denen anzunehmen ist, daß alle Mitarbeiter Aussagen dazu machen können aufgrund der Relevanz oder Präsenz dieses Stimulus im Arbeitsalltag. So könnten z.B. in einem Call Center auf den Namen des vertriebenen Produktes hin die gängigen Verkaufsargumente erwartet werden, welche den Mitarbeitern i.d.R. in Produktschulungen nahe gebracht werden. Als höherer Schwierigkeitsgrad wäre hingegen ein Stimulus einzuschätzen, der im Beispiel des Call Centers eine seltene oder spezifische Kundenanfrage repräsentiert, z.B. einen Störungszustand des Produktes beschreibt. In diesem Sinne schwierige Stimuli könnten also besondere Experten in der interessierenden Domäne identifizieren und dadurch mehr Streuung in den individuellen Assoziationskennwerten erzeugen. Allerdings sollten solche Stimuli, bei denen erwartet wird, daß nur einige Spezialisten darauf flüssig assoziieren können, sehr sparsam eingesetzt werden, um das Erfolgserleben der Probanden und damit auch die Motivation für die weitere Teilnahme am Wortassoziationsversuch nicht zu beeinträchtigen. Letztlich aber fehlen weiterhin systematische empirische Belege dafür, welche Merkmale von Stimuli und deren Anordnung das Assoziationsverhalten von Probanden möglicherweise beeinflussen, um daraus Regelmäßigkeiten zur Auswahl von Stimuli und zur Konstruktion eines Wortassoziationsversuchs ableiten zu können.

9. Gütekriterien des Wortassoziationsversuchs

Die vorliegenden empirischen Befunde konnten zur Erhellung der Gütekriterien des Wortassoziationsversuchs beitragen. Dabei erwies sich die Integration von Studien aus verschiedenen Untersuchungsfeldern als vorteilhaft hinsichtlich der Generalisierbarkeit der Ergebnisse (z.B. Bungard, Holling und Schultz-Gambard, 1996; Häcker, Leutner und Amelang, 1998; Rosenthal und DiMatteo, 2001). Im Zentrum der empirischen Auseinandersetzung stand die Bestimmung der Validität des Wortassoziationsversuchs, diese wurde in Hinblick auf ihre verschiedenen Aspekte und unter Heranziehung verschiedener Kriterien über mehrere Studien hinweg betrachtet. Neben den Kriterien der Reliabilität und der Objektivität wurden zudem auch verschiedene Nebengütekriterien z.T. empirisch belegt, z.T. besprochen. Im Folgenden sollen die

wesentlichen Erkenntnisse hinsichtlich der Verfahrensgüte des Wortassoziationsversuchs kurz zusammengefaßt und diskutiert werden.

9.1. Zusammenfassung der empirischen Befunde

Zur Bestimmung der Validität (Fragestellung 3) konnten insgesamt 15 Studien herangezogen werden, die einen empirischen Beitrag dazu leisteten. Dabei stand die Kriteriumsvalidität (mit den Kriterien Klausurleistung sowie Geschäftsführerurteil) im Vordergrund, des weiteren die Konstruktvalidität, welche verschiedene konstruktverwandte Wissenstests heranzog sowie die Abgrenzung zum Konstrukt der allgemeinen Wortflüssigkeit thematisierte. Als Kennwert des Wortassoziationsversuchs diente dabei die mittlere Anzahl von Assoziationen pro Stimulus.

Die **Inhaltsvalidität** des Wortassoziationsversuchs kann aufgrund der intensiven Vorbereitung der Erhebung durch den Versuchsleiter als gegeben angenommen werden. Arbeitsbeobachtungen, Dokumentenanalysen, Experteninterviews oder ähnliche Zugänge ermöglichen es dem Untersucher, die Repräsentativität der gewählten Wissensdomäne sowie der daraus abgeleiteten Stimuli abzusichern.

Zur empirischen Ermittlung der **kriteriumsbezogenen Validität** standen zwei Kategorien von Kriterien zur Verfügung: zum einen die Klausurleistung von Probanden, welche als ein Maß für das Wissen interpretiert werden kann, zum anderen die Beurteilung durch den Geschäftsführer (Fachkompetenz sowie Gesamtbeurteilung), welche als ein Maß für die Arbeitsleistung gelten kann. Da Arbeitsleistung i.S. von Performanz außer durch das fachspezifische Wissen auch noch durch viele andere Faktoren determiniert wird, ist zu erwarten, daß deren Zusammenhänge zur Variable Wissen (anhand des Assoziationskennwertes) weniger stark ausfallen als die Beziehung zwischen Klausur und Wortassoziationsversuch. In den meisten der neun Studien, in denen der Wortassoziationsversuch anhand der Klausurleistung validiert werden konnten, ergaben sich signifikante Korrelationen geringer bis mittlerer Stärke zwischen Klausur und Assoziationskennwert, eine Integration dieser Befunde bestätigt dies. In zwei weiteren Studien konnten Leistungsurteile der Geschäftsführer als Kriterium zur Verfahrensvalidierung herangezogen werden. Hier fielen die korrelativen Zusammenhänge in den einzelnen Studien nicht ganz so eindeutig aus, die Integration der einzelnen Befunde erbringt einen schwachen korrelativen Zusammenhang zwischen den Variablen Geschäftsführerurteil und Assoziationskennwert, welche als kleine bis mittlere Effekte bewertet werden können. Interessanterweise läßt sich die Gesamtbeurteilung des Geschäftsführers etwas besser anhand des Assoziationskennwertes vorhersagen als das separierte Urteil für die Fachkompetenz. Dies ist insofern bemerkenswert, als daß in die Gesamtbeurteilung noch viele andere Aspekte einfließen, die im Grunde nichts mit fachbezogenem Wissen zu tun haben. Insgesamt gesehen zeigen sich zur Vorhersage der Klausurleistung anhand des Assoziationskennwertes erwartungsgemäß größere Effekte als zur Vorhersage der Geschäftsführerbeurteilung.

Die **Konstruktvalidierung** erfolgte sowohl unter dem Aspekt der konvergenten als auch der divergenten Validierung. Als konstruktverwandte Tests standen dabei verschiedene andere wissensdiagnostische Verfahren zur Verfügung: die Strukturlegetechnik, der Assoziations-Struktur-Test, Multiple-Choice-Tests, ein Lückentext sowie ein Tiefeninterview. Die Parameter des

Strukturlegens weisen ebenso wie des Assoziations-Struktur-Tests schwache bis mittlere korrelative Beziehungen zum Assoziationskennwert auf und bezeichnen damit insgesamt recht große Effekte. Dabei ist allerdings zu berücksichtigen, daß diese Verfahren z.T. unmittelbar zusammenhängen: beim Assoziations-Struktur-Test rekrutieren sich die verwendeten Konzepte direkt aus dem vorangegangenen Wortassoziationsversuch eines Probanden. Bei der Strukturlegetechnik ist dies auch möglich, aber in zwei der drei hier herangezogenen Studien flossen die individuellen Assoziationen zumindest in das hypothetische Wissensmodell ein, aus welchem schließlich die im Netz zu legenden Knoten abgeleitet wurden. Hohe Zusammenhänge zwischen den Parametern der Strukturlegetechnik sowie des Assoziations-Struktur-Tests mit dem Kennwert des Wortassoziationsversuchs waren also auch zu erwarten. Für die Leistung in einem Multiple-Choice-Test hingegen erwies sich der Assoziationskennwert als ein schlechter Prädiktor. In den zwei hier vorgestellten Studien dazu konnten nur sehr schwache bis gar keine Korrelationen gefunden werden. Vermutlich ist bei derart geschlossenen Fragen der Einfluß des Ratens Ursache dafür, daß der Multiple-Choice-Test ein verzerrtes Bild vom Wissensstand der Probanden zeichnet. Das Tiefeninterview, welches in einer der empirischen Studien herangezogen werden konnte, weist ebenfalls keine nennenswerte korrelative Beziehung zum Assoziationskennwert auf. Der in einer weiteren Studie eingesetzte Lückentext hingegen erbringt eine korrelative Beziehung mittlerer Stärke zum Assoziationskennwert. Zusammenfassend ist bezüglich der konvergenten Konstruktvalidität des Wortassoziationsversuchs zu vermerken: zu einigen wissensdiagnostischen Verfahren bestehen sehr enge Beziehungen, wobei aber gleichsam eine große konzeptuelle Nähe zwischen den Verfahren besteht, bei anderen Verfahren hingegen lassen sich kaum bedeutsame Beziehungen zum Wortassoziationsversuch ausmachen. Allerdings ist anzumerken, daß alle hier herangezogenen wissensdiagnostischen Instrumente selbst nicht hinreichend abgesichert sind hinsichtlich ihrer Gütekriterien und zudem – wie auch der Wortassoziationsversuch selbst – für jede Wissensdomäne neu konstruiert werden müssen. Über die konvergente Konstruktvalidität hinaus war es aufgrund der Untersuchungsdesigns in einigen der Studien möglich, die inkrementelle Validität zu Außenkriterien zu prüfen. Dabei zeigte sich insbesondere, daß das Tiefeninterview die Geschäftsführerbeurteilung deutlich zuverlässiger vorhersagen konnte als es der Assoziationskennwert vermochte.

Die divergente Konstruktvalidität wurde in drei Studien geprüft, indem neben dem Wortassoziationsversuch ein Test zur Erfassung der allgemeinen Wortflüssigkeit eingesetzt wurde. In zwei dieser Studien konnte kein Zusammenhang zwischen Assoziationskennwert und dem Kennwert für die Wortflüssigkeit nachgewiesen werden. In der dritten Erhebung deutet sich eine solche Beziehung an, allerdings trifft dies bei differenzierterer Betrachtung in erster Linie für die Stimuli aus dem Bereich des Allgemeinwissens zu. Beschränkt man sich hingegen nur auf diejenigen Teile des Assoziationsversuches, die Stimuli aus einem Spezialwissensbereich verwenden, so ist ein solcher Zusammenhang nicht mehr nachweisbar. Insgesamt weisen die empirischen Befunde also darauf hin, daß bei der Methode der freien Assoziation mit Stimuli aus einer fachspezifischen Domäne tatsächlich auch der individuelle Wissensbestand erfaßt wird und nicht die generelle Fähigkeit, sich flüssig zu äußern.

Die **Reliabilität** des Wortassoziationsversuchs konnte anhand einer Studie mit Testwiederholung belegt werden (Fragestellung 4). Danach weisen die Assoziationskennwerte zwischen beiden

Erhebungszeitpunkten einen hohen signifikanten korrelativen Zusammenhang auf. Selbst wenn über diesen quantitativen Kennwert hinaus die Inhalte der Assoziationen betrachtet werden, zeigt sich eine große Übereinstimmung der Assoziationen über beide Erhebungszeitpunkte hinweg: fast die Hälfte der Assoziationsinhalte der ersten Erhebung wurde auch bei der Wiederholungsmessung benannt. Die individuellen Assoziationsleistungen sind also zwischen den beiden Erhebungen vergleichbar, so daß von einer Reliabilität des Wortassoziationsversuchs ausgegangen werden kann.

Die **Objektivität** des Wortassoziationsversuchs kann als gegeben hingenommen werden (Fragestellung 5). Die Durchführungsobjektivität kann gewährleistet werden durch die Verwendung einer festgelegten Instruktion, welche die Interaktion zwischen Proband und Versuchsleiter standardisiert. Bei schriftlicher Durchführung des Wortassoziationsversuchs ist dies noch einfacher zu realisieren. Die computergestützte Durchführung des Versuchs kommt dem ebenso entgegen und wirkt sich zudem vorteilhaft aus auf die Auswerteobjektivität, da die dadurch erhaltene Liste von Assoziationen das Auszählen erleichtert. Die Interpretationsobjektivität kann ebenfalls als gegeben hingenommen und zudem weiter abgesichert werden durch die Nutzung eines zuvor erstellten hypothetischen Wissensmodells.

Im Rahmen der Validierung wurde auch thematisiert, inwieweit der verwendete und leicht zu erhebende Assoziationskennwert (als mittlere Anzahl von Assoziationen pro Stimulus) angemessen zur Beschreibung des individuellen Wissensbesitzes ist oder ob weitergehende **Modifikationen** dieses Kennwertes notwendig sind (Fragestellung 6). Dazu wurden zwei Ansätze herangezogen: zum einen die Gewichtung qualitativ unterschiedlicher Assoziationsinhalte (Fragestellung 6a), zum anderen die Referenz auf die innerhalb der Stichprobe hochfrequenten Assoziationen (Fragestellung 6b). Zusammenfassend ist festzuhalten, daß die auf diese Weise vorgenommenen Modifikationen des Assoziationskennwertes wenig dazu beitragen konnten, die Bewertung der Assoziationsleistung eines Probanden grundlegend zu verändern. Die Verteilung der Kennwerte wurde zwar innerhalb der Stichprobe auseinandergelöst, die Rangreihe der Probanden änderte sich dadurch aber kaum. Auch i.S. einer inkrementellen Validität erbrachten die modifizierten Kennwerte keinen Mehrertrag zur Vorhersage eines Außenkriteriums gegenüber dem einfachen quantitativen Assoziationskennwert. Allerdings sind die vorgenommenen Auswertestrategien jenseits der Modifikation der individuellen Parameter sehr geeignet, um den Wissensbestand der Stichprobe auf einer qualitativen Ebene zu beschreiben, so daß sie unter diesem Aspekt ihre Berechtigung besitzen.

Neben den soeben dargestellten Hauptgütekriterien setzte sich die vorliegende Arbeit auch mit einer Reihe von Nebengütekriterien auseinander. So wurde in drei Studien die **Veränderungssensitivität** des Wortassoziationsversuchs geprüft (Fragestellung 7). Es konnte übereinstimmend gezeigt werden, daß der Wortassoziationsversuch in der Lage ist, Veränderungen des individuellen Wissensbesitzes im Ergebnis einer Lehrveranstaltung abzubilden. Dies betrifft sowohl die quantitativen Kennwerte (also die Anzahl assoziierter Begriffe) als auch qualitative Aspekte. So konnte anhand von Prä-Post-Designs illustriert werden, daß infolge der Lehrveranstaltungen nicht nur mehr assoziiert wurde, sondern zudem auch Inhalte und Begrifflichkeiten aus dem Lehrstoff Eingang in den individuellen Wissensbesitz fanden.

Die **Akzeptanz** des Wortassoziationsversuchs durch die Probanden konnte ebenfalls empirisch belegt werden (Fragestellung 8). Zwar berichten Versuchsleiter immer wieder von vereinzelt Schwierigkeiten, insbesondere unter dem Aspekt von Leistungsdruckgefühlen sowie

Motivationsdefiziten, dennoch zeigen die empirischen Befunde aus Akzeptanz-Befragungen in zwei Studien eine insgesamt positive Einstellung der Probanden gegenüber dem Verfahren. Hinsichtlich des Durchführungsmodus des Wortassoziationsversuchs bevorzugten die Probanden zumeist das schriftliche Format, da sie dabei weniger Zeit- und Leistungsdruck erlebten. Im Vergleich mit einem Lückentext präferierten die Probanden den Wortassoziationsversuch aufgrund seiner großen Freiheitsgrade bei der Wiedergabe des eigenen Wissens. Insgesamt also kann die Akzeptanz des Wortassoziationsversuchs durch die Probanden als gegeben hingenommen werden.

Als für die arbeitspsychologische Praxis sehr bedeutsamer Aspekt wurde auch die **Ökonomie** des Wortassoziationsversuchs empirisch beleuchtet (Fragestellung 9). Dazu wurde in zwei Studien die Durchführungsökonomie des Wortassoziationsversuchs derjenigen eines anderen wissensdiagnostischen Verfahrens gegenüber gestellt. Im Kontrast zur Strukturlegetechnik erweist sich der Wortassoziationsversuch als vergleichsweise ökonomisch: die Probanden benötigen deutlich weniger Zeit für das Assoziieren (und können dabei einen größeren Ausschnitt aus der interessierenden Wissensdomäne bearbeiten), ebenso verhält es sich hinsichtlich der Auswerteökonomie für den Versuchsleiter, auch die materiellen bzw. räumlichen Anforderungen fallen für das Assoziieren deutlich günstiger aus. Gegenüber einem Lückentext jedoch erbringt der Wortassoziationsversuch keine ökonomischen Vorteile, selbst wenn beide Verfahren den identischen Ausschnitt aus der Wissensdomäne thematisieren. Hinsichtlich der verschiedenen Durchführungsvarianten des Wortassoziationsversuchs selbst ist folgendes zu bedenken: ein mündlich durchgeführter Wortassoziationsversuch erfordert für sich genommen weniger Zeit als ein schriftlicher, ist aber nur in der Einzeltestung möglich, ist also für den Versuchsleiter in der Summe zeitaufwendiger. Auf der anderen Seite erfordert das Aufbereiten der Daten eines mündlichen Assoziationsversuches zusätzlichen Aufwand, da diese zunächst transkribiert werden müssen. Insgesamt gesehen aber kann der Wortassoziationsversuch als ein ökonomisches Verfahren charakterisiert werden, lediglich die Auswertung der erhobenen Daten erfordert einigen Aufwand, v.a. wenn man qualitative Aussagen treffen will.

Die **Nützlichkeit** des Wortassoziationsversuchs kann anhand der vorgestellten Studien belegt werden (Fragestellung 10). Dem praktischen Bedürfnis nach einem ökonomischen und sensitiven Instrument zur Diagnose des fachspezifischen Wissens kann der Wortassoziationsversuch entsprechen. So wurde er beispielsweise zur Evaluation von Weiterbildungsveranstaltungen erfolgreich eingesetzt werden oder im Rahmen von Wissensmanagementprozessen zur Bestimmung eines hypothetischen Wissensmodells herangezogen.

Einzig die Nebengütekriterien der **Normierung** (Fragestellung 11) sowie der **Vergleichbarkeit** (Fragestellung 12) kann der Wortassoziationsversuch nicht erfüllen. Dies liegt in erster Linie darin begründet, daß der Wortassoziationsversuch eher ein generelles Vorgehen beschreibt, weniger einen Test mit festgeschriebenen Items im klassischen Sinne. Vielmehr wird für jedes Untersuchungsfeld ein neuer Satz an Items (d.h. Stimuli) zusammengestellt, welche die interessierende Wissensdomäne möglichst repräsentativ abbilden sollen. Von der Auswahl der Stimuli hängt zentral ab, was und wieviel daraufhin assoziiert werden kann, daher kann es keine allgemeingültigen Normen geben, anhand derer die Assoziationsleistung eines Probanden beurteilt werden kann. Statt dessen kann „nur“ die Referenz zur Assoziationsleistung innerhalb der Stichprobe oder zum hypothetischen Wissensmodell herangezogen werden. Aus dem gleichen Grunde können auch keine vergleichbaren

Parallelformen eines Wortassoziationsversuchs erstellt werden: die Spezifität der Stimulusinhalte und deren jeweiliger Bedeutungshof lassen dies nicht zu.

9.2. Diskussion der Befunde

Die Reliabilität des Wortassoziationsversuchs konnte anhand eines Retest-Designs empirisch belegt werden. Damit ist eine wichtige Voraussetzung gegeben, um die Validität des Verfahrens absichern zu können. Die von Strube (1984) im Zusammenhang mit Allgemeinwissen gefundene Tendenz, daß bei Testwiederholung i.S. eines Übungseffekts zunehmend mehr assoziiert wird, konnte für den Bereich des berufsbezogenen Fachwissens nicht bestätigt werden. Eine semantische Übereinstimmung der Assoziationsinhalte über beide Erhebungszeitpunkte hinweg würde man nicht fordern – sie ist nicht zu erwarten, da aufgrund der theoretischen Annahme eines semantischen Netzwerkes von einem Stimulus aus mehrere Kanten mit gleicher Bindungsstärke abzweigen können. Je größer der Bedeutungshof eines Stimulus und je elaborierter das semantische Netzwerk eines Probanden, desto unwahrscheinlicher ist es, daß der Proband bei Testwiederholung wieder genau die identischen Pfade seines Netzwerkes aktiviert. Um so bemerkenswerter ist die hier gefundene Übereinstimmung der Assoziationsinhalte über die beiden Erhebungszeitpunkte hinweg: im Mittel wurde knapp die Hälfte der Assoziationen in beiden Versuchen genannt. Es ist zu erwarten, daß dieser Befund noch deutlicher würde, wenn die vorgelegten Stimuli in noch stärkerem Maße spezialisiert wären als es in der vorgestellten Studie der Fall war. Wirtz und Caspra (2002) sowie Stracke (2004) interpretieren die Retest-Reliabilität als ein Maß für die zeitliche Stabilität des untersuchten Merkmals (hier also des Fachwissens) – dies kann also für den Wortassoziationsversuch sowohl hinsichtlich der Menge als auch der Inhalte der assoziierten Wissenseinheiten als gegeben angenommen werden. Die quantitativen und insbesondere die qualitativen Befunde zur Reliabilität des Wortassoziationsversuchs stützen zudem die Aussagen hinsichtlich der Veränderungssensitivität (Mazur, 2006, charakterisiert die Reliabilität als unabdingbare Voraussetzung für die Sensitivität): nur, wenn das Instrument bei mehrmaligen Erhebungen und ohne, daß Einfluß auf das Merkmal Wissen genommen wurde, zu vergleichbaren Ergebnissen kommt, lassen sich Veränderungen im Prä-Post-Vergleich mit zwischengeschaltetem Treatment sinnvoll interpretieren und auf das Treatment zurückführen anstatt auf zufällige Merkmalschwankungen (z.B. Stracke, 2004).

Die Objektivität des Wortassoziationsversuchs kann aufgrund der technischen Realisierungen als gegeben hingenommen werden (siehe auch Rothe, 2003). Bortz und Döring (2003) halten die Objektivität für ein sehr leicht zu realisierendes Kriterium, bei geschulten Versuchsleitern sowie einer standardisierten Testanweisung lassen sich insbesondere quantitative Parameter (wie in diesem Falle der Assoziationskennwert als Anzahl von Assoziationen) unproblematisch erheben und interpretieren. Die einzige Einschränkung betrifft die Auswertobjektivität: im Idealfall verhält sich der Proband instruktionsgemäß und antwortet nur in einzelnen Worten, in diesem Falle werden die Begriffe einfach ausgezählt, und verschiedene Auswerter sollten dabei zum gleichen Ergebnis kommen. Verhält sich der Proband hingegen nicht gemäß der Instruktion und antwortet eher in ganzen Sätzen (insbesondere bei mündlicher Durchführung scheint dies i.S. einer natürlichen Gesprächssituation nahezuliegen), ist dieses einfache Auszählen einer Liste von Begriffen erschwert. Die Befunde von Hauser, Spada, Rummel und Meier (2006) aber legen (wie bereits an

früherer Stelle beschrieben) nahe, daß auch bei komplexeren freien Antworten mehrere Auswerter durchaus zum gleichen Ergebnis kommen können. Eine a priori festgelegte Auswerteinstruktion unterstützt dabei die Objektivität, wie auch Bühner (2006) empfiehlt.

Im Zusammenhang mit der Qualifikation des Auswerter sei an dieser Stelle nochmals kurz auf die Konstruktion eines Wortassoziationsversuchs eingegangen. Wie beispielsweise von Pokrandt (2004), Ruprecht (2004) oder Liebs (2003) angemerkt, ist die inhaltliche Vorbereitung für den Wortassoziationsversuch relativ aufwendig: der Versuchsleiter muß zunächst durch Interviews und Dokumentenanalysen den Wissensgegenstand identifizieren und eingrenzen, sich anschließend intensiv mit diesem Wissensgegenstand auseinandersetzen. Das Wissen, das der Versuchsleiter selbst über die interessierende Wissensdomäne erwirbt, muß genügen, um annähernd die Relevanz eines möglichen Stimulus für die Domäne beurteilen und in etwa antizipieren zu können, welche Begriffe auf diesen Stimulus möglicherweise genannt würden. Nicht zuletzt in der Phase der Auswertung muß der Versuchsleiter beurteilen können, wie sinnvoll oder gewichtig eine Assoziation ist oder, ob es sich um einen gebräuchlichen Fachbegriff dieser Domäne handelt. Allerdings ist zu sagen, daß das Durchführen von Wissensdiagnosen immer in irgendeiner Weise eine Auseinandersetzung mit der Wissensdomäne voraussetzt, unabhängig vom diagnostischen Instrument. Die Entscheidung für oder gegen ein Diagnoseinstrument kann nicht ohne Kenntnis des Arbeitsgegenstandes gefällt werden. Um diese Kenntnis zu erlangen, können beispielsweise Dokumentenanalysen oder Beobachtungsverfahren eingesetzt werden, welche nach der Unterscheidung von Rothe (1994a) eher der Wissenserfassung als der Wissensdiagnose i.e.S. dienen, also dem Untersucher Grundkenntnisse über die Domäne verschaffen sollen, daher bewertet Rothe (2006) diese Analysetechniken auch als anforderungsfrei gegenüber dem Vorwissen des Versuchsleiters ein. Hinsichtlich der Diagnoseinstrumente i.e.S. stellt beispielsweise die Konstruktion eines Multiple-Choice-Test mit seinen geschlossenen Fragen die Anforderung, sinnvolle Antwortalternativen zu generieren, welche vom Versuchsleiter eindeutig formuliert und auch hinsichtlich ihres Schwierigkeitsgrades abgeschätzt werden müssen. Dies kann nicht ohne eingehende Kenntnisse über die Wissensdomäne oder zumindest intensive Konsultation von entsprechenden Experten geschehen. Beim Wortassoziationsversuch hingegen als ein Verfahren mit offenen Antworten präsentiert der Versuchsleiter gegenüber den Probanden zunächst nur die Eingrenzung des Wissensgegenstandes. Konkrete Erwartungen, welche Inhalte auf die Stimuli assoziiert werden könnten, sollte der Versuchsleiter a priori haben, muß sie aber gegenüber den Probanden nicht offenlegen. Vielmehr können diese in das hypothetische Wissensmodell einfließen und so die Auswertobjektivität unterstützen. Zusammenfassend ist also festzustellen: die intensive Auseinandersetzung mit der Wissensdomäne im Vorfeld der Untersuchungen bzw. bei der Konstruktion eines Wortassoziationsversuchs trägt zur Qualifizierung des oder der Versuchsleiter bei, welche schließlich in der Auswertungsphase ermöglicht, mehrere Auswerter zu vergleichbaren Ergebnissen kommen zu lassen.

Die Hauptproblematik für die Interpretationsobjektivität des Wortassoziationsversuchs besteht in der nach oben offenen Menge an möglichen Assoziationen, welche die Festlegung eines Solls erschwert. Die Inhaltsspezifität der jeweiligen Stimuli macht die Normierung des Wortassoziationsversuchs und damit auch die Vergleichbarkeit einzelner Wortassoziationsversuche miteinander unmöglich. Es läßt sich also anhand eines einzelnen individuellen

Assoziationskennwertes keine Aussage darüber treffen, wie „gut“ oder „schlecht“ ein Proband hinsichtlich seines Wissensbesitzes zu bewerten ist. Selbst für das Leistungsniveau einer gesamten Stichprobe ist eine solche Einschätzung nicht möglich. Statt dessen bieten sich zur Interpretation der Assoziationskennwerte zwei Strategien an: zum einen der Intragruppenvergleich, der v.a. auf der quantitativen Ebene die Unterscheidung in grobe Leistungsgruppen ermöglicht. Letztlich zielen alle in der einschlägigen Literatur zum Wortassoziationsversuch beschriebenen Auswertevarianten entweder auf den Vergleich einzelner Probanden innerhalb der Stichprobe oder aber gänzlich auf Einzelfallanalysen (siehe Hiltmann und Luhr, 1982; Strube, 1984; Müller, 1998; Rothe, 2006). Zum anderen ermöglicht ein auf Grundlage von Experteninterviews und / oder Literaturrecherchen erstelltes hypothetisches Wissensmodell eine stärker qualitativ orientierte Interpretation: auf Basis des Modells könnten die Assoziationen der Probanden auf das Vorhandensein der als relevant erachteten Begriffe und Inhalte bewertet und auch quantifiziert werden, wie es beispielsweise bei der Auswertung der Tiefeninterviews in der Studie von Rothe und Hinnerichs (2005) umgesetzt wurde. Beide Ansätze bieten aber Bezugspunkte für die Interpretation der individuellen Assoziationskennwerte. Letztlich sind also Einschränkungen für die Interpretierbarkeit des Assoziationskennwertes zu bemerken, aber bei der Interpretation individueller Kennwerte im Intragruppenvergleich bzw. in Relation zu einem hypothetischen Wissensmodell ist anzunehmen, daß verschiedene Untersucher dennoch zu vergleichbaren Aussagen gelangen.

Die aufgezeigten empirischen Befunde der vorliegenden Arbeit sprechen insgesamt für eine Validität des Wortassoziationsversuchs. Zu bedenken ist dabei, daß die dargestellten korrelativen Zusammenhänge zwischen Assoziationskennwert und diversen Kriterien nur sinnvoll zu interpretieren sind, wenn man eine lineare Beziehung zwischen Assoziationsleistung und Wissen bzw. Leistung annimmt. Es gibt aber berechtigte Hinweise darauf, daß gerade im hohen Expertisebereich die gezeigte Arbeitsleistung nicht immer konform mit dem Wissensstand geht. So bewertet beispielsweise Hacker (2005b) die Beziehung zwischen Wissen und Leistung als nicht eindeutig, da ein umfassendes Wissen nicht zwangsläufig zu einer höheren Leistung führt, er verweist dabei v.a. auf das handlungsleitende vs. das träge Wissen (Hacker, 1999, 2005b). Auch Rothe (2006) sieht noch weiteren Aufklärungsbedarf hinsichtlich des Zusammenhangs zwischen fachspezifischem Wissen und Arbeitsleistung. Dennoch: folgt die Beziehung zwischen Wissen und Leistung tatsächlich einem kurvilinearem Zusammenhang, so sind die hier gefundenen Belege unter der Annahme eines linearen Zusammenhangs erstaunlich deutlich ausgefallen. Will man den Wortassoziationsversuch als ein Screening-Instrument verstehen, so ist die Annahme eines linearen Zusammenhangs hilfreicher und zielführender, denn: unterstellt man beispielsweise einen umgekehrt U-förmigen Zusammenhang, so könnte das Instrument nicht unterscheiden, ob ein niedriger Assoziationskennwert einen Probanden am unteren oder aber am oberen Ende der Verteilung beschreibt – zusätzliche Erhebungen müßten zur Klärung herangezogen werden.

Hinsichtlich der konvergenten Validität sind die herangezogenen alternativen wissensdiagnostischen Verfahren zumindest kritisch zu betrachten. Alles in allem ist ein unterschiedlicher Aufwand für Vorbereitung, Durchführung und Auswertung der eingesetzten wissensdiagnostischen Instrumente anzumerken, der nicht mit Prädiktorqualität in Hinblick auf die Beurteilung der Arbeitsleistung einherzugehen scheint. Insbesondere der Multiple-Choice-Test als quantitatives Maß für den Wissensbestand erweist sich als problematisch, da die Möglichkeit von

Mehrfachantworten auch immer die Gefahr des Ratens birgt, Seiwald (2003) sieht darin einen gravierenden Nachteil für die Validität von Leistungstests. Auffällig waren die Befunde hinsichtlich des Tiefeninterviews: es fanden sich keine Zusammenhänge zwischen dieser Methode und dem Wortassoziationsversuch. Während i.S. der inkrementellen Validität der Kennwert des Wortassoziationsversuchs kaum in der Lage war, das Außenkriterium Geschäftsführerurteil vorherzusagen, ermöglichten die Ergebnisse des Tiefeninterviews dies hingegen sehr gut (beschreibbar durch eine hohe signifikante Korrelation). Dies ist insofern nachvollziehbar, als daß ein Wortassoziationsversuch den Probanden eben sehr frei antworten läßt, sowohl hinsichtlich der assoziierten Inhalte als auch mit Blick auf den Umfang der Assoziationskette. Beim Tiefeninterview hingegen, so wie es in der vorliegenden Studie eingesetzt wurde, wird der Proband in deutlich stärkerem Maße gelenkt durch die Fragen des Interviewers. Damit kann sich das Interview deutlich stärker am Soll-Wissen orientieren. Hinzu kommt, daß dieses Soll-Wissen ebenso die Vorstellungen und Ansprüche des Geschäftsführers aufgreift wie es bei der hier herangezogenen Leistungsbeurteilung der Fall ist – beide Variablen entspringen also in gewisser Weise der gleichen Quelle, nämlich der Person des Geschäftsführers. Unter diesem Aspekt ist also die hohe Korrelation zwischen beiden Variablen (Tiefeninterview und Geschäftsführerurteil) nachvollziehbar.

Im Zuge der Validierung mit verschiedenen Außenkriterien wurden die Zusammenhänge zwischen den Variablen anhand von Extremgruppenvergleichen illustriert. Dabei wurde anschaulich, daß die extrem guten Assoziierer i.d.R. auch eine hohe Ausprägung im jeweiligen Kriterium aufwiesen und sich dabei als recht homogene Gruppe zeigten. In der Interpretation bedeutet dies: jemand, der viel assoziiert hat, weiß viel über die Domäne, und zeigt folglich auch eine gute Leistung im Kriterium. Man kann nur dann viele Assoziationen produzieren, wenn ein entsprechend umfangreicher Gedächtnisbesitz vorhanden ist. Der Umkehrschluß ist aber offenbar nicht ohne weiteres möglich, denn nicht immer zeigten Probanden mit niedrigem Assoziationskennwert auch einen niedrigen Wert im Kriterium. Wenn ein Proband wenig assoziiert hat, kann ein geringer Umfang des jeweiligen fachspezifischen Wissens ursächlich dafür sein. Es wäre aber auch denkbar, daß in diesem Falle andere Erklärungsmodelle herangezogen werden müssen, beispielsweise mangelnde Kooperation des Probanden bei der Erhebung. Es sei an dieser Stelle auf Rothe (1994a) und Mazur (2006) verwiesen, wonach der Rückschluß vom Nicht-Reden auf ein Nicht-Wissen unangemessen sei. Insbesondere die motivationale Komponente bedarf einer näheren Betrachtung. Meyer (2007) hat beispielsweise in seiner Verfahrensvalidierung des Assoziations-Struktur-Tests nachgewiesen, daß Probanden signifikant schwächere (numerische) Zusammenhänge zwischen Begriffspaaren angeben, wenn sie diese benennen müssen (diese Aufgabe entfällt allerdings, wenn der Zusammenhang mit Null benannt wird) als Probanden, welche die Art der Zusammenhänge nicht verbal beschreiben müssen. Der von der Versuchsperson wahrgenommene Aufwand als Konsequenz einer Entscheidung beeinflusst also ihr Entscheidungsverhalten. Ähnliches kann auch für den Wortassoziationsversuch gemutmaßt werden: Probanden assoziieren so lange, wie es Spaß macht oder zumindest keine Anstrengung verursacht. Es ist zu vermuten, daß das Ausmaß der Phase ohne Anstrengungsgefühle in entscheidendem Maße dadurch determiniert wird, wie leicht die Assoziationen aus dem individuellen semantischen Netz abrufbar sind, was wiederum davon abhängt, wie elaboriert der Wissensbesitz ist und wie gebahnt die Verbindungen zwischen den Knoten sind. Es ist in weiteren (experimentell angelegten) Studien zu prüfen, inwieweit der Wortassoziationsversuch dieser Problematik begegnen kann.

Im Rahmen der Untersuchungen zur Akzeptanz des Wortassoziationsversuchs sowie auch in anderen Studien benannten Probanden immer wieder das Gefühl, aufgrund der Testsituation so etwas wie Leistungsdruck zu erleben und sich deshalb unwohl zu fühlen. Dazu ist zu sagen, daß eigentlich immer, wenn es um die Erhebung von Wissen geht, so etwas wie eine Prüfungssituation geschaffen wird, egal, mit welcher Methode das Wissen erhoben werden soll (es sei denn, der Proband bleibt völlig im Unklaren über die Zielstellung der Methode). Im Kontext von Unternehmen wird dieser Effekt noch verstärkt, da Beschäftigte einer solchen Wissensanalyse, die von der Geschäftsführung gebilligt wurde, in aller Regel Skepsis entgegen bringen und oftmals die Beurteilung der eigenen Leistung und entsprechende Konsequenzen fürchten, Bungard, Holling und Schultz-Gambard (1996) beschreiben dies als Auftraggebereffekt. Eine geringfügige Abmilderung dieser Situation wäre möglicherweise dadurch zu erreichen, indem Begriffe wie „Wissenstest“ gegenüber den Probanden vermieden werden und statt der Quantität (wieviel weiß man) eher die Qualität (was weiß man) der Assoziationen betont würde. Allerdings gibt es bisher in der einschlägigen Literatur keine Aussagen dazu, inwieweit diese Verlagerung der Instruktion die Assoziationsleistung beeinflusst und ob der quantitative Assoziationskennwert dann überhaupt noch angemessen ist.

Insgesamt bleibt festzuhalten, daß der Wortassoziationsversuch ein ausgeglichenes Kosten-Nutzen-Verhältnis vorweisen kann. Die Kosten bestehen für den Versuchsleiter zunächst in einer intensiven Erarbeitung der Wissensdomäne, um den Wortassoziationsversuch konstruieren zu können (wie bereits beschrieben ist dies letztlich allen Verfahren zur Wissensdiagnose eigen), sowie in einer u.U. zeitintensiveren Auswertephase, insbesondere wenn man an einer qualitativen Analyse interessiert ist. In der Durchführung selbst ist der Wortassoziationsversuch mit seiner Option zur Gruppentestung, der einfachen Handhabung, der materiellen Ausstattung und zeitlichen Inanspruchnahme der Probanden äußerst ökonomisch (z.B. gemäß der Definition nach Lienert und Raatz, 1998). Gerade der Aspekt der Durchführungsökonomie dürfte in der arbeitspsychologischen Feldforschung ein gewichtiges Argument sein bei der Auswahl eines Diagnoseinstrumentes, sofern nicht inhaltliche Gründe schwerer wiegen (z.B. wenn in erster Linie prozedurales Wissen erhoben werden soll). Aufgrund der hohen Testökonomie können Abstriche bei den anderen Gütekriterien in Kauf genommen werden (Woike, 2003), obwohl diese im Allgemeinen recht erfreulich ausgeprägt sind, selbst die Prüfung der Validität des Wortassoziationsversuchs erbringt zumindest befriedigende Befunde. Um den Nutzen des Wortassoziationsversuchs besser beurteilen zu können, sollen im nachfolgenden Kapitel die Koeffizienten der Gütekriterien einer differenzierten Bewertung unterzogen werden.

9.2.1. Anforderungen an die Testgüte

Wie bereits beschrieben, konnte die Reliabilität des Verfahrens im Rahmen eines Retest-Desings nachgewiesen werden, der sehr signifikante Korrelationskoeffizient von $r = .849$ ($p = .000$) illustriert den Zusammenhang zwischen den Assoziationskennwerten beider Erhebungszeitpunkte. Nach Bortz und Döring (2003) sollte ein Test, der nicht nur explorativen Zwecken dient, eine Reliabilität von über $r = .80$ aufweisen, wobei ein solcher Wert dann als immer noch mittelmäßig interpretiert werden würde. Dies erfüllt der empirisch ermittelte Reliabilitätskoeffizient des Wortassoziationsversuchs, da dem Verfahren jedoch i.S. eines Screening-Verfahrens eher ein explorativer Charakter zuzuschreiben ist, kann die gefundene Reliabilität als vergleichsweise hoch

bewertet werden. Lienert und Raatz (1998) fordern für Tests, die Gruppendifferenzen beurteilen sollen, mindestens eine Reliabilität von $r \geq .5$, für die Beurteilung individueller Differenzen sehen sie $r \geq .7$ als gerade noch ausreichend an. Auch diesen Anforderungen genügt die empirisch ermittelte Reliabilität des Wortassoziationsversuchs. Nach Lienert und Raatz (1998) spricht ein hoher Retest-Koeffizient auch für eine hohe Merkmalskonstanz, in diesem Falle also des fachbezogenen Wissens.

Lienert und Raatz (1998) binden die zu fordernde Höhe des Validitätskoeffizienten an den Verwendungszweck eines Tests, machen aber keine Aussagen darüber, welche Höhe des Koeffizienten sie bei Screening-Verfahren für angemessen halten. Vielmehr stehe die praktische Bedeutsamkeit des Verfahrens im Vordergrund, die allerdings bei einer Validität von $r \leq .3$ auszuschließen sei (Lienert und Raatz, 1998). Danach sollte die Validität eines Tests mindestens ein $r \geq .7$ aufweisen, wenn er zur individuellen Begutachtung von Personen eingesetzt wird; sofern für die individuelle Eignungsbeurteilung noch andere Informationen herangezogen werden, könne – je nach Gewichtung – auch schon ein $r \geq .5$ ausreichend sein (Lienert und Raatz, 1998). Bortz und Döring (2003) interpretieren einen Validitätskoeffizienten zwischen $r = .4$ und $r = .6$ als mittelmäßig. Diesen Forderungen wird der Wortassoziationsversuch nur mühsam gerecht, die Befundintegrationen zu den Validitätskriterien Klausurleistung ($r = .436$; $p = .000$ für die Klausurpunkte) sowie Geschäftsführerurteil ($r = -.349$; $p = .000$ für die Gesamtbeurteilung) erbringen nur schwache bis mittlere signifikante Korrelationskoeffizienten. Da sich der Wortassoziationsversuch aber eher als ein Screening-Instrument versteht, sollten aus den individuellen Assoziationsleistungen auch keine weitreichenden Konsequenzen für die Probanden abgeleitet werden (wie es z.B. bei Eignungsuntersuchungen der Fall wäre). Unabhängig von numerischen Richtwerten zur Höhe des Validitätskoeffizienten machen Lienert und Raatz (1998) aber auch darauf aufmerksam, daß ein gefundener Validitätskoeffizient auch danach zu beurteilen ist, ob das durch den Test geprüfte Merkmal ohne die Verwendung von Testverfahren leicht oder schwer zu erfassen ist. Unter diesem Aspekt können die hier dargestellten Befunde zur Validität des Wortassoziationsversuchs anhand der verschiedenen Kriterien als hinreichend angesehen werden, denn die Erfassung des Konstrukts des fachbezogenen Wissens ist (wie bereits in Kapitel 2 beschrieben) nicht ohne Schwierigkeiten.

Lienert und Raatz (1998) fordern außerdem, daß die Validität eines Testes umgekehrt proportional zu seiner Ökonomie sein soll. Auch Bortz und Döring (2003) betonen das angemessene Verhältnis von Informationsgewinn und dem mit dem Test verbundenen Aufwand – der Einsatz eines Tests ist demnach immer dann zu rechtfertigen, wenn die Entscheidungen, die auf Grundlage des eingesetzten Verfahrens getroffen werden, tauglicher sind, als es ohne den Test möglich wäre. Der Wortassoziationsversuch kann eine hohe Ökonomie für sich in Anspruch nehmen. Dies wird v.a. in Kontrast zu denjenigen Verfahren deutlich, welche sich intensiver mit der Wissensstruktur auseinandersetzen (Strukturlegetechnik, Assoziations-Struktur-Test). So kann im Falle des Wortassoziationsversuchs eine „nur“ ausreichende Validität durch eine hohe Ökonomie des Verfahrens kompensiert werden.

Neben den gesicherten Erkenntnissen von Lienert und Raatz (1998) soll eine weitere Referenz zur Beurteilung der Verfahrensgüte herangezogen werden: die DIN EN ISO 10075-3 (DIN, 2004), welche die psychischen Belastungen und Beanspruchungen am Arbeitsplatz thematisiert, formuliert

für entsprechende Erhebungsinstrumente Standards sowie methodische Anforderungen. Da dies die einzige bekannte Quelle für die Festlegung methodischer Standards ist, soll sie hier zumindest Erwähnung finden und als Orientierung dienen, auch wenn der in der Norm behandelte Gegenstandsbereich nicht auf Wissensdiagnosen ausgelegt ist. Die Berechtigung solcher Festlegungen ist ohnehin heftig umstritten: Oesterreich (2005) weist darauf hin, daß solche festgeschriebenen Grenzwerte, wie sie die Norm proklamiert (DIN, 2004), in keiner einschlägigen Veröffentlichung zu finden sind und damit nicht dem wissenschaftlichen Standard entsprechen, v.a. da diese Grenzwerte die Art und Methodik des Erhebungsinstrumentes völlig unberücksichtigt lassen. Im Zentrum der Norm steht die Klassifikation von drei Präzisionsstufen (DIN, 2004): die größte Stufe der Analyse bildet die sog. orientierende Messung, bei der zumindest eine Identifikation von negativen Effekten (von Arbeitsbelastungen) möglich ist. Etwas präziser erfaßt die Stufe des Screenings Ausprägungen des Untersuchungsgegenstandes und sollte gemäß der Autoren der Norm (DIN, 2004) dazu in der Lage sein anzuzeigen, an welcher Stelle weitere Erhebungen notwendig sind. Den höchsten Präzisionsgrad repräsentiert die Ebene der akkuraten Messung, welche einen differenzierten Erkenntnisgewinn erbringt und unmittelbar die Ableitung von Interventionen ermöglicht. Anhand der in der Norm festgesetzten quantitativen Mindestwerte kann sich der Wortassoziationsversuch gut behaupten: die für Screening-Instrumente geforderte Reliabilität von $r \geq .8$ erfüllt er ebenso wie die geforderte Validität von $r \geq .4$ (siehe Befundintegration zum Kriterium der Klausurleistung). Hinsichtlich der Sensitivität, welche in der DIN (2004) verstanden wird als der Grad, in dem das Verfahren zwischen verschiedenen Abstufungen des untersuchten Konstruktes unterscheiden kann, sind mindestens drei verschiedene Abstufungen gefordert. Bei der Vorstellung der hier herangezogenen empirischen Studien konnte gezeigt werden, daß der Wortassoziationsversuch zumindest in der Lage ist, zwischen extremen Leistungsgruppen (hinsichtlich der Assoziationsleistung) zu differenzieren, somit würde er zumindest die Anforderungen an orientierende Verfahren erfüllen.

Alles in allem sind die empirisch ermittelten Gütekriterien (Reliabilität und Validität) des Wortassoziationsversuchs als durchaus akzeptabel zu charakterisieren. Selbst die Referenz auf die festgelegten Mindestkoeffizienten in der DIN-Norm (2004), so umstritten sie auch sein mögen (und zudem auf einen anderen Inhaltsbereich zielen als auf Wissensdiagnosen), gereicht dem Assoziationsversuch zu einer zufriedenstellenden Bilanz hinsichtlich seiner Verfahrensgüte.

9.2.2. Was kann der Wortassoziationsversuch leisten, was kann er nicht leisten?

Hinsichtlich der verschiedenen Klassen von Wissen lassen sich für den Wortassoziationsversuch folgende Ableitungen treffen: die Erfassung *impliziten* Wissens ist mittels der Assoziationstechnik grundsätzlich nicht möglich, da das Assoziieren und dessen Verbalisierung bereits eine Explikation möglicherweise impliziter Wissensbestände bedeuten würde (siehe z.B. Anderson, 2001). Vielmehr zielt der Wortassoziationsversuch ausdrücklich auf die Erfassung des *expliziten* Wissens ab. Da das explizite Wissen sowohl deklarative als auch prozedurale Anteile umfassen kann (Büssing, Herbig und Latzel, 2004), ist auch der Wortassoziationsversuch in Grenzen in der Lage, *prozedurales* Wissen zu erfassen. Allerdings ist dies nur zu erwarten, wenn die betreffenden Arbeitstätigkeiten noch nicht zu stark vom einzelnen Beschäftigten prozeduralisiert worden sind und reflektiert werden können (z.B. bei Berufseinsteigern) bzw. wenn die Arbeitsaufgabe ein Reflektieren oder Verbalisieren der Arbeitshandlungen erfordert (z.B. im technischen Support). Der Fokus des

Wortassoziationsversuchs jedoch liegt eindeutig auf der Erfassung *deklarativer* Wissensinhalte, da diese umfassend verbalisierbar sind (Hacker, 2005b). Daß der Wortassoziationsversuch in der Lage ist, im Bereich des Allgemeinwissens individuelle Wissensbestände zu diagnostizieren, hat z.B. Strube (1984) ausführlich dargestellt. Die Angemessenheit dieses Instrumentes zur Erfassung von *berufsbezogenem* Wissen konnte in zahlreichen Studien (z.B. Rothe, Timpe und Warning, 1991; Rothe, 1994a; Rothe und Schindler, 1996; Rothe und Timpe, 1997, Rothe und Hinnerichs, 2005) und nicht zuletzt durch die vorliegende Arbeit ebenso belegt werden. Folglich kann der Wortassoziationsversuch für ein Untersuchungsdesign in Betracht gezogen werden, wenn das Ziel der Untersuchung die Erhebung expliziten, deklarativen Fachwissen ist.

Der Wortassoziationsversuch kann als Screening-Instrument verstanden werden. Damit einher geht die Tatsache, daß sich die Probanden einer Untersuchungsstichprobe eher grob durch die Parameter des Verfahrens charakterisieren lassen, eine umfassende und eindeutige Beschreibung des individuellen Wissensbestandes ist nicht möglich. Dabei betrifft die Unsicherheit v.a. die schon mehrmals angesprochene Beziehung zwischen dem Nicht-Verbalisieren und dem Nicht-Wissen: bedeutet der Umstand, daß der Proband ein für die Domäne zentrales Wissensselement nicht benannt hat, tatsächlich, daß er es nicht weiß? Daß der Proband von ihm benannte Assoziationen auch in einen sinnvollen Zusammenhang bringen und anwenden kann, bleibt zunächst auch nur eine Unterstellung, könnte aber mit weiterführenden Methoden aufgeklärt werden (z.B. Strukturlegetechnik, Analogiebildung), nach Kubinger (2006) steht dies auch in Einklang mit der Funktion eines Screening-Verfahrens. Die günstig ausfallende Ökonomie des Wortassoziationsversuchs ermöglicht es, ggf. größere Probandengruppen zu testen. Nach Woike (2003) sowie Amelang und Schmidt-Atzert (2006) stellt dies das zentrale Kriterium für den Praxiseinsatz eines Screening-Verfahrens dar. Auch nach dem Verständnis der DIN (2004) entsprechen die empirisch ermittelten Gütekriterien des Wortassoziationsversuchs den Anforderungen eines Screening-Instrumentes.

Neben dieser konzeptuellen Bewertung des Wortassoziationsversuchs sollen nun verschiedene arbeitspsychologische Anwendungsbereiche von Wissensdiagnosen im Zusammenhang mit dem Assoziieren diskutiert werden. Verschiedene Autoren (z.B. Reinmann-Rothmeier und Mandl, 2000; Stracke, 2004) halten die Aufdeckung von individuellen Wissenslücken im Rahmen von Wissensmanagementprozessen für zentral. Diesem Anspruch kann der Wortassoziationsversuch nur bedingt gerecht werden. Als Haupthindernis dabei erweist sich das freie Antwortformat, welches den Probanden kaum eingrenzt. Ein Rückschluß von einem Nicht-Verbalisieren auf eine Wissenslücke ist daher unzulässig (siehe auch Mazur, 2006). Der Wortassoziationsversuch kann also das Wissen erfassen, das vorhanden ist, nicht aber jenes, welches nicht im individuellen Wissensbesitz verankert ist. Darüber hinaus ist wie bereits besprochen eine Beurteilung der individuellen Assoziationsinhalte in Richtig oder Falsch nicht möglich aufgrund der fehlenden Kenntnis über die relationalen Beziehungen zwischen den assoziierten Konzepten. Eine Aufdeckung von Fehlkonzepten ist also allein anhand des Wortassoziationsversuchs nicht möglich, sondern kann erst durch weiterführende Analysen wie z.B. die Strukturlegetechnik (gemäß Rothe, 2003, 2006) geleistet werden. Trotz dieser Einschränkungen erweist sich der Wortassoziationsversuch als durchaus brauchbares, wenn auch nicht alleiniges Instrument, um Maßnahmen zur Sicherung des Unternehmenserfolges zu planen. Die Studie von Rothe und

Hinnerichs (2005) beispielsweise illustriert eindrucksvoll, wie der Wortassoziationsversuch als Teil einer Methodenbatterie zur Abbildung des relevanten arbeitsbezogenen Wissens und schließlich zur Ableitung geeigneter Maßnahmen zum individuellen und organisationalen Wissensmanagement herangezogen werden konnte.

Evaluationsmaßnahmen dienen der Beurteilung, ob eine durchgeführte Interventionsmaßnahme den gewünschten Erfolg gebracht hat, zumeist wird dies in Zusammenhang mit Weiterbildungsmaßnahmen geprüft. Nach Bergmann (1999) kann die Wirkungsweise einer Intervention anhand der Veränderung gegenüber dem Ausgangsniveau beschrieben werden (idealerweise sollten zudem die Zielgrößen bestimmt werden, so daß auch der Grad der Zielerreichung bestimmt werden kann). Da ein solcher Wirkungsnachweis ein Prä-Post-Design erfordert, stellt sich für das eingesetzte Diagnoseinstrument die Forderung nach Reliabilität und Veränderungssensitivität, um die durch die Maßnahmen verursachten Veränderungen auch tatsächlich abbilden zu können. Diesen Ansprüchen kann der Wortassoziationsversuch gerecht werden, beide Forderungen erfüllt er. Der große Gewinn, den der Wortassoziationsversuch beispielsweise gegenüber traditionellen Wissenstests im Multiple-Choice-Format erbringt, ist über die Darstellung des Wissensumfangs hinaus auch die Abbildung sich veränderter Wissensinhalte. Gleiches gilt für die Ermittlung von Ausbildungsinhalten.

Im Zusammenhang mit Personalentscheidungen scheint der Wortassoziationsversuch nur bedingt einsetzbar. Lienert und Raatz (1998) fordern, daß der Validitätskoeffizient um so höher sein sollte, je mehr Konsequenzen für die einzelne Person davon abhängen. Da Personalentscheidungen immer konsequenzenreich für den Betroffenen sind, sind die in der vorliegenden Arbeit berichteten Validitätskoeffizienten des Wortassoziationsversuchs wohl als unzureichend zu beurteilen. Zwar sind Wissenstests an sich gemäß der Metaanalysen von Schmidt und Hunter (1998, 2000) grundsätzlich durchaus geeignet, um künftig berufliche Leistung prognostizieren zu können, aber auch in diesem Anwendungsfeld erweist sich das offene Antwortformat als problembehaftet. Vermutlich werden die Assoziationen eines Kandidaten in einer für ihn angespannten Situation weniger frei, sondern vielmehr strategisch selektiert sein. Es ist denkbar, daß soziale Erwünschtheit oder Bewertungsangst eher diejenigen Assoziationen bzw. deren Nennung zulassen, auf die der Kandidat im Bedarfsfall auch einen tiefergehenden Diskurs führen könnte (wie z.B. im Falle von mündlichen Prüfungen oftmals anzutreffen). Diese Umstände beeinflussen vermutlich sowohl Menge als auch Qualität der Assoziationen. Eine mögliche Lösung wäre, den Wortassoziationsversuch in eine umfangreiche Methodenbatterie einzubetten (z.B. im Rahmen von Assessment Centern) und diesem Instrument so ein wenig an Gewicht zu nehmen. Eine schriftliche Durchführung wäre vermutlich ebenso hilfreich, da der Kandidat in diesem Fall weniger erwarten muß, daß seine Assoziationen „Aufhänger“ für ein nachfolgendes Gespräch werden könnten. Nichts desto trotz kann der Wortassoziationsversuch auch im Rahmen von Personalentscheidungen interessante Beiträge liefern, z.B. wenn ein Kandidat sehr spezifische oder aktuelle Begriffe assoziiert. Allerdings ist in diesem Rahmen lediglich eine qualitative Betrachtung der Assoziationsergebnisse (i.S. von Einzelfallanalysen) angemessen, ein Intragruppenvergleich der Assoziationsquantität schließt sich aufgrund der o.g. Einwände aus.

10. Ausblick und offene Fragen

Für das letztendliche Verständnis des Wortassoziationsversuchs als wissensdiagnostisches Instrument für den Einsatz in arbeitspsychologischen Kontexten bleiben nach wie vor einige Fragen offen, die v.a. Determinanten der Versuchsdurchführung sowie das Verhalten der Probanden betreffen. So ist beispielsweise hinsichtlich der Konstruktion eines Wortassoziationsversuchs unklar, in welchem Verhältnis die **Länge der Stimulusliste** zur Repräsentativität des Wissensbesitzes steht. Läßt sich eine generelle Aussage darüber treffen, wie viele Stimuli mindestens notwendig sind, um eine sichere Aussage über den Leistungsstand eines Probanden zu treffen? Oder hängt die Anzahl dargebotener Stimuli einzig von der jeweiligen Domäne ab und der Zentralität oder Repräsentativität der Stimuli? In jedem Falle scheint eine intensive Auseinandersetzung des Versuchsleiters mit der Wissensdomäne unerlässlich, um abschätzen zu können, wie gut die gewählten Stimuli die Domäne abdecken, durch das Hinzuziehen eines betrieblichen Experten wird die Sicherheit dieser Auswahl weiter gestärkt. Dennoch bleibt die Tatsache bestehen, daß durch einen Stimulus immer nur eine Stichprobe aus dem individuellen Wissensbesitz gezogen werden kann, der Anspruch auf Vollständigkeit bei Abbildung des Wissens kann ohnehin nie erfüllt werden. Je mehr Teilstichproben aus der Gesamtmenge des individuellen Wissens gezogen werden (also je mehr Stimuli bearbeitet werden), desto wahrscheinlicher nähert man sich – theoretisch – dem „wahren Wert“ des individuellen Wissensbesitzes an. Auf der anderen Seite ist jedoch zu bedenken, daß mit der Länge der Stimulusliste auch der zeitliche, kognitive und v.a. auch volitionale Aufwand für den Probanden steigt. Es gilt also, das richtige Maß zu halten zwischen einer minimal hinreichenden und maximal notwendigen Anzahl von Stimuli – leider gibt es über diese beiden Punkte noch keine Erkenntnisse.

Es sollten weitere, u.U. experimentelle Untersuchungen herangezogen werden, um die Frage der **Instruktion** näher zu beleuchten. Bei Felduntersuchungen konnte beobachtet werden, daß die Verwendung der Phrase „Wissen erfassen“ einige Probanden verschreckte und sofort Leistungsdruck aufbaute. Auf der anderen Seite verleiten Formulierungen wie „alles nennen, was Ihnen dazu einfällt“ offenbar auch dazu, sehr frei zu assoziieren und sich deutlich aus dem Bereich des Fachwissens hinaus zu bewegen (wenn z.B. emotionale Befindlichkeiten in Zusammenhang mit dem Stimulus geäußert werden). Es gilt, bei der Formulierung der Instruktion das Gleichgewicht zu finden zwischen exakter Handlungsanweisung und der Sensibilität der Probanden (in Situationen, in denen im Unternehmen individuelles Wissen erfaßt werden soll). In diesem Zusammenhang ist auch zu prüfen, inwieweit sich in der Instruktion eine Betonung der Quantität oder aber der Qualität der zu assoziierenden Konzepte auf die individuelle Assoziationsleistung auswirkt.

Weiterhin ist ungeklärt, wie **motivationale Aspekte** das Verhalten der Probanden im Wortassoziationsversuch beeinflussen. Als Verfahren der freien Reproduktion ist ein Raten wie in anderen (klassischen) Wissenstests kaum möglich – es kann nur das produziert werden, was auch im individuellen Wissensbesitz repräsentiert ist. Daher ist zu erwarten, daß (zunächst) diejenigen Wissensinhalte im Rahmen des Wortassoziationsversuchs benannt werden, welche vom Probanden ohne größere Anstrengungen auf den Stimulus hin aktiviert wurden. Fraglich ist dabei, was sich an diese Phase der Anstrengungslosigkeit anschließt: bricht der Proband das Assoziieren an dieser Stelle ab oder setzt er neue Anstrengungen ein, um seinen Gedächtnisbesitz weiter nach Inhalten abzusuchen. Somit können sich motivationale Einflüsse v.a. auf die Länge der Assoziationsliste

auswirken, bisweilen auch auf die Qualität der Beiträge. Es ist zu prüfen, inwieweit Konzepte wie das der Leistungsmotivation die Aussagekraft des Assoziationskennwertes systematisch beeinflussen. Grundsätzlich ist arbeitspsychologische Forschung im Feld immer angewiesen auf die Akzeptanz und Mitarbeit der Beschäftigten, nach Bungard, Holling und Schultz-Gambard (1996) müssen diese auch ihre Rolle als Proband übernehmen wollen und sich schließlich aktiv in die Untersuchung einbringen. Gerade bei Wissensanalysen aber ergeben sich gemäß Sonntag und Schaper (1997) hohe motivationale und kognitive Beanspruchungen für die Probanden, da sie ihre eigene Tätigkeit aus einer ungewohnten Perspektive heraus reflektieren müssen. Zudem unterstellen Mitarbeiter meist der Untersuchung, im Sinne der Unternehmensleitung zu agieren (Bungard, Holling und Schultz-Gambard, 1996, bezeichnen dies als Auftraggebereffekt) und reagieren daher grundsätzlich mit Skepsis. Bei einem vergleichsweise brisanten Thema wie Fachwissen sind die Vorbehalte um so größer, da dies als Leistungsbewertung wahrgenommen werden kann, welche u.U. Konsequenzen für den eigenen Arbeitsplatz nach sich ziehen könnte. Die Folgen für die wissensdiagnostische Erhebung könnten entweder Überengagement sein oder aber Boykott.

Des weiteren ist ungeklärt, inwieweit **Priming-Effekte** (also die Vor-Aktivierung bestimmter Gedächtnisbereiche, siehe z.B. Anderson, 2001) die Aussagefähigkeit des Assoziationskennwertes verzerren. Da im Zuge der Suche nach geeigneten Stimuli für den Wortassoziationsversuch eine Wissensdomäne gewählt wird, welche zentral für die zu untersuchende Arbeitstätigkeit ist, haben diese Inhalte und Begriffe auch eine erhöhte Auftretenswahrscheinlichkeit im Arbeitsalltag. Verändern sich Assoziationsmengen und -inhalte wesentlich, wenn ein Proband einen Stimulus zufällig kürzlich z.B. in einem Kundengespräch thematisiert hatte? Zur Überprüfung solcher möglichen Priming-Effekte durch zeitnah stattgefundene inhaltliche Auseinandersetzungen wäre ein experimenteller Ansatz denkbar, in dem zunächst bestimmte Wissensdomänen mit definierten Informationen bearbeitet werden (z.B. durch Lesen von Fachtexten, Lösen kleiner Arbeitsaufgaben) und anschließend im Wortassoziationsversuch die Wiederkehr genau dieser Informationen geprüft wird.

Hinsichtlich der Interpretation des Wortassoziationsversuchs ist weiterhin die Frage der **relationalen Beziehung** zwischen Stimulus und Assoziationen offen. Ist die Art der Relationen zwischen Konzepten eine bedeutsame Determinante der Assoziationsstärke? Werden Konzepte, die z.B. in Unterbegriffs- oder Attributrelation zum Stimulus stehen, leichter bzw. schneller assoziiert als Konzepte in Kontrast- oder Kausalrelationen? Bisher liegen dazu keine Erkenntnisse vor, lediglich das Erkennen von assoziativen Verbindungen, die in bestimmten Relationen zu einander stehen, wurde bisher thematisiert (z.B. Klix, 1984), nicht aber das Reproduzieren. Die Beantwortung dieser Frage erbringt zunächst nur einen Erkenntnisgewinn für die Grundlagenforschung, denn – wie bereits an mehreren Stellen angemerkt – kann der Wortassoziationsversuch nicht zwischen richtigen und falschen Wissensrepräsentationen unterscheiden. Auch, wenn sich bestimmte Relationen mit einer erhöhten Wahrscheinlichkeit am Beginn der Assoziationskette wiederfinden würden, so erlaubt dies dem Versuchsleiter dennoch nicht, vom Probanden geäußerte Assoziationen als richtig oder falsch zu bewerten, ohne Kenntnis der vom Probanden tatsächlich repräsentierten Relationen. Diese können nur aufgedeckt werden durch Verfahren wie der Strukturlegetechnik, welche genau diese Beziehungen zwischen Konzepten und Stimulus expliziert. Für die Praxis ist diese Problematik insofern interessant, als daß

von einigen Forschern (z.B. Rothe, 2003) vorgeschlagen wird, die Position von Assoziationen innerhalb von Assoziationsketten als Auswertestrategie heranzuziehen. Die Kenntnis über etwaige Regelmäßigkeiten im Zusammenhang zwischen semantischer Relation und Assoziationsstärke könnte dazu nützliche Interpretationsansätze liefern, wenngleich auch hier die oben beschriebenen Einwände zu berücksichtigen sind. Zudem könnten experimentelle Untersuchungen dieser Art das theoretische Gebilde eines semantischen Netzwerkes weiter stützen und einen wertvollen Beitrag zur Konstruktvalidierung leisten.

Ebenso würde eine Analyse der **Reihenfolge der Assoziationen in der Assoziationskette** zur Aufklärung des theoretischen Ansatzes des Wortassoziationsversuchs beitragen sowie Erkenntnisse zur Interpretation der Ergebnisse liefern. Wird die Position eines Konzeptes innerhalb einer Assoziationskette als Auswertestrategie herangezogen (z.B. Rothe, 2003), so kann man dabei unterstellen, daß die zuerst genannten Konzepte eine besonders starke Verbindung zum Stimulus aufweisen und diejenigen am Ende der Assoziationskette eine schwächere Beziehung (Strube, 1984). Woodworth und Schlosberg (1965) weisen allerdings zu Recht darauf hin, daß der Proband seine Assoziationen immer nur sukzessive benennen kann, selbst wenn alle gleichzeitig angeregt wurden. Für die Interpretation von Assoziationsketten ist jedoch bedeutsam, inwieweit der Zugriff auf die Konzepte während des Assoziationsprozesses einer Regelmäßigkeit folgt. So wäre denkbar, daß ein Proband zunächst alle unmittelbar um den Stimulus herumliegenden Knoten aufruft, da sie einen direkten Pfad zum Stimulusbegriff aufweisen, und dann zur nächst niedrigeren Ebene in der Begriffshierarchie übergeht. Es wäre allerdings auch vorstellbar, daß zunächst ein naheliegendes Konzept assoziiert wird und der thematisch dazugehörige Ast mit all seinen Unterebenen „abgearbeitet“ wird, bevor ein weiterer Begriff aus der unmittelbaren Nähe des Stimulus genannt und damit ein neuer Ast mit all seinen Teilästen aktiviert wird. Zur Aufklärung dieser Problematik bietet sich an, im Anschluß an den Wortassoziationsversuch die Strukturlegetechnik einzusetzen, bei der die Knoten von den Probanden selbst assoziiert wurden (nach Rothe, 2003, 2006). Die Reihenfolge der Assoziationen in der Assoziationskette kann a posteriori im gelegten Netz abgetragen werden und ggf. Systematiken sichtbar machen. In diesem Zusammenhang ist die Analyse der Begriffscluster interessant, die während des Assoziationsprozesses entstehen. Da man annimmt, daß innerhalb der Cluster ein hoher Zusammenhang zwischen den Konzepten besteht, sollte sich diese Nähe auch im Wissensnetz der Strukturlegetechnik abbilden. Zur Bestimmung der Cluster kann beispielsweise das computergestützte Tool Assoziations-Struktur-Test von Meyer (2007) herangezogen werden, welches ermöglicht, auf Grundlage der Analyse von Pausenzeiten zwischen den Eingaben einzelner Assoziationen inhaltliche Cluster zu bestimmen. Zur weiteren Absicherung (insbesondere mit Blick auf die inhaltliche Dimension der Cluster) bietet sich die strukturdimensionale Analyse von Huth (2000) an, bei der Begriffspaare vom Probanden durch eine hierarchische Ähnlichkeitsskalierung in Clustern zusammengefaßt werden (siehe auch Lander und Huth, 1999). Die differenzierte Analyse der Reihenfolge von Assoziationsketten, Clustern in den Ketten sowie deren Entsprechung in strukturabbildenden Verfahren wie der Strukturlegetechnik oder dem Assoziations-Struktur-Test würden das Verständnis für den Assoziationsprozeß weiter befördern und damit helfen, die Interpretationsansätze der Assoziationsleistungen abzusichern.

11. Fazit der vorliegenden Arbeit

Trotz weiterhin offener Fragen hinsichtlich der Determinanten von Assoziationsleistungen sowie der damit verbundenen Interpretationsmöglichkeiten konnte anhand vieler arbeitspsychologischer Felduntersuchungen die Brauchbarkeit des Wortassoziationsversuchs als wissensdiagnostisches Instrument belegt werden. Es wurde gezeigt, daß der Wortassoziationsversuch in vielen verschiedenen arbeitspsychologischen Untersuchungsfeldern und Wissensdomänen einsetzbar ist (siehe Kapitel 4). Dabei ließ sich das grundsätzliche Vorgehen bei der Konstruktion eines Assoziationsversuches hinsichtlich Inhalt, Umfang und Durchführungsmodus flexibel auf das jeweilige Untersuchungsfeld anpassen und integrierte sich damit gut in den organisatorischen Rahmen des Feldes. Die individuellen Assoziationskennwerte (also die Anzahl der produzierten Assoziationen) waren dabei durchaus in der Lage, zwischen Probanden verschiedener Leistungsstärke (i.S. von Wissensumfang) zu differenzieren.

Die Bestimmung der Haupt- und Nebengütekriterien erbrachte eine zufriedenstellende Bilanz (siehe Kapitel 6 und 7). Die Reliabilität des Wortassoziationsversuchs konnte belegt werden, somit ist eine wichtige Voraussetzung zur Beurteilung der Validität sowie der Veränderungssensitivität gegeben. Auch die Prüfung der Validität anhand der Außenkriterien Geschäftsführerbeurteilung sowie Klausurleistung erbrachte zufriedenstellende Koeffizienten und kann daher ebenfalls als belegt angesehen werden. Ebenso konnte i.S. der diskriminanten Validierung gezeigt werden, daß mittels der Assoziationstechnik tatsächlich das Konstrukt des Fachwissens und nicht der generellen Fähigkeit zur Wortflüssigkeit erfaßt wird. Damit sind wichtige Voraussetzungen erfüllt, um den Wortassoziationsversuch im Rahmen wissensdiagnostischer Erhebungen einzusetzen. Die Objektivität von Durchführung, Auswertung und Interpretation kann ebenfalls im Wesentlichen als gegeben hingenommen werden. Auch die Beurteilung der Nebengütekriterien der Ökonomie sowie der Akzeptanz durch die Probanden fällt positiv aus. Insgesamt erfüllt das Verfahren das Kriterium der Nützlichkeit. Lediglich die Kriterien der Normierung sowie der Vergleichbarkeit kann der Wortassoziationsversuch aufgrund seiner Inhaltsspezifität nicht erfüllen. Durch eine qualitativ ausgerichtete Analyse der Assoziationen kann – über die Anzahl assoziierter Konzepte hinaus – nicht nur der individuelle Wissensbesitz, sondern auch die interessierende Wissensdomäne anschaulich illustriert werden. Zu beachten ist dennoch, daß der Geltungsbereich eines Wortassoziationsversuchs immer eng umgrenzt werden muß: die jeweiligen Befunde beziehen sich auf die Untersuchungsstichprobe mit ihren konkreten Arbeitsanforderungen sowie auf die Auswahl der thematisierten Wissensdomäne und der daraus abgeleiteten Stimuli.

Für die Anwendung des Wortassoziationsversuchs in der arbeitspsychologischen Praxis läßt sich aufgrund der bisherigen Erfahrungen festhalten, daß dieses Verfahren auf jeden Fall dann ein geeignetes Instrument zur Wissensdiagnose darstellt ist, wenn die jeweiligen Arbeitsanforderungen den kognitiven Anforderungen des Wortassoziationsversuchs ähneln, d.h. wenn fachbezogenes Wissen im Zuge der Arbeitsaufgabe verbalisiert werden muß (z.B. im Servicebereich oder bei Kundenkontakt). Mindestvoraussetzung sollte aber sein, daß die betreffende Arbeitstätigkeit überhaupt die Anwendung von (explizitem) Fachwissen erfordert und der Einsatz dieses Wissens eine wesentliche Größe für die berufliche Leistung darstellt – wenn Arbeitshandeln weitestgehend durch Variablen wie z.B. Schnelligkeit oder Wahrnehmungsleistungen determiniert wird (z.B.

Müllsortierung am Fließband), wäre eine auf Verbaldaten beruhende Analyse des Fachwissens, wie sie der Wortassoziationsversuch vornimmt, doch eher unangemessen.

Der Wortassoziationsversuch ist ein geeignetes Verfahren zur Erfassung expliziten, überwiegend deklarativen arbeitsbezogenen individuellen Wissens. Insgesamt konnte gezeigt werden, daß der Wortassoziationsversuch nicht nur ein für die arbeitspsychologische Praxis taugliches Instrument zur *Beschreibung* der interessierenden Wissensdomäne ist, sondern auch die *Diagnose* des individuellen Fachwissens leisten kann, wenngleich er als alleinig eingesetztes Instrument nur vergleichsweise grobe Aussagen i.S. eines Screening-Verfahrens treffen kann.

Teil F - LITERATURVERZEICHNIS

- Ackermann, P.L. & Beier, M.E. (2006). Methods for Studying the Structure of Expertise: Psychometric Approaches. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Anderson, J.R. (2001). *Kognitive Psychologie*. Berlin: Spektrum Akademischer Verlag.
- Ballstaedt, S.-P. (1994). Dokumentenanalyse. In: H. Mandl & G.L. Huber (Hrsg.). *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*. Weinheim: Beltz PVU.
- Banse, R., Seise, J. & Zerbes, N. (2001). Implicit Attitudes towards Homosexuality: Reliability, Validity and Controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, 48, 2, 145-160.
- Beatty, I. D., Gerace, W. J. & Dufresne, R. J. (2002). *Measuring and modeling physics students' conceptual knowledge structures through term association times*. [Online preprint].
- Beatty, I.D. & Gerace, W.J. (2002). Probing physics students' conceptual knowledge structures through term association. *American Journal of Physics*, 70 (7), 750-758.
- Bergmann, B. (1999). Die Beurteilung von Interventionseffekten. In: E. Witruk und H.-J. Lander (Hrsg.). *Informationsverarbeitungsanalysen. Kognitionspsychologische und meßmethodische Beiträge*. Leipzig: Leipziger Universitätsverlag GmbH.
- Bergmann, B. (2000). Kompetenzentwicklung im Arbeitsprozeß. *Zeitschrift für Arbeitswissenschaft*, 2, 138-144.
- Bergmann, B. (2001). Berufliche Kompetenzentwicklung. In: R.K. Silbereisen & M. Reitzle (Hrsg.). *Bericht über den 42. Kongreß der Deutschen Gesellschaft für Psychologie in Jena 2000*. Lengerich: Pabst Science Publishers.
- Birkhan, G. (1992). Die (Un)Brauchbarkeit der klassischen Testgütekriterien für Dialog-Konsens-Verfahren. In: B. Scheele (Hrsg.) *Struktur-Lege-Verfahren als Dialog-Konsens-Methodik. Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung Subjektiver Theorien*. Münster: Aschendorff Verlag.
- Bonato, M. (1990). *Wissensstrukturierung mittels Strukturlegetechniken. Eine graphentheoretische Analyse von Wissensnetzen*. Frankfurt am Main: Peter Lang.
- Bortz, J. & Döring, N. (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J. & Lienert, G.A. (2003). *Kurzgefaßte Statistik für die klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*. Berlin: Springer.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Berlin: Springer-Verlag.
- Bredenkamp, J. & Erdfelder, E. (1996). Methoden der Gedächtnispsychologie. In: D. Albert & K.-H. Stapf (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 4: Gedächtnis*. Göttingen: Hogrefe.

- Buchanan, B.G., Davis, R. & Feigenbaum, E.A. (2006). Expert Systems: A Perspective from Computer Science. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bungard, W., Holling, H. & Schultz-Gambard, J. (1996). *Methoden der Arbeits- und Organisationspsychologie*. Weinheim: Psychologie Verlags Union.
- Burgert, M. (1992). Indikation hinsichtlich Personmerkmalen: Möglichkeiten und Grenzen der Verfahrensadaptation. In: B. Scheele (Hrsg.) *Struktur-Lege-Verfahren als Dialog-Konsens-Methodik. Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung Subjektiver Theorien*. Münster: Aschendorff Verlag.
- Büssing, A. & Herbig, B. (2003). Implizites Wissen und Wissensmanagement - Schwierigkeiten und Chancen im Umgang mit einer wichtigen menschlichen Ressource. *Zeitschrift für Personalpsychologie*, 2 (2), 51-65.
- Büssing, A., Herbig, B. & Ewert, T. (2001). Implizites und explizites Wissen - Einflüsse auf Handeln in kritischen Situationen. *Zeitschrift für Psychologie*, 209, No. 2, 174-200.
- Büssing, A., Herbig, B. & Ewert, T. (2002). Implizites Wissen und erfahrungsgelitetes Arbeitshandeln. Entwicklung einer Methode zur Explikation in der Krankenpflege. *Zeitschrift für Arbeits- und Organisationspsychologie*, 46, 1, 2-21.
- Büssing, A., Herbig, B. & Latzel, A. (2004). Explikation impliziten Wissens – Verändert sich das Handeln? *Sonderdruck aus: Zeitschrift für Psychologie*, 212 (2), 87–106.
- Ceglarek, P. & Rothe, H.-J. (2002). *Wissensmanagement im Call Center. CCall Report 16*. Hamburg: Verwaltungsberufsgenossenschaft.
- Ceglarek, P. & Rothe, H.-J. (2003a). *Analyse der psychischen Belastungen der Mitarbeiter der Berliner Stoffdruckerei GmbH und Ableitung von Empfehlungen für arbeitsgestalterische Maßnahmen. Abschlußbericht zum Projekt 22, bearbeitet im Rahmen der Initiative „Neue Qualität der Arbeit“ (INQA), gefördert durch das Bundesministerium für Wirtschaft und Arbeit*. Potsdam: unveröff. Projektbericht.
- Ceglarek, P. & Rothe, H.-J. (2003b). Wissensmanagement im Call Center – personale Aspekte. In: H.-G. Giesa, K.-P. Timpe & U. Winterfeld (Hrsg.) *Psychologie der Arbeitssicherheit und Gesundheit, 12. Workshop 2003*. Heidelberg: Asanger Verlag.
- Ceglarek, P. (2007). Das Wissen Studierender über Arbeits- und Gesundheitsschutz. In: P. Bärenz, A.-M. Metz, H.-J. Rothe (Hrsg.). *Psychologie der Arbeitssicherheit und Gesundheit. Arbeitsschutz, Gesundheit und Wirtschaftlichkeit. 14. Workshop 2007*. Heidelberg: Asanger Verlag.
- Ceglarek, P., Meyer, B., Kermas, M. & Rothe, H.-J. (2006). Verfahren zur Diagnose von (strukturellem) Wissen im Vergleich. In: H. Hecht, S. Berti, G. Meinhardt & M. Gamer (Hrsg.). *Beiträge zur 48. Tagung experimentell arbeitender Psychologen, 26. März - 29. März 2006*. Lengerich: Pabst Science Publishers.
- Chi, M.T.H. (2006). Laboratory Methods for Assessing Experts' and Novices' Knowledge. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.

- Cianciolo, A.T., Matthew, C., Sternberg, R.J. Wagner, R.K. (2006). Tacit Knowledge, Practical Intelligence and Expertise. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Cranach, M.v. & Bangerter, A. (2000). Wissen und Handeln in systemischer Perspektive: Ein komplexes Problem . In: H. Mandl & J. Gerstenmaier (Hrsg.). *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze*. Göttingen: Hogrefe-Verlag.
- Dann, H.D. (1992). Variation von Lege-Strukturen zur Wissensrepräsentation. In: B. Scheele (Hrsg.) *Struktur-Lege-Verfahren als Dialog-Konsens-Methodik. Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung Subjektiver Theorien*. Münster: Aschendorff Verlag.
- Davis, M., Curtis, M.B. & Tschetter, J.D. (2003). Evaluating cognitive training outcomes: validity and utility of structural knowledge assessment. *Journal of Business and Psychology, Vol. 18, 2*, 191-206.
- DIN (2004). *DIN EN ISO 10075-3: Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastungen. Teil 3: Grundsätze und Anforderungen an Verfahren zur Messung und Erfassung psychischer Arbeitsbelastungen*. Berlin: Beuth.
- Ebbinghaus, H. (1885 / 1992). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft. [Bibliothek klassischer Texte]
- Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computerunterstützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs-Technik (NET)*. Lengerich: Pabst Science Publishers.
- Eckert, A. (2000). Die Netzwerk-Elaborierungs-Technik (NET) – Ein computerunterstütztes Verfahren zur Diagnose komplexer Wissensstrukturen. In: H. Mandel & F. Fischer (Hrsg.) *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Ericsson, K.A. (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on representative Tasks. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Gäßler, B. (1994). *Psychotherapeuten als Experten. Gedächtnis und Informationsverarbeitung*. Regensburg: S. Roderer Verlag.
- Gawronski, B. & Conrey, F.R. (2004). Der Implizite Assoziationstest als Maß automatisch aktivierter Assoziationen: Reichweite und Grenzen. *Zeitschrift für Klinische Psychologie und Psychotherapie, 32 (4)*, 307–314, / *Psychologische Rundschau, 55 (3)*, 118–126.
- Gegner, U. (1993). Zur Strukturierung von Wortfeldern bei Hörgeschädigten – ein Assoziationsexperiment. *Sprache & Kognition, 12*, Heft 3, 157-169.
- Grabowski, J. (1999). Verbale Wissensdiagnose: High-Level-Prozesse der Sprachproduktion. In: I. Wachsmuth & B. Jund (Hrsg.). *KogWiss99. Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft. Bielefeld, 28. Sept. – 1. Okt. 1999*. St. Augustin: infix.
- Grabowski, J. (2005). Der Schriftlichkeitsüberlegenheitseffekt: Sprachproduktionsprozesse bei der verbalen Wissensdiagnose. Sonderdruck aus: *Zeitschrift für Psychologie, 213 (4)*, 193–204.
- Greenwald, A.G., McGhee, D.E. & Schwartz, J.K.L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

- Groeben, N. (1992). Die Inhalts-Struktur-Trennung als konstantes Dialog-Konsens-Prinzip?! In: B. Scheele (Hrsg.) *Struktur-Lege-Verfahren als Dialog-Konsens-Methodik. Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung Subjektiver Theorien*. Münster: Aschendorff Verlag.
- Großmann, N. & Teske-El Kodwa, S. (1995). Ziele und Probleme der Gestaltung wissensorientierter Unterstützungssysteme. In: W. Hacker, H.-J. Rothe, H. Wandke & J. Ziegler (Hrsg.). *Entwicklung und Einsatz wissensorientierter Unterstützungssysteme*. Bremerhaven: Wirtschaftsverlag.
- Gruber, H. & Mandl, H. (1996). Das Entstehen von Expertise. In: J. Hoffmann u. W. Kintsch (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 7: Lernen*. Göttingen: Hogrefe.
- Gruber, H. (1996). Wissensdiagnostik. In: G. Strube (Hrsg.). *Wörterbuch der Kognitionswissenschaft*. Stuttgart: Klett-Cotta.
- Häcker, H., Leutner, D. & Amelang, M. (1998). Standards für pädagogisches und psychologisches Testen. *Diagnostika und Zeitschrift für Differentielle und Diagnostische Psychologie, Supplementum 1/1998*.
- Häcker, H.O. & Stapf, K.-H. (2004). *Dorsch Psychologisches Wörterbuch*. Bern: Verlag Hans Huber.
- Hacker, W. (1992). *Expertenkönnen. Erkennen und Vermitteln*. Stuttgart: Verlag für Angewandte Psychologie.
- Hacker, W. (1993). Methodische Impulse aus dem Spannungsfeld zwischen grundlagenwissenschaftlicher Orientierung und angewandter Forschung – Das Beispiel Anforderungsanalyse. In: W. Bungard & T. Herrmann (Hrsg.). *Arbeits- und Organisationspsychologie im Spannungsfeld zwischen Grundlagenorientierung und Anwendung*. Bern: Verlag Hans Huber.
- Hacker, W. (1995). Diagnose von Expertenwissen. Von Abzapf (broaching)- zu Aufbau- ((re)construction)- Konzepten ?. In: K. Pawlik (Hrsg.). *Bericht über den 39. Kongreß der Deutschen Gesellschaft für Psychologie in Hamburg 1994*. Göttingen: Hogrefe.
- Hacker, W. (1996). *Diagnose von Expertenwissen. Von Abzapf- (broaching-) zu Aufbau- ([re-]construction-) Konzepten*. Berlin: Akademie-Verlag. (Sitzungsberichte der Sächsischen Akademie der Wissenschaften, Philologisch-historische Klasse, Bd. 134, Heft 6)
- Hacker, W. (1999). Informationsökonomie aus psychologischer Sicht. *Forschungsberichte Band 70, August 1999*. Dresden: Technische Universität Dresden.
- Hacker, W. (2005a). Experimentelles über die Assoziation von Vorstellungen. In: T. Rammsayer & S. Troche (Hrsg.). *Reflexionen der Psychologie. 100 Jahre Deutsche Gesellschaft für Psychologie*. Göttingen: Hogrefe Verlag.
- Hacker, W. (2005b). *Allgemeine Arbeitspsychologie. Psychische Regulation von Wissens-, Denk- und körperlicher Arbeit*. Bern: Verlag Hans Huber.
- Hasebrook, J. (2003). Computergestütztes Lernen in der Arbeit. *Wirtschaftspsychologie, Heft 1*, 5-8.
- Hauser, S., Spada, H., Rummel, N., & Meier, A. (2006). Expertise development in clinical psychology. In R. Sun, N. Miyake & C. D. Schunn (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. (pp. 1458-1463). Mahwah: Erlbaum.

- Henning, H.J (1999). Begriffliche Wissensverarbeitung in der Psychologie – Exploration von Daten mit Hilfe der Formalen Begriffsanalyse. In: W. Hacker & M. Rinck (Hrsg.). *Bericht über den 41. Kongreß der Deutschen Gesellschaft für Psychologie in Dresden 1998*. Lengerich: Pabst Science Publishers.
- Hiltmann, H. & Luhr, R. (1982). Wortassoziation und verbale Ergänzung. In: K.-J. Groffmann u. L. Michel (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie II: Psychologische Diagnostik, Band 3: Persönlichkeitspsychologie*. Göttingen: Verlag für Psychologie.
- Hinnerichs, L. & Rothe, H.-J. (2002). *Untersuchungen zu den Tatsachen modernen Wissens- und Kompetenzmanagements*. Universität Potsdam: unveröff. Abschlußbericht.
- Hoffmann, J., Zießler, M. & Grosser, U. (1984). Psychologische Gesetzmäßigkeiten der begrifflichen Klassifikation von Objekten. In: F. Klix (Hrsg.). *Gedächtnis, Wissens, Wissensnutzung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Hoffmann, R.R. & Lintern, G. (2006). Eliciting and Representing the Knowledge of Experts. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Hron, A. (1994). Interview. In: H. Mandl & G.L. Huber (Hrsg.). *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*. Weinheim: Beltz PVU.
- Hron, J., Lauche, K. & Schultz-Gambard, J. (2000). Training im Qualitätsmanagement: Eine Interventionsstudie zur Vermittlung von Qualitätswissen und handlungsleitenden Kognitionen. *Zeitschrift für Arbeits- und Organisationspsychologie, Vol. 44, No. 4*, 192-201.
- Huth, M. (2000). *Ein struktur- und dimensionsanalytischer Ansatz zur Erfassung und zum Vergleich begrifflich repräsentierten Wissens im menschlichen Langzeitgedächtnis. Ein Beitrag zur differentiellen Analyse von Wissensstrukturen*. Universität Leipzig: unveröff. Dissertation.
- Immenroth, M. (2003). *Mentales Training in der Medizin. Anwendung in der Chirurgie und Zahnmedizin*. Hamburg: Verlag Dr. Kovač.
- Jäger, A.O. & Althoff, K (1994). *Der WILDE-Intelligenz-Test (WIT). Ein Strukturdiagnostikum. Handanweisung*. Göttingen: Hogrefe, Verlag für Psychologie.
- Janetzko, D. & Strube, G. (2000). Knowledge Tracking – Eine neue Methode zur Diagnose von Wissenstrukturen. In: H. Mandel & F. Fischer (Hrsg.) *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Jordan, S. (2004). *Erfassung und Modellierung von Managementwissen mit Hilfe der Struktur-Ge-Technik im Rahmen einer Fort- und Weiterbildungsmaßnahme der Sparkassenakademie OSGV*. Universität Potsdam: unveröff. Diplomarbeit.
- Keckeis, M. (2005). *Lernunterstützung durch Mapping-Techniken. Ein Vergleich von Struktur-Ge-Technik und Mind-Mapping bezüglich Wissenszuwachs und Einstellung*. Universität Potsdam: unveröff. Diplomarbeit.
- Kermas, M. (2006). *Ermittlung qualitativer Kenngrößen von gelegten Wissensstrukturen*. Universität Potsdam: unveröff. Diplomarbeit.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau, 57 (4)*, 243–253.
- Klix, F. (1984). Über Wissensrepräsentation im menschlichen Gedächtnis. In: F. Klix (Hrsg.). *Gedächtnis, Wissens, Wissensnutzung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Klix, F. (1988). Gedächtnis und Wissen. In: H. Mandel & H. Spada (Hrsg.) *Wissenspsychologie*. Weinheim: Psychologie Verlags Union.

- Klix, F. (1991). Über notwendige Kriterien eines psychologisch relevanten Modells der Wissensrepräsentation. *Zeitschrift für Psychologie, Supplement 11*, 175-186.
- Klix, F. (1992). *Die Natur des Verstandes*. Göttingen: Hogrefe.
- Klix, F. (1996). Lernen und Denken. In: J. Hoffmann u. W. Kintsch (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 7: Lernen*. Göttingen: Hogrefe.
- Klocke, U. (2004). *Folgen von Machtausübung und Einflußnahme für Wissenszuwachs und Effektivität in Kleingruppen*. Berlin: dissertation.de.
- Kluwe, R. H. (1988). Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen. In: H. Mandel & H. Spada (Hrsg.) *Wissenspsychologie*. Weinheim: Psychologie Verlags Union.
- Knopf, M., Mack, W. & Kressley-Mba, R. (2005). Wissen und Erinnern. Zur Genese des episodischen Gedächtnisses bei Säuglingen und präverbalen Kindern. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 32 (4), 307–314, / *Psychologische Rundschau*, 56 (2), 113–122.
- Kubinger, K.D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kühnen, U., Schießl, M., Bauer, N., Paulig, N., Pöhlmann, C. & Schmidhals, K. (2001). How Robust is the IAT? Measuring and Manipulating Implicit Attitudes of East- and West-Germans. *Zeitschrift für Experimentelle Psychologie*, 48, 2, 135-144.
- Lander, H.-J. & Huth, M. (1999). Ein Verfahren zur Ermittlung der Strukturierung und Dimensionierung begrifflich repräsentierten Wissens im Langzeitgedächtnis. In: E. Witruk und H.-J. Lander (Hrsg.). *Informationsverarbeitungsanalysen. Kognitionspsychologische und methodische Beiträge*. Leipzig: Leipziger Universitätsverlag GmbH.
- Lang-von Wins, T. & Rosenstiel, L.v. (2000). Potentialfeststellungsverfahren. In: M. Kleinmann, & B. Strauß (Hrsg.). *Personalfeststellung und Personalentwicklung*. Göttingen: Hogrefe.
- Lang-von Wins, T. (2003). Die Kompetenzhaltigkeit von Methoden moderner psychologischer Diagnostik- Personalauswahl- und Arbeitsanalyseverfahren sowie aktueller Management-Diagnostik-Ansätze. In: J. Erpenbeck & L.v.Rosenstiel (Hrsg.). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
- Lewin, M. (1986). *Psychologische Forschung im Umriß*. Berlin: Springer-Verlag.
- Ley, T. & Albert, D. (2003). Kompetenzmanagement als formalisierbare Abbildung von Wissen und Handeln für das Personalwesen. *Wirtschaftspsychologie, Heft 3*, 86-93.
- Liebs, N. (2003). *Zum Zusammenhang zwischen individuellem Wissen, Leistung, Belastungserleben und Verhaltensdispositionen. Eine Felduntersuchung*. Universität Potsdam: unveröff. Diplomarbeit.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Mandl, H. & Ballstaedt, S.-P. (1986). Assessment of concept-building in text comprehension. In: F. Klix & H. Hagendorf (Eds.). *Human memory and cognitive capabilities. Mechanisms and performances*. Amsterdam: North-Holland.
- Mandl, H. & Fischer, F. (2000). Mapping-Techniken und Begriffsnetze in Lern- und Kooperationsprozessen. In: H. Mandel & F. Fischer (Hrsg.) *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.

- Mandl, H. & Huber, G.L. (1994). *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*. Weinheim: Beltz PVU.
- Mandl, H. (1997). Eröffnungsvortrag. In: H. Mandl (Hrsg.). *Schwerpunktthema Wissens und Handeln: Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie und München 1996*. Göttingen: Hogrefe.
- Marsh, K.L., Johnson, B.T. & Scott-Sheldon, L.A.J. (2001). Heart versus Reason in Condom Use: Implicit versus Explicit Attitudinal Predictors of Sexual Behavior. *Zeitschrift für Experimentelle Psychologie*, 48, 2, 161-175.
- Mazur, J.E. (2006). *Lernen und Verhalten*. München: Pearson Studium.
- Meer, E. van der (1984). Die Verfügbarkeit semantischer Relationen als differentialdiagnostisches Kriterium. In: F. Klix (Hrsg.). *Gedächtnis, Wissens, Wissensnutzung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Meyer, B. (2007). *The effects of computer-elicited structural knowledge and knowledge distribution in a group on complex problem solving performance*. Humboldt-Universität zu Berlin: unveröff. Dissertation.
- Meyer, B., Ceglarek, P., Kermas, M., Lattke, T. & Rothe, H.-J. (2006). AST: Ein computerbasiertes Verfahren zur Wissensdiagnose und seine Anwendung. In: H. Hecht, S. Berti, G. Meinhardt & M. Gamer (Hrsg.). *Beiträge zur 48. Tagung experimentell arbeitender Psychologen, 26. März - 29. März 2006*. Lengerich: Pabst Science Publishers.
- Meyer, B., Scholl, W., & Zhang, Z. (2007). Predicting task performance with elicitation of non-explicit knowledge. In N. Gronau (Ed.), *Proceedings of the 4th conference on professional knowledge management, 28th-30th March, 2007* (Vol. 2, pp. 303-311). Berlin: GITO.
- Michelsen, U.A. & Cordes, M. (2005). Ein praktikables Verfahren zur Bestimmung der Wissensquote und der Ratequote bei Mehrfachwahlaufgaben mit Bestantwort. *Diagnostica*, 51, Heft 3, 156-166.
- Mieg, H.A. (2006). Social and Sociological Factors in the Development of Expertise. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Müller, W.G. (1998). Ein interkultureller Vergleich affektiver und assoziativer Wortbedeutungen. *Sprache & Kognition*, 17, Heft 1-2, 73-88.
- Neumann, R., Ebert, M., Gabel, B., Gülsdorff, J., Krannich, H., Lauterbach, C. & Wiedl, K. (1998). Vorurteile zwischen Bayern und Norddeutschen: Die Anwendung einer neuen Methode zur Erfassung evaluativer Assoziationen in Vorurteilen. *Zeitschrift für Experimentelle Psychologie*, 45, 2, 99-108.
- Oeser, R. (1996). Begriff. In: G. Strube (Hrsg.). *Wörterbuch der Kognitionswissenschaft*. Stuttgart: Klett-Cotta.
- Oesterreich, R. (2005). Standards für Untersuchungsinstrumente zur psychischen Belastung und Beanspruchung in der ISO 10075-3 und Standards in der Wissenschaft. *Zeitschrift für Arbeits- und Organisationspsychologie*, 49, 3, 149-153.
- Opwis, K. & Luer, G. (1996). Modelle der Repräsentation von Wissen. In: D. Albert & K.-H. Stapf (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 4: Gedächtnis*. Göttingen: Hogrefe.
- Oswald, M. & Gadenne, V. (1984). Wissen, Können und künstliche Intelligenz. Eine Analyse der Konzeption des deklarativen und prozeduralen Wissens. *Sprache & Kognition*, 3, 173-183.

- Pleiss, C. & Oesterreich, R. (2003). Wissensdivergenz als Anforderung kooperativer Arbeit. *Wirtschaftspsychologie, Heft 3*, 66-71.
- Plötzer, R. (1996). Aktivationsausbreitung. In: G. Strube (Hrsg.). *Wörterbuch der Kognitionswissenschaft*. Stuttgart: Klett-Cotta.
- Pokrandt, S. (2004). *Fachspezifisches Wissen, Arbeitsleistung, Arbeitszufriedenheit und Commitment - Eine empirische Zusammenhangsanalyse*. Universität Potsdam: unveröff. Diplomarbeit.
- Puppe, F. (1996). Expertensystem. In: G. Strube (Hrsg.). *Wörterbuch der Kognitionswissenschaft*. Stuttgart: Klett-Cotta.
- Reinmann-Rothmeier, G. & Mandl, H. (2000). *Individuelles Wissensmanagement. Strategien für den persönlichen Umgang mit Informationen und Wissen am Arbeitsplatz*. Bern: Verlag Hans Huber.
- Rosenstiel, L.v. (2000). Wissen und Handeln in Organisationen . In: H. Mandl & J. Gerstenmaier (Hrsg.). *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze*. Göttingen: Hogrefe-Verlag.
- Rosenstiel, L.v. (2004). Arbeits- und Organisationspsychologie – Wo bleibt der Anwendungsbezug? *Zeitschrift für Arbeits- und Organisationspsychologie, 48, 2*, 87-94.
- Rosenthal, R. & DiMatteo, M.R. (2001). Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Reviews Psychology, 52*, 59-82.
- Rothe, H.-J. & Ceglarek, P. (2007). Komplexe Arbeitsanalyse in einer mittelständischen Stoffdruckerei. In: E. Schäfer, M. Buch, I. Pahls & J. Pfitzmann (Hrsg.) *Arbeitsleben! Arbeitsanalyse – Arbeitsgestaltung – Kompetenzentwicklung*. Kassel: kassel university press GmbH.
- Rothe, H.-J. & Hinnerichs, L. (2005). Wissens- und Kompetenzmanagement – verhaltensbeeinflussende subjektive und organisationale Bedingungen. In: Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung e.V. (Hrsg.). *Kompetenzmessung im Unternehmen. Lernkultur- und Kompetenzanalysen im betrieblichen Umfeld*. Münster: Waxmann.
- Rothe, H.-J. & Schindler, M. (1996). Expertise und Wissen. In: H. Gruber & A. Ziegler (Hrsg.). *Expertiseforschung. Theoretische und methodische Grundlagen*. Opladen: Westdeutscher Verlag.
- Rothe, H.-J. & Timpe, K.-P. (1997). Wissensanforderungen bei der Störungsdiagnose an CNC-Werkzeugmaschinen. In: K. Sonntag & N. Schaper (Hrsg.). *Störungsmanagement und Diagnosekompetenz. Leistungskritisches Denken und Handeln in komplexen Systemen*. Zürich: vdf Hochschulverlag.
- Rothe, H.-J. & Warning, J. (1991). Zur Identifikation von semantischen Relationen im fachspezifischen Gedächtnisbesitz von Experten mittels Analogiebildung. *Zeitschrift für Psychologie, Supplement 11*, 375-384.
- Rothe, H.-J. (1994a). Erfassung und Modellierung von Fachwissen als Grundlage für den Aufbau von Expertensystemen: Teil 1 – Ausgangsbedingungen und Konzeption. *Zeitschrift für Psychologie, 202 (3)*, 201-215.
- Rothe, H.-J. (1994b). Erfassung und Modellierung von Fachwissen als Grundlage für den Aufbau von Expertensystemen: Teil 2 – Methodenkritische Analysen. *Zeitschrift für Psychologie, 202 (4)*, 321-348.

- Rothe, H.-J. (1995). Wissensorientierte Unterstützungssysteme. In: W. Hacker, H.-J. Rothe, H. Wandke & J. Ziegler (Hrsg.). *Entwicklung und Einsatz wissensorientierter Unterstützungssysteme*. Bremerhaven: Wirtschaftsverlag.
- Rothe, H.-J. (2003). Wissensdiagnose auf Basis von Assoziieren und Strukturlegen. In: J. Erpenbeck & L.v.Rosenstiel (Hrsg.). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
- Rothe, H.-J. (2006). Wissensdiagnose mittels freier Assoziation. In: E. Witruk & H.-J. Lander (Hrsg.). *Kognitionspsychologische Analysen von Gedächtnisprozessen. Festschrift zum 75. Geburtstag von Prof. Dr. Hans-Jürgen Lander*. Aachen: Shaker.
- Rothe, H.-J., Timpe, K.-P. & Warning, J. (1991). Methodenkritische Analyse der Erfäßbarkeit von Fachwissen. In: F. Klix, E. Roth & E. van der Meer (Hrsg.). *Kognitive Prozesse und geistige Leistung*. Berlin Deutscher Verlag der Wissenschaften.
- Rothermund, K. & Wentura, D. (2001). Figure-Ground Asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie*, 48, 2, 94-106.
- Ruprecht, U. (2004). *Arbeits- und Wissensanalyse in einem Telekommunikationsunternehmen*. Universität Potsdam: unveröff. Diplomarbeit.
- Sarris, V. & Reiß, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie*. München: Pearson Studium.
- Schaper, N. (2003). Arbeitsproben und situative Fragen zur Messung arbeitsplatzbezogener Kompetenzen. In: J. Erpenbeck & L.v.Rosenstiel (Hrsg.). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
- Schaper, N. (2004). Fach-, Methoden- und Sozialkompetenz durch arbeitsplatzbezogenes Lernen in der betrieblichen Ausbildung. In: B. Wiese (Hrsg.). *Individuelle Steuerung beruflicher Entwicklung – Kernkompetenzen in der modernen Arbeitswelt*. Frankfurt: Campus.
- Scheele, B. & Groeben, N. (1984). *Die Heidelberger Struktur-lege-Technik (SLT)*. Weinheim: Beltz.
- Scheele, B. (Hrsg.) (1992). *Struktur-lege-Verfahren als Dialog-Konsens-Methodik. Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung Subjektiver Theorien*. Münster: Aschendorff Verlag.
- Schenk, J. & Ganschow, E. (2002). *Begriffe und Kategorien zur Beschreibung von Musik. Empirisches Projekt*. Humboldt-Universität zu Berlin, Naturwissenschaftliche Fakultät II, Institut für Psychologie.
- Schermer, F.J. (1991). *Lernen und Gedächtnis*. Stuttgart: Kohlhammer.
- Schindler, D. (2002). *Erfassung, Modellierung und Evaluation von Managementwissen als Grundlage für ein web based training von Führungskräften der DaimlerChrysler Vertriebsorganisation Deutschland*. Universität Potsdam: unveröff. Diplomarbeit.
- Schmidt, F.L. & Hunter, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology. *Psychological Bulletin*, Vol. 124, Nr. 2, 262-274.
- Schmidt, F.L. & Hunter, J.E. (2000). Meßbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In: M. Kleinmann, & B. Strauß (Hrsg.). *Personalfeststellung und Personalentwicklung*. Göttingen: Hogrefe.

- Schober, B. (2003). Ökologische Validität. In: K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim Beltz PVU.
- Schraagen, J.M. (2006). Task Analysis. In: K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffmann (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Schrameier, A. (1990). *Wortbedeutung im Gedächtnis*. Göttingen: Hogrefe Verlag für Psychologie.
- Schuler, H. (1989). Leistungsbeurteilung. In: H. Schuler u. A. Weinert (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich D: Praxisgebiete, Serie II: Wirtschafts- Organisations- und Arbeitspsychologie, Band 3: Organisationspsychologie*. Göttingen: Verlag für Psychologie.
- Schüppel, J. (1996). *Wissensmanagement. Organisatorisches Lernen im Spannungsfeld von Wissens- und Lernbarrieren*. Wiesbaden: Deutscher Universitäts-Verlag.
- Schwarzer, R. (1983). Befragung. In: H. Feger & J. Bredekamp (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich BD: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 2: Datenerhebung*. Göttingen: Verlag für Psychologie.
- Schweizer, K. & Koch, W. (2002). Fluide Intelligenz und mittelfristig erworbenes Wissen in einer spezifischen Wissensdomäne. *Zeitschrift für Psychologie*, 210 (3), 111-121.
- Sedlmeier, P. & Renkewitz, F. (2007). *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson Studium.
- Seiwald, B.B. (2003). Antwortformat. In: K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim Beltz PVU.
- Sonntag, K. & Schaper, N. (1997). Aufgaben- und Wissensanalysen zur Ermittlung des Lernbedarfs bei komplexen Diagnoseaufgaben. In: K. Sonntag & N. Schaper (Hrsg.). *Störungsmanagement und Diagnosekompetenz. Leistungskritisches Denken und Handeln in komplexen Systemen*. Zürich: vdf Hochschulverlag.
- Sonntag, K. & Schaper, N. (2006). Wissensorientierte Verfahren der Personalentwicklung. In: H. Schuler (Hrsg.). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.
- Sonntag, K., Rothe, H.-J. & Schaper, N. (1994). Wissenserfassung bei diagnostischen Tätigkeiten in komplexen Fertigungssystemen als Grundlage für die Gestaltung beruflichen Trainings. *Unterrichtswissenschaft*, 22 (3), 215-232.
- Steffens, M.C. & Plewe, I. (2001). Item's Cross-Category Associations as a Confounding Factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, 48, 2, 123-134.
- Stracke, I. (2004). *Einsatz computerbasierter Concept Maps zur Wissensdiagnose in der Chemie. Empirische Untersuchungen am Beispiel des Chemischen Gleichgewichts*. Münster: Waxmann.
- Strube, G. & Schlieder, C. (1998). Wissensrepräsentation im Symbolverarbeitungsansatz. In: F. Klix & H. Spada (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Praxisgebiete, Serie II: Kognition, Bd. 6: Wissen*. Göttingen: Hogrefe.
- Strube, G. (1984). *Assoziation. Der Prozeß des Erinnerns und die Struktur des Gedächtnisses*. Berlin: Springer-Verlag.
- Tergan, S.-O. (1988). Qualitative Wissensdiagnose – Methodische Grundlagen. In: H. Mandel & H. Spada (Hrsg.) *Wissenspsychologie*. Weinheim: Psychologie Verlags Union.
- Tergan, S.-O. (1989a). Psychologische Grundlagen der Erfassung individueller Wissensrepräsentationen. Teil 1: Grundlagen der Wissensmodellierung. *Sprache und Kognition*, 8, Heft 3, 152-165.

- Tergan, S.-O. (1989b). Psychologische Grundlagen der Erfassung individueller Wissensrepräsentationen. Teil 2: Methodologische Aspekte. *Sprache und Kognition*, 8, Heft 4, 193-202.
- Timpe, K.-P. & Rothe, H.-J. (1997). Wissenspsychologische Beiträge zur rechnerunterstützten Störungsdiagnose. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 92 (1997), 5, 243-245.
- Timpe, K.-P., Rothe, H.-J. & Seifert, R. (1994). Wissenspsychologische Beiträge zur Entwicklung von Unterstützungssystemen für die Störungsdiagnose. In: Bergmann & P. Richter (Hrsg.). *Die Handlungsregulationstheorie. Von der Praxis einer Theorie*. Göttingen: Hogrefe.
- Troche, S., Rammstedt, B. & Rammsayer, T. (2002). Vergleich einer Papier-Bleistift- und einer computergestützten Version des Leistungsprüfsystems (LPS). *Diagnostica*, 48, Heft 3, 115–120.
- Ulrich, R., Stapf, K.-H. & Giray, M. (1996). Faktoren und Prozesse des Einprägens und Erinnerns. In: D. Albert & K.-H. Stapf (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 4: Gedächtnis*. Göttingen: Hogrefe.
- Wagener, D. (2003). Der Computerwissenstest CWIS-4: Befunde zur Reliabilität und Validität. *Zeitschrift für Personalpsychologie*, 2 (4), 169-181.
- Wegwarth, O. (2002). *Evaluation der Effizienz des Nachwuchsführungskräfte-Moduls "Management und Führung" an der Ostdeutschen Sparkassenakademie*. Universität Potsdam: unveröff. Diplomarbeit.
- Weidle, R. & Wagner, A.C. (1994). Die Methode des Lauten Denkens. In: H. Mandl & G.L. Huber (Hrsg.). *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*. Weinheim: Beltz PVU.
- Weinert, F.E. (2001). Concept of Competence: A conceptual clarification. In: D. Rychen & L. Salganik (Eds.). *Defining and selecting key competencies*. Göttingen: Hogrefe & Huber.
- Weltgesundheitsorganisation (1999). *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); klinisch-diagnostische Leitlinien*. Bern: Verlag Hans Huber.
- Wender, K.F. (1988). Semantische Netze als Bestandteil gedächtnispsychologischer Theorien. In: H. Mandel & H. Spada (Hrsg.) *Wissenspsychologie*. Weinheim: Psychologie Verlags Union.
- Wessells, M.G: (1994). *Kognitive Psychologie*. München: Ernst Reinhardt Verlag.
- Wetzenstein, E., Nitschke, J, Polkehn, K. & Wandke, H. (2003). Assistenzsysteme für Unterhaltungselektronik im Haus und im Pkw. *Wirtschaftspsychologie*, 1, 221-224.
- Wirtz, M.A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Woike, J.K. (2003). *Screening*. In: K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim Beltz PVU..
- Woodworth, R.S. & Schlosberg, H. (1965). *Experimental psychology*. London: Methuen.
- Wreschner, A. (1904 / 2004). Experimentelles über die Assoziation von Vorstellungen. In: F. Schumann (Hrsg.). *Bericht über den 1. Kongress für experimentelle Psychologie in Gießen vom 18. bis 21. April 1904. Wiederabdruck der Erstausgabe Leipzig 1904 im Auftrag der Deutschen Gesellschaft für Psychologie*. Göttingen: Hogrefe Verlag.
- Yousfi, S. (2005a). Mythen und Paradoxien der klassischen Testtheorie (I). Testlänge und Gütekriterien. *Diagnostika*, 51, Heft 1, 1–11.

- Yousfi, S. (2005b). Mythen und Paradoxien der klassischen Testtheorie (II). Trennschärfe und Gütekriterien. *Diagnostika*, 51, Heft 2, 55–66.
- Ziegler, J. (1995). Unterstützung neuer Arbeitsformen durch wissensorientierte Informationssysteme. In: W. Hacker, H.-J. Rothe, H. Wandke & J. Ziegler (Hrsg.). *Entwicklung und Einsatz wissensorientierter Unterstützungssysteme*. Bremerhaven: Wirtschaftsverlag.
- Zöfel, P. (2003). *Statistik für Psychologen*. München: Pearson Studium.

Teil G - DANKSAGUNG

Da die vorliegende Arbeit die Integration vieler Befunde aus verschiedenen Studien zum Ziel hat, sollen an erster Stelle jene Diplomanden, Kollegen und Freunde genannt werden, welche mir nicht nur die von ihnen erhobenen Daten zur Verfügung stellten, sondern auch immer für Rückfragen und Diskussionen zur Verfügung standen. Ohne sie wäre meine Datenlage bei weitem nicht so umfangreich und abwechslungsreich. Mein herzlicher Dank gilt daher (in alphabetischer Reihenfolge):

Liane Hinnerichs

Nadine Liebs

Sascha Jordan

Stephanie Pokrandt

Michaela Keckeis

Ute Ruprecht

Maik Keramas

Odette Wegwarth

Mein besonderer Dank gilt dabei Dr. Bertolt Meyer, der mir nicht nur seine Studienergebnisse zur Verfügung stellte, sondern mir außerdem auch die Nutzung des von ihm entwickelten Verfahrens Assoziations-Struktur-Test ermöglichte – dessen Einsatz als computerbasiertes Assoziationstool vereinfachte die Erfassung der assoziierten Inhalte erheblich und ersparte mir das mühsame Transkribieren oder Abschreiben. In diesem Zusammenhang sei ebenso Tobias Lattke gedankt, der jederzeit für alle technischen Probleme im Umgang mit dem Assoziations-Struktur-Test ansprechbar war und Rat wußte.

Ebenso möchte ich meinem Doktorvater Prof. Dr. Heinz-Jürgen Rothe danken sowie Prof. Dr. Anna-Marie Metz. Sie gaben mir nicht nur die Gelegenheit, auch ohne Arbeitsstelle Kontakt zum Lehrstuhl zu halten, sondern sie waren mir jederzeit sowohl in fachlicher als auch persönlicher Hinsicht sehr angenehme Gesprächspartner. Vielen Dank für die herzliche Atmosphäre!

Nicht zuletzt möchte ich herzlichst meinen Eltern sowie meinem Mann Philipp danken, die mich über die gesamte Phase der Promotion begleitet und dabei in aller erdenklichen Form unterstützt haben, alle Höhen und Tiefen dieser Zeit auffingen und dabei immer an mich geglaubt haben. Danke!

Teil H - ANHANG

Anhang 1: Untersuchungsmaterial in Studie 1 (Modus) – Einführung, Instruktion zum Wortassoziationsversuch, Beispiel für einen Stimulus im schriftlichen Modus, Akzeptanz-Fragebogen zum Wortassoziationsversuch	252
Anhang 2: Untersuchungsmaterial in Studie 2 (psychotherapeutische Ambulanz) zum ersten Meßzeitpunkt – Einführung, Instruktion zum Wortassoziationsversuch, Akzeptanz-Fragebogen zum Wortassoziationsversuch.....	257
Anhang 3: Untersuchungsmaterial in Studie 2 (psychotherapeutische Ambulanz) zum zweiten Meßzeitpunkt – Akzeptanz-Fragebogen zum Wortassoziationsversuch, Lückentext, Akzeptanz-Fragebogen zum Lückentext	260
Anhang 4: Untersuchungsmaterial der Studien 6 bis 9 (Klausuren in verschiedenen Fächern der Psychologie) am Beispiel der Klinischen Psychologie – Deckblatt mit Instruktion, Wortassoziationsversuch – Beispiel für einen Stimulus.....	265
Anhang 5: Studie 15 (Wissen über Arbeitsschutz) – Multiple-Choice-Test (Wissenstest) des Online-Fragebogens	267
Anhang 6: Studie 15 (Wissen über Arbeitsschutz) – Items des Wortassoziationsversuchs in der Online-Ansicht einschließlich Instruktion	269
Anhang 7: Studie 2 (psychotherapeutische Ambulanz) – statistische Kennwerte (Mittelwert und Standardabweichung) des Assoziationskennwertes im Vergleich zwischen den Formen A und B, über die verschiedenen Abstraktionsebenen der Stimuli zu T1	269
Anhang 8: Studie 2 (psychotherapeutische Ambulanz) – statistische Kennwerte (Mittelwert und Standardabweichung) des Assoziationskennwertes im Vergleich zwischen den Formen A und B, über die verschiedenen Abstraktionsebenen der Stimuli zu T2	269

Anhang 1: Untersuchungsmaterial in Studie 1 (Modus) – Einführung, Instruktion zum Wortassoziationsversuch, Beispiel für einen Stimulus im schriftlichen Modus, Akzeptanz-Fragebogen zum Wortassoziationsversuch



UNIVERSITÄT POTSDAM

Campus II, Karl-Liebnecht-Str. 24-25, 14 476 Potsdam-Golm
Humanwissenschaftliche Fakultät, Institut für Psychologie, Lehrstuhl
Arbeits-, Betriebs- und Organisationspsychologie

Sehr geehrte Damen und Herren!

Vielen Dank für Ihre Teilnahme an dieser kleinen Untersuchung, bei der es um die verschiedenen Durchführungsbedingungen eines Verfahrens zur Wissenserhebung – dem sog. Wortassoziationsversuch - geht!

Die Untersuchung besteht aus drei Teilen:

- WIT-Test (dazu wird Ihnen die Untersuchungsleiterin noch detaillierte Instruktionen geben),
- schriftlicher Wortassoziationsversuch (anschließend einige Fragen dazu, wie Sie mit dieser Methode zurecht kamen),
- mündlicher Wortassoziationsversuch (anschließend einige Fragen dazu, wie Sie mit dieser Methode zurecht kamen).

Bitte blättern Sie erst, wenn Sie dazu aufgefordert werden!

Bitte noch einige Angaben zu Ihrer Person:

Alter: _____

Geschlecht: weiblich männlich

Vielen Dank!

Dipl.Psych. Petra Ceglarek

Vpn-Nr. _____

Form B

(Fortsetzung)

Wortassoziationsversuch

Ihre Aufgabenstellung:

Ich möchte erfahren, was Sie über die folgenden Themenbereiche aus dem Alltagsleben wissen.

Auf den folgenden Seiten werden Sie jeweils ein Stichwort lesen. Bitte schreiben Sie spontan alles auf, was Ihnen zu diesem Begriff einfällt, aber bitte antworten Sie nur in Stichworten oder kurzen Phrasen und kennzeichnen Sie diese jeweils durch Anstriche! Wenn Ihnen zu einem Begriff nichts mehr einfällt, gehen Sie bitte zum nächsten Begriff weiter!

Vielen Dank für Ihre Teilnahme an der Untersuchung!

Bitte tragen Sie hier die Uhrzeit ein, zu der Sie begonnen haben!

Zeit bei Beginn der Bearbeitung:

.....

(Fortsetzung)

40. **Saldo**

Zeit bei Ende der Bearbeitung:

.....

(Fortsetzung)

Im Folgenden bitte ich Sie um eine kurze Einschätzung des Verfahrens als *schriftliche Variante*, die Sie gerade bearbeitet haben. Bitte kreuzen Sie an, inwiefern die Aussagen auf Sie zutreffen, und begründen Sie Ihre Entscheidungen!

1. Die Aufgabenstellung / Instruktion war gut verständlich.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht überhaupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

2. Nur in Stichworten oder kurzen Phrasen zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

3. Schriftlich zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

(Fortsetzung)

Nach dem mündlichen Modus:

Im Folgenden bitte ich Sie um eine kurze Einschätzung des Verfahrens als *mündliche Variante*. Bitte kreuzen Sie an, inwiefern die Aussagen auf Sie zutreffen, und begründen Sie Ihre Entscheidungen!

1. Nur in Stichworten oder kurzen Phrasen zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

2. Mündlich zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

3. Die Begriffe nicht nur zu hören, sondern auch zu sehen, hat mir das Antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...erschwert	... etwas erschwert	... weder erschwert noch erleichtert	... etwas erleichtert	... erleichtert

weil:

.....

.....

Bitte vergleichen Sie nun abschließend die beiden Verfahrensformen miteinander:

4. Welche Variante würden Sie bevorzugen?

- die mündliche Form, *weil*
 - die schriftliche Form, *weil*
 - keine – ich finde beide Formen vergleichbar , *weil*
-

Anhang 2: Untersuchungsmaterial in Studie 2 (psychotherapeutische Ambulanz) zum ersten Meßzeitpunkt – Einführung, Instruktion zum Wortassoziationsversuch, Akzeptanz-Fragebogen zum Wortassoziationsversuch



UNIVERSITÄT POTSDAM

Campus II, Karl-Liebknecht-Str. 24-25, 14 476 Potsdam-Golm

Humanwissenschaftliche Fakultät, Institut für Psychologie, Lehrstuhl
Arbeits-, Betriebs- und Organisationspsychologie

Sehr geehrte Damen und Herren!

Vielen Dank für Ihre Teilnahme an dieser kleinen Untersuchung, bei der es um die Durchführungsbedingungen eines Verfahrens zur Wissenserhebung – dem sog. Wortassoziationsversuch - geht!

Die Untersuchung findet an zwei Terminen statt und besteht aus folgenden Teilen:

- WIT-Test
- Wortassoziationsversuch
- Fragen dazu, wie Sie mit dieser Methode zurecht kamen sowie
- Fragen zu psychologischen Fachthemen.

Bitte blättern Sie erst, wenn Sie dazu aufgefordert werden!

☞ Bitte noch einige Angaben zu Ihrer Person:

Alter: _____

Geschlecht: weiblich männlich

Ausbildung: *(bitte den höchsten bisher erreichten Status angeben)*

cand.Psych. Dipl.Psych. KJP in Ausbildung KJP / PP (approbiert)

Vielen Dank!

Dipl.Psych. Petra Ceglarek

Vpn-Nr. ____

(Fortsetzung)

Wortassoziationsversuch

Ihre Aufgabenstellung:

Bitte gehen Sie auf folgende Internetseite:

<http://abulifa.wiwi.hu-berlin.de:8080/ast/>

und geben Sie in das sich öffnende Fenster folgenden Code ein:

MTcuMDE3LjA=

Ich möchte erfahren, was Sie über die folgenden Themenbereiche aus dem Fachwissensbereich wissen. Im folgenden werden Sie jeweils ein Stichwort lesen. Bitte schreiben Sie spontan alles auf, was Ihnen zu diesem Begriff einfällt, aber bitte antworten Sie nur in Stichworten oder kurzen Phrasen und trennen Sie diese durch die Enter-Taste! Nennen Sie bitte alles, was Ihnen zu einem Begriff einfällt – selbst dann, wenn Sie das Gleiche bei einem vorangegangenen Begriff bereits genannt haben (Wiederholungen sind ausdrücklich gestattet)! Wenn Ihnen zu einem Begriff nichts mehr einfällt, gehen Sie bitte zum nächsten Begriff weiter!

Tragen Sie bitte die Uhrzeit bei Beginn Ihrer Bearbeitung in das untenstehende Kästchen ein und folgen Sie nun den Anweisungen auf dem Bildschirm!

Vielen Dank für Ihre Teilnahme an der Untersuchung!

Bitte tragen Sie hier die Uhrzeit ein, zu der Sie mit dem Wortassoziationsversuch beginnen bzw. beenden!

Zeit bei Beginn der Bearbeitung:

.....

Zeit bei Ende der Bearbeitung:

.....

(Fortsetzung)

Im Folgenden bitte ich Sie um eine kurze Einschätzung des Verfahrens, das Sie gerade bearbeitet haben. Bitte kreuzen Sie an, inwiefern die Aussagen auf Sie zutreffen, und begründen Sie Ihre Entscheidungen!

1. Die Aufgabenstellung / Instruktion war gut verständlich.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht überhaupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

2. Nur in Stichworten oder kurzen Phrasen zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

Anhang 3: Untersuchungsmaterial in Studie 2 (psychotherapeutische Ambulanz) zum zweiten Meßzeitpunkt – Akzeptanz-Fragebogen zum Wortassoziationsversuch, Lückentext, Akzeptanz-Fragebogen zum Lückentext

Im Folgenden bitte ich Sie um eine kurze Einschätzung des Wortassoziationsversuchs, den Sie gerade bearbeitet haben. Bitte kreuzen Sie an, inwiefern die Aussagen auf Sie zutreffen, und begründen Sie Ihre Entscheidungen!

1. Die Aufgabenstellung / Instruktion war gut verständlich.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht überhaupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

2. Nur in Stichworten oder kurzen Phrasen zu antworten ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

3. Während des Bearbeitens hatte ich das Gefühl, daß der Wortassoziationsversuch in der Lage ist, meinen eigenen Leistungsstand (bzgl. des behandelten Themas) gut abzubilden.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht überhaupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

weil:

.....

.....

(Fortsetzung)

Es folgt ein Lückentext zu einigen Störungsbildern, wie sie im ICD-10 verstanden werden. Bitte ergänzen Sie die freien Stellen im Text jeweils durch Wörter oder kurze Phrasen, so daß der Satz inhaltlich richtig vervollständigt ist. Sollte Ihnen zu einem Satz keine sinnvolle Ergänzung einfallen, gehen Sie bitte zur nächsten Freistelle weiter.

F8 Entwicklungsstörungen

Der Beginn von Entwicklungsstörungen (F8) liegt ausnahmslos _____ .
Betroffen sind z.B. die Funktionsbereiche _____ oder

_____ .
Die meisten Störungen dieser Gruppe treten bei Jungen _____ auf als bei Mädchen

F80 UES des Sprechens und der Sprache

Bei UES des Sprechens und der Sprache sind die normalen Muster des Spracherwerbs gestört. Einer Sprachentwicklungsstörung folgen oft Schwierigkeiten z.B. im Bereich _____

_____ .
Für die Differentialdiagnose sind zwei der vier Hauptkriterien _____
und _____ .

Schwierig ist die Abgrenzung von z.B. einer _____ .

F80.1 expressive Sprachstörung

In der normalen Sprachentwicklung gibt es beträchtliche individuelle Unterschiede. Deutliche Hinweise auf eine Verzögerung sind das Nichtbeherrschen von _____ im Alter von 2 Jahren und das Unvermögen, einfache Zweiwortsätze im Alter von 3 Jahren zu bilden.

Häufig sind Beeinträchtigungen der gesprochenen Sprache begleitet von leicht verzögerten oder auffälligen Wort – Laut – Produktionen. Die Diagnose darf nur gestellt werden, wenn die Schwere der Entwicklungsverzögerung bezüglich der expressiven Sprache _____ der Grenzen der Varianz der Norm für das Alter des Kindes liegen, die rezeptiven Sprachfertigkeiten jedoch _____ der normalen Grenzen liegen.

Ein Ausschluß sollte z.B. bzgl. folgender Störung geprüft werden: _____ .

F81 UES schulischer Fertigkeiten

Die umschriebenen Entwicklungsstörungen schulischer Fertigkeiten umfassen Gruppen von Störungen mit spezifischen und deutlichen Beeinträchtigungen des Erlernens schulischer Fertigkeiten. Sie sind nicht direkte Folge anderer Krankheiten (wie z.B. _____), aber sie können zusammen mit diesen auftreten.

Dieses Krankheitsbild tritt bei Jungen _____ auf als bei Mädchen.

Eine Grundbedingung zur Diagnostik einer UES ist das Vorliegen einer klinisch eindeutigen Beeinträchtigung spezieller schulischer Fertigkeiten. Zur Beurteilung kann z.B. herangezogen werden:

_____ .

(Fortsetzung)

Die Beeinträchtigung muß in dem Sinne spezifisch sein, daß sie nicht allein durch allgemeine _____ erklärbar ist. Zur Prüfung sind standardisierte Testverfahren einzusetzen.

F81.0 Lese- und Rechtschreibstörung

Das Hauptmerkmal der Lese- und Rechtschreibstörung ist eine umschriebene und eindeutige Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein erklärbar ist durch z.B. _____.

In den frühen Stadien des Erlernens einer alphabetischen Schrift kann es Schwierigkeiten geben, das Alphabet aufzusagen, die Buchstaben korrekt zu benennen, einfache Wortreime zu bilden und bei der Analyse oder der Kategorisierung von Lauten (trotz normaler Hörschärfe). Später können dann Fehler beim Vorlesen auftreten, die sich zeigen z.B. als _____ oder _____.

Als diagnostische Leitlinie müssen die Leseleistungen des Kindes _____ dem Niveau liegen, das aufgrund des Alters, der allgemeinen Intelligenz und der Beschulung zu erwarten ist.

F84 tiefgreifende Entwicklungsstörungen

Die Gruppe der tiefgreifenden Entwicklungsstörungen (F84) ist charakterisiert durch _____.

In den meisten Fällen sind die Störungen seit _____ Lebensjahren manifest.

Eine Intelligenzminderung liegt bei tiefgreifenden Entwicklungsstörungen _____ vor.

In einigen Fällen gehen die Störungen mit bestimmten somatischen Krankheitsbildern einher und sind ihnen möglicherweise zuzuschreiben: z.B. infantile Zerebralparese, angeborene Röteln, tuberöse Sklerose, zerebrale Lipoidose und das Syndrom des fragilen X-Chromosoms. Treten solche begleitenden somatischen Störungen auf, ist bei der Vergabe der Diagnose Folgendes zu berücksichtigen: _____.

F84.0 frühkindlicher Autismus

Frühkindlicher Autismus ist eine tiefgreifende Entwicklungsstörung, die durch eine abnorme oder beeinträchtigte Entwicklung definiert ist und sich _____ Lebensjahr manifestiert.

In jedem Fall finden sich qualitative Beeinträchtigungen in den sozialen Interaktionen. Sie zeigen sich z.B. in Form von _____.

Ebenso sind qualitative Beeinträchtigungen der Kommunikation allgemein anzutreffen. Diese zeigen sich z.B. als _____.

Außerdem ist charakteristisch eine gestörte Funktionsfähigkeit im Bereich _____.

F90 hyperkinetische Störungen

Ein Hauptmerkmal von hyperkinetischen Störungen ist _____.

Der Beginn hyperkinetischer Störungen liegt _____.

(Fortsetzung)

Eine begleitende Störung kann sein _____.

Die Beziehungen zu anderen Kindern oder Erwachsenen ist oft _____.

F90.0 einfache Aktivitäts- und Aufmerksamkeitsstörung

Untersuchungen zeigen, daß der Verlauf der einfachen Aktivitäts- und Aufmerksamkeitsstörung bis ins Adoleszenz- und Erwachsenenalter stark davon beeinflusst wird, ob z.B. _____ oder _____ begleitend vorhanden ist oder nicht.

F90.0 soll verwendet werden, wenn die allgemeinen Kriterien für eine hyperkinetische Störung (F90) erfüllt sind, die Kriterien für _____ jedoch nicht.

Ein dazugehöriger Begriff zu F90.0 ist _____.

F91 Störungen des Sozialverhaltens

Störungen des Sozialverhaltens sind durch ein sich wiederholendes und andauerndes Muster von _____ Verhalten charakterisiert.

Beispiele für Verhaltensweisen, welche die Diagnose begründen, sind

_____ oder _____.

Jedes dieser Beispiele ist bei erheblicher Ausprägung ausreichend für die Diagnose; isolierte dissoziale Handlungen genügen dagegen nicht. Es wird empfohlen, diese Diagnose nur dann zu stellen, wenn die Dauer des oben beschriebenen Verhaltens _____ Monate oder länger beträgt.

F91.2 Störung des Sozialverhaltens bei vorhandenen sozialen Bindungen

Die Kategorie F91.2 Störung des Sozialverhaltens bei vorhandenen sozialen Bindungen umfaßt Störungen des Sozialverhaltens mit andauerndem dissozialem oder aggressivem Verhalten bei Kindern, die allgemein gut in ihrer Altersgruppe eingebunden sind. Die Kriterien für F91 sind erfüllt und es besteht nicht nur oppositionelles, aufsässiges und trotziges Verhalten. Hauptdifferenzierungsmerkmal ist _____.

Oft besteht die Bezugsgruppe aus _____, das ist jedoch keine notwendige Bedingung für die Diagnose.

Beziehungen zu Autoritätspersonen sind häufig _____.

Ein dazugehöriger Begriff ist _____.

F93.1 phobische emotionale Störung des Kindesalters

Kinder können wie Erwachsene Befürchtungen vor verschiedensten Objekten oder Situationen entwickeln. Einige dieser Befürchtungen (oder Phobien) sind nicht Bestandteil einer normalen psychosozialen Entwicklung, beispielsweise die Agoraphobie. Andere Befürchtungen zeigen eine deutliche Spezifität für bestimmte Entwicklungsphasen und treten (in gewissem Grade) bei der Mehrheit der Kinder auf, beispielsweise die Angst vor _____.

Verglichen mit der Altersstufe ist der Beginn der Störung F93.1 _____, das Ausmaß der Angst ist dabei _____.

Die Angst ist nicht Teil einer _____.

(Fortsetzung)

Im Folgenden bitte ich Sie um eine kurze Einschätzung des Lückentextes, den Sie gerade bearbeitet haben. Bitte kreuzen Sie an, inwiefern die Aussagen auf Sie zutreffen, und begründen Sie Ihre Entscheidungen!

1. Die Aufgabenstellung / Instruktion war gut verständlich.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht über- haupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

2. Das Antworten in Stichworten oder kurzen Phrasen innerhalb der Sätze ...

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...fiel mir schwer	... fiel mir eher schwer	... fiel mir weder schwer noch leicht	... fiel mir eher leicht	... fiel mir leicht

weil:

.....

.....

3. Während des Bearbeitens hatte ich das Gefühl, daß der Lückentext in der Lage ist, meinen eigenen Leistungsstand (bzgl. des behandelten Themas) gut abzubilden.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
trifft nicht über- haupt nicht zu	trifft eher nicht zu	teils-teils	trifft etwas zu	trifft völlig zu

weil:

.....

.....

Bitte vergleichen Sie nun abschließend den Wortassoziationsversuch und den Lückentext miteinander:

Welche Variante würden Sie bevorzugen?

den Wortassoziationsversuch , *weil*

den Lückentext, *weil*

keine – ich finde beide Methoden vergleichbar , *weil*

.....

Anhang 4: Untersuchungsmaterial der Studien 6 bis 9 (Klausuren in verschiedenen Fächern der Psychologie) am Beispiel der Klinischen Psychologie – Deckblatt mit Instruktion, Wortassoziationsversuch – Beispiel für einen Stimulus



UNIVERSITÄT POTSDAM

Campus II, Karl-Liebknecht-Str. 24-25, 14 476 Potsdam-Golm

Humanwissenschaftliche Fakultät, Institut für Psychologie,
Lehrstuhl Arbeits-, Betriebs- und Organisationspsychologie

Sehr geehrte Damen und Herren!

Vielen Dank für Ihre Teilnahme an dieser Untersuchung, bei der es um ein Verfahren zur Wissenserhebung – dem sog. Wortassoziationsversuch – geht! Die Untersuchung dient der Validierung des Wortassoziationsversuchs an einem Außenkriterium, in diesem Falle der Leistung in der Klausur.

Um beide Variablen miteinander in Beziehung setzen zu können, werden Angaben zur Person benötigt. Daher bitte ich Sie um folgende persönliche Angaben:

☞ **Ihr Name:**

☞ **Ihre
Matrikelnummer:**

Diese Angaben werden vertraulich behandelt, und sobald die Zuordnung von Klausurleistung und Assoziationsleistung geschehen ist, wird dieses Blatt und damit auch Ihre persönlichen Daten vernichtet!

Mit freundlichen Grüßen

Dipl.Psych. Petra Ceglarek

Ihre Aufgabenstellung:

Ich möchte erfahren, was Sie über die folgenden Begriffe aus dem Bereich der Klinischen Psychologie wissen. Nachfolgend werden Sie jeweils ein Stichwort lesen. **Bitte schreiben Sie spontan alles auf, was Sie zu diesem Begriff wissen**, aber bitte antworten Sie nur **in Stichworten oder kurzen Phrasen!**

Bitte bearbeiten Sie die Begriffe nacheinander, so wie sie im folgenden präsentiert werden. Nennen Sie bitte alles Wissen, was Ihnen zu einem Begriff einfällt! Wenn Ihnen zu einem Begriff nichts mehr einfällt, gehen Sie bitte zum nächsten Begriff weiter!

Vielen Dank für Ihre Mithilfe!

Anhang 5: Studie 15 (Wissen über Arbeitsschutz) – Multiple-Choice-Test (Wissenstest) des Online-Fragebogens

Es folgt ein kleiner Wissenstest zum Thema Arbeitsschutz und Unfallprävention**B11 Stell Dir vor, Dir würde an der Universität ein Unfall passieren ...**

... bei welcher Institution bist du unfallversichert?	Ja	Nein	Weiß nicht
eigene Krankenkasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Versicherung des Unfallgegners	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eigene Haftpflichtversicherung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haftpflicht des Unfallverursachers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Landesunfallkasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uni haftet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
private Unfallversicherung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
variiert mit der Unfallart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

B12 ... wen mußt Du nach dem Unfall verständigen?

	Ja	Nein	Weiß nicht
Dozent der nächsten Lehrveranstaltung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Abteilung für Sicherheitswesen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dekan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rektor der Uni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sicherheitsbeauftragten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Krankenversicherung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
private Unfallversicherung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Landesunfallkasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

B13 ... welche Leistungen kannst Du im Falle einer Verletzung vom Versicherer erwarten?

	Ja	Nein	Weiß nicht
Krankengeld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unfallrente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rehabilitationsmaßnahmen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schmerzensgeld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Erstattung der Behandlungskosten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Erstattung von Verdienstaussfällen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ersatz von beschädigten Sachmitteln	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(Fortsetzung)

B14 Welche studentischen Aktivitäten sind unfallversichert?		Ja	Nein	Weiß nicht
	Besuch von Lehrveranstaltungen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Weg zur Uni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Umweg über den Supermarkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Weg zur Immatrikulation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Studium zu Hause	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Exkursionen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Hochschulsport	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Praktikum	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Job innerhalb der Uni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Job außerhalb der Uni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Auslandssemester	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Arbeit im Fachschaftsrat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Chor/Orchester der Uni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Semesterfeier des Fachschaftsrates	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Recherche in der Stadtbibliothek	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Recherche in der Universitätsbibliothek	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Besuch eines Gastvortrages in Berlin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Absolventenfeier	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Mensaparty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
B15 Wer ist an der Uni für den Arbeitsschutz verantwortlich?		Ja	Nein	Weiß nicht
	der Rektor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Abteilung für Sicherheitswesen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Du als Student	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	der Betriebsarzt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	die Sicherheitsfachkraft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	das Studentenwerk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	die Landesregierung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	der Asta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
B16 Welche Pflichten hat die Uni zur Verhütung von Unfällen?		Ja	Nein	Weiß nicht
	Gefährdungsanalyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Aushang einer Auflistung potentieller Gefahren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Unterweisung der Studierenden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Aufsicht bei besonderen Gefahren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Bestellung von Sicherheitsbeauftragten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Bestellung eines Betriebsarztes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Studierende mit Sicherheitseinrichtungen vertraut machen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Erste Hilfe Material zur Verfügung stellen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang 6: Studie 15 (Wissen über Arbeitsschutz) – Items des Wortassoziationsversuchs in der Online-Ansicht einschließlich Instruktion

A12 Im Folgenden werden Dir einzelne Begriffe genannt. Wir sind daran interessiert, was Du zu jedem dieser Begriffe weißt. Bitte schreib also alles auf, was Dir dazu einfällt! (Bitte nur Stichworte, einzelne Begriffe/ Aussagen durch ENTER trennen.)

Arbeitsschutz:

Gesundheitsschutz:

Arbeitsunfall:

Anhang 7: Studie 2 (psychotherapeutische Ambulanz) – statistische Kennwerte (Mittelwert und Standardabweichung) des Assoziationskennwertes im Vergleich zwischen den Formen A und B, über die verschiedenen Abstraktionsebenen der Stimuli zu T1

	<i>Thema F8</i>										<i>Thema F9</i>							
	<i>gesamt</i>		<i>alle</i>		<i>Oberbegriff</i>		<i>Zwischenbegriffe</i>		<i>Unterbegriffe</i>		<i>alle</i>		<i>Oberbegriff</i>		<i>Zwischenbegriffe</i>		<i>Unterbegriffe</i>	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
<i>M</i>	19,77	14,59	20,19	14,17	24,00	16,90	20,00	11,97	19,10	15,47	19,35	15,00	11,50	21,80	20,73	14,43	20,60	13,30
<i>SD</i>	11,90	6,21	11,92	8,87	16,53	18,31	11,72	8,83	14,25	6,64	12,86	4,47	4,06	12,39	14,31	5,05	15,29	4,53
Minimum	10,79	6,79	11,29	4,71	9,00	2,00	11,33	2,33	10,00	6,67	9,14	6,57	6,00	6,00	8,67	7,33	7,67	6,00
Maximum	50,79	26,36	49,71	31,43	67,00	57,00	48,00	28,67	57,33	28,33	51,86	21,29	16,00	45,00	55,3	23,67	60,33	21,67

Anhang 8: Studie 2 (psychotherapeutische Ambulanz) – statistische Kennwerte (Mittelwert und Standardabweichung) des Assoziationskennwertes im Vergleich zwischen den Formen A und B, über die verschiedenen Abstraktionsebenen der Stimuli zu T2

	<i>Thema F8</i>										<i>Thema F9</i>							
	<i>gesamt</i>		<i>alle</i>		<i>Oberbegriff</i>		<i>Zwischenbegriffe</i>		<i>Unterbegriffe</i>		<i>alle</i>		<i>Oberbegriff</i>		<i>Zwischenbegriffe</i>		<i>Unterbegriffe</i>	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
<i>M</i>	16,94	15,86	15,90	14,43	16,30	15,80	16,40	12,97	15,27	15,43	17,79	17,29	18,11	20,20	18,77	19,50	16,70	14,10
<i>SD</i>	7,28	5,18	7,86	6,98	8,79	13,48	9,18	7,60	7,67	6,27	7,58	5,22	13,87	9,16	7,18	3,87	7,47	7,13
Minimum	9,00	11,43	8,14	5,14	7,00	4,00	8,00	4,33	7,00	6,33	9,71	12,86	6,00	9,00	11,00	13,00	8,67	3,67
Maximum	32,64	27,71	30,14	26,57	37,00	47,00	35,00	26,67	31,33	26,33	35,14	28,86	52,00	35,00	30,67	27,00	34,00	28,67